



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Predicting tick-borne disease risk: improving the distribution mapping of ticks and tick bite risk in Scotland**

Rita Cláudia Cardoso Ribeiro



This thesis is presented for the degree of Doctor of  
Philosophy at The University of Edinburgh

## Declaration

I declare that this thesis has been composed entirely by the candidate, Rita Ribeiro. This work has not previously been submitted for a Doctor of Philosophy, a degree or any professional qualification. I have done all the work described except where acknowledged otherwise. All sources of information have been acknowledged.

Some material has been prepared in collaboration with co-authors as follows:

**Chapter 2:** Tick distribution data were provided by Professor Lucy Gilbert (University of Glasgow), Dr William Wint (Environmental Research Group Oxford (ERGO)) and Dr Jolyon Medlock (Public Health England (PHE)).

**Chapter 3:** Professor Finn Lindgren (University of Edinburgh) provided statistical advice and help.

**Chapters 4 and 5:** The internal database was managed by Dr Aaron Reeves and Dr Andrew Duncan from the Scotland's Rural College (SRUC); the website was developed by Dr Andrew Duncan; the mobile application (app) was developed by a freelancer.

**Chapters 2, 3 and 5:** Moderate Resolution Imaging Spectroradiometer (MODIS) data were provided by ERGO.

Rita Ribeiro

January 2021

To the pillars of my life, my family

## Abstract

Maps of *Ixodes ricinus* distribution and maps depicting tick bite risk are essential for government and health organisations to target prevention and control strategies for tick-borne diseases. However, researchers often lack robust long-term and geographically extensive tick distribution data, and information about human exposure to ticks to measure risk. Citizen science projects, through the collective effort of many volunteers, have the potential to provide valuable data on tick bite risk and tick distribution, but are often based on opportunistic submission of reports.

The overarching aim of this study was to assess methodologies to improve public health decision-making through distribution mapping of ticks and tick bite risk for Scotland. Research was undertaken to: compare the quality and robustness of predictive mapping with the three types of tick data most often used for predictive mapping; use statistical approaches to improve the quality of predictions of the distribution of *I. ricinus* in Scotland, including the predicted uncertainty; assess whether questing tick surveys reflect human-tick encounter risk; and finally, test the feasibility of a new citizen science approach to assess human risk of tick encounters.

Analysis of the three existing datasets with *I. ricinus* distribution in Scotland showed that whereas data from questing tick surveys generate detailed predictive maps at local scale, at the country level, predictions were affected by poor data coverage. Additionally, dissimilarities in the predicted distribution pattern of *I. ricinus* between data from passive submission and from questing tick surveys were identified. This suggests the need for data from public submissions to gather information on absences and to account for volunteer effort. A predictive map of *I. ricinus* distribution in Scotland developed with a sophisticated Bayesian statistical technique (the stochastic partial differential equation (SPDE)) which accounted for several sources of variation was

successful in improving the predictions in areas with poor data coverage, and the associated uncertainty.

The relationship between questing tick surveys and human tick bite rate was then assessed. Questing tick surveys were carried out whilst collecting contemporaneous data on tick encounters from orienteers running the same areas in 11 events at world, national, regional and local orienteering events. This novel approach found that questing tick surveys are a good indicator of tick bite risk. Also, the number of people multiplied by the hours of exposure is the most meaningful denominator for human exposure to tick bites (correlation coefficient with questing tick abundance of 0.8,  $p=0.0052$ ). From 340 reports from orienteers recorded across all events, a mean incidence of 409 tick bites per 1,000 person-hours exposure was recorded. Significant correlations were found between tick bite rate and temperature on the event day, the proportion of pastures around the track used by orienteers and the start time of the activity.

A citizen science project was implemented in Scotland between May and November 2018 and again between March and November 2019. The project used a novel approach that included collection of denominator data (number of people exposed, and time spent) and additionally asked people to report both when they did, and importantly, when they did not encounter ticks. Tick bite and tick encounter rates calculated from participant reports were compared with predictions of questing tick abundance in two study areas, Lochaber and the Cairngorms using data collected from questing tick surveys. A total of 1,914 reports from 65 volunteers were received, with 231 and 118 reports received, respectively, from the Cairngorms and Lochaber areas. On average, the Cairngorms area registered 0.083 tick bites per person per hour of activity and 0.268 tick encounters per person per hour. Lochaber area registered 0.018 tick bites and 0.028 tick encounters per person per hour. Tick bite and tick encounter rates in the Cairngorms correlated better with predicted tick abundance in the area (correlation coefficient of 0.27 and 0.31,

respectively) compared to Lochaber (correlation coefficient of 0.15 in both cases). Tick bite and tick encounter rates were found to depend both on questing tick abundance, and on factors related to human activity and behaviour. Type of human activity explained more variation in tick bite rate than questing tick abundance. Tick bite and tick encounter rates were quantified by activity type. These findings are valuable in identifying high risk activities and targeting public health messages.

This study resulted in new methodologies to improve predictive mapping of ticks, and better understanding of tick bite risk and the factors that drive it, with the overall aim of improving control and prevention of tick-borne diseases.

## Lay Summary

In Europe, the bite of infected ticks called *Ixodes ricinus* can transmit the bacteria causing Lyme disease. Adequate knowledge of *I. ricinus* distribution and the risk of tick bite are essential for government and health organisations to implement prevention and control strategies for diseases transmitted by ticks. This, in turn, can help citizens in their decision-making to prevent tick bites. However, it is difficult to develop accurate maps of tick distribution and tick bite risk. One of the limitations that researchers often face is the lack of robust tick distribution data, particularly for making predictions at national level. To overcome this problem, tick data that were not collected for the purpose of predictive mapping are frequently used. This lack of appropriate data negatively affects the quality of the predictions. In addition, these maps are normally developed with data where ticks are collected from the vegetation when waiting to attach to a host, but the relationship between these data from surveys of the vegetation and the actual risk of tick bites to people is not fully understood. Through participation in citizen science projects, people can submit information associated with their encounters with ticks, hence gathering data that can be used by researchers to estimate tick bite risk and tick distribution. However, these reports of tick sightings often lack detailed information on people's exposure to ticks (such as the number of people exposed, the time of exposure or the distance covered through risky habitats), which therefore introduces uncertainty into the estimates of risk.

This study investigated methodologies to improve predictive maps of tick distribution and tick bite risk. The research was conducted with tick data gathered in Scotland, a country with a high number of cases of Lyme disease (from 2008 to 2013 the estimated average annual human incidence was 6.8 per 100,000). To begin with, the quality of three types of data that are often used to develop tick distribution maps was compared. Results revealed that the tick data collected from the vegetation using standard methods were the most useful for predictions at local scale. However, at national level, the low



number of data points in some areas made the predictions less accurate. Additionally, differences in the predicted distribution pattern of *I. ricinus* between people's submissions and from surveys in the vegetation were identified. This may be because tick sightings do not provide information on tick absence and lack information about the time people are exposed and on people's level of commitment to submit when they find ticks. Subsequent work applied a specific modern statistical technique (the stochastic partial differential equation (SPDE)) which considers several sources of variation in the tick data. Using this approach, it was possible to make more accurate predictions of *I. ricinus* distribution in Scotland.

Finally, this project explored the components that explain the risk of getting tick bites. Initially, factors that affect tick activity (such as weather) and human behaviour (such as type of activity undertaken) were kept constant. This was achieved by recruiting volunteers who were orienteers, who reported information on their encounters with ticks, time spent in, and route through, tick areas after having competed in a defined location. The results were then compared with tick abundance estimated using data from surveys of the vegetation in the same area and time. Results indicated that tick abundance in an area is an important factor to explain the risk of tick bites. This study then tested the advantages of a novel citizen science approach. People carrying out different outdoor activities were asked to submit information on tick encounters, even when they did not find any ticks, and information on the time spent and route taken. These data were then used to look again at the relationship between tick abundance data from surveys of the vegetation, and tick bite risk. Results showed that the risk of getting a tick bite depends both on the abundance of ticks in the vegetation and on human behaviour. Human behaviour, such as type of activity undertaken, was the factor that most explained human exposure to ticks and tick bites. This study also generated measures of tick bite risk for different outdoor activities, which is important in identifying high risk groups and targeting public health messages.

This work presents new methodologies for tick distribution and for tick bite risk mapping and suggests what additional information needs to be collected to improve our understanding of the complexity of human-tick encounters.

## Acknowledgements

This PhD was a once-in-a-lifetime journey, a scientific adventure which I did not make alone. Therefore, I would like to express my gratitude (unfortunately, very briefly) to everyone who has supported me throughout this unforgettable journey.

Firstly, I would like to thank to my five supervisors: from SRUC, Professor George Gunn and Dr Jude Eze; from the University of Edinburgh, Professor Alastair Macrae; and from the University of Glasgow, Dr Harriet Auty and Professor Lucy Gilbert. Collectively, thank you for all the supervision, support and dedication. Harriet, I don't have enough words to express my gratitude. Thank you for always being there for me, for sharing your enthusiasm with me regarding this research, for the work with project volunteers, for the ticks we collected together, for patiently correcting my writing, and for always encouraging me to think further. Jude, you have the biggest heart in the world. Whenever I shared the happy or sad and stressful moments of this PhD with you, you always had something kind and positive to say to me. Thank you for the statistical support and for introducing me to the world of Bayesian modelling. Alastair, thank you so much for always keeping me on track to write a thesis to meet the high standards of the University of Edinburgh, for your dedication to this work, and for making the effort to travel north to Inverness so many times. Lucy, thank you so much for showing me the tick ecology side of my research, to help me formulate valuable scientific questions, and for your great enthusiasm. George, thank you for always believing in my work, and for all your professional and personal unwavering support. Perhaps, if it was not you, *maintenant j'étais en France en essayant de défendre mon PhD des Culicoides*.

I would like to acknowledge that this PhD was funded by a SRUC Postgraduate Studentship, without which this work would not have been possible. I am also

grateful for the good advice of my thesis committee members, Dr James Prendergast and Professor William Mackaness.

My thanks to all my colleagues from the Epidemiology Research Unit (SRUC), for all these years of friendship, brainstorming moments and fruitful discussions, runs, walks, cakes, WhatsApp groups, and for believing in me: Carla Gomes, Judith Evans, Catriona Webster, Ian Hutchinson, Jane Brennan, Jo Baughan, Kate Stephen, Aaron Reeves, Sue Tongue, Julie Stirling, Hannah Bishop, Roger Humphry, Andrew Duncan, Michael Pearce, Madeleine Henry, Jason Millar, Maria Costa, and Annette Boerlage (and the canine friends, Flo, Riley, Oliver and Diesel). I am particularly grateful for the support and dedication of Jo for correcting my texts and emails, for helping me to improve my English skills, and for collecting ticks with me (Jo, there is not enough gingerbread in this world to thank you); Jane for helping me to organise my field trips and orienteering events; Judith and Catriona for helping me to organise the (nearly 10,000) ticks in the lab freezer; Aaron for all his support with the app and the database; and Andrew for all the help with the database, R code, website and app.

This research has been possible thanks to the land managers and landowners who allowed me to carry out questing tick surveys on their properties, and to the project volunteers. I am grateful for their interest in this research and for their additional effort in reporting negative tick encounters. A special thank you goes to the following organisations: Scottish woodlands, Glenmore lodge, Abernethy Trust, and the University of the Highlands and Islands. I particularly want to thank all the orienteers that contributed to this research, particularly the clubs INVOC, MAROC, MOR and BASOC. Individually I would like to acknowledge Colin Matheson, Sheila Reynolds and Juliette Maire.

I also would like to thank all the expert advice and help of Professor Finn Lindgren (University of Edinburgh), Dr Paul Bessell (Roslin Institute), and Dr William Wint and Neil Alexander (ERGO). I am also very grateful to Dr Roger

Evans and Dr Sally Mavin (from the Scottish Lyme Disease and Tick-borne infections Reference Laboratory) for helping me at the world orienteering event and in analysing some of the ticks. Thank you to the students Hannah Mylin and Amy Cooper.

Finally, thank you to my family and friends who have been with me on this journey and kept me going with love, friendship, and encouragement. Above all, my parents (Ana e Ilídio) and sister Mafalda, who I admire the most and have always been my biggest support. Finally, to my partner David for all his love and unconditional support during these years, and for reminding me that there's more to life than research.

# Table of contents

<b>Declaration</b> .....	<b>i</b>
<b>Abstract</b> .....	<b>iii</b>
<b>Lay Summary</b> .....	<b>vi</b>
<b>Acknowledgements</b> .....	<b>ix</b>
<b>Table of contents</b> .....	<b>xii</b>
<b>List of figures</b> .....	<b>xx</b>
<b>List of tables</b> .....	<b>xxiv</b>
<b>List of abbreviations and acronyms</b> .....	<b>xxvi</b>
<b>Chapter 1 : Introduction</b> .....	<b>1</b>
1.1. The public health importance of vector-borne diseases.....	1
1.1.1. Tick-borne diseases in Europe.....	3
1.1.1.1. Lyme borreliosis.....	3
Genetic diversity of <i>Borrelia burgdorferi</i> ( <i>sensu lato</i> ) .....	3
Public health relevance of Lyme borreliosis .....	4
1.1.1.2. Other tick-borne diseases vectored by <i>I. ricinus</i> .....	5
1.1.2. Biology and ecology of <i>I. ricinus</i> .....	7
Infection of <i>I. ricinus</i> with <i>Borrelia burgdorferi</i> ( <i>s.l.</i> ) .....	9
Phenology of <i>I. ricinus</i> .....	10
1.1.2.1. Human-tick interactions: risk factors for human-tick encounters .....	12
1.1.2.2. Other tick species present in the United Kingdom .....	13
1.2. Modelling and mapping vectors for public health decision-making .	14
1.2.1. The public health importance of maps of disease vectors.....	14
1.2.2. Species distribution models.....	15
1.2.2.1. Approaches used and examples for disease vectors.....	15
Vector distribution data used in SDMs .....	19

1.2.3.	Uncertainty and other knowledge gaps .....	20
	Mapping <i>I. ricinus</i> in Scotland .....	21
	Relationship between questing tick surveys and tick bite risk.....	21
1.3.	Citizen science: a powerful tool for data collection .....	25
1.3.1.	Challenges of citizen science datasets and possible solutions .	26
1.3.2.	Value of citizen science to study disease vectors.....	28
1.3.2.1.	Citizen science projects about tick vectors of diseases .....	29
	Current challenges and knowledge gaps .....	31
1.4.	Thesis aim and objectives.....	34
<b>Chapter 2 : Using imperfect data in predictive mapping of vectors: a regional example of <i>Ixodes ricinus</i> distribution.....</b>		<b>36</b>
2.1.	Introduction .....	36
2.2.	Materials and Methods.....	39
2.2.1.	Tick data.....	39
	Dataset 1 - <i>I. ricinus</i> abundance data from questing tick surveys ...	40
	Dataset 2 - <i>I. ricinus</i> public tick submissions.....	40
	Dataset 3 - <i>I. ricinus</i> combined dataset.....	41
2.2.2.	Georeferenced environmental data and variable selection .....	42
2.2.3.	Statistical model, model validation and predictive map .....	43
2.2.4.	Kernel density analysis, model and predictive and uncertainty maps for a sub-sample of Dataset 1 .....	46
2.3.	Results.....	46
	Model 1 (tick relative abundance, using Dataset 1: questing tick surveys) .....	47
	Model 2 (tick presence-absence, using Dataset 2: tick information from public submissions) .....	49

Model 3 (tick presence-absence, using Dataset 3: combined dataset)	49
Model and predictive map for the subset of Dataset 1	54
2.4. Discussion	57
2.4.1. Predictors for <i>I. ricinus</i> relative abundance and presence-absence	58
2.4.2. Use of <i>I. ricinus</i> abundance data from questing tick surveys (Dataset 1)	60
2.4.3. Use of <i>I. ricinus</i> presence data obtained through public submissions plus absence points (Dataset 2)	62
2.4.4. Use of <i>I. ricinus</i> combined datasets (Dataset 3)	63
2.4.5. Predictive <i>I. ricinus</i> maps for Scotland	64
2.5. Conclusions	65
<b>Chapter 3 : Improving predictive mapping of <i>Ixodes ricinus</i> distribution in Scotland by accounting for spatial correlation using the Stochastic Partial Differential Equation (SPDE)</b>	<b>67</b>
3.1. Introduction	67
The need to account for spatial correlation when modelling tick distribution	67
The Bayesian approach to deal with spatial correlation	68
The stochastic partial differential equation	70
Study objectives and hypothesis	71
3.2. Materials and Methods	72
3.2.1. The inlabru R package	72
3.2.2. Quantitative <i>I. ricinus</i> survey data: data management	73
3.2.3. Environmental covariates	74
3.2.4. The mesh and the SPDE model	75



3.2.5. Modelling and predictions.....	77
3.3. Results.....	83
3.3.1. The constrained refined Delaunay triangulation: mesh .....	83
3.3.2. Assessment of model predictive quality: comparison between Models 1 and 2.....	83
3.3.2.1. Predicted posterior tick density and predicted uncertainty .	87
3.3.3. New predictive models with the random field .....	88
3.3.4. Assessment of the extra variation due to the drag and site effects, and the ZIP process in model predictions .....	92
3.4. Discussion .....	93
3.4.1. The accuracy and robustness of model validation and selection ..	94
3.4.2. The value in adding the spatial random field to improve model predictions and the predicted uncertainty .....	96
3.4.3. The value of adding the random effects of drag and site, and the ZIP process, in improving model uncertainty.....	98
3.4.4. Inlabru vs INLA/SPDE .....	98
3.5. Conclusions .....	99

**Chapter 4 : Linking human tick bite risk with tick abundance in the environment: a survey of orienteers in Scotland.....101**

4.1. Introduction.....	101
4.2. Materials and Methods.....	104
4.2.1. Development and testing of the questionnaire and website for data collection .....	105
4.2.2. Engagement with the orienteering community and selection of events .....	107
4.2.3. Data collection at orienteering events .....	107

4.2.4.	Questing tick surveys: data collection .....	109
4.2.5.	Questing tick surveys: data management .....	111
4.2.5.1.	Calculating the proportion of land cover at sampling sites and event .....	111
4.2.5.2.	Calculating the mean of questing ticks at sampling sites and event .....	111
4.2.6.	Orienteering data: management and analysis.....	112
4.2.6.1.	Calculating the distance run .....	112
4.2.6.2.	Calculating the proportion of land cover from tracking data... ..	112
4.2.7.	Statistical analysis .....	113
4.2.7.1.	Assessment of the relationship between questing tick surveys and human tick bite risk .....	113
Comparing habitat and weather conditions between questing tick surveys and orienteering event days .....		113
Selection of the appropriate denominator for tick bite and tick encounter rates.....		114
The effect of the variable 'group of orienteers'.....		115
4.2.7.2.	Determination of the environmental risk factors for tick bite risk to humans.....	116
4.2.7.3.	Estimation of the mean incidence of tick bites and tick encounters of orienteers .....	117
4.3.	Results.....	117
4.3.1.	Descriptive analysis of data collected at orienteering events ..	117
4.3.2.	Descriptive analysis of the tick data collected during questing tick surveys .....	120
4.3.3.	Relationship between questing tick surveys and tick bite risk .	121

Comparing habitat and weather conditions between questing tick surveys and orienteering event days .....	121
Selection of the appropriate denominator for tick bite and tick encounter rates.....	122
The effect of ‘group of orienteers’ .....	126
4.3.4. Environmental risk factors for human tick bite risk .....	127
4.3.5. Mean incidence of tick bites and tick encounters among orienteers .....	131
4.4. Discussion .....	131
4.4.1. Relationship between questing tick surveys and the actual tick bite risk to humans .....	132
4.4.2. Environmental risk factors for tick bite risk to humans.....	135
4.4.3. Tick bite and tick encounter rates among orienteers .....	138
4.4.4. Contributions and study limitations.....	141
4.5. Conclusions .....	142
<b>Chapter 5 : Use of a novel citizen science approach to assess human risk of tick bites.....</b>	<b>144</b>
5.1 Introduction .....	144
5.2. Materials and Methods.....	147
5.2.1. Engagement with volunteers .....	147
5.2.2. The citizen science project ‘TickApp’.....	148
5.2.2.1. Citizen science data management.....	150
5.2.3. Definition of the two study areas .....	152
5.2.3.1. Identification of habitats for questing tick surveys in the two study areas .....	154
5.2.3.2. Data collection during questing tick surveys .....	155
5.2.4. Statistical analysis .....	157

5.2.4.1.	Descriptive analysis .....	157
5.2.4.2.	Relationship between predicted questing tick abundance and human tick bite and tick encounter rates in the Cairngorms and Lochaber .....	157
	Mapping tick bite and tick encounter rates in the Cairngorms and Lochaber.....	157
	Predictive model and map of questing tick abundance in the Cairngorms and Lochaber .....	158
	Georeferenced environmental data and variable selection.....	158
	Statistical models and predictive maps for questing tick abundance in the Cairngorms and Lochaber .....	159
	Correlation between predicted tick abundance from questing tick surveys and tick bite and tick encounter rates in Cairngorms and Lochaber.....	160
5.2.4.3.	Identification of risk factors for tick bite and tick encounter rates .....	161
5.3.	Results.....	163
5.3.1.	Descriptive analysis of the CS data.....	163
	Descriptive analysis of the risk factors for tick bite and tick encounter rates.....	167
	Descriptive analysis of questing tick data .....	170
5.3.2.	Relationship between predicted questing tick abundance and human tick bite and tick encounter rates in the Cairngorms and Lochaber .....	172
	Mapping tick bite and tick encounter rates in the Cairngorms and Lochaber.....	172
	Predictive model and map of questing tick abundance in the Cairngorms and Lochaber .....	175

Correlation between predicted tick abundance from questing tick surveys and tick bite and tick encounter rates in the Cairngorms and Lochaber.....	178
5.3.3. Identification of risk factors for tick bite and tick encounter rates ..	179
5.4. Discussion .....	181
5.4.1. Relationship between predicted questing tick abundance and human tick bite and tick encounter rates .....	182
Citizen science data.....	184
Predictive maps of questing tick abundance.....	187
5.4.2. Risk factors for tick bite and tick encounter rates .....	189
5.4.3. Implications of this work and future research priorities.....	192
5.5. Conclusions .....	195
<b>Chapter 6 : Final discussion .....</b>	<b>196</b>
6.1. Connecting the dots and interpreting results.....	197
6.2. Main contributions and implications: what gaps have the results filled?	202
6.2.1. Scientific contributions.....	202
6.2.2. Contributions for policy decision-making .....	204
6.2.3. Public health contributions .....	206
6.3. Study limitations and future research.....	207
<b>References.....</b>	<b>213</b>
<b>Appendix A: Supplementary material for Chapter 1 .....</b>	<b>255</b>
<b>Appendix B: Supplementary materials for Chapter 2 .....</b>	<b>258</b>
<b>Appendix C: Supplementary materials for Chapter 4 .....</b>	<b>278</b>
<b>Appendix D: Supplementary materials for Chapter 5 .....</b>	<b>280</b>

## List of figures

<b>Figure 1.1:</b> The three active life stages of <i>I. ricinus</i> tick.....	8
<b>Figure 1.2:</b> <i>I. ricinus</i> life cycle.....	8
<b>Figure 2.1:</b> Spatial distribution of the three datasets used in the models. ...	47
<b>Figure 2.2:</b> Predictive maps of <i>I. ricinus</i> distribution in mainland Scotland with respective uncertainty. ....	51
<b>Figure 2.3:</b> Interquartile range of the covariates used in the three models..	53
<b>Figure 2.4:</b> Predictive map of <i>I. ricinus</i> distribution in Aberdeenshire with respective uncertainty. ....	56
<b>Figure 2.5:</b> Interquartile range of the covariates used in the model for Aberdeenshire.....	57
<b>Figure 3.1:</b> Distribution of the training and the testing datasets in mainland Scotland.....	74
<b>Figure 3.2:</b> The two-dimensional constrained refined Delaunay triangulation (mesh) used in this work to define the spatial dependency structure.....	83
<b>Figure 3.3:</b> (a) Model standardised residuals (y-axis) for the training dataset for Models 1 and 2; (b) model standardised residuals (y-axis) for the testing dataset for Models 1 and 2. ....	85
<b>Figure 3.4:</b> Prediction of the posterior density surface of counts of <i>I. ricinus</i> nymphs per drag using the grid of triangulations covering the domain and respective uncertainty (the SD) for Models 1 and 2. ....	88
<b>Figure 3.5:</b> Prediction of the posterior density surface of counts of <i>I. ricinus</i> nymphs per drag using the grid of triangulations covering the domain and uncertainty (the SD) for Model 3. ....	91
<b>Figure 3.6:</b> Comparison between the predicted uncertainty from Model 3 with and without the variations due to the random effects and the ZIP process..	93
<b>Figure 4.1:</b> The ‘tick tent’ at two orienteering events.....	108
<b>Figure 4.2:</b> Distribution of the number of event competitors who did and did not participate in the study. ....	118
<b>Figure 4.3:</b> Distribution of the number of tick bites and tick encounters, per person per report, in each event. ....	119

<b>Figure 4.4:</b> Distribution of the number of questing ticks per blanket drag (10m <sup>2</sup> ) per event area.....	120
<b>Figure 4.5:</b> Correlation coefficient (as R) and respective P-value, between the average proportion of land cover types around the tracks of orienteers and the average proportion of land cover types at the sites of questing tick surveys, in each of the 11 event areas. ....	121
<b>Figure 4.6:</b> Correlation coefficient (as R) and respective P-value, between the average number of ticks collected per drag (10m <sup>2</sup> ) per event area (three sites), and tick bite rate.....	124
<b>Figure 4.7:</b> Correlation coefficient (as R) and respective P-value between the average number of ticks collected per drag (10m <sup>2</sup> ) per event area (three sites), and tick encounter rate. ....	124
<b>Figure 4.8:</b> Distribution of the number of questing ticks per blanket drag (10m <sup>2</sup> ) per 'group of orienteers'.....	126
<b>Figure 4.9:</b> Correlation coefficient (as R) and respective P-value between tick bite and tick encounter rates (per person per hour) and temperature, rain and proportion of forest covered. ....	129
<b>Figure 5.1:</b> The two study areas where questing tick surveys were carried out and volunteers reported to the CS project; 'A' the Cairngorms, and 'B' Lochaber.....	153
<b>Figure 5.2:</b> Practices during questing tick surveys. ....	156
<b>Figure 5.3:</b> Citizen science data submitted to the project 'TickApp' in Scotland, between May and November 2018 and March and November 2019. ....	163
<b>Figure 5.4:</b> Citizen science data submitted to the project 'TickApp' from the two study areas, between May and November 2018 and March and November 2019.....	164
<b>Figure 5.5:</b> Frequency and seasonal distribution of the reports submitted by activities. ....	166
<b>Figure 5.6:</b> Percentage of reports submitted by type of path used. ....	166
<b>Figure 5.7:</b> Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per type of activity, in the Cairngorms and Lochaber. ....	167

<b>Figure 5.8:</b> Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per type of path, in the Cairngorms and Lochaber.	168
<b>Figure 5.9:</b> Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per season, in the Cairngorms and Lochaber. ....	169
<b>Figure 5.10:</b> Mean (with SE bars) of tick bite and tick encounter rates (per person per hour) between March to October/or November using combined data from 2018 and 2019.....	170
<b>Figure 5.11:</b> Average number (with SE bars) of questing ticks (nymphs and adult ticks) collected per 10m <sup>2</sup> , per land cover type and season, from questing tick surveys carried out in 2019, in the Cairngorms and Lochaber.....	171
<b>Figure 5.12:</b> Average number per drag (with SE bars) of questing ticks (nymphs and adult ticks) collected per 10m <sup>2</sup> in each month, in (a) the Cairngorms and (b) Lochaber. ....	171
<b>Figure 5.13:</b> (a) Number of people (volunteers and other members of the groups) participating in the project in the Cairngorms, between May to November 2018 and March to November 2019 per 1 km <sup>2</sup> pixel; (b) tick bite and (c) tick encounter rates (per person per hour) in each 1 km <sup>2</sup> pixel.....	173
<b>Figure 5.14:</b> (a) Number of people (volunteers and other members of the groups) participating in the project in Lochaber, between May to November 2018 and March to November 2019 per 1 km <sup>2</sup> pixel; (b) tick bite and (c) tick encounter rates (per person per hour) in each 1 km <sup>2</sup> pixel. ....	174
<b>Figure 5.15:</b> Predictive map of tick abundance and uncertainty map for the Cairngorms. ....	176
<b>Figure 5.16:</b> Predictive map of tick abundance and uncertainty map for Lochaber.....	177
<b>Figure 5.17:</b> Correlation coefficient (R) and significance of the correlation between the observed number of questing ticks collected per 10m <sup>2</sup> , visit and site, and the predicted questing tick abundance at (a) Cairngorms and (b) Lochaber.....	177
<b>Figure 5.18:</b> Correlation coefficient (R) and significance of the correlation between the predicted questing tick abundance and: (a) tick bite and (c) tick	



encounter rates, in the Cairngorms; and (b) tick bite and (d) tick encounter rates, in Lochaber. .... 178

**Figure 6.1:** The components of tick bite rate, according to the studies carried out in Chapters 4 and 5..... 204

## List of tables

<b>Table 1.1:</b> Non-exhaustive list of published work with predictions of the distribution of <i>I. ricinus</i> in Europe. ....	23
<b>Table 1.2:</b> Non-exhaustive list of citizen science projects implemented in Europe. ....	33
<b>Table 2.1:</b> Georeferenced environmental variables used in the study. ....	42
<b>Table 2.2:</b> Results for Model 1 (Dataset 1). ....	48
<b>Table 2.3:</b> Results for Model 2 (Dataset 2) and Model 3 (Dataset 3). ....	50
<b>Table 2.4:</b> Model results for Aberdeenshire area. ....	55
<b>Table 3.1:</b> Description of Models 1 and 2. ....	84
<b>Table 3.2:</b> S_DS and S_LOG scores for Models 1 and 2. ....	86
<b>Table 3.3:</b> Differences in the score mean, and respective P-value, between Model 1 and Model 2, for the training and testing datasets. ....	87
<b>Table 3.4:</b> Description of Models 3, 4 and 5. ....	89
<b>Table 3.5:</b> S_DS and S_LOG scores for Models 3, 4 and 5. ....	90
<b>Table 3.6:</b> Differences in the score mean, and respective P-value, between Models 2, 4 and 5 and Model 3 (used as reference model), for the training and testing datasets. ....	90
<b>Table 3.7:</b> Results of Model 3. ....	92
<b>Table 4.1:</b> Results from four negative binomial models for the number of tick bites and tick encounters. ....	123
<b>Table 4.2:</b> Results from six univariable negative binomial models for tick bite and tick encounter rates. ....	125
<b>Table 4.3:</b> Results from the multivariable negative binomial models for tick bite and tick encounter rates (per person per hour). ....	127
<b>Table 4.4:</b> Results from the final multivariable negative binomial models for tick bite and tick encounter rates (per person per hour) and the studied environmental covariates. ....	130
<b>Table 5.1:</b> Georeferenced environmental variables used in the study. ....	158
<b>Table 5.2:</b> Summary statistics of the number of reports submitted to the CS project in the Cairngorms and Lochaber. ....	165

<b>Table 5.3:</b> Final model outputs for questing tick abundance in the Cairngorms and Lochaber, for parameterising respective predictive maps.....	175
<b>Table 5.4:</b> Results of the final model to identify risk factors for tick bite rate. ....	180
<b>Table 5.5:</b> Results of the final model to identify risk factors for tick encounter rate.....	181
<b>Table 6.1:</b> Appropriate tick data to use to predict and map tick distribution and/or human risk of tick bites at local, regional or country levels. ....	206

## List of abbreviations and acronyms

App: Application

BRT: Boosted regression trees

CPO: Conditional predictive ordinate

CS: Citizen science

DIC: Deviance information criteria

GAM: Generalised additive model

GARP: Genetic algorithm for rule set production

GIS: Geographic information system

GLM: Generalised linear model

GLMM: Generalised linear mixed model

GPS: Global positioning system

HERC: Human (Research) Ethical Review Committee

ID: User identification

INLA: Integrated nested Laplace approximation

LB: Lyme borreliosis

LST: Land surface temperature

MaxEnt: Maximum entropy model

MCMC: Markov chain Monte Carlo

MODIS: Moderate resolution imaging spectroradiometer

NDVI: Normalised difference vegetation index

PHE: Public Health England

RDSVS: Royal (Dick) School of Veterinary Studies

RF: Random forest

SD: Standard deviation

SDM: Species distribution model

S\_DS: Dawid-Sebastiani score

SE: Standard error

s.l.: sensu lato

S\_LOG: Logarithmic score

SOA: Scottish orienteering association

SPDE: Stochastic partial differential equation

SRUC: Scotland's Rural College

s.s.: sensu stricto

TBD: Tick-borne disease

TBE: Tick-borne encephalitis

TBEV: Tick-borne encephalitis virus

UK: United Kingdom

US: United States

VBD: Vector-borne disease

WAIC: Watanabe-Akaike information criteria

ZIP: Zero-inflated Poisson

# Chapter 1: Introduction

Since Lyme borreliosis (LB) was identified in the 1970s (Steere et al., 1977), its importance for public health has become clear, and it is the most common tick-borne disease (TBD) of humans in the Northern Hemisphere (O'Connell et al., 1998; van den Wijngaard et al., 2015; Stone, Tourand and Brissette, 2017). Mapping and control of LB is particularly relevant in Scotland, where the incidence of LB is one of the highest in Europe (Ling *et al.*, 2000). Therefore, the overall focus of this thesis is to investigate new methodologies to improve predictive mapping of ticks, and to provide a better understanding of tick bite risk and the factors that drive it, with the overall aim of improving control and prevention of TBDs.

## 1.1. The public health importance of vector-borne diseases

Vector-borne diseases (VBDs) are a group of infectious diseases caused by bacteria, viruses and parasites, which are transmitted to hosts (humans and/or animals) by a vector – such as mosquitoes, ticks, triatomines, sandflies, blackflies, aquatic snails, fleas, lice and tsetse flies. This group of diseases, of which some are zoonotic, pose a major threat to public and animal health, and an economic burden to society. Vector-borne diseases account for more than 17% of all human infectious diseases, and result in about 700,000 deaths every year (WHO, 2017, 2020). Malaria, the most lethal VBD (WHO, 2017), was found to be associated with slower economic development (Gallup and Sachs, 2001), and just in 2016, \$4.3 billion was spent on malaria worldwide (Haakenstad *et al.*, 2019). The World Health Organisation lists 12 VBDs as being of major importance for public health, and three of them are transmitted by ticks, namely LB or Lyme disease, tick-borne encephalitis (TBE) and Crimean-Congo haemorrhagic fever (WHO, 2017, 2020). For animal health, there is a long list of important VBDs, several of which are also transmitted by ticks such as babesiosis, theileriosis and anaplasmosis (OIE, 2020).

The occurrence of VBDs used to be limited to certain geographic areas of the world. However climate change has resulted in increases in the distribution and abundance of disease vectors, and the rapid unplanned urbanisation, deforestation, intensive farming, international trade and travel are increasing the risk of human-vector contact. Therefore, all of these factors have contributed to the emergence and re-emergence of vector species and outbreaks of VBDs into new areas (Hay *et al.*, 2002; Purse *et al.*, 2005; Gray *et al.*, 2009; Medlock *et al.*, 2012, 2018; WHO, 2014, 2017; Medlock and Leach, 2015; Müller *et al.*, 2019). In the last two decades, the world has faced the unprecedented global spread of dengue, chikungunya and yellow fever (WHO, 2017), and outbreaks of West Nile (Nash *et al.*, 2001; Hadler *et al.*, 2015) and Zika virus (Musso, Cao-Lormeau and Gubler, 2015). Europe, in particular, was afflicted with outbreaks of West Nile (Bisanzio *et al.*, 2011; Napp, Petrić and Busquets, 2018), chikungunya (Venturi *et al.*, 2017) and dengue viruses (Lourenço and Recker, 2014), and with the return of local malaria transmission (Andriopoulos *et al.*, 2013). There were also outbreaks of bluetongue and Schmallenberg viruses affecting ruminants (Purse *et al.*, 2005; Hoffmann *et al.*, 2012; Ribeiro *et al.*, 2015).

In the last two decades, changes in distribution and incidence have also been reported for TBDs. Several countries in Europe have reported an increase in reported cases and an expansion of the range of LB and TBE (e.g. Daniel *et al.*, 2003; Lindgren and Jaenson, 2006; Lukan, Bullova and Petko, 2010; Jaenson, Hjertqvist, *et al.*, 2012; Hofhuis *et al.*, 2015; Mavin, Watson and Evans, 2015; Tulloch *et al.*, 2019). One of the reasons for this is the expansion of the geographical range of the main vector, *Ixodes ricinus* (e.g. Tälleklint and Jaenson, 1998; Lindgren and Gustafson, 2001; Daniel *et al.*, 2003; Cadenas *et al.*, 2007; Gray *et al.*, 2009; Jaenson, Hjertqvist, *et al.*, 2012; Laaksonen *et al.*, 2017).

These examples of emergence and re-emergence of VBDs demonstrate the need to better understand disease carrying vectors and their distribution, as a necessary step to decrease the global burden of VBDs (WHO, 2017).

### **1.1.1. Tick-borne diseases in Europe**

In Europe, TBDs are the most common VBDs (ECDC, 2012). *I. ricinus* (Linnaeus, 1758) is the most abundant and widespread tick species (Gray, 1998; Tälleklint and Jaenson, 1998; Gritsun, Lashkevich and Gould, 2003; Gassner *et al.*, 2011; Estrada-Peña *et al.*, 2013; Oechslin *et al.*, 2017), and also the species most often biting humans (Robertson, Gray and Stewart, 2000; Bartosik *et al.*, 2011; Cull *et al.*, 2019; Lernout *et al.*, 2019). *I. ricinus* is medically highly relevant as it is the dominant vector of LB and TBE (Gray, 1998; Perret, Rais and Gern, 2004).

#### **1.1.1.1. Lyme borreliosis**

##### **Genetic diversity of *Borrelia burgdorferi* (*sensu lato*)**

Lyme borreliosis is caused by the *Borrelia burgdorferi* (*sensu lato*) complex of spirochete bacteria (Burgdorfer *et al.*, 1983; Ling *et al.*, 2000). *B. burgdorferi* (*s.l.*) comprises at least 20 genospecies globally (Stone, Tourand and Brissette, 2017), which vary in pathogenicity and cause different symptoms. They also differ in geographic distribution, Ixodidae vector and transmission host(s) (van Dam *et al.*, 1993; Stone, Tourand and Brissette, 2017; Mysterud *et al.*, 2019). From this group, four genospecies have been reported in ticks in Scotland, namely *B. afzelii* (the most common genospecies), *B. garinii*, *B. burgdorferi* (*sensu strictu*) and *B. valaisiana* (James *et al.*, 2012, 2014; Millins *et al.*, 2016). *B. afzelii* and *B. garinii* are most frequently associated with clinical disease in Europe (Stanek and Reiter, 2011), but all four genospecies can cause human disease (Rijpkema *et al.*, 1997; Diza *et al.*, 2004; Stone, Tourand and Brissette, 2017). Although these genospecies exhibit some differences in clinical presentation (e.g. *B. garinii* is neurotropic and *B. afzelii* has been mostly associated with skin manifestations), associated symptoms and signs have not



been fully characterised due to difficulty in diagnosis (Rijpkema *et al.*, 1997; Stone, Tourand and Brissette, 2017).

The genospecies of the *B. burgdorferi* (s.l.) complex infect a wide range of vertebrate hosts. In Europe, reservoirs for *B. afzelii* are mainly small mammals such as the wood mouse, voles, shrews, red and grey squirrels (Hanincová, Schäfer, *et al.*, 2003; James *et al.*, 2014; Millins *et al.*, 2015; Mysterud *et al.*, 2019). *B. garinii* is mainly transmitted by birds, although it is also found in red and grey squirrels (Hanincová, Taragelova, *et al.*, 2003; James *et al.*, 2014; Millins *et al.*, 2015). *B. valaisiana* is associated with birds (Hanincová, Taragelova, *et al.*, 2003), and *B. burgdorferi* (s.s.) is generalist, able to infect both mammals and birds (Millins *et al.*, 2016). Prevalence of infection with *B. burgdorferi* (s.l.) in ticks is related to the abundance of competent hosts, and also with the burden of ticks on transmission hosts (Ostfeld and Keesing, 2000; James *et al.*, 2012; Millins *et al.*, 2016). Therefore, LB risk depends on the density of infected questing ticks, which depends on both the density of questing ticks ('questing' is the technique *I. ricinus* ticks use to find a host: ticks climb up the vegetation and wait for a passing host) (Mejlon and Jaenson, 1997) and the presence of competent reservoir hosts such as rodents and birds, as well as incompetent hosts such as deer which host ticks but do not infect them with the pathogen (James *et al.*, 2012; Millins *et al.*, 2016).

### **Public health relevance of Lyme borreliosis**

Although most cases that are treated promptly are relatively mild, LB can have serious consequences for infected people, leading to debilitating disease with significant morbidity if left untreated (Mavin, Watson and Evans, 2015; Steere *et al.*, 2016). Human infection with *Borrelia burgdorferi* (s.l.) can have three stages: first the appearance of a skin lesion 'erythema migrans' (stage 1), which if untreated, can be followed by the disseminated infection, particularly affecting the nervous system (stage 2), and by late infection, which often involves arthritis or acrodermatitis chronica atrophicans (stage 3) (Steere *et al.*, 2016). Therefore LB poses a large burden on health services (Lohr *et al.*,

2015). For example, in the Netherlands, the total burden of LB was estimated in 2010 at 10.55 disability-adjusted life years per 100,000 population (van den Wijngaard *et al.*, 2015).

In Europe, there are an estimated 85,000 reported cases of LB each year (ECDC, 2012), however reporting in the affected countries is inconsistent, and many cases are undiagnosed (Lindgren and Jaenson, 2006; Cairns *et al.*, 2019). A recent study for Western Europe has estimated a population-weighted average incidence rate of 22.04 cases per 100,000 person years (Sykes and Makiello, 2017). In the United Kingdom (UK), LB is not a notifiable disease, but laboratory confirmed cases of infection with *Borrelia* spp. are notifiable. A recent publication indicated that the incidence of LB in England and Wales has increased since 1997, reaching 1.95 cases per 100,000 in 2016 (Tulloch *et al.*, 2019). Scotland is one of the European countries with the highest LB incidence rate, especially in the Highlands (Ling *et al.*, 2000). Since 1996, the number of reported cases of LB in Scotland has increased, and from 2008 to 2013 the estimated average annual incidence was 6.8 per 100,000 (44.1 per 100,000 in the Highlands). However since 2010 is when LB stopped being a notifiable disease in Scotland, data since then are not comparable with earlier data (Mavin, Watson and Evans, 2015).

There is no vaccine available to help preventing infection. Therefore prevention is based on minimising the risk of exposure to ticks and tick bites, rapid removal of attached ticks and early recognition of symptoms to initiate prompt treatment (Piesman and Dolan, 2002; Kilpatrick *et al.*, 2017; Sprong *et al.*, 2018; Gillingham *et al.*, 2020).

#### **1.1.1.2. Other tick-borne diseases vectored by *I. ricinus***

Besides LB, there are other TBDs whose pathogens are transmitted by *I. ricinus*. These include TBE, a well-established disease in central Europe, and emerging tick-borne zoonoses, such as babesiosis.

Tick-borne encephalitis is a severe encephalitis in humans caused by tick-borne encephalitis virus (TBEV), a virus which belongs to the flavivirus group, family Flaviviridae and genus Flavivirus (Gritsun, Lashkevich and Gould, 2003; Lindquist and Vapalahti, 2008). There are other viruses antigenically related to TBEV (Porterfield, 1975; Calisher, 1988) such as louping ill virus, which is also transmitted by *I. ricinus*, but causes disease in sheep, cattle, grouse and other species, only rarely affecting humans (Macleod and Gordon, 1932; Jeffries *et al.*, 2014; Gilbert, 2016). This similarity between TBEV and louping ill virus can cause cross-reactivity in serological tests and therefore diagnostic challenges, such as indicated in a recent study in the UK (Kreusch *et al.*, 2019).

Although most TBE cases are asymptomatic and vaccination is available (Demicheli, Debalini and Rivetti, 2009), in case of infection, the development of disease can cause permanent neurological sequelae and even death (Gritsun, Lashkevich and Gould, 2003; Lindquist and Vapalahti, 2008; Šmit and Postma, 2015). In Europe in 2017, 2,550 confirmed cases of TBE were reported, including nine deaths. The Czech Republic, Germany and Lithuania were the European countries with most reported confirmed cases (ECDC, 2019b). However, it is considered that between 10,000 to 12,000 cases of TBE occur annually in Europe and Asia (Gritsun, Lashkevich and Gould, 2003; WHO, 2017). Until 2019, TBEV was not considered to occur in the UK. However recent research in England identified the first possible human case of TBE (Kreusch *et al.*, 2019), and the occurrence of infected ticks with TBEV in some local areas (Holding *et al.*, 2020; Holding, Dowall and Hewson, 2020). The origin for this first appearance of TBEV in UK is not yet known, but may be associated with the transport of infected ticks by migratory birds (PHE, 2019).

In addition to *Borrelia burgdorferi* (s.l.) and TBEV, *I. ricinus* can also transmit several other microorganisms, including *Anaplasma phagocytophilum* (Macleod and Gordon, 1933) causing human granulocytic anaplasmosis (Dumler *et al.*, 2001); *Rickettsia helvetica* (Nilsson, Lindquist and Pålsson,

1999) and *Rickettsia monacensis* (Jado *et al.*, 2007) causing spotted fever rickettsiosis; *Babesia divergens*, *Babesia microti* and *Babesia venatorum*, causing zoonotic babesiosis (Zintl *et al.*, 2003; Hildebrandt, Gray and Hunfeld, 2013; Westblade *et al.*, 2017; Gray *et al.*, 2019); *Borrelia miyamotoi*, a relapsing fever *Borrelia* which can induce symptoms similar to LB (Fukunaga *et al.*, 1995; Franck *et al.*, 2020); *Candidatus Neoehrlichia mikurensis*, causing an inflammatory disease in humans (Fehr *et al.*, 2010; Portillo *et al.*, 2018), and *Francisella tularensis* causing the zoonotic disease tularaemia (McCoy and Chapin, 1912; Ellis *et al.*, 2002; Gehringer *et al.*, 2012). With the exception of *Francisella tularensis*, the role of these pathogens in causing human disease has not yet been defined (Azagi *et al.*, 2020). In Europe, most clinical cases are related with co-infections or occur in immunocompromised patients. Therefore, infections may often be neglected or misdiagnosed (Hildebrandt, Gray and Hunfeld, 2013; Sprong *et al.*, 2018; Azagi *et al.*, 2020). Nonetheless, as the number of reported cases is increasing, these microorganisms are known as emerging infections in humans (Hildebrandt, Gray and Hunfeld, 2013; Oechslin *et al.*, 2017; Gray *et al.*, 2019; Franck *et al.*, 2020). For example in the UK, the recent identification of *B. miyamotoi* in *I. ricinus* ticks (Hansford *et al.*, 2015) and *B. venatorum* in sheep (Gray *et al.*, 2019) have raised concerns about their potential public health importance.

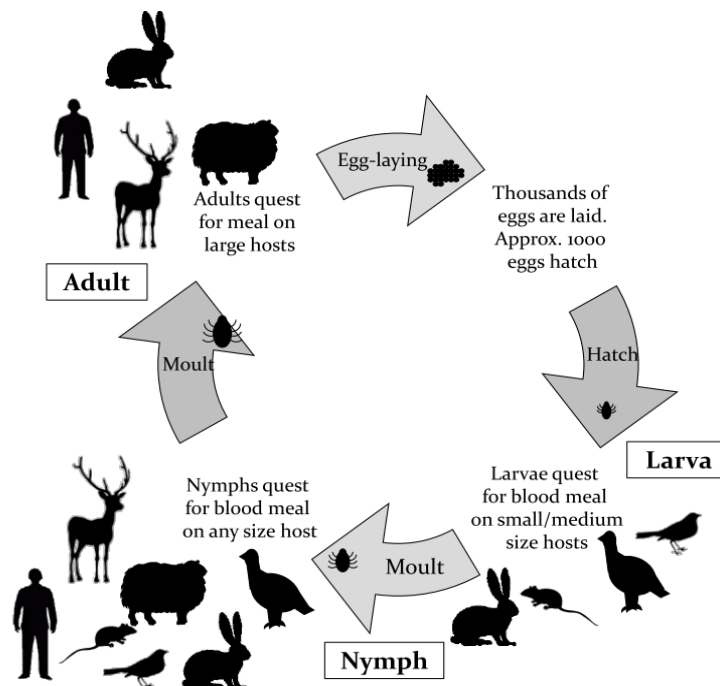
### **1.1.2. Biology and ecology of *I. ricinus***

*I. ricinus* belongs to the family Ixodidae. Ixodid ticks have four life stages: egg, larvae, nymphs and adults (Gray, 1998; Anderson and Magnarelli, 2008). Larvae, nymphs and adults only require one blood meal: the two immature stages require a blood meal before developing to the next stage, and the adult female requires a blood meal before producing eggs (Figures 1.1 and 1.2) (Randolph, 1998; Anderson and Magnarelli, 2008; Gassner *et al.*, 2011). Each tick stage feeds on a different host individual for a few days, then detaches, drops to the ground, and develops in the vegetation to the next stage (Gray, 1998; Randolph, 1998). Adult ticks mate on the host. After feeding, females

lay a batch of eggs (about 3000 or more eggs) and the cycle is completed (Anderson and Magnarelli, 2008).



**Figure 1.1:** The three active life stages of *I. ricinus* tick. From top, anticlockwise, one adult female, two larvae, and one nymph (reproduced from Lindquist and Vapalahti, 2008).



**Figure 1.2:** *I. ricinus* life cycle. *I. ricinus* life cycle with the preferred hosts for each of the three active life stages (reproduced from Norman, Worton and Gilbert, 2016).

*I. ricinus* ticks have low mobility and their dispersion is dependent on host movement (Randolph, 1998). The three activity stages quest at different vegetation heights, which can also be related to differences in host preference (Mejlon and Jaenson, 1997). Larvae emerge from the eggs and quest near the ground, where the relative humidity is higher (Mejlon and Jaenson, 1997; Gray, 1998). Ticks are vulnerable to dehydration whilst questing; when waiting for a host, questing ticks can leave their questing site and move to the litter zones to rehydrate (Perret, Rais and Gern, 2004).

*I. ricinus* feed on a wide range of vertebrate hosts including reptiles, birds and mammals (MacLeod, 1936; Perret, Rais and Gern, 2004). Although the most important hosts for feeding larvae and nymphs are birds, small mammals such as rodents (James *et al.*, 2012; Cayol *et al.*, 2017) and deer (Mysterud, Hatlegjerde and Sørensen, 2014), the immature stages are found on hosts of all sizes (including humans), whilst adult stages feed more exclusively on larger hosts such as sheep and deer (Figure 1.2) (James *et al.*, 2012; Norman, Worton and Gilbert, 2016). Deer are often the most important host for adult female ticks, and are often termed 'tick reproduction hosts' (Gray, 1998). Therefore, deer presence or density is positively associated with tick densities (Gilbert *et al.*, 2012; Hofmeester *et al.*, 2017; Dickinson, Millins and Biek, 2020). Although deer play an important role as hosts of *I. ricinus*, deer are incompetent hosts of *B. burgdorferi* (*s.l.*) (Jaenson and Tälleklint, 1992; Rosà *et al.*, 2003; Mysterud *et al.*, 2019).

### **Infection of *I. ricinus* with *Borrelia burgdorferi* (*s.l.*)**

The transmission cycle of *B. burgdorferi* (*s.l.*) to *I. ricinus* is complex and dependent on the host competence for the bacteria (Randolph, 1998; James *et al.*, 2012), where host competence is defined as the duration and probability of effectively transmitting the bacteria to a feeding tick (particularly important if it is a genospecies likely to cause disease in humans) (Kilpatrick *et al.*, 2017). Measuring host competence is challenging since it requires trapping and monitoring of wild caught individuals, or experimental infections. Competence

also depends on the different *Borrelia* genospecies and varies among both individuals and populations of a host species. Therefore, it can also vary between regions (Kilpatrick *et al.*, 2017). Competent hosts for *B. burgdorferi* (*s.l.*) in Europe were listed previously in Section 1.1.1.1.

The most significant mode of transmission is the transstadial infection of larvae and nymphs feeding on competent reservoir hosts (Randolph, 1998, 2004). Transovarial infection via a female tick to its offspring can occur, but this is rare (Hubálek and Halouzka, 1998), and in general, less than 1% of the questing larvae are infected (Gray, 1998; Wagemakers *et al.*, 2015). Additionally, uninfected ticks may also acquire a non-systemic infection by co-feeding close to infected ticks (Gern and Rais, 1996; Ogden, Nuttall and Randolph, 1997; Randolph, 2009). Although adult ticks have higher prevalence of infection than nymphs (e.g. Kowalec *et al.*, 2017), nymphs are most important in the transmission of *B. burgdorferi* (*s.l.*) to humans, as nymphs are the stage that most frequently bites humans (Robertson, Gray and Stewart, 2000).

### **Phenology of *I. ricinus***

Understanding tick phenology, the seasonal population dynamics of different tick life stages, is essential to understand the transmission of tick-borne pathogens (Perret, Rais and Gern, 2004; Levi *et al.*, 2015). The life cycle of *I. ricinus* is generally completed in three years, but it may vary from two to six years in warmer or colder climates. The phenology therefore varies considerably in the different regions that this tick is present, depending on variations in biotic (host species, host density and behaviour, and vegetation structure) and abiotic factors (climate) (Gray, 1991).

Climate can influence the abundance and distribution of *I. ricinus*, its life cycle processes, seasonal activity, population dynamics, survival, development time and individual tick behaviour (Pietzsch *et al.*, 2005; James *et al.*, 2012). This major influence of climate results from the direct influence of temperature and relative humidity on *I. ricinus* development, survival and questing. In fact,

temperature and relative humidity are thought to be the principal factors limiting the geographic range of *I. ricinus* (MacLeod, 1936; Lindgren, Talleklint and Polfeldt, 2000; Perret *et al.*, 2000; Randolph *et al.*, 2002).

Temperature directly influences *I. ricinus* interstadial development rate, oviposition rate and egg development rates (Randolph *et al.*, 2002). In warmer climates, this process may allow a tick to emerge as the next stage within the same season (Dobson and Randolph, 2011), whilst in colder climates (e.g. Scotland), most ticks will not emerge as the next stage until the following year. Higher temperatures also induce questing behaviour and extend questing season in cooler climates (Randolph, 2004; Hancock, Brackley and Palmer, 2011; Gilbert, Aungier and Tomkins, 2014; Tomkins *et al.*, 2014). In contrast, tick activity is inhibited by cold temperatures, and cold temperatures and/or short day length in winter induce tick diapause (dormancy) (Randolph *et al.*, 2002). Field surveys in Scotland, Switzerland and northern Italy have estimated that a weekly average maximum daily temperature of seven/eight degrees Celsius is needed for *I. ricinus* nymphs to emerge from their overwintering diapause and starting questing for hosts in spring (Gilbert, Aungier and Tomkins, 2014; Tomkins *et al.*, 2014). Ticks are also vulnerable to desiccation during questing, when developing to the next instar (moulting) and, regarding adult females, when laying eggs. To survive desiccation, it is critical that the relative humidity of their microclimate does not fall below 80%, which is found in areas with a good vegetation cover and a mat of decaying vegetation (Gray, 1998; Gassner *et al.*, 2011).

These requirements for temperature, humidity, vegetation cover and hosts mean that ticks have preferences for certain habitats, such as forested areas which have canopies and understory vegetation that create a microclimate for ticks, maintaining suitable conditions of temperature and humidity, and are also home of hosts for ticks (Gray *et al.*, 1998; Estrada-Peña, 2001).



In recent decades, several countries (such as Norway, Sweden, Germany, Czech Republic and UK, including Scotland), have reported an increasing abundance and distribution range of *I. ricinus*, spreading into higher latitudes and altitudes (Tälleklint and Jaenson, 1998; Daniel *et al.*, 2003; Kirby *et al.*, 2004; Materna, Daniel and Danielová, 2005; Scharlemann *et al.*, 2008; Jore *et al.*, 2011; Jaenson, Jaenson, *et al.*, 2012; Schwarz *et al.*, 2012; Hvidsten *et al.*, 2020). The expansion of *I. ricinus* is not completely understood, but is attributed to abiotic environmental changes, such as climate change (Lindgren, Talleklint and Polfeldt, 2000; Gern, Cadenas and Burri, 2008; Gilbert, Aungier and Tomkins, 2014; Jore *et al.*, 2014; Alkische, Peterson and Samy, 2017; Hvidsten *et al.*, 2020), and also changes in host animal abundance and distribution (for example in Scotland, the increase in deer abundance and distribution (Clutton-Brock, Coulson and Milner, 2004; Gilbert *et al.*, 2012)), which relate to landscape structure and land use (Gilbert, 2013; Ehrmann *et al.*, 2017).

#### **1.1.2.1. Human-tick interactions: risk factors for human-tick encounters**

An understanding of *I. ricinus* ecology helps in identifying spatial and temporal patterns and risk factors for human-tick encounters, including tick bites. *I. ricinus* activity is seasonal, and ticks are active and questing for hosts when temperature and humidity conditions are suitable during the year; from early spring to autumn. Human tick bites can occur during all periods of tick activity. In UK, human tick bites were reported from March to November, with peaks in the summer months of June and July (Cull *et al.*, 2019). This can be explained by the influence of temperature on both tick activity and human behaviour, by favouring outdoor activities and the use of light clothing (Keukeleire *et al.*, 2015).

Habitat type is also very important in determining the human risk of tick bites because it influences both tick abundance and survival, and the likelihood of human-tick contact (Tack *et al.*, 2012). Risky habitats for tick bites include

habitats that support high tick populations, such as forested areas, particularly mixed and deciduous forests, which are also common areas for recreational activities (Robertson, Gray and Stewart, 2000; Bartosik *et al.*, 2011; Mulder *et al.*, 2013; Keukeleire *et al.*, 2015; Garcia-Martí *et al.*, 2018). However, ticks are also found in suburban and urban environments, including green spaces used for recreational activities (Maetzel, Maier and Kampen, 2005; Hansford *et al.*, 2017; Oechslin *et al.*, 2017; Jore *et al.*, 2020), and tick bites are reported to occur in private and public gardens (Mulder *et al.*, 2013). Therefore, people are exposed to ticks and tick bites through recreational and occupational activities that bring them into areas where ticks are present (Bartosik *et al.*, 2011; Faulde *et al.*, 2014; Keukeleire *et al.*, 2015; Hall *et al.*, 2017; Fernandez *et al.*, 2019).

#### **1.1.2.2. Other tick species present in the United Kingdom**

There are 20 identified species of tick endemic in the UK (Jameson and Medlock, 2011). Human and veterinarian reports indicate that the most frequently reported tick species in the UK are *I. ricinus*, followed by *I. hexagonus*, *I. canisuga* and *I. frontalis* (Ogden *et al.*, 2000; Jameson and Medlock, 2011; Abdullah *et al.*, 2016; Cull *et al.*, 2018). However, *I. ricinus* is the species that most commonly bites humans (Jameson and Medlock, 2011; Cull *et al.*, 2018). The fact that *I. ricinus* is a generalist feeder and awaits a host on the vegetation, whilst the other species are specialists that commonly live in the nests of their host species, means that contact between *I. ricinus* and humans is more frequent (Piesman and Gern, 2004; Jameson and Medlock, 2011). In Scotland, all reported questing tick surveys found that 100% of the ticks were *I. ricinus* (James *et al.*, 2012; Millins *et al.*, 2015; Gandy, 2020), whilst all surveys of human tick bites have also found all ticks to be *I. ricinus* (Hall *et al.*, 2017).

## **1.2. Modelling and mapping vectors for public health decision-making**

### **1.2.1. The public health importance of maps of disease vectors**

Since most vectors are strongly influenced by their environment, they usually have heterogeneous distributions. Therefore, maps of the distribution of disease vectors have been extensively used in research and policy. These outputs constitute an essential tool to help preparedness for VBDs outbreaks, and assist public health organisations, government and other stakeholders to set public health guidelines and policies (De Roeck *et al.*, 2014; Estrada-Peña, Alexander and Wint, 2016; ECDC and EFSA, 2018; ECDC, 2019a). For example, the European project VectorNet has a network of experts for gathering, sharing and analysing data of the distribution of several vectors of diseases, including ticks (ECDC and EFSA, 2014), to inform European level decision-making.

Maps with the distribution of vectors are important tools for vector surveillance and monitoring, allowing identification of areas of vector occurrence, and areas with high potential for infection introduction and transmission (Guerra *et al.*, 2002; Bisanzio *et al.*, 2011; Kraemer *et al.*, 2015; Boehnke *et al.*, 2015; Alkiske, Peterson and Samy, 2017; Santos and Meneses, 2017; Kamal *et al.*, 2018; Domşa, Mihalca and Sándor, 2018; Ducheyne *et al.*, 2018; Ferro e Silva *et al.*, 2018; Simons *et al.*, 2019; ECDC, 2019a; Hönig *et al.*, 2019; Aguilar-Vega *et al.*, 2020). These maps increase the understanding of environmental factors which influence vector distribution and seasonality (Eisen, Eisen and Lane, 2006; Ribeiro *et al.*, 2015; Diarra *et al.*, 2018; Ferro e Silva *et al.*, 2018) and help identify areas of future expansion or contraction of vector distribution, to then inform adaptation efforts (Porretta *et al.*, 2013; Boeckmann and Joyner, 2014; Feria-Arroyo *et al.*, 2014; Kraemer *et al.*, 2019). Maps with the distribution of vectors also help identify the seasonal and spatial variation of the risk of human-vector contact (Moiroux *et al.*, 2013; Swart *et al.*, 2014; Garcia-Martí *et al.*, 2018; Kjær *et al.*, 2019a). In addition, these maps help to make surveillance more cost-effective, directing efforts to target areas (Sallam

*et al.*, 2016; Cuéllar *et al.*, 2018) and can be used as tools to help communicate and raise awareness for local risks (Vourc'h *et al.*, 2016; Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017).

### **1.2.2. Species distribution models**

Most vector distribution maps are based on predicted distributions. Predictive maps of a species' distribution are the outcome of species distribution models (SDMs), also known as environmental or ecological niche models. This big group of models relate species distribution data with environmental information and geographic space data, to make predictions about species distribution (Elith and Leathwick, 2009; Miller, 2010). Epidemiological data on disease distribution can also be incorporated into these models and maps to produce risk maps for disease distribution (Kitron, 2000).

Species distribution models have evolved over the years together with the development of new mathematical and statistical techniques, computing power and storage capacity, and with the availability of extensive environmental datasets (Elith and Leathwick, 2009; Miller, 2010; Braks *et al.*, 2016). The choice of technique depends on the question to be studied (Elith and Leathwick, 2009), the type of response variable and covariates available (Guisan and Thuiller, 2005), and the sample size, and can be influenced by the complexity of the modelling technique and computational resources (Mengersen *et al.*, 2017).

#### **1.2.2.1. Approaches used and examples for disease vectors**

Species distribution models can be classified into two general approaches, mechanistic or process-based, and empirical-statistical models (Braks *et al.*, 2016). Mechanistic models quantify the relationship between the target variable and environmental attributes based on previous knowledge and proven causality of the aspects of the process to be modelled. This is achieved through the simulation of the mechanisms considered to be responsible for the observed correlation between target and environmental variables (Beerling,

Huntley and Bailey, 1995). Therefore, this type of model is often used to describe patterns of disease dynamics (Lloyd-Smith *et al.*, 2009; Xiao and Zou, 2014; Heesterbeek *et al.*, 2015), and not just vector distribution. In the field of TBDs, mechanistic models have been used to investigate the spatial-temporal pattern of infected tick abundance and the influence of temperature on LB risk (Dobson and Randolph, 2011; Dobson, Finnie and Randolph, 2011; Hancock, Brackley and Palmer, 2011; Li *et al.*, 2016; Worton, 2016).

Empirical-statistical models (also called correlative environmental models), quantify the correlation between species distribution records and environmental covariates (Braks *et al.*, 2016). This group of models can be classified into categories of methods to deal with presence-only data, regression and machine learning methods (according to Miller, 2010). One of the most commonly used methods for using presence-only data to describe the habitat suitability of a species is ecological niche factor analysis (Miller, 2010). This is a multivariable analysis which compares the geographical predicted distribution for the locations with presence data, with the predicted distribution of the whole area, and generates a habitat suitability index for each cell (or pixel) in the area of study (Hirzel, Helfer and Metral, 2001; Hirzel *et al.*, 2002). This method has been used to describe the habitat suitability of *I. ricinus* (Table 1.1, which shows examples of methodological approaches, including data types used, that have been reported for *I. ricinus*).

Regression methods establish a relationship between vector presence-absence or abundance, and a set of environmental covariates, and generally assume that the response variable (vector distribution) can be modelled using additive combinations of predictors (Elith and Leathwick, 2009). The statistical algorithm can then be used to predict vector distribution (Lehmann, Overton and Leathwick, 2002). The most frequently used regression methods are generalised linear models (GLMs) and generalised additive models (GAMs) (Thuiller, 2003). These methods can deal with presence-absence or abundance data by using a different link function (e.g. binomial, Poisson,

negative binomial or zero inflated). Although GLMs are not able to deal with complex response curves, GAMs describe non-linear responses and can provide more flexibility for fitting ecologically realistic relationships (Yee and Mitchell, 1991; Lehmann, Overton and Leathwick, 2002; Elith and Leathwick, 2009). Regression methods have been frequently applied to study vector distribution, with examples using data from *Culicoides* midges, which are vectors of bluetongue virus (Ribeiro *et al.*, 2015; Ramilo *et al.*, 2017; Diarra *et al.*, 2018), mosquitos transmitting malaria (Moiroux *et al.*, 2013) or West Nile virus (Bisanzio *et al.*, 2011), *I. scapularis* (the main vector of LB in America and Canada) (Guerra *et al.*, 2002; Brownstein, Holford and Fish, 2003; Diuk-Wasser *et al.*, 2010, 2012; Koffi *et al.*, 2012) and *I. ricinus* (Table 1.1).

Latent Gaussian models are a subset of structured additive regression models. Other examples of structured additive regression models include, among others, GLMs, GAMs, smoothing spline models, state space models, semiparametric regression, spatial and spatiotemporal models, log-Gaussian Cox processes and geostatistical and geosadditive models (Rue, Martino and Chopin, 2009). Latent Gaussian models can be implemented within a Bayesian framework. Bayesian inference is based upon Thomas Bayes' theorem and considers that just one form of uncertainty exists, which is described by probability distributions, and combines two types of information to derive the posterior distribution of parameters and predictions - the prior and the (current) data model (Bernardo and Smith, 2000). Bayesian inference has been seen as a potential tool to improve SDMs, due to the integration of prior information (Bernardo and Smith, 2000), providing better ways to deal with errors and uncertainty (Bernardo and Smith, 2000; Guisan and Thuiller, 2005), the ability to account for spatial and temporal dependency, and the ability to develop complex hierarchical models (Rue, Martino and Chopin, 2009; Lindgren and Rue, 2015). Spatial and spatial-temporal latent Gaussian models which account for dependency in the data can be easily computed with the integrated nested Laplace approximation (INLA) and the stochastic partial differential equation (SPDE) (Lindgren, Rue and Lindström, 2011). Bayesian

models using INLA and SPDE have been developed for some disease vectors, such as *Culicoides* midges (Kifle, Hens and Faes, 2017), *Culex pipiens* (Myer, Campbell and Johnston, 2017) and tsetse flies (Stanton *et al.*, 2018) but not yet for ticks. The value of a Bayesian approach for modelling and mapping species distribution will be described in detail in Chapter 3.

Machine-learning methods are a category of more flexible ‘data-driven’ methods, which are not based on specific distribution functions, and do not require *a priori* model specification (Miller, 2010). Although machine learning methods can be complex and difficult to interpret (sometimes compared to a ‘black box’), and models can be demanding in terms of running time and computation resources (Elith, Leathwick and Hastie, 2008), they have been widely applied to model vector distribution. Examples include *Culicoides* midges (Cuéllar *et al.*, 2018; Diarra *et al.*, 2018; Aguilar-Vega *et al.*, 2020); mosquito vectors of Zika, dengue and chikungunya (Kraemer *et al.*, 2015, 2019; Santos and Meneses, 2017; Ducheyne *et al.*, 2018; Kamal *et al.*, 2018; Richman *et al.*, 2018; Tiffin *et al.*, 2019) and mosquito vectors of malaria (Moffett, Shackelford and Sarkar, 2007; Sinka *et al.*, 2012), triatomine vectors of Chagas disease (Sarkar *et al.*, 2010; Ferro e Silva *et al.*, 2018); *I. scapularis* (Atkinson *et al.*, 2014; Feria-Arroyo *et al.*, 2014; Johnson *et al.*, 2016; Soucy *et al.*, 2018) and *I. ricinus* (Table 1.1). From this group of methods the ones most commonly adopted for modelling and mapping *I. ricinus* distribution are random forest (RF) (Breiman, 2001; Prasad, Iverson and Liaw, 2006); boosted regression trees (BRT) (Friedman, 2001; Elith, Leathwick and Hastie, 2008) and maximum entropy models (MaxEnt) (Phillips, Anderson and Schapire, 2006). Random forest and BRT can be used with abundance and presence-absence data, and are ensemble model techniques thereby fitting many models and combining them for prediction (Elith, Leathwick and Hastie, 2008; Leathwick *et al.*, 2008). MaxEnt is a method able to deal with presence-only data, however it does require the use of ‘background’ information (Phillips, Anderson and Schapire, 2006; Phillips *et al.*, 2009).

However, different model techniques vary in their ability to capture all the variations present in the data, and therefore predictions of species distributions can vary widely among modelling approaches. Therefore, to decrease the predictive uncertainty of single models, some researchers opt to combine predictions from multiple models (from different techniques) into an 'ensemble', obtaining a 'consensus model' (Araújo and New, 2007). This approach has been adopted for studies into the predicted distribution of mosquito vectors of Rift Valley Fever, sandfly vectors of Leishmania, triatomine vectors of Chagas disease, *Culicoides* midges and ticks, using ensembles of models using presence-only data, regression and machine learning methods (Carvalho *et al.*, 2015; Chalghaf *et al.*, 2018; Ferro e Silva *et al.*, 2018; Ciss *et al.*, 2019; ECDC, 2019a; Simons *et al.*, 2019).

### **Vector distribution data used in SDMs**

Vector data used in SDMs comprises three types: presence-only data; presence-absence; and abundance data (Table 1.1). Presence-only data are frequently gathered from public submissions, citizen science projects, history museums and data curators (Estrada-Peña and Venzal, 2006; Estrada-Peña, Venzal and Acedo, 2006; Porretta *et al.*, 2013). One major limitation of presence-only data is that it lacks information on where the species was absent, which increases the risk of bias, and limits the type of questions that can be asked and the predictive power of the models (Pearce and Boyce, 2006; Bird *et al.*, 2014).

Presence-absence data may be collected in this form or can be generated from abundance data, and from the addition of pseudo-absence points to presence-only data. The generation of pseudo-absence points are based on information of unsuitable habitats for the vector (Pearce and Boyce, 2006), and is the approach frequently adopted by VectorNet to model *I. ricinus* distribution on a continental scale (ECDC, 2019a).



Vector abundance data are gathered from standardised methods of collection, where vectors are collected using traps with an attractant (carbon dioxide, light, odour, sound, animal or human bait) or by dragging or flagging (ECDC and EFSA, 2018). Questing ticks are generally collected by sweeping a flag (small piece of blanket), or by dragging a blanket over the vegetation, for a certain distance or time (Gray and Lohan, 1982), during questing tick surveys (thereafter in this thesis this method will be referred to as questing tick surveys). Although considered as standard techniques to collect ticks in the vegetation (Gilbert, 2010; ECDC and EFSA, 2018), flagging or dragging are not suitable means to estimate the density or abundance of ticks in the population (as they fail to collect ticks in diapause, moulting or attached to a host) but only the density or abundance of ticks that are questing (Perret, Rais and Gern, 2004). Additionally, the resources required for questing tick surveys mean that the data are often not available at larger geographical and seasonal scales (Li *et al.*, 2016).

### **1.2.3. Uncertainty and other knowledge gaps**

Predictive mapping of vector distribution, including *I. ricinus*, is a complex process, and models and correspondent predictive map outputs have uncertainty (Elith and Leathwick, 2009). This uncertainty results from data deficiencies due to data quality and coverage (for example small sample size, presence of bias and/or lack of absence points in distribution data), missing covariates (due to a lack of understanding on vector ecology), and from the model technique applied, which will be able to capture or not all the variation present (Barry and Elith, 2006; Estrada-Peña, Alexander and Wint, 2016). Therefore, although predictive maps of vector distribution are frequently published, including the plethora of predictive maps for *I. ricinus* (Table 1.1), very few studies address the uncertainty of the predictions (which is different from model sensitivity analysis), and consider uncertainty when using maps for decision-making (Elith and Leathwick, 2009). Addressing the predicted uncertainty is essential to assess the accuracy and robustness of the predictions, and is important to understand data and model limitations, which

can help to identify steps for data and model improvement (Heesterbeek *et al.*, 2015; Braks *et al.*, 2016). Additionally, it is useful to adopt robust and sophisticated modelling techniques which can account for data variability, and which give information on the predictive uncertainty (Heesterbeek *et al.*, 2015).

### **Mapping *I. ricinus* in Scotland**

Although LB is a public health concern in Scotland, very few maps with *I. ricinus* predicted distribution have been published, particularly at an appropriate scale for national and local decision-making (some examples include Braga, 2012; Li *et al.*, 2016).

Current *I. ricinus* distribution data available for Scotland comes from three different sources: presence-only data from public submissions to Public Health England (PHE), relative abundance data from questing tick surveys (from Professor Lucy Gilbert and co-workers), and a composite dataset which combines the presence-only data with data from questing tick surveys and the generation of pseudo-absence points (ECDC and EFSA, 2018; ECDC, 2019a). Although data from questing tick surveys are generally considered to be the gold standard, the data available for Scotland were collected by aiming to answer specific ecological hypotheses, rather than conducted for the purpose of mapping tick distribution. Therefore, there is a lack of reliable information on tick distribution in Scotland. The question of whether these three datasets can be used for developing tick distribution models, in terms of data quality, coverage and predicted uncertainty, will be addressed in Chapters 2 and 3, using *I. ricinus* distribution data for Scotland.

### **Relationship between questing tick surveys and tick bite risk**

Many of the published models and maps of tick distribution rely on data from questing tick surveys. These predicted outputs have been used to answer several questions regarding questing tick ecology, and to make inferences about human risk and exposure to ticks and TBDs (e.g. Eisen *et al.*, 2006; Eisen, Eisen and Lane, 2006; Diuk-Wasser *et al.*, 2010; Boehnke *et al.*, 2015;

Brugger *et al.*, 2016; Vourc'h *et al.*, 2016; Hönig *et al.*, 2019). However, studies that directly assess how well results from questing tick surveys explain human exposure and risk to ticks are rare. Additionally, there are other studies that use tick data from passive submissions to model tick distribution, without testing the quality of these data for robust inference, and/or gathering additional information to improve data quality. These questions will be addressed in Chapters 4 and 5.

**Table 1.1:** Non-exhaustive list of published work with predictions of the distribution of *I. ricinus* in Europe. The table describes the location, scale and spatial resolution used, the objective of the predictions, the response variable and source of data, the modelling approach, and if the predicted uncertainty was assessed.

Location, scale and resolution	Objective	Response variable, source	Method	Uncertainty	Reference
Europe, continental scale, resolution of 8 km <sup>2</sup>	Model and map current and future potential distribution	Presence-only data, different sources	Machine learning (GARP)	No	(Boeckmann and Joyner, 2014)
France, local scale, no information about resolution	Map human risk of infection with <i>B. burgdorferi</i> (s.l.)	Abundance data (flagging); density of infected nymphs	Regression (GLM)	No	(Vourc'h <i>et al.</i> , 2016)
Germany, local scale, resolution of 0.5 km <sup>2</sup>	Map spatial distribution	Abundance data (flagging)	Regression (GLM)	No	(Boehnke <i>et al.</i> , 2015)
Norway, local scale, no information about resolution	Determine the contribution of red deer space use on the spatial distribution of <i>I. ricinus</i>	Abundance data (flagging)	Regression (GLMM and GAM)	No	(Qviller <i>et al.</i> , 2016)
Europe, continental scale, resolution of 0.1 degree (approximately 11.1 km <sup>2</sup> )	Identify long-term trends of habitat suitability	Presence-only data (different sources)	Ecological Niche Factor Analysis	No	(Estrada-Peña and Venzal, 2006; Estrada-Peña, Venzal and Acedo, 2006)
Iceland, country scale, no information about resolution	Identify areas for future tick establishment	Presence-only data (different sources: flagging, ticks collected from mammals and birds, and from passive surveillance; and same number of pseudo-absences)	Machine learning (BRT)	No	(Alfredsson <i>et al.</i> , 2017)
Romania, country scale, resolution of 0.6 km <sup>2</sup>	Map spatial distribution	Presence-only data (flagging and published literature)	Machine learning (MaxEnt)	No	(Domşa, Mihalca and Sándor, 2018)
Eurasian region, continental scale, resolution of 5 km <sup>2</sup>	Predict current and future (2050 and 2080) climatic niche	Presence-only data (multiple sources: published literature and curated collections)	Machine learning (MaxEnt)	No	(Porretta <i>et al.</i> , 2013)
Italy, local scale, resolution of 1 km <sup>2</sup>	Identification of a cost-effective sampling strategy for local scale surveillance	Presence-only data (dragging)	Machine learning (MaxEnt)	No	(Signorini <i>et al.</i> , 2019)

**Table 1.1:** Continued

Location, scale and resolution	Objective	Response variable, source	Method	Uncertainty	Reference
Czech Republic and Germany, regional scale (areas from the two countries), resolution of 250 m <sup>2</sup>	Development of an acarological risk model	Abundance data (flagging) and prevalence with pathogens	Regression (GLM)	No	(Hönig <i>et al.</i> , 2019)
Germany, country scale, resolution of 100 m <sup>2</sup>	Map spatial distribution	Abundance data (flagging)	Regression (GLM)	No	(Brugger <i>et al.</i> , 2016)
Scotland, country scale; resolution of 1 km <sup>2</sup>	Develop a mechanistic, agent-based model to study the temperature-driven seasonality of <i>I. ricinus</i> ticks and transmission of <i>B. burgdorferi</i> ( <i>s.l.</i> )	Density of questing infected nymphs	Mechanistic agent-based model	No	(Li <i>et al.</i> , 2016; Worton, 2016)
Europe and North Africa, continental scale, about 20 km <sup>2</sup> resolution	Predict current and future (2050 and 2070) spatial distribution	Presence-only data (multiple sources: Global Biodiversity Information Facility, VectorMap, and scientific literature)	Machine learning (MaxEnt)	Yes	(Alkische, Peterson and Samy, 2017)
Scandinavia, regional scale, 1 km <sup>2</sup> resolution	Predict presence-absence and abundance	Presence-absence and abundance data (dragging)	Machine learning (BRT)	No	(Kjær <i>et al.</i> , 2019a, 2019b)
The Netherlands, country scale; resolution of 1 km <sup>2</sup>	Modelling and mapping tick dynamics	Abundance data (dragging and flagging by volunteers)	Machine learning (RF)	Yes	(Garcia-Martí, Zurita-Milla, van Vliet, <i>et al.</i> , 2017)
Scotland, country scale, resolution of 1 km <sup>2</sup>	Predict current and future (2050) nymph abundance in Scotland	Abundance data (dragging)	Regression (GLMM)	No	(Braga, 2012)
Europe, continental scale; resolution of 1 km <sup>2</sup>	Map spatial distribution	Presence-absence data (multiple sources: presence and absence from dragging or flagging, presence-only and pseudo-absences)	Machine learning (BRT and RT)	No	(ECDC, 2019a)

Abbreviations: BRT, boosted regression trees; GAM, generalised additive model; GARP, genetic algorithm for rule set production; GLM, generalised linear model; GLMM, generalised linear mixed model; MaxEnt, maximum entropy model; RF, random forest

### **1.3. Citizen science: a powerful tool for data collection**

The term 'citizen science' was introduced by the British sociologist Alan Irwin in 1995, to describe a more participatory science (in Cooper, 2017). A more complete definition is that citizen science (CS) refers to the engagement and active participation of common citizens in scientific research, with the aim of generating new scientific knowledge (Miller-Rushing, Primack and Bonney, 2012). Citizen science projects involve, to a varying degree, the collaboration of citizens and scientists and can be classified into three categories: 1) contributory projects where participants only contribute data; 2) collaborative projects where participants, besides contributing with data collection, help in other steps of the project (refine project design, analyse data or disseminate project outputs); and 3) co-created projects, where citizen scientists participate actively in the entire scientific project (Bonney *et al.*, 2009).

Citizen science is not a new subject, and citizens have been actively recording their observations of the natural world for centuries (Bonney *et al.*, 2009; Silvertown, 2009; Dickinson *et al.*, 2012; Miller-Rushing, Primack and Bonney, 2012). However, the advance of communications and computing technology has contributed to the increase in the number of CS projects (Silvertown, 2009; Miller-Rushing, Primack and Bonney, 2012; Bonney *et al.*, 2014). Nowadays, citizens can be powered with high levels of knowledge, internet access, mobile phones and other systems that allow them to easily receive and share data, anytime and anywhere (Bonney *et al.*, 2014; August *et al.*, 2015; Guilbaud and Guilbaud, 2017; Lwin *et al.*, 2017; Palmer *et al.*, 2017; Hamer, Curtis-Robles and Hamer, 2018). This advance and availability of new technologies has led to a new recognition of the potential of CS for collecting data at larger temporal and spatial scales, and allowing science to answer questions that could not have previously been addressed (Tregidgo, West and Ashmore, 2013; Palmer *et al.*, 2017).

Citizen science has been used to make numerous contributions to science, covering a broad range of research areas (Bonney *et al.*, 2014; Cooper, 2017).

In the last decades, there has been an increased scientific enthusiasm for CS. When using the term 'citizen science' in the search engine Web of Science, the number of outputs increased from nine items between 1990 and 2000, to 111 between 2001 and 2010, and to 4,300 between 2011 and 2020 (date of search: 23/09/2020). Citizen science has been successfully used to monitor nitrogenous air pollution (Tregidgo, West and Ashmore, 2013); to identify invasive plants (Gallo and Waitt, 2011); to understand long-term changes in plant (Miller-Rushing and Primack, 2008) and bird phenology (Ellwood, Primack and Talmadge, 2010) responses to climate change; to monitor biodiversity and natural resources (MacKechnie *et al.*, 2011; Danielsen *et al.*, 2014; Embling, Walters and Dolman, 2015; Chandler *et al.*, 2017); to monitor the population changes of bird and mammalian species across the UK (Harris *et al.*, 2016; Massimino, Harris and Gillings, 2018); to estimate climate niches and species distribution (Tiago, Pereira and Capinha, 2017); and for tracking disease-carrying mosquitoes (Palmer *et al.*, 2017; Eritja *et al.*, 2019).

Additionally, CS has educational and social objectives. Participants can increase their understanding about the subject that is being studied and improve their scientific literacy (Bonney *et al.*, 2009; Silvertown, 2009; MacKechnie *et al.*, 2011), and can participate in management, surveillance, control and prevention efforts in the place they live (Bonney *et al.*, 2009; Silvertown, 2009; Tregidgo, West and Ashmore, 2013; Danielsen *et al.*, 2014; Kelling *et al.*, 2015; Palmer *et al.*, 2017).

### **1.3.1. Challenges of citizen science datasets and possible solutions**

Although CS shows great potential for collection of datasets over space and time, issues of data quality can be a limitation of CS datasets, particularly for presence-only data derived from opportunistic data collection or observation efforts (Bird *et al.*, 2014; Mengersen *et al.*, 2017; Callcutt, Croft and Smith, 2018; Hamer, Curtis-Robles and Hamer, 2018). Citizen science data can contain high levels of variability because CS projects involve a large number of participants who have different skills and motivation, which can lead to

decreased precision in measurements such as measurement error or bias, and variation in reporting rates associated with observer expertise and effort (Dickinson, Zuckerberg and Bonter, 2010; Bird *et al.*, 2014; Embling, Walters and Dolman, 2015; Johnston *et al.*, 2017; Palmer *et al.*, 2017).

Common reasons for bias in CS datasets are related to the presence of spatially and temporally correlated factors, as previous exposure to information, areas that are more or less attractive, or media attention, which may increase project participation in certain areas and times (Boakes *et al.*, 2010; Crall *et al.*, 2011; Bird *et al.*, 2014; Johnston *et al.*, 2017). When CS datasets are characterised by spatial data (also called volunteer geographic information (Goodchild, 2007)), opportunistic efforts in collection or reporting will translate into spatial bias, which limits the representativeness of the spatial data (the degree to which a sample of spatial observations can represent the population) and decreases the quality and accuracy of the data (Goodchild and Li, 2012; Zhang and Zhu, 2018). Consequently, the statistical assumptions, including predictive maps of species distribution, may not be valid (MacKechnie *et al.*, 2011; Callcutt, Croft and Smith, 2018).

Historical datasets are frequently used as additional sources of CS data. However, these datasets often have incomplete information, and lack detail on how the data were collected, and who was/were the observer(s). It is therefore difficult to be sure if the patterns identified are due to real trends in the data or caused by changes in the observer or collection methods (Miller-Rushing, Primack and Bonney, 2012). Another issue with CS datasets is that methodologies are often simplified to make them more accessible to citizen scientists, which can result in a decreased sensitivity of the method used (Tregidgo, West and Ashmore, 2013).

However, the benefits of CS approaches have motivated researchers to find ways to improve the quality of data collected, database management and data analysis (Dickinson, Zuckerberg and Bonter, 2010; Wiggins *et al.*, 2011; Bird



*et al.*, 2014). For example, verification studies can be conducted to ensure that the data collected by participants is accurate and reliable (Crall *et al.*, 2011; Jordan *et al.*, 2012; Tregidgo, West and Ashmore, 2013; Danielsen *et al.*, 2014). Several studies have demonstrated that with appropriate protocols, training of citizen scientists and iterative project development, data validation steps (such as expert validation and/or the use of new technologies with automatic validation tools and verification platforms), accounting for volunteer expertise or effort, and the application of sophisticated statistical techniques to account for the high variability in CS datasets, volunteers can collect data of quality equal or similar to those collected by experts, and with larger geographical coverage (Gallo and Waitt, 2011; Jordan *et al.*, 2012; Tregidgo, West and Ashmore, 2013; Danielsen *et al.*, 2014; Kosmala *et al.*, 2016; Johnston *et al.*, 2017; Palmer *et al.*, 2017; Callcutt, Croft and Smith, 2018; Massimino, Harris and Gillings, 2018).

### **1.3.2. Value of citizen science to study disease vectors**

Citizen science has proved to be an efficient tool to collect data on, and to study, disease vectors such as triatomines, mosquitoes and ticks. Triatomine bugs occur in the United States (US) and Latin America, and are vectors of *Trypanosoma cruzi*, the agent of Chagas disease in humans and animals (Bern *et al.*, 2011; Wozniak *et al.*, 2015). Citizen science projects have been particularly important to study these insects, since there is no reliable standardised collection method for triatomines (Hamer, Curtis-Robles and Hamer, 2018). The datasets generated through CS allowed scientists to study the spatial and temporal distribution of these insects, and policy decision makers to improve target interventions of vector control. These CS projects increased community and political awareness regarding Chagas disease, and public health protection through the education of citizens on how to prevent encounters with the vector (Yoshioka, 2013; Curtis-Robles *et al.*, 2015, 2018). Curtis-Robles *et al.*, (2015) indicated that their CS project contributed to an increased strength of the relationship between researchers, public health agencies, clinical veterinarians, medical practitioners and the general public.

The recent outbreaks of Zika, chikungunya, dengue and other diseases demonstrated the need for a better understanding of mosquito behaviour and geographical distribution (Switters and Osimo, 2019). An initiative at global scale, the Global Mosquito Alert Consortium, was created to build CS capacity for surveillance and control of disease-vector mosquitos (Tyson *et al.*, 2018). Citizen science projects for mosquito surveillance at national or larger scales have been carried out in developed and developing countries (e.g. Maki and Cohnstaedt, 2015; Palmer *et al.*, 2017; Murindahabi *et al.*, 2018; Asingizwe *et al.*, 2019; Eritja *et al.*, 2019; Switters and Osimo, 2019). The CS project in Spain, 'Mosquito Alert', generates large datasets of the distribution of the Asian tiger mosquito (*Aedes albopictus*) and other species. Data analysis confirmed that this CS project provided accurate early warning information about mosquito invasion in Spain, and that CS can constitute a scalable and cheap tool for vector monitoring and surveillance of endemic and invasive species, helping governments and public health agencies in their decision-making (Palmer *et al.*, 2017; Bartumeus, Oltra and Palmer, 2018; Eritja *et al.*, 2019). Citizen science projects about mosquitoes have also demonstrated their strong social and educational objectives, as citizens learn about mosquito habitats, behaviour, diseases they may transmit, and about control and prevention measures (Palmer *et al.*, 2017; Asingizwe *et al.*, 2019). Recognising this benefit of CS projects, public health agencies are including programmes in their strategies for vector management that depend on citizen education and change of behaviours (vonHedemann *et al.*, 2016).

#### **1.3.2.1. Citizen science projects about tick vectors of diseases**

There are several examples of CS projects to study ticks and TBDs, particularly in Europe, US and Canada, where LB is a major public health concern, and where CS helps public engagement in the prevention of tick bites and to increase awareness of TBDs (Seifert *et al.*, 2016; Sakamoto, 2018). Examples for Europe are described in Table 1.2, and for the US and Canada in Appendix A: Supplementary material for Chapter 1.

Citizen science projects for ticks and TBDs vary in their objectives, and therefore in their complexity and number and type of questions asked to the volunteers (Table 1.2). Some of the projects comprise long-term national tick surveillance projects, for example the Tick Surveillance Scheme in the UK aims to improve the knowledge of the distribution of endemic and exotic tick species in the British Isles (Pietzsch *et al.*, 2005; Jameson and Medlock, 2011; Cull *et al.*, 2018, 2019; Hansford *et al.*, 2018). Other projects have been implemented during a certain period to answer specific scientific questions, such as the project in Finland which led to the development of the 'Tick Bank', the vast collection of tick species in Finland (Laaksonen *et al.*, 2017, 2018). Projects such as 'Tekenradar' in the Netherlands and 'TekenNet' in Belgium, besides educating and informing citizens about the risk of tick bites, also collect thousands of ticks and/or datapoints on human tick bites. This has enabled researchers to identify factors that determine tick bite risk (Mulder *et al.*, 2013; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017), to map human exposure to tick bites (Garcia-Martí *et al.*, 2018) and to study the prevalence of pathogens in ticks which have bitten people (Lernout *et al.*, 2019).

Citizen science projects about ticks have generated large datasets with spatial and temporal information of the vector distribution and the pathogens they transmit (e.g. Laaksonen *et al.*, 2017, 2018; Cull *et al.*, 2018; Nieto *et al.*, 2018). As an example, a CS project in Finland received about 20,000 ticks collected from humans and animals, which revealed the northward shift of *I. ricinus* in Finland, as well as enabling assessment of the prevalence of *Borrelia burgdorferi* (s.l.), *B. miyamotoi* and TBEV (Laaksonen *et al.*, 2017). The same outputs would have been impossible to obtain by the scientific community alone using the standard methods of collection, dragging and flagging.

Another value of CS projects is that this approach enables the collection of information on human-tick encounters, such as tick bites (Nieto *et al.*, 2018). In fact, collecting citizen reports of tick bites is seen as a complementary strategy for LB surveillance, particularly if considered periodically and locally

(van den Wijngaard *et al.*, 2017). The epidemiological importance of a tick bite has incentivised several CS projects to gather information on tick bites. Citizen science projects with ticks bites resulted in the identification of hotspot areas and risk factors for tick bites, spatial and seasonal patterns of human exposure of tick bites, and to gather data on the prevalence and dynamics of pathogenic agents (Mulder *et al.*, 2013; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017; Garcia-Martí *et al.*, 2018; Nieto *et al.*, 2018; Porter *et al.*, 2019; Salkeld *et al.*, 2019; Chauhan *et al.*, 2020; Jore *et al.*, 2020).

### **Current challenges and knowledge gaps**

As described, CS can be a valuable approach for collecting tick distribution data and information on human-tick encounters. However, although CS projects have led to some important scientific discoveries about ticks and TBDs, there remain some challenges and knowledge gaps.

For public health decision-making, it is necessary to assess the risk of a tick bite (Porter *et al.*, 2019), where risk is the product of the density of infected ticks in the environment multiplied by the human exposure (Ginsberg, 1993; Eisen and Eisen, 2016). However, as human exposure is difficult to assess, often studies do not estimate the risk but instead make inferences based on the density of questing infected ticks (or as a proxy, density of questing ticks) from questing tick surveys (such as Schwarz *et al.*, 2009; Swart *et al.*, 2014; Boehnke *et al.*, 2015; Brugger *et al.*, 2016; Li *et al.*, 2016; Vourc'h *et al.*, 2016). However, as discussed in Section 1.2.3, it is not clear how well questing ticks in the environment reflect the actual risk of tick bites.

Additionally, it would be relevant to get further information on human exposure to ticks and tick bites, such as the number of people in a certain area, the time spent doing certain outdoor activities or the distance covered. Most published CS projects on ticks and tick bites are based on the passive and opportunistic submission of reports (e.g. Laaksonen *et al.*, 2017; Garcia-Martí *et al.*, 2018; Nieto *et al.*, 2018), and rarely gather information on human exposure. Thus,

the generated CS datasets lack information on absences and contain a great level of variability and bias. When using these datasets, inference is only available for the subset of citizens that had a tick encounter or a tick bite, which will result in incomplete information to assess human exposure and to estimate risk. Therefore, including reports of absence data is essential for making robust inferences about human tick bite risk. However, CS approaches that also record when tick bites or tick encounters did not occur may be difficult to implement. Nonetheless, this approach has never been assessed. In addition, although volunteer effort has been taken into account in studies with mosquitos to decrease associated bias (Palmer *et al.*, 2017), volunteer effort has never been accounted in CS projects with ticks. These knowledge gaps will be addressed in thesis Chapters 4 and 5.

**Table 1.2:** Non-exhaustive list of citizen science projects implemented in Europe.

The projects are based mainly on *I. ricinus*, the dominant vector species in Europe. This table includes information about the project, the data collected, and a link for the study and/or published references.

Country	Project name and founders	Description of the project and data collected	Link for the study and/or references
<b>United Kingdom</b>	Tick Surveillance Scheme, Public Health England	Project set up in 2005. Citizens can send the ticks they collect from themselves, pets and other animals, which are then identified and tested by experts.	<a href="https://www.gov.uk/guidance/tick-recording-scheme">https://www.gov.uk/guidance/tick-recording-scheme</a> (Pietzsch <i>et al.</i> , 2005; Jameson and Medlock, 2011; Cull <i>et al.</i> , 2018, 2019; Hansford <i>et al.</i> , 2018)
<b>The Netherlands</b>	'Tekenradar', National Institute for Public Health and the Environment and Wageningen University	'Tekenradar' website platform was launched in 2012. This platform informs citizens about the risk and prevention of tick bites and at the same time, collects data on tick bites. Citizens record the date, location (geographical coordinates), age of the person bitten and photo of the tick and/or bite. Additionally, this platform records information on Lyme borreliosis occurrence from individuals.	<a href="https://www.tekenradar.nl/">https://www.tekenradar.nl/</a> (Mulder <i>et al.</i> , 2013; Antonise-Kamp <i>et al.</i> , 2017; Garcia-Martí, Zurita-Milla, Swart, <i>et al.</i> , 2017; Garcia-Martí, Zurita-Milla, van Vliet, <i>et al.</i> , 2017; Garcia-Martí <i>et al.</i> , 2018; Garcia-Martí, 2019)
<b>Finland</b>	Tick collection campaign, University of Turku	The tick collection campaign was implemented between April to November 2015. Citizens were asked to send ticks (biting or crawling) collected from humans and animals, and in addition information on the collection site, date, and the tick host.	(Laaksonen <i>et al.</i> , 2017, 2018)
<b>Belgium</b>	'TekenNet', Epidemiology of Infectious Diseases Services, Belgian Institute of Public Health	'TekenNet' platform (website and mobile application) was created in 2015. It provides information on ticks and Lyme borreliosis and allows people to report a tick bite and the development of disease, together with the geographical coordinates of the location where they were bitten.	<a href="https://tekennet.wiv-isp.be">https://tekennet.wiv-isp.be</a> (Lernout <i>et al.</i> , 2019)
<b>Switzerland</b>	Tick prevention, National Reference Centre for Tick-Borne Diseases and other partners	The mobile application collects tick bites reports from citizens from Switzerland and Liechtenstein, which then are used to develop a map of tick bites.	<a href="https://zecke-tique-tick.ch/en/tickbite-map-switzerland/">https://zecke-tique-tick.ch/en/tickbite-map-switzerland/</a>
<b>France</b>	Signalement-Tique, Institut National de la recherche agronomique and others	The website and mobile application, launched in 2017, were developed to gather data from tick bites of humans and pets, including geolocation, date, photos of the ticks.	<a href="https://www.citique.fr/signalement-tique/">https://www.citique.fr/signalement-tique/</a>

## 1.4. Thesis aim and objectives

The aim of this thesis is to investigate methodologies to improve predictive distribution mapping of ticks and tick bite risk in Scotland, to better inform public health decision-making. This aim will be achieved by applying modern statistical techniques to the current available data on tick distribution in Scotland, to understand the value and limitations of different tick distribution datasets, and by generating new data on the rate of tick bites and tick encounters to humans. Scotland is a country where *I. ricinus* is endemic, where LB is well-established, and along with the emergence of other pathogenic microorganisms, this gives rise to public and media concerns. Often policy makers require reliable maps of *I. ricinus* distribution to help with decision-making. Although some models and maps have been published for Scotland, there are inherent challenges in modelling and mapping tick distribution, which are associated with the lack of robust long-term and geographically extensive tick data.

Therefore, the objectives of this PhD are:

- 1) to test the quality and robustness for predictive mapping and public health decision-making of the three most extensive and up-to-date datasets of *I. ricinus* distribution in mainland Scotland, by assessing model results, biological plausibility, predictive and uncertainty maps and covariate coverage (Chapter 2);
- 2) to improve the predictive map of *I. ricinus* questing nymph distribution in mainland Scotland obtained in Chapter 2 using data collected during questing tick surveys, including the estimated uncertainty in the predictions, by applying a sophisticated Bayesian statistical methodology (the Stochastic Partial Differential Equation) (Chapter 3);

3) to assess how data collected during questing tick surveys compare with the real risk of tick bites to humans, by quantifying the relationship between relative abundance of questing ticks collected from the vegetation and tick bite rate of exposed orienteers (Chapter 4);

4) to investigate the feasibility of CS approaches that include collection of data on tick absence and volunteer effort, to assess how citizen science-based tick reporting data compares with scientific surveys of questing ticks in the environment, to identify the risk factors for human tick bite and encounter rates and finally, to compare the relative contribution of human behavioural risk factors and questing tick abundance in explaining the variation of tick bite and tick encounter rates (Chapter 5).



## Chapter 2: Using imperfect data in predictive mapping of vectors: a regional example of *Ixodes ricinus* distribution

The work described in this Chapter 2 has been published (please see a copy of the published manuscript in Appendix B: Supplementary materials for Chapter 2 (S1)): **Ribeiro R**, Eze JI, Gilbert L, Wint GRW, Gunn G, Macrae A, Medlock JM, Auty H (2019). Using imperfect data in predictive mapping of vectors: a regional example of *Ixodes ricinus* distribution. *Parasites & Vectors* 12(536), pp 1-13. doi.org/10.1186/s13071-019-3784-1

### 2.1. Introduction

*I. ricinus* is the most abundant and widespread tick species in Western Europe. As well as *B. burgdorferi* (*s.l.*) causing LB (Gray, 1998), it transmits other pathogens responsible for causing diseases of humans and animals. This species is now found at higher northern latitudes and higher altitudes than previously reported (such as in Norway and Sweden) (Tälleklint and Jaenson, 1998; Jore *et al.*, 2011; Hvidsten *et al.*, 2020) and is more abundant in some areas (as reported in Great Britain and in Germany) (Scharlemann *et al.*, 2008; Schwarz *et al.*, 2012). Understanding the drivers of the distribution and abundance of *I. ricinus* is one of the critical steps in assessing the risk of TBDs and informing policy on awareness, surveillance and control strategies (Estrada-Peña, 2001; Schulz, Mahling and Pfister, 2014; Kjær *et al.*, 2019a). Reliable maps of *I. ricinus* distribution are essential to understand and identify changes in the pattern of *I. ricinus* and diseases it transmits (Medlock *et al.*, 2013), and to identify hot-spots of vector occurrence. This is necessary to help inform policy makers in allocating resources to high risk areas, including targeting education and preventive measures (Lindgren and Jaenson, 2006) or management of important tick population hosts such as deer (Gilbert *et al.*, 2012), and to guide citizens on their decisions regarding preventive measures (Kjær *et al.*, 2019b).

Predictive maps are created based on spatial models which allow extrapolation of predictions for areas where no records are available (Braks *et al.*, 2016). *I. ricinus* distribution models can be developed using two different approaches, correlative environmental models (empirical-statistical models) or mechanistic or process-based models (Braks *et al.*, 2016). Correlative environmental models are based on the observed correlations between species distribution records (*I. ricinus* presence-only, presence-absence or quantitative data on abundance) and environmental predictor variables. In contrast, mechanistic models use detailed knowledge of *I. ricinus* ecology to simulate the mechanisms considered to underlie the observed correlations with environmental attributes (Beerling, Huntley and Bailey, 1995).

Two important features of distribution maps which influence the process of decision-making are the geographic scale and the resolution. The purpose of the study will determine the geographical scale of the map and the resolution will determine the degree of precision, realism and applicability of the models and maps (Kitron, 2000). Therefore, if the objective is to make decisions at country or regional levels, finer resolution maps can detect high variability in tick distribution patterns and provide more effective guidance in the management strategy to use (Braks *et al.*, 2016).

Many *I. ricinus* distribution models and maps have been produced and published in the scientific literature or websites, aiming to predict current and future distribution of *I. ricinus* on different geographic scales, ranging from European level (Estrada-Peña and Venzal, 2006; Estrada-Peña, Venzal and Acedo, 2006; Beugnet, Chalvet-Monfray and Loukos, 2009; Boeckmann and Joyner, 2014; Alkische, Peterson and Samy, 2017; ECDC, 2019a), to regions within Europe (Kjær *et al.*, 2019a, 2019b), to country (Jore *et al.*, 2011; Swart *et al.*, 2014; Brugger *et al.*, 2016; Li *et al.*, 2016; Alfredsson *et al.*, 2017; Domşa, Mihalca and Sándor, 2018), regional (Boehnke *et al.*, 2015; Hönig *et al.*, 2019) or local levels (Medlock *et al.*, 2008; Schwarz *et al.*, 2009; Qviller *et al.*, 2016; Vourc'h *et al.*, 2016; Signorini *et al.*, 2019).

However, predicting *I. ricinus* distribution and abundance is challenging due to the complex ecology of *I. ricinus* (with multiple tick stages and multiple hosts), the limited availability of detailed, long-term and geographically extensive tick distribution data, and a wide range of environmental variables that may influence tick distribution. Reliable data on *I. ricinus* presence and absence or abundance can be collected during questing tick surveys which use standardised sampling methods, such as the blanket dragging technique (James *et al.*, 2012; ECDC and EFSA, 2018). However, the resources required for field sampling (trained personnel, cost and time required) mean that data are often not available at meaningful spatial and seasonal scales. Other sources of data that were not collected with the specific purpose of predictive mapping are therefore often used instead. Data submitted by the public (citizens, medical and veterinary health agencies, wildlife groups and amateur entomologists) can be used to improve the knowledge of *I. ricinus* distribution (Jameson and Medlock, 2011) but usually comprise presence-only data so are subject to biases. An alternative approach, often undertaken by large-scale projects such as VectorNet (ECDC and EFSA, 2018), is to combine available data sources into one composite dataset.

Although LB is an important public health concern in Scotland (Mavin, Watson and Evans, 2015), published predictive maps of *I. ricinus* distribution in Scotland are limited, particularly at an appropriate scale for national and local decision-making. Although some (as yet unpublished) predictive maps have been made (Braga, 2012; Worton, 2016), the only peer reviewed publication is a mechanistic model predicting the current distribution of infected *I. ricinus* nymphs and under climate warming (Li *et al.*, 2016). Large-scale presence-absence maps at the European level at coarse resolutions (such as 20km<sup>2</sup>) (Alkishe, Peterson and Samy, 2017) or using polygon data of the administrative units (the nomenclature of units for territorial statistics, level 3 (ECDC, 2019a)) do not have sufficient resolution for targeting public health resources within Scotland, where *I. ricinus* is endemic.

The objective of this work was to compare the performance of three datasets to predict *I. ricinus* distribution in Scotland, in order to produce predictive maps for use by decision-makers. Model, map and uncertainty outputs of predicted tick abundance and distribution over Scotland were generated from three datasets: a) *I. ricinus* abundance data from questing tick surveys; b) *I. ricinus* presence-only data obtained from public submissions plus absence points; and c) a composite dataset that combines presence data from public submissions, presence and absence from questing tick surveys, literature reviews and expert opinion, and absence from a habitat suitability mask for *I. ricinus*. These datasets, which comprise the only data available on tick distributions at a national scale for Scotland, also represent three data types commonly used in mapping tick distributions (i.e. tick data from questing tick surveys; tick data from public submissions; and composite datasets). The outputs derived from these different inputs were assessed to highlight the strengths and limitations of each data type and the performance of these different types of data in predicting tick distribution in order to make recommendations for future tick mapping for use in a public health context.

The issue of using imperfect data is common across the board with vector data, as is the challenge of trying to compare predictive maps in the absence of gold standards. This work therefore provides a useful exemplar illustrating the challenges of identifying an appropriate *I. ricinus* predictive distribution map for Scotland and identifying lessons relevant to other studies.

## **2.2. Materials and Methods**

### **2.2.1. Tick data**

Three datasets with information on *I. ricinus* occurrence and abundance in Scotland were used. As is often the case with predictive mapping exercises, none of these datasets were collected with the main objective of predicting tick distribution at the national level, but they represent the most extensive datasets currently available for mainland Scotland. Details on the availability

of the three datasets can be seen in the published manuscript (Appendix B: Supplementary materials for Chapter 2 (S1)).

### **Dataset 1 - *I. ricinus* abundance data from questing tick surveys**

Dataset 1 (Figure 2.1a) is tick abundance data and consists of counts of questing *I. ricinus* ticks (nymphs and adults) in sampled environments in mainland Scotland between 2006 and 2017. Questing ticks were sampled using the standard technique of dragging a white blanket of 1 m<sup>2</sup> across the ground vegetation area of 10 x 10 m, with an average of approximately 15 drags per site (Gilbert, 2010; James *et al.*, 2012; Millins *et al.*, 2016). During this eleven-year period, 687 sites were visited, with varying frequency (one to four visits), and a total of 10,611 drags were performed.

### **Dataset 2 - *I. ricinus* public tick submissions**

Dataset 2 (Figure 2.1b) comprised 325 tick submissions by the public to PHE, between 1998 and 2016 in mainland Scotland. Data were recorded from the Biological Records Centre, Monks Wood and, since 2005, made through the Tick Surveillance Scheme. This scheme focuses predominantly on tick surveillance in England and Wales, so whilst ticks are also submitted from Scotland, the numbers of submissions are relatively small. The number of ticks (adults and nymphs) submitted per geographical location were transformed to presence-only data comprising 198 data points. Both adults and nymphs were included; 60% of submissions reported adult ticks only. To address the inherent limitations of modelling presence-only data, a similar number of data points on absence were added. Therefore, 200 absence and pseudo-absence points from Dataset 3 were randomly selected to include in Dataset 2. It is recognised that other methods could be applied in the selection of the pseudo-absence points as described by Barbet-Massin *et al.*, (2012). However, this work adopted a sample of absence points for Scotland that was validated by a group of entomologists and public health experts within VectorNet project (ECDC, 2019a) as contained in Dataset 3 (details below).

### **Dataset 3 - *I. ricinus* combined dataset**

Dataset 3 (Figure 2.1c) consists of presence and absence records of *I. ricinus* (adult and nymphs) in Scotland and is part of a large dataset of *I. ricinus* records for all Europe, produced for VectorNet project (ECDC and EFSA, 2018) by a team of tick experts (a network of entomologists and public health professionals) supported by the European Centre for Disease Control and Prevention. This dataset covers a period of 10 years (between 2006 and 2016). The full methodology was described by ECDC (2019a). In VectorNet project, tick records were assembled from different sources, from public submissions (including presence-only data from Dataset 2 for Scotland) and from questing tick surveys (including Dataset 1 for Scotland) and then validated. Due to the small amount of absence data in comparison with presence data, additional absence points were assigned using a mask of suitable habitats for *I. ricinus*. The habitat suitability mask was defined by the same tick experts in VectorNet project as primary, secondary and unsuitable habitat types (land classes where a species is unlikely to be found except in exceptional circumstances such as continuous and discontinuous urban fabric, industrial or commercial units, port areas, post-flooding or irrigated croplands (or aquatic) or closed to open (> 15%) broad-leaved forest regularly flooded) as depicted in two land cover maps: CORINE 2006 and GLOBCOVER 2009 (Olivier *et al.*, 2012; EEA, 2014) and by adding additional information about *I. ricinus* environmental limits (e.g. the fact that *I. ricinus* is only present in areas with less than 150 days of snow cover per year and where the vegetation period is greater than 145 days). Inferred absences were then extracted from unsuitable areas defined by the habitat (Braks *et al.*, 2016; ECDC, 2019a). The dataset used in this study, after data management, included 1,102 presence points and 1,058 absence points.

For consistency between datasets, only mainland Scotland was considered. The extraction of points in mainland Scotland and the random selection of 200 absence points for Dataset 2 were conducted using ArcGIS version 10.2.2 (ESRI, 2012).

## 2.2.2. Georeferenced environmental data and variable selection

Ecologically relevant climatic, topographic, land cover and host-related variables for *I. ricinus* occurrence and abundance were selected. These variables were collated as geographic information system (GIS)-based raster maps (Table 2.1).

**Table 2.1:** Georeferenced environmental variables used in the study.

Variables		Resolution
Moderate resolution imaging spectroradiometer (MODIS) variables (monthly averages, 2001–2013)	Normalised difference vegetation index (NDVI)	1 km <sup>2</sup>
	Land surface temperature (LST)	1 km <sup>2</sup>
	Cumulative land surface temperature (end of May 2010–2012, MODIS derived data)	0.01 * 0.01 degrees (~1.1 km <sup>2</sup> )
	Length of vegetation growth period (2008–2014)	0.01 * 0.01 degrees
Topographic variables	Elevation above sea level	90 m <sup>2</sup>
Long-term average climate data from UK Met Office (from 1981–2010)	Monthly maximum, mean and minimum temperature (monthly average)	5 km <sup>2</sup>
	Number of consecutive dry days (annual average)	5 km <sup>2</sup>
	Extreme temperature range (annual average)	5 km <sup>2</sup>
	Rainfall (monthly total precipitation)	5 km <sup>2</sup>
	Days of air and ground frost (monthly average)	5 km <sup>2</sup>
Host-related variables	Mean relative humidity (monthly average)	5 km <sup>2</sup>
	Index of presence of roe (Alexander <i>et al.</i> , 2014)	0.008333 * 0.008333 degrees (~1 km <sup>2</sup> )
	Index of presence of red deer (Wint <i>et al.</i> , 2014)	0.008333 * 0.008333 degrees
Land Cover 2006	Red deer density for 2016 (Albon <i>et al.</i> , 2017)	1 km <sup>2</sup>
	44 land cover types	1 km <sup>2</sup>

Monthly derived variables were extracted from each month. For data extraction compatibility and modelling purposes, all variables were converted to a standardised extent (mainland Scotland), format (.tif), spatial resolution (1 km<sup>2</sup>, using interpolation and the ‘resample’ function) and projection (British National Grid). Environmental data were extracted for each of the sites of tick collection and reporting [687 sites with counts of *I. ricinus* (Dataset 1); 398 presence-absence points (Dataset 2); and 2,160 presence-absence points (Dataset 3)] using the tool ‘extract multiple values to points’ in ArcGIS version 10.2.2 (ESRI, 2012).

Before model implementation, a correlation analysis and univariable regression analysis were performed for each response variable. If two variables were strongly correlated (correlation coefficient higher than 0.6), one of them was dropped. Variables with correlation coefficients between 0.5 and 0.6 were kept for analysis but under observation for possible interactions. Following the univariable analysis, biologically relevant variables with a P-value less than 0.10 were considered as model candidates. In general, due to issues of autocorrelation and collinearity, satellite-derived variables were preferred when compared with similar interpolated climatic variables (such as LST from MODIS compared with long-term average temperature from UK Met Office, Table 2.1) (Estrada-Peña, Estrada-Sánchez and Estrada-Sánchez, 2015).

### **2.2.3. Statistical model, model validation and predictive map**

Models were fitted using the integrated nested Laplace approximation (INLA) R package. This Bayesian approach was selected due to its ability to account for irregular sampling intensity, spatial dependency and to quantify uncertainty in data and variables, attributing to each variable a distribution of values (Rue, Martino and Chopin, 2009). It is recognised that other methods could be used, but the objective was not to compare different modelling techniques for species distribution models, but instead compare dataset types using the same modelling technique.

The response variables were the count of *I. ricinus* ticks (nymphs and adults) per drag, visit and site of collection, and *I. ricinus* (nymphs and adults) presence and absence. A model for predicting tick relative abundance was first created considering just the counts of nymphs per drag, visit and site because nymphs of *I. ricinus* pose the greatest risk of tick bites of humans (Robertson, Gray and Stewart, 2000). However, for consistency with Datasets 2 and 3 which include reports of adult ticks, it was decided to model *I. ricinus* relative abundance considering the total count of adult ticks and nymphs per drag, visit and site. This model did not differ significantly from the model using nymphs



only. The fixed effects were the previously selected set of most suitable environmental covariates, including the spatial location of the data (as an interaction term between latitude and longitude). A zero-inflated Poisson (ZIP) distribution was chosen to model *I. ricinus* abundance (Model 1) due to the skewed distribution of the number of ticks counted per drag, with a majority of drags counting zero ticks. Presence and absence of *I. ricinus* (Model 2 and Model 3) were modelled using a binomial distribution.

The selected model for predicting *I. ricinus* relative abundance (Model 1) had two random effects: a) the effect of the site to capture the unstructured heterogeneity in the distribution of tick abundance among sites (i.e. to account for the unstructured spatial dependency), and b) the effect of each data point (each drag) in order to account for overdispersion not captured by the ZIP distribution, and also to account for possible serial correlation in the data arising due to repeated sampling or drags in each site. Tick presence and absence (Model 2 and Model 3) was modelled without random effects, because the inclusion of random effects did not improve model fit and predictive power.

The models are described by the equations below:

1: Zero-inflated Poisson model

$$\text{Log}(M_{ij}) \sim \alpha + \beta_1 \text{LST} + \beta_2 \text{Roedeer} + \beta_3 \text{Deciduous} + \dots \beta_k \text{Lat: Long} \\ + f(\text{Drag}_{ij}) + f(\text{site}_j)$$

Where  $M_{ij}$  is the mean tick abundance for the  $i$ th drag in site  $j$ ;  $\alpha$  is the intercept;  $\beta_k$  are the measure of covariate effects; and  $f()$  denote the random effects due to drags and sites, respectively.

2: Presence-absence model

$$\text{Log}(O_j) \sim \alpha + \beta_1 \text{NDVI} + \beta_2 \text{Rain} + \beta_3 \text{Frost} + \dots \beta_k \text{Moorland}$$

Where  $O_j$  is the odds of tick presence in a given point;  $\alpha$  is the intercept, and  $\beta_k$  are the measure of covariate effects.

The models were evaluated using the deviance information criteria (DIC) as a measure for goodness of fit and a parameter from the cross-validation leave-one-out, namely the negative of the sum of the log-conditional predictive ordinate score (CPO) as a measure for the predictive quality of the model (Gelman, Hwang and Vehtari, 2013). A backward stepwise procedure was used to select the most parsimonious model. For all three datasets, the most suitable models were selected based on the lowest values of DIC and CPO, amongst competing models with various covariate combinations. The model posterior means were used to produce the predictive maps of *I. ricinus* abundance (Model 1/Dataset 1) and presence-absence (Model 2/Dataset 2 and Model 3/Dataset 3). The predicted uncertainty was calculated using two approaches. The first approach used was to rescale the range between the 2.5% and the 97.5% quantiles of predicted values to a 0-1 scale (i.e. rescaled uncertainty). The second approach was to transform the standard deviation of the predictive values to logarithm base 10. The advantage of the rescaled approach (i.e. 0-1 scale) is that it brings the uncertainties from the three models in the same range of values. However it does not reflect the absolute values of the uncertainties. The resolution of all maps was 1 km<sup>2</sup> per pixel.

A matrix of boxplots, comparing the interquartile range of the model covariates over mainland Scotland with the interquartile range of the same covariates covered by the data points in each model, was developed to visualise the degree to which the three datasets cover the range of the covariates used in the models. Descriptive analyses, plots, models and maps were made using R software version 3.4.4 (R Core Team, 2013).

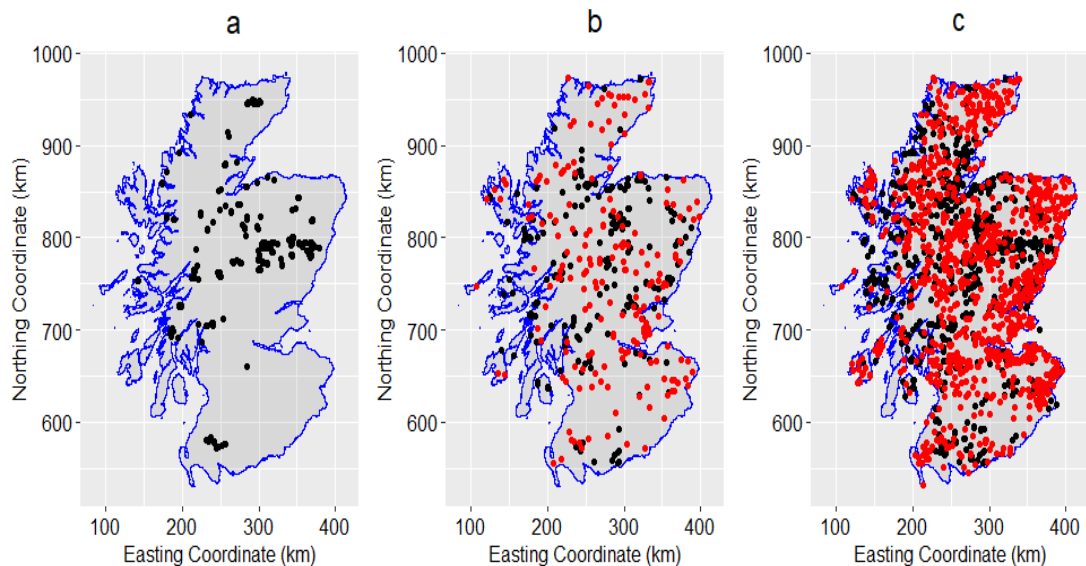
#### **2.2.4. Kernel density analysis, model and predictive and uncertainty maps for a sub-sample of Dataset 1**

After the analysis of the predictive map from Model 1 (Dataset 1), in order to assess the value of *I. ricinus* abundance data (from questing tick surveys) for predictive mapping at local scales, a second model with a subset from Dataset 1 was developed. Therefore, the council area in mainland Scotland with the best coverage of questing tick surveys was selected to develop a predictive model and map. To select the council area with the best coverage of questing tick surveys, a kernel density analysis was used to calculate the density of sites and drags per site around each neighbourhood to build a surface that accurately reflected the areas with higher density of tick sampling.

Dataset 1 and all the covariates were extracted for the selected council area using a shapefile of the area. ArcGIS version 10.2.2 (ESRI, 2012) was used for the kernel density analysis and data extraction. Due to the presence of zero inflation in the counts of questing ticks collected in the selected area, the ZIP model with site and drag random effects used to model Dataset 1 was adopted. Variable selection, model development, validation and prediction followed the same procedure as presented in Section 2.2.3.

### **2.3. Results**

Dataset 1 has an uneven distribution of tick collection sites over mainland Scotland, with concentration of collection sites in the east, particularly Aberdeenshire, and in contrast a lack of sampling points on the west coast (Figure 2.1a). The distribution of data points in Dataset 2 (Figure 2.1b) is sparse compared to Dataset 3 (Figure 2.1c).



**Figure 2.1:** Spatial distribution of the three datasets used in the models. (a) Distribution of sites of questing tick surveys in mainland Scotland (Dataset 1); (b) Distribution of sites of presence-only reports (black dots) and absences of *I. ricinus* (red dots) (Dataset 2); (c) Distribution of combined presence of *I. ricinus* from field surveys and public submissions (black) and absences (red dots) (Dataset 3).

**Model 1 (tick relative abundance, using Dataset 1: questing tick surveys)**

A spatial model of the count of ticks (adult and nymphs) per drag, visit and site was run initially (DIC of the most suitable spatial model was 29,786.66, CPO was 20,427.23). Subsequently, month was added in the model as a categorical variable, improving model predictive power (DIC 29,774.49; CPO 19,686.78). The results of Model 1 are presented in Table 2.2, the plots of model fitted and observed values can be seen in Appendix B: Supplementary materials for Chapter 2 (S2), and the predictive map for the month with the highest effect on questing tick abundance (April) is presented in Figure 2.2a.

**Table 2.2:** Results for Model 1 (Dataset 1).

Posterior mean, SD, 2.5% and 97.5% quantiles and estimates (logarithm) of fixed effects for the model of tick abundance.

Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Intercept	-150.7016	32.7201	-214.9447	-86.5187
April	2.3606	0.3390	1.7198	3.0520
May	1.9424	0.3174	1.3467	2.5944
June	1.8192	0.3178	1.2227	2.4718
July	1.2388	0.3149	0.6485	1.8863
August	1.3438	0.3151	0.7530	1.9916
September	1.4308	0.3186	0.8325	2.0850
LST in July	0.0103	0.0022	0.0059	0.0147
Number of days of frost in September	-0.4035	0.0954	-0.5910	-0.2167
Roe deer presence	0.0096	0.0034	0.0030	0.0163
% cover of deciduous woodland	2.5341	0.7380	1.0837	3.9806
% cover of coniferous woodland	0.9053	0.2138	0.4848	1.3240
Interaction between latitude and longitude	0.0010	0.0018	-0.0026	0.0045

Abbreviations: SD, standard deviation; LST, land surface temperature. In the final model the month 'April' was considered as a dummy variable.

Land surface temperature in July (posterior mean = 0.0103), presence of roe deer (posterior mean = 0.0096) and deciduous (posterior mean = 2.5341) and coniferous forest (posterior mean = 0.9053) were associated with an increase in *I. ricinus* questing tick abundance, whilst a higher number of frost days in September (posterior mean = -0.4035) lead to a decrease in tick abundance. Questing tick abundance was highest in April (Table 2.2). The interaction term between latitude and longitude was included in the model because, although not significant, it decreased model residual variance and can help account for spatial effects in questing tick abundance, such as spatial autocorrelation. The predictive map of tick abundance (Figure 2.2a) shows that tick abundance increases from the north and west to the south and east of Scotland, with highest predicted tick abundance in Aberdeenshire and the central belt. Areas of average to high uncertainty are present over all the east, centre and south

of Scotland, whereas all the west shows less uncertainty in predicted values (Figure 2.2d-g).

### **Model 2 (tick presence-absence, using Dataset 2: tick information from public submissions)**

Model 2 (DIC of 388.61, CPO of 195.81) fits the data well (Appendix B: Supplementary materials for Chapter 2 (S2)). Presence of *I. ricinus* was correlated with an increase in NDVI (posterior mean = 0.1373) and some measures of habitat composition. An increased number of days with air frost in November (posterior mean = -0.1729) and increased precipitation in April (posterior mean = -0.0148) were associated with tick absence. The site location of tick submission was important (Table 2.3). The predicted map for Model 2 (Figure 2.2b) does not capture areas of lower probability of tick presence well and shows high levels of uncertainty for most of Scotland (Figure 2.2e-h).

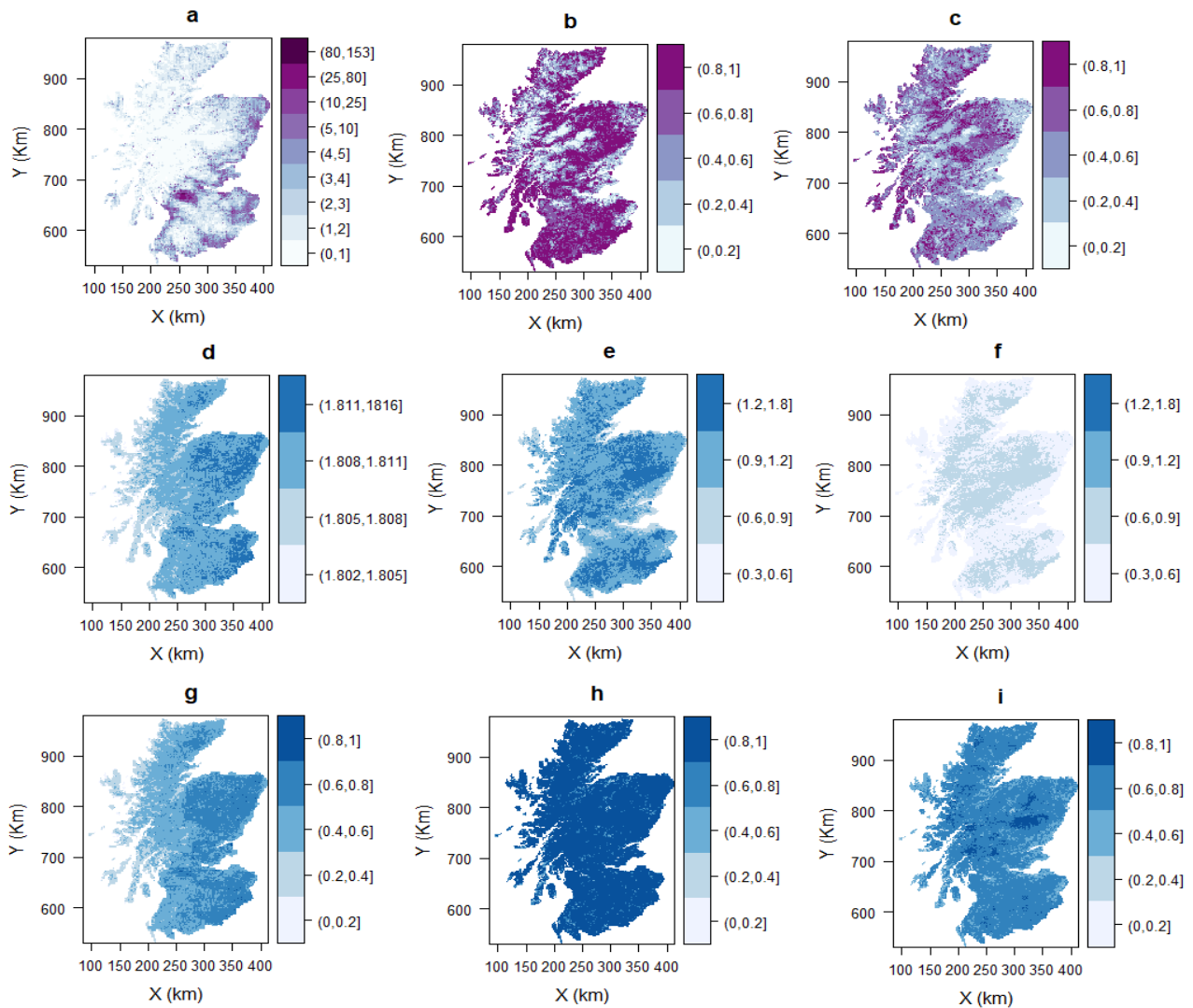
### **Model 3 (tick presence-absence, using Dataset 3: combined dataset)**

The adopted model (Model 3) gave the lowest values of DIC of 2,614.61 and a CPO of 1,307.74 (plot of model goodness of fit in Appendix B: Supplementary materials for Chapter 2 (S2)). Model 3 presented very similar covariates as Model 2, but deciduous forest and deer density became significant predictors, likely due to the increased number of points used to model tick presence-absence (Table 2.3). Figure 2.2c shows a similar pattern of *I. ricinus* probability of presence as Figure 2.2b, but the predictive map using Model 3 (Figure 2.2c) has more detailed definition. Uncertainty is lower for the north, east, and centre of Scotland (Figure 2.2f-i).

**Table 2.3:** Results for Model 2 (Dataset 2) and Model 3 (Dataset 3). Posterior mean, SD, 2.5% and 97.5% quantiles for the binomial models of tick presence-absence with the data from public submissions (Dataset 2) and the combined dataset (Dataset 3). The estimates of the fixed effects are presented in the logarithm form.

<b>Model</b>	<b>Fixed effects</b>	<b>Posterior mean</b>	<b>SD</b>	<b>2.5% quantile</b>	<b>97.5% quantile</b>
Model 2: Presence-absence model with presence points from public submissions plus absence points	Intercept	-6.2657	1.0232	-8.3326	-4.3135
	NDVI August <sup>a</sup>	0.1373	0.0176	0.1040	0.1732
	Number of days of air frost November	-0.1729	0.0521	-0.2784	-0.0738
	Rain April	-0.0148	0.0053	-0.0255	-0.0045
	% cover of coniferous woodland	5.1989	1.2015	3.0921	7.8095
	% cover of moorland	2.2180	0.5656	1.1499	3.3725
	Interaction between latitude and longitude	0.0053	0.0036	-0.0017	0.0123
Model 3: Presence-absence model with combined dataset	Intercept	-3.4700	0.4771	-4.4160	-2.5424
	NDVI August	0.0005	0.0001	0.0004	0.0006
	Red deer density	0.0336	0.0100	0.0139	0.0533
	Number of days of air frost November	-0.0527	0.0207	-0.0936	-0.0122
	Rain April	-0.0123	0.0020	-0.0163	-0.0085
	% cover of moorland	1.3920	0.1640	1.0726	1.7161
	% cover of deciduous woodland	3.1762	0.6757	1.9203	4.5770
	% cover of coniferous woodland	2.1861	0.2128	1.7753	2.6100
Interaction between latitude and longitude	-0.0029	0.0013	-0.0054	-0.0004	

<sup>a</sup>The posterior mean of NDVI was divided by 100. Abbreviations: SD, standard deviation; NDVI, normalised difference vegetation index



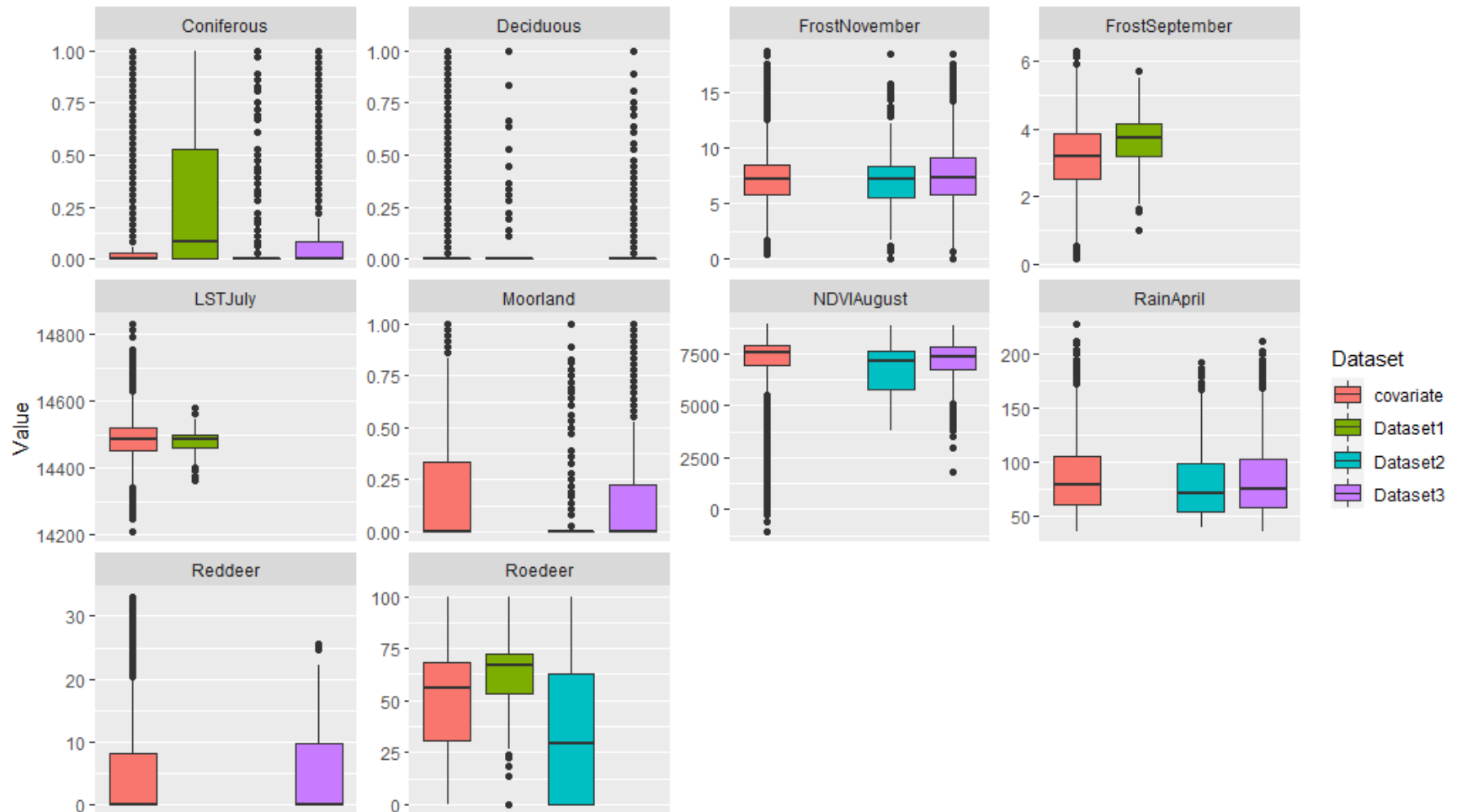
**Figure 2.2:** Predictive maps of *I. ricinus* distribution in mainland Scotland with respective uncertainty.

a) Predictive map of *I. ricinus* abundance in April (Dataset 1) and uncertainty maps (d) log<sub>10</sub> of the SD and (g) rescaled uncertainty; (b) predictive map of probability of presence of *I. ricinus* using presence-only data from public submissions and absence points (Dataset 2) and (e) uncertainty map as log<sub>10</sub> SD and (h) rescaled uncertainty; (c) predictive map of probability of presence of *I. ricinus* using the combined dataset (Dataset 3), (f) uncertainty as log<sub>10</sub> SD and (i) rescaled uncertainty. Darker areas of blue have higher uncertainty. X is the Easting coordinate and Y is the Northing coordinate.



In order to explore the validity of the predictions, the degree to which the three datasets cover the range of the covariates used in the models was assessed. The interquartile range of each covariate in mainland Scotland was compared to the interquartile range of each covariate in the models for the data points (sites/drags) included (Figure 2.3). The predictions of the three models were associated with uncertainty that was not captured in the uncertainty measures in Figure 2.2, because the tick data did not cover all the range of the covariates used. Dataset 3 covered the covariate range used in the predictions better than Dataset 1 or 2 (Figure 2.3). Dataset 1 was collected in predominantly forest areas.

Several models were considered to try to improve the fit of the model for Dataset 2, with alternatives shown in Appendix B: Supplementary materials for Chapter 2 (S3). The index of presence of roe deer was found to be important for *I. ricinus* presence in Dataset 2, and it was included in the first selected Model 2 (Appendix B: Supplementary materials for Chapter 2 (S3, Table B-S3.1)). However as can be seen in Figure 2.3, the range of the index of roe deer presence is not well covered by Dataset 2, contributing to higher uncertainty in the predictions (Appendix B: Supplementary materials for Chapter 2 (S3, Figure B-S3.1a-b)). The covariate roe deer was therefore removed from the final model (Table 2.2). Using all the covariates of Model 3 for fitting a model with Dataset 2 helped to corroborate how the covariates (type and range) are important in predictive mapping and can be a source of error for model predictions (also show in Appendix B: Supplementary materials for Chapter 2 (S3, Table B-S3.1, Figure B-S3.1c-d)). Neither of these models were selected as the final model for the predictions with Dataset 2. The final model is shown in Table 2.3.



**Figure 2.3:** Interquartile range of the covariates used in the three models. Facet-wrap of boxplots showing the interquartile range (y-axis) of the covariates over mainland Scotland and compared with the range of the same covariates covered by the data points in each model (Dataset 1, Dataset 2 and Dataset 3).

### **Model and predictive map for the subset of Dataset 1**

A map of the council areas in mainland Scotland with the sites for questing tick surveys, and the output heat map from the kernel density analysis can be seen in Appendix B: Supplementary materials for Chapter 2 (S4). Aberdeenshire was the council area with the highest density of questing tick surveys and was therefore selected for analysis. In total, 163 sites were surveyed (23.8% out of 686 sites over mainland Scotland) and 4,561 drags were performed over the years (43% out of 10,611 drags carried out over mainland Scotland). For Aberdeenshire, a spatial model of the count of ticks (adult and nymphs) per drag, visit and site was run initially (DIC of the most suitable spatial model was 13,099.6, CPO was 7,744.5). A model without the interaction term between latitude and longitude was selected because the model predictive power increased when removing these covariates and their interaction term. Subsequently, month was added in the model as a categorical variable, improving model fit and predictive power (DIC 12,858.2; CPO 7,411.1). The results of this model are presented in Table 2.4 (plot of model goodness of fit in Appendix B: Supplementary materials for Chapter 2 (S5)). Maps for the month with the highest questing tick abundance (April) and corresponding predictive uncertainty were created (Figure 2.4), and the matrix of boxplots showing the interquartile range of the covariates over Aberdeenshire and over the data points is shown in Figure 2.5.

Land surface temperature in July (posterior mean = 5.868), presence of roe deer (posterior mean = 0.010), and NDVI in April (posterior mean = 0.031) were associated with an increase of tick abundance in Aberdeenshire. Rain in July (posterior mean = -0.035) was associated with a decrease in tick abundance. April had the highest effect on questing tick abundance (Table 2.4). A model using the same covariates as Model 1 was also applied to this subset of data from Dataset 1, but the output demonstrated that this was not the most suitable model to explain tick abundance in Aberdeenshire. However, there are some similarities between Model 1 and the model for Aberdeenshire, namely the fact that April was the month with highest effect on tick abundance,

and land surface temperature and presence of roe deer were also identified as positive predictors of tick abundance. Although the covariates deciduous and coniferous woodland were not selected in this model, NDVI, which is an index of green vegetation and used as a proxy of dense forest areas, was identified as a good predictor.

**Table 2.4:** Model results for Aberdeenshire area.

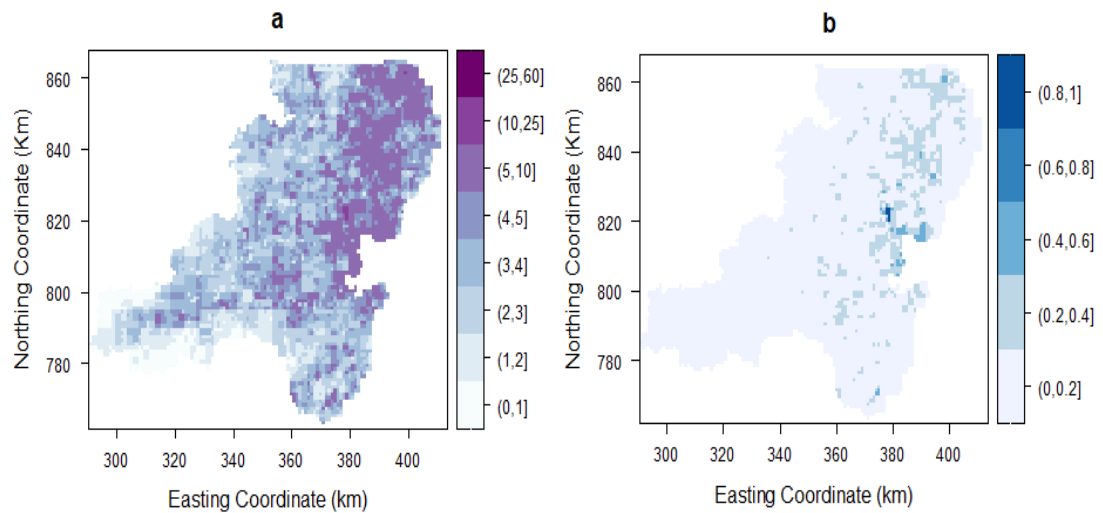
Posterior mean, SD, 2.5% and 97.5% quantiles for the model of relative tick abundance in Aberdeenshire (subset of Dataset 1). The estimates of the fixed effects are presented in the logarithm form.

Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Intercept	-85.856	78.324	-239.630	67.801
April	1.848	0.755	0.368	3.331
May	1.578	1.555	-1.475	4.629
June	0.758	0.734	-0.681	2.199
July	0.237	0.739	-1.211	1.688
August	0.330	0.736	-1.113	1.776
September	-0.043	0.738	-1.489	1.407
LST in July <sup>a</sup>	5.868	5.407	-4.748	16.475
Rain in July	-0.035	0.014	-0.062	-0.008
Roe deer presence	0.010	0.005	-0.001	0.021
NDVI in April <sup>b</sup>	0.031	0.020	-0.008	0.071

<sup>a</sup>In this model the posterior mean of LST in July was divided by 1000. <sup>b</sup>In this model the posterior mean of NDVI was divided by 100. Abbreviations: SD, standard deviation; LST, land surface temperature; NDVI, normalised difference vegetation index

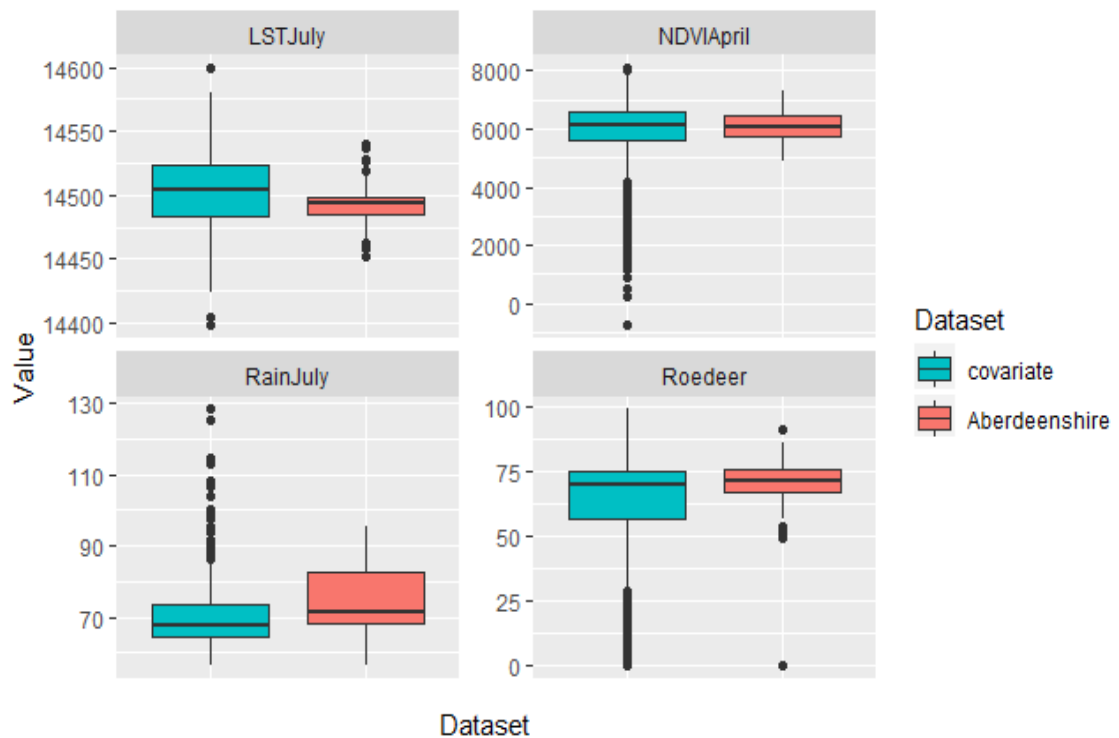
The predictive map of tick abundance in Aberdeenshire (Figure 2.4a) is detailed enough to identify areas with low and high predicted tick abundance, and shows a similar distribution trend when compared with the predictive map on Figure 2.2a for the same area. Although Aberdeenshire was the area of mainland Scotland with high sampling intensity, there was an uneven distribution of sampling sites within Aberdeenshire with more sites concentrated in the south-west part and no sampling performed on the north-east coast. Even so, both Model 1 and the present model identify the north-east coast as an area with probability of high questing tick abundance. The

predicted uncertainty map for Aberdeenshire (Figure 2.4b) identifies the area where the predicted tick abundance is high but no sampling was performed as an area with higher uncertainty in the predictions. According to Figure 2.5, the clustering of sampling sites in Aberdeenshire resulted in the covariate range in this area still not being well represented by the points of sampling.



**Figure 2.4:** Predictive map of *I. ricinus* distribution in Aberdeenshire with respective uncertainty.

(a) Predictive map of *I. ricinus* questing tick abundance in April in Aberdeenshire and (b) uncertainty map. The uncertainty maps were calculated from the range of 95% confidence intervals of predicted values and rescaled to a 0-1 scale. Darker areas of blue have higher uncertainty.



**Figure 2.5:** Interquartile range of the covariates used in the model for Aberdeenshire. Facet-wrap of boxplots showing the interquartile range (y-axis) of the covariates over Aberdeenshire and compared with the range of the same covariates covered by the data points in the model.

## 2.4. Discussion

Predictive maps of tick distribution are essential for understanding human disease risk and allocating resources for prevention and control. However, they require extensive data on tick distribution, and robust long term and geographically extensive datasets are often difficult to obtain. Therefore, datasets are often used that were not collected for this specific purpose, or do not cover the entire extent of the required predicted distribution. The main aim of this study was to compare the performance of three different types of data to predict *I. ricinus* distribution in Scotland. The predicted tick distributions from the three datasets were critically assessed to deduce how useful each dataset may be to inform policy, future data collection and risk mapping, both for Scotland and more widely.

#### **2.4.1. Predictors for *I. ricinus* relative abundance and presence-absence**

The predictors identified in the models are consistent with the current knowledge on *I. ricinus* ecology. *I. ricinus* activity is seasonal; in Scotland activity peaks in April and then generally declines over the season as ticks die or find a host, as observed in Model 1 and model for Aberdeenshire. It is not surprising that roe deer presence and red deer density are correlated with both tick abundance and presence, as deer are important hosts maintaining *I. ricinus* populations (Gilbert *et al.*, 2012). Many other studies have found deer abundance as a predictor of tick presence and/or abundance (Ruiz-Fons and Gilbert, 2010; Gilbert *et al.*, 2012; James *et al.*, 2012). Temperature affects tick behaviour, interstadial development rate, fertility, survival and the proportion of active ticks (Randolph *et al.*, 2002; Gilbert, Aungier and Tomkins, 2014; Tomkins *et al.*, 2014). Warmer climates are frequently associated with tick presence and abundance (e.g. Gilbert, 2010). Land surface temperature was also used in other studies to estimate *I. ricinus* presence-absence and abundance (Swart *et al.*, 2014; Estrada-Peña, Estrada-Sánchez and Estrada-Sánchez, 2015; Hönig *et al.*, 2019; Kjær *et al.*, 2019b). In accordance, this study found that areas with warmer climate and lower number of frost days (minimum temperature below zero degree Celsius) were associated with higher tick abundance, consistent with previous studies (Braga, 2012; Paul *et al.*, 2016).

Habitats that provide a sheltered canopy over the ground, characterised by thick ground vegetation or shrub and deep leaf litter layers, such as mixed and deciduous forest, and habitats with bracken and deep heather in moorland, provide moist and mild microclimates which aid tick survival and activity. Forests are also a source of food and shelter for many species of tick hosts (Gray *et al.*, 1998). NDVI, which quantifies the level of photosynthetic activity of the vegetation, has been previously identified as being an important parameter that correlates with *I. ricinus* presence and abundance (Estrada-Peña, Estrada-Sánchez and Estrada-Sánchez, 2015; Hönig *et al.*, 2019; Kjær *et al.*, 2019b). More questing *I. ricinus* are usually predicted in areas with

warmer climate and higher rainfall or higher humidity (Ruiz-Fons *et al.*, 2012). However, the negative effect of annual precipitation found in this study was also found by James *et al.* (2012) in Scotland and Schulz *et al.* (2014) in Germany. These findings suggest that the wet conditions in Scotland are probably too wet for *I. ricinus* to quest, while drier areas of Scotland are still wet enough for good survival rates. Therefore, future models of tick distribution could consider non-linear relationships between rainfall and tick distribution data.

For all three datasets, plausible predictors were identified, and predictive maps were created. Although the models fitted the data well, the three predictive maps do not present a consistent pattern of *I. ricinus* distribution, and the predictions are associated with a large amount of uncertainty particularly for Model 2 (Dataset 2 – public submissions) and to a lesser degree for Model 1 (Dataset 1 – questing tick surveys). The predicted uncertainty was lower when using a subset of Dataset 1 to predict *I. ricinus* distribution in the area that had a better coverage of questing tick surveys (model for Aberdeenshire). These conclusions were made considering the outputs of the rescaled uncertainty (the difference between the 97.5% and the 2.5% quantiles rescaled for 0-1 scale). In this study, the approach of rescaling the uncertainty was used (also adopted by Alkische, Peterson and Samy, 2017) because the response data type in the three models are different (abundance ranging from zero to 109 versus presence-absence ranging from zero to one). To analyse the uncertainty of the predicted values from the three models it is necessary to generate comparable values. However, whichever approach is used (i.e. the difference between 2.5% and 97.5% of predicted values or the SD of the predictions) the predicted uncertainty is correlated with the model posterior mean for each covariate and does not include uncertainty associated with extrapolation outside the covariate range. Therefore, this study explored factors that could influence the validity of the predictions for each dataset and identified when each dataset might be most appropriately used.



#### **2.4.2. Use of *I. ricinus* abundance data from questing tick surveys (Dataset 1)**

The relative abundance of questing *I. ricinus* ticks is generally measured by dragging a blanket over the vegetation during questing tick surveys. This technique does not measure the absolute density of the whole tick population in an area, because it does not count ticks that are moulting, resting, feeding or in diapause (Swart *et al.*, 2014). It is also worth noting that the blanket drag method efficiency is affected by ground vegetation height and density (Ruiz-Fons and Gilbert, 2010). Ground vegetation height and density should be included in statistical models, and repeated sampling of a site is recommended due to the impact of weather conditions on tick activity on the day of sampling. Another limitation from this technique is that tick absence data cannot be considered completely free of error because some of the zeros could in reality indicate very low tick densities rather than true absolute absence, due to the finite number of blanket drag transects per site and conditions on the day of surveying (Swart *et al.*, 2014). However, this is a standard scientific technique which provides a useful comparable index of abundance of questing ticks between sites (ECDC and EFSA, 2018).

Conducting blanket drags is time-consuming which makes it resource-intensive to perform large-scale and/or long-term field studies using this technique. However, estimating questing *I. ricinus* relative abundance gives more information about this species distribution when compared with presence-only and presence-absence data. Abundance data are necessary to calculate the density of infected ticks, which is important in estimating disease risk (Swart *et al.*, 2014), as well as providing more information on *I. ricinus* dynamics. Abundance data also improve model predictive performance and ability to discriminate trends at finer scales, compared to presence-absence data. This improvement is particularly important for species of high abundance compared to 'rare' species (Howard *et al.*, 2014). When the objective is to create a predictive map for a country such as Scotland where *I. ricinus* is endemic, abundance models will provide more meaningful distribution maps.

The predictive map of questing *I. ricinus* relative abundance (Figure 2.2a) clearly identifies areas with high and low tick abundance and shows an increasing trend of tick abundance from the west to the east coast of Scotland. However, although the predictions of relative abundance of questing ticks had lower uncertainty compared to both of the predictive maps of *I. ricinus* probability of presence, it is clear that the sample sites are clustered and do not cover all of mainland Scotland, and that the covariate range covered by mainland Scotland is not fully represented in the data. Since there is an ecological gradient from the west (higher temperatures, higher rainfall) to the east, it is concerning that there may be insufficient data for accurate predictions in the west. Dataset 1 was collected mainly (although not exclusively) in forest areas, reflecting the data collection which was aimed at specific ecological studies (James *et al.*, 2012; Millins *et al.*, 2016). Hence the dataset does not have good coverage for some areas of mainland Scotland, and therefore the reliability of the predictions is likely to be lower outside the core survey areas. These issues of low coverage are common to these types of data, due to the resources needed to conduct questing tick surveys over a wide scale.

The predictive model created using the subset of Dataset 1 with the highest questing tick survey coverage resulted in a detailed predictive map with less uncertainty (when uncertainty of the predictions of both models was comparable), corroborating that tick abundance data from questing tick surveys can be applied at regional and local level decision-making (Figure 2.4). However, because the selection of sites for questing tick surveys was not stratified for important covariates (such as habitat type or elevation) and still showed clustering, there was also some uncertainty in the predictions. Some parts of the north-east coast where no sampling was performed were identified as having high tick abundance. Although these areas could have suitable conditions for ticks according to model results, such as temperature and humidity and presence of hosts, they may in fact represent habitats not suitable for ticks (such as urban areas or pastures). Two approaches could have been performed to overcome this issue: a) to use a mask of unsuitable habitats for

ticks (ECDC, 2019a) or b) to only include in the model the area with the sampling points, and better cover the range of the covariates.

Quantitative survey data are often considered the 'gold standard', but this is only true for models that make predictions in the same geographical area and covariate range from which the surveys were conducted.

#### **2.4.3. Use of *I. ricinus* presence data obtained through public submissions plus absence points (Dataset 2)**

Presence-only data, generated from submissions by the public and often obtained from CS studies, are frequently used to map species distribution. These data usually require fewer resources to collect than questing tick surveys but include random error associated with uncertainty in the location of ticks were collected, variability in sampling (e.g. variability between different people reporting) and in effort (e.g. some people contribute more data than others and effort can also change over time) (Bird *et al.*, 2014; Palmer *et al.*, 2017). Bias is also associated with the fact that people report from places that are visited frequently or are more accessible (Boakes *et al.*, 2010). This type of data lacks information on where the species is absent, which limits the predictive power of the inference and also restricts the type of questions that can be asked (Pearce and Boyce, 2006). In this study, information of where the tick was absent (true absences from Dataset 1 and pseudo-absence points from habitat unsuitability mask from Dataset 3) was added to the presence-only records from submissions to improve the predictive power of the model. This process is not free of error since *I. ricinus* is not confirmed to be absent at all the points used as absence (Pearce and Boyce, 2006). A general disadvantage of this type of distribution data is that all presences are treated as equal, regardless of the abundance of *I. ricinus* ticks that the habitat supports, which may not provide enough information to enable the model to differentiate a scarce habitat classified as having the species present from a habitat where the species is in fact established (Howard *et al.*, 2014).

The predictive map resulting from Model 2 indicates high probability of *I. ricinus* presence over much of mainland Scotland, and does not reflect the known vector habitat preferences, as presence is predicted in some unsuitable areas. This does not provide particularly useful information for targeting public health interventions and illustrates the challenges of using sparse presence-absence data in areas where ticks are endemic. The predictive map presents high uncertainty (based on the rescaled approach) demonstrating low confidence in the predictions, likely due to the small sample size. This is also evidenced by the high absolute uncertainty compared to Model 3. In addition, there is uncertainty relating to the predictions as Dataset 2 does not cover all the covariate range.

Although the potential biases of submission data are common to similar studies, they can often be minimised if sufficient sample sizes are obtained. The dataset used in this survey was not collected for the purpose of mapping tick distribution, so the sample size was low (~200). In England and Wales where the submission scheme has been promoted, over 4,000 data points were collected for the same period, giving more capacity for predictive mapping. These results should not rule out the use of data from public submissions that can be used to infer range limits of *I. ricinus* after careful analysis to account for adventitious ticks dispersed by hosts (ECDC, 2019a).

#### **2.4.4. Use of *I. ricinus* combined datasets (Dataset 3)**

The predictive model based on a dataset that combined data from questing tick surveys, public submission data and absences of *I. ricinus* increased the spatial coverage of the data in mainland Scotland (Figure 2.1c) and produced a more detailed predictive map. In addition, this dataset had the best coverage for the covariate range used in predictions (Figure 2.3). It is not surprising that the spatial trend of the predictions from both presence-absence models were comparable. However, the model developed using the combined data (Model 3) provides a better description of the presence and absence of *I. ricinus*, not only because of the higher number of points but also because it includes

presence and absence data from questing tick surveys (Dataset 1). This method of adding information from different datasets can be more easily applied at country and continental levels to obtain distribution maps. However, because composite datasets combine different types of data, it is more challenging to understand how the different errors, bias and limitations of each dataset might affect the model outputs and the predicted uncertainty.

#### **2.4.5. Predictive *I. ricinus* maps for Scotland**

The three datasets used in this study are the only *I. ricinus* datasets that are available at a national scale in Scotland (as far as the author is aware). As discussed above, although data from questing tick surveys are usually regarded as the gold standard, the data used in this study did not have good coverage, both geographically and over the covariate range, for the whole of Scotland. This dataset is appropriate for making decisions that require detailed distribution data only in areas where the coverage is good. Outputs from Dataset 2, comprising public submission data, were limited by the small sample sizes in this dataset, which gave high model uncertainty. Therefore Dataset 3, which uses data from multiple sources, provides the most convincing predictive map and is recommended for decision-making at national scale.

It is conceivable that any of these maps could be used alone for decision-making, without further consideration of the limitations of the data inputs. The differences between the three maps highlight the importance of exploring sources of uncertainty in models and in predictions and presenting this alongside predictive maps. Although there are a high number of published papers on *I. ricinus* predicted distribution, uncertainty is rarely presented (rare examples are Alkische, Peterson and Samy, 2017; Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017). For other vectors, when uncertainty is considered, the uncertainty metric used in this study is commonly reported (such as Lord *et al.*, 2018), but its value is limited because the uncertainty values correlate with the posterior mean. In addition, this measure does not include uncertainty

associated with extrapolation outside the covariate range. Further development of methodological approaches to quantify this uncertainty, such as statistical tools for the diagnosis of model prediction reliability or to limit predictions to the range and covariates encountered during surveys, would be beneficial as also highlighted by Conn, Johnson and Boveng (2015). Although this exercise was conducted with the aim of improving tick distribution and LB management, the findings are relevant to other VBD systems for animal and human health.

## **2.5. Conclusions**

The choice of the most suitable predictive map of *I. ricinus* distribution in Scotland depends on the objective. For local level decision-making, Dataset 1 (*I. ricinus* data from questing tick surveys) and map 1 are more appropriate, with good coverage for some areas of Scotland. For decision-making at national level, using the combined Dataset 3 (map 3) provides better coverage of the country and the range of the covariates. Although questing tick surveys provide detailed data on questing tick relative abundance, the resources required often limit the number of areas that can be sampled, which makes it challenging to make predictions for extensive areas. If available at larger spatial and temporal resolution, relative abundance data will result in finer scale maps that are more effective for risk management and communication at national and regional levels. The analysis in this study highlights the need for additional surveying in areas with poor previous coverage.

Future maps of *I. ricinus* abundance could be improved by adding *a priori* information of habitat preferences into the model structure (Conn, Johnson and Boveng, 2015). For large-scale mapping at lower resolution, or if there are few tick data from quantitative surveys, data on *I. ricinus* presence-only should be combined with data from field surveys and absence data for modelling presence-absence. To overcome the problems inherent in the use of presence-only data from public submissions, it is necessary to decrease associated errors and bias by accounting for observer effort and expertise

(Palmer *et al.*, 2017) or to find approaches by which absence data are also reported (Swart *et al.*, 2014). When predictive maps are needed for public health decision-making, such as allocation of resources for awareness campaigns, information on uncertainty should be included with vector distribution maps. However, because map uncertainty reflects a single source of uncertainty (the spatial model), improved statistical techniques are required to quantify uncertainties relating to predictions.

## **Chapter 3: Improving predictive mapping of *Ixodes ricinus* distribution in Scotland by accounting for spatial correlation using the Stochastic Partial Differential Equation (SPDE)**

### **3.1. Introduction**

#### **The need to account for spatial correlation when modelling tick distribution**

Predictive mapping of tick species such as *I. ricinus* is a difficult process, affected not only by the complexity of tick ecology, but also by the quality and coverage of data inputs, which will affect the quality, extent and resolution of the predicted map and by the selected model technique, which will vary in the degree it can capture all the variability present (Ribeiro *et al.*, 2019). One additional feature about predicting *I. ricinus* distribution that was not addressed in the previous chapter, is the presence of spatial correlation (or dependency) in observed and in unobserved explanatory variables (variables that have not been identified or are unmeasured). This spatial dependency needs to be accounted for in the inferential process to get reliable predictions from the spatial models. *I. ricinus* ticks have limited dispersion which depends predominantly on their hosts' movement (Randolph, 1998), and therefore exhibits a spatial structure in explanatory variables (Beguin *et al.*, 2012). Additionally, spatial correlation results from the methods of data collection used. *I. ricinus* data are often collected during questing tick surveys by dragging or flagging (Gilbert, 2010; James *et al.*, 2012; ECDC and EFSA, 2018). When dragging or flagging in a site, multiple transects are performed, and questing ticks are counted in each transect. The outcome is tick abundance or presence-absence of questing ticks per unit of time or area, observed in each site location and consequently, tick data is characterised by a hierarchy of drags or flags within sites.



Analysis of species distributions such as *I. ricinus* is sensitive to spatial dependency in model residuals, and not accounting for spatial correlation can reduce model performance (Latimer *et al.*, 2006) and lead to underestimation of prediction errors (Gelfand *et al.*, 2006), and incorrect results. Therefore, it requires a model that incorporates the information that an observation is more correlated with an observation collected at the same or a neighbouring location, than with another observation that is collected from a faraway location. This is based on the first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). However, limitations of most common SDMs (such as GLMs and environmental envelopes) are the inability to account for spatial correlation or irregular sampling intensity, and to adequately quantify uncertainty of all model parameters (Latimer *et al.*, 2006). Although not exclusive of Bayesian inference, this approach brings several advantages in modelling ecological data with spatial and/or temporal correlation, by integrating the concept of hierarchical spatial structure and accounting for similarities based on the neighbourhood between sites of collection.

### **The Bayesian approach to deal with spatial correlation**

Bayesian inference brings several advantages to improve SDMs specifically: a) the model parameters are considered as probabilistic variables with joint probability density functions, and incorporate external information (prior information) to derive the posterior distribution (Bernardo and Smith, 2000); b) the hierarchical structure in the data and/or parameters is easily specified, which make predictions for new observations and missing data possible (Blangiardo *et al.*, 2013; Poggio *et al.*, 2016); c) uncertainty is implicitly represented (Bernardo and Smith, 2000; Poggio *et al.*, 2016); d) the posterior distribution provides a more intuitive and interpretable quantity than a frequentist P-value (Blangiardo *et al.*, 2013) and e) the final output specifies the properties of the posteriors, such as means, standard deviations and quantiles (Fuglstad and Beguin, 2018). With all these characteristics, Bayesian inference has advanced the study of the uncertainty associated with predictive

maps, a common feature due to the different errors associated with modelling for predictive mapping, providing information on the reliability of predictions and helping the process of decision-making (Poggio *et al.*, 2016).

However, the major drawback of Bayesian analysis that made their application unattractive previously, was their complexity and long computation time. Bayesian implementation of spatially explicit models has developed greatly, and started to be used more widely in research with the arrival of Markov chain Monte Carlo methods (MCMC) (Brooks *et al.*, 2011) and the WinBUGS software (Spiegelhalter, Thomas and Best, 1999; Stensgaard *et al.*, 2011). However, although MCMC algorithms are very flexible, there are still problems related to convergence, processing time and implementation (Beguin *et al.*, 2012; Blangiardo *et al.*, 2013). MCMC methods involve computationally and time-intensive simulations to obtain the posterior distribution for the parameters. For fitting large spatial datasets and/or complex models, this results in very long running times (Rue, Martino and Chopin, 2009; Eidsvik *et al.*, 2012; Blangiardo *et al.*, 2013; Poggio *et al.*, 2016).

An alternative approach to MCMC to obtain the posterior distributions is the integrated nested Laplace approximation (INLA) (Rue, Martino and Chopin, 2009). INLA is a deterministic algorithm that applies both analytical approximation and numerical integration to implement approximate Bayesian inference for a large class of Bayesian hierarchical models, called latent Gaussian models (e.g. linear regression models, additive models, hierarchical spatial models), which can account for hierarchical structure, non-Gaussian errors, and spatial and spatial-temporal correlation (Rue, Martino and Chopin, 2009; Rue *et al.*, 2013; Lindgren and Rue, 2015). INLA is a fast, computationally efficient alternative to MCMC because it uses deterministic approximations instead of random simulations (Beguin *et al.*, 2012), and performs as well as MCMC (Rue, Martino and Chopin, 2009). INLA also enables a higher degree of process automation (Beguin *et al.*, 2012) and can be easily implemented in R software, using R-INLA package (Rue *et al.*, 2013).

## **The stochastic partial differential equation**

To improve the process of modelling spatial data and account for spatial correlation, INLA was joint with the stochastic partial differential equation (SPDE) approach (Lindgren, Rue and Lindström, 2011). The SPDE approach establishes a link between Gaussian fields and Gaussian Markov random fields, taking the advantages from the two fields and dealing with the spatial correlation in a very efficient way.

In modelling hierarchical data, spatial dependency among observations is accounted for by introducing in the model Gaussian fields, which model the spatial signal in the observations that cannot be accounted for by covariates (Musenge *et al.*, 2013; Bachl *et al.*, 2019). Gaussian fields are spatially continuous random processes in which variables observed at any point are normally distributed, and are correlated with variables observed at other points in space according to a continuous correlation process (Bachl *et al.*, 2019). Although very important in spatial statistical modelling, including in geostatistics, Gaussian fields are characterised by dense covariance matrices, requiring a long computational time to run (sometimes infeasible, particularly when a large number of geolocations are used), called the 'big n problem' (Lindgren, Rue and Lindström, 2011). The SPDE approach transforms by triangulation the continuously indexed Gaussian field with Matérn covariance function, by a neighbourhood structure and a sparse precision matrix of a Gaussian Markov random field (Rue and Held, 2005). The SPDE approach integrates the good computational properties of Gaussian Markov random fields, reducing the usual computational burden (Cameletti *et al.*, 2012; Lindgren and Rue, 2015), and the computationally effective approximations of INLA algorithm (Rue, Martino and Chopin, 2009), providing a flexible structure for spatial-temporal random fields (Lindgren, Rue and Lindström, 2011).

The use of the INLA-SPDE approach has shown advantages in building complex hierarchical spatial-temporal models with large datasets, when controlling for several confounding factors and at the same time, reducing the

normally required computational time; in modelling sparse datasets, accounting for the uncertainty in the estimates; in improving the development of high resolution maps, including predictions for areas with no data; and as an efficient method to quantify and reduce bias due to spatial and temporal heterogeneity in the data. Consequently, INLA-SPDE approach has been used in several fields including environmental modelling (Eidsvik *et al.*, 2012; Poggio *et al.*, 2016; Fuglstad and Beguin, 2018), for modelling and mapping spatial-temporal disease and mortality for public health decision-making (Musenge *et al.*, 2013; Núñez *et al.*, 2016; Godana, Mwalili and Orwa, 2019; Utazi *et al.*, 2019), for health economic evaluations (Heath, Manolopoulou and Baio, 2016), and for predictive mapping of vectors and VBDs (Kifle, Hens and Faes, 2017; Myer, Campbell and Johnston, 2017; Stanton *et al.*, 2018; Chiaravalloti-Neto *et al.*, 2019). This approach has not yet been used for modelling tick distribution, but clearly has the potential to address some of the current limitations.

### **Study objectives and hypothesis**

The objective of this study was to improve on the predictive map of *I. ricinus* questing nymph distribution in mainland Scotland obtained in Chapter 2, including the estimated uncertainty in the predictions, by applying the SPDE method to the data collected during questing tick surveys (Dataset 1 in Chapter 2). This was achieved by predicting the posterior density of *I. ricinus* nymphs (counts per drag) in mainland Scotland, and estimating the uncertainty of the predictions, by accounting for the spatial correlation through the introduction of a random field component. Additionally, model predictions accounted for extra sources of variability, namely the extra variation of site and drag random effects, and for the ZIP process. It was hypothesised that the introduction of the spatial random field would improve the predicted posterior density and the uncertainty in predicted values when compared with a model without this component. Finally, it was hypothesised that adding the extra variation of the drag and site effects, and the ZIP process in model predictions, would generate a more informative uncertainty map. Complex predictive models and maps

were created using inlabru R package (Bachl *et al.*, 2019), and the predictive performance of the different models was assessed using scoring rules.

## **3.2. Materials and Methods**

### **3.2.1. The inlabru R package**

Inlabru R package (Bachl *et al.*, 2019) was the central package for this work, used to apply the SPDE and create a smooth spatial density surface for *I. ricinus* questing nymphs, based on data collected during questing tick surveys. This package was initially developed as part of a research project to implement novel methods to model spatial distribution from ecological survey data and is continuously being updated into new versions.

Inlabru was built on INLA R package. Inlabru has a simpler syntax and provides easier access to Bayesian inference to fit realistically complex spatial and spatial-temporal models in an accurate and fast way, accounting for spatial and temporal correlation (Bachl *et al.*, 2019). Inlabru package supports the following R packages: a) the sp package (Pebesma and Bivand, 2005) to deal with spatial data as points, lines and polygons ('Spatial Points Data Frame', 'Spatial Lines Data Frame' and 'Spatial Polygons Data Frame'); b) INLA package (Rue *et al.*, 2013) to approximate continuous space with the random field; and c) ggplot2 (Wickham, Chang, *et al.*, 2020) and ggmap (Kahle and Wickham, 2013) packages for data visualisation tools (plus specific inlabru functions such as 'gg' and 'gm' to extend their functionality). Other R packages were used in this work, namely raster (Hijmans *et al.*, 2020) and rgdal (Bivand, Keitt, *et al.*, 2020), to deal with raster spatial data (.tif); dplyr (Wickham, François, *et al.*, 2020) to write the required code; tidyverse (Wickham, 2019) for clean coding and RcolorBrewer (Neuwirth, 2014) for colour ranges. All the modelling procedures were done in R version 3.5.3 (R Core Team, 2019a), using inlabru development version.

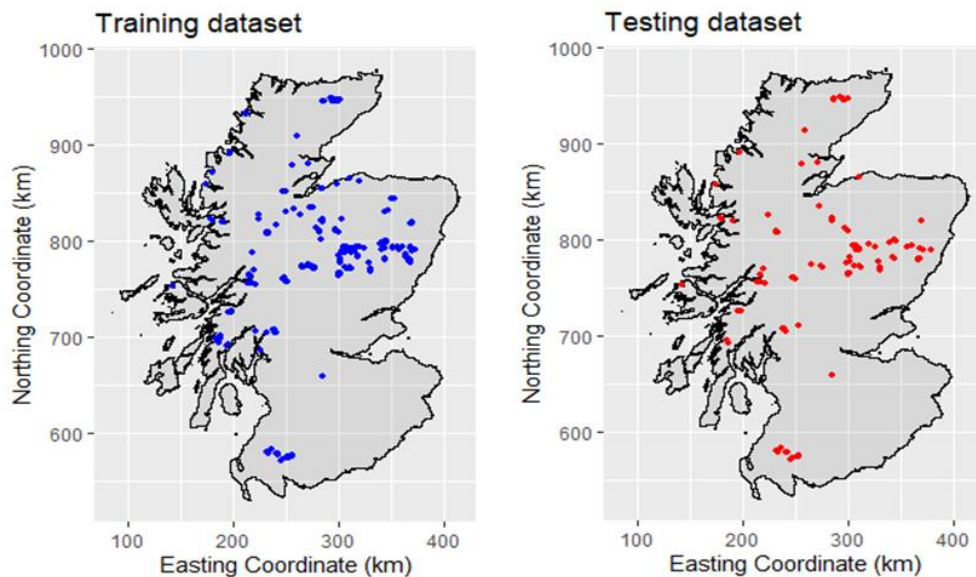
### 3.2.2. Quantitative *I. ricinus* survey data: data management

The data used for analyses comprised *I. ricinus* counts from questing tick surveys, as described in Chapter 2 (Dataset 1). However, only data on nymphs was used since nymphs are the instar most often involved in human tick bites (Robertson, Gray and Stewart, 2000), and represent a higher risk of transmitting pathogenic agents (such as the bacteria that causes LB (Robertson, Gray and Stewart, 2000; James *et al.*, 2012; Li *et al.*, 2016)). During questing tick surveys, nymphs were collected in higher numbers than adult ticks (94.2% of ticks collected were nymphs compared to 5.8% adult ticks). Therefore, data on nymph and adult ticks combined show a similar distribution to data on nymphs only. Data were collected from 2006 to 2017; 687 sites were visited and a total of 10,611 drags were performed. The domain for the analysis was defined as mainland Scotland.

To achieve the aim of this study to improve predictive mapping of *I. ricinus* questing nymphs in mainland Scotland, the candidate model predictive ability was evaluated on independent data, to better assess the model performance across the domain. The full dataset with the counts of questing nymphs per drag, visit and site of collection was split into two datasets: a) a training dataset for model training and b) a testing dataset for model validation. Therefore in order to test model predictive performance for different locations in the domain (mainland Scotland), the training and testing datasets were defined based on a random selection of sites, with 80% of the sites for the training sample and 20% of the sites for the testing sample. Due to the clustered nature of sites of questing tick surveys in Scotland, this procedure was done by considering the total number of sites in each area (council area) and selecting proportionally (and randomly) distinct sites per area, to guarantee that all areas of the domain had sites in the training and testing datasets.

The training dataset consisted of 8,510 drags in 547 sites (each drag being a row), and the testing dataset consisted of 139 sites with 2,101 drags. Spatial points from each dataset, training and testing, were stored together with their

attributes as spatial points data frame structures. Then to each spatial point the same projection (British National Grid) and extent as the spatial polygon object of mainland Scotland were attributed. Figure 3.1 shows the spatial distribution of training and testing data in mainland Scotland.



**Figure 3.1:** Distribution of the training and the testing datasets in mainland Scotland.

The blue dots represent the sites of the training dataset and the red dots the sites of the testing dataset.

### 3.2.3. Environmental covariates

Relevant environmental covariates (climatic, topographic, land cover and host-related variables) that might explain tick distribution in Scotland were selected for this work. To predict nymph distribution over the domain, these covariates were obtained as GIS-based raster maps. Ten covariates were identified as suitable for the modelling procedure: LST in July (1 km<sup>2</sup> resolution, average from 2001-2013, from MODIS); elevation above sea level (90 m<sup>2</sup> resolution); rainfall (total precipitation) in July and days of ground frost in September (long-term average climate data from UK Met Office (from 1981–2010, 5 km<sup>2</sup> resolution)); an index of presence of roe (Alexander *et al.*, 2014) and red deer (Wint *et al.*, 2014) (both at a resolution of 0.008333 \* 0.008333 degrees (~1

km<sup>2</sup>); the proportion of deciduous and of coniferous forest (from Land Cover 2006 (1 km<sup>2</sup> resolution)) and finally, latitude and longitude for each site. Covariates were defined as spatial pixels data frame objects ('Spatial Pixels Data Frame'), with the same extent (mainland Scotland) and projection (British National Grid). The final resolution of all covariates was 1 km<sup>2</sup> per pixel.

To be able to fit the SPDE model in continuous space, it was necessary to evaluate the SPDE everywhere in the domain. Because this procedure includes an infinite number of points, covariates needed to be specified as functions. Using specific `inlabru` functions, covariates, as spatial pixel data frame objects, were then standardised and defined as continuous covariates over space. However, the current stable version of `inlabru` does not require these steps because they are included already in the internal code, and covariates just need to be defined as spatial pixels data frame objects.

#### **3.2.4. The mesh and the SPDE model**

Spatially structured random effects were included in the model by using the SPDE approach. The first step to create the SPDE model was to define the 'mesh', or constrained refined Delaunay triangulation (Figure 3.2), a finite grid of triangulations of the spatial domain, which will add smooth random effects to the model (Lindgren and Rue, 2015). The mesh divides the domain into a set of non-intersecting triangles (which may be irregular), where any two triangles meet in an edge or corner (Lindgren, Rue and Lindström, 2011). The mesh can be designed from different principles; one approach often used to define this triangulation is to use the locations of the observations as initial triangle vertices, and then add further vertices to satisfy the triangulation quality constraints (such as Lindgren, Rue and Lindström, 2011; Cameletti *et al.*, 2012; Musenge *et al.*, 2013; Poggio *et al.*, 2016; Silva *et al.*, 2017). Alternatively, the mesh can be designed based on the domain boundaries (Núñez *et al.*, 2016; Krainski *et al.*, 2017). The latter approach was done in this study because a mesh created based on the boundaries of the domain can be used in future research, for predictions over the domain using other datasets.



Therefore, a spatial polygon data frame of the boundaries of mainland Scotland was used, and the function 'inla.mesh.2d()' used to create a two-dimensional mesh. The quality of the mesh was assessed, and the mesh was refined using the shiny application (app) 'meshbuilder()', an interactive graphical tool available in INLA package, to assess the approximation errors and to build a refined mesh of the domain. The final mesh consisted of 2,626 vertices, and the same projection as the previously referred spatial elements was defined (the lengths have the same units as the coordinate system).

The parameters for defining the mesh in the function 'inla.mesh.2d()' were the following (Krainski *et al.*, 2017):

- a) the boundary was created using the function 'as.inla.mesh.segment()' to extract the boundary of the spatial polygon data frame object of the domain;
- b) the maximum edge length sets the triangle structure by specifying the maximum allowed triangle edge lengths in the inner domain and outer extension. This was defined as 5 km inside the inner domain and 45 km in the outer extension;
- c) the minimum angle sets the triangle structure in addition to the maximum edge by specifying the minimum internal angles of the triangles on the inner domain and outer extension. This was set as 30 for the inner domain and 21 for the outer extension (values up to 21 guarantee the convergence of the algorithm);
- d) the maximum n and maximum n strict were used to avoid large meshes. The argument n is the initial number of points on the extended boundary. This was defined as 'max.n'=c(48,000, 16,000) and maximum n strict as 'max.n.strict'=c(128,000, 128,000);
- e) the cut-off is the minimum distance between mesh vertices and provides further control to the shape of the triangles, meaning that points closer than the supplied distance are replaced by a single vertex. This parameter was used to filter away adjacent points and avoid

building small triangles around clustered locations. This was set as 5 km;

f) the offset specifies the size of the inner and outer extensions (defines how much the inner domain and the outer extension should be extended). This was defined as 10 km for the inner and 60 km for the outer extensions.

The Matérn correlation model was defined using the function 'inla.spde2.pcmatern', which considers the mesh and the hyperparameters, the priors for range (practical range) and sigma (the marginal standard deviation). Priors can be specified as informative or left as the default (non-informative) (Poggio *et al.*, 2016). The priors selected in this work are sensible, but not universal, default priors, called PC-priors (Fuglstad *et al.*, 2019), which are available in INLA and inlabru. By using non-informative priors, little *a priori* knowledge of the values and variability that model parameters might take was assumed, therefore this allowed the properties of the data to inform the posterior distributions (Myer, Campbell and Johnston, 2017).

### **3.2.5. Modelling and predictions**

The response variable was the count of nymphs collected per drag, visit and site of collection, and models were fitted using the 'bru()' function (Bachl *et al.*, 2019). Models were designated by specifying a formula for the linear predictor that defined the log density function, the components of the predictor (intercept and covariates, with and without the SPDE), and the observed variable distribution (Bachl *et al.*, 2019). To account for the variability of the random effects of site and drag on the posterior density and predicted uncertainty, PC-priors for site and drag were defined. Due to the excess of zeros in the response variable (52% of the drags are zero counts of nymphs) it was assumed (as done in Model 1 in Chapter 2) that the probability density function of the response variable was defined by a ZIP (Type 1) process (as defined in page 80). The components of the linear predictor - the intercept, the previously

identified set of continuous covariates, and the SPDE model, were stored in the same object.

Model specification:

$$\text{Log}(M_{ij}) \sim \alpha + \beta_1 \text{Rain} + \beta_2 \text{Reddeer} + \beta_3 \text{Deciduous} + \dots + \beta_k \text{Long} + f(\text{Drag}_{ij}) + f(\text{site}_j) + f(\mu_j)$$

Where  $M_{ij}$  is mean tick abundance for  $i$ th drag in site  $j$ ;  $\alpha$  is the intercept;  $\beta_k$  are the measure of covariate effects; and  $f()$  denote the random effects due to drags and sites respectively, and to the spatially structured effect  $\mu_j$ . The specification of the zero-inflated Poisson (Type 1) model can be seen in page 80.

Using the training dataset, the first step was to run a hierarchical model of the counts of nymphs per drag as a function of the set of covariates used in Model 1 in Chapter 2 (LST in July, ground frost in September, proportion of forest that was deciduous, proportion of forest that was coniferous, probability of roe deer presence, latitude and longitude). This was also used in Model 1 in this chapter. Secondly, a new model was created by adding the random field to Model 1 and named Model 2. Predictions for the training and testing datasets for Model 1 and Model 2 were generated using the `inlabru predict` function based on 1,000 Monte Carlo samples.

The model predictive quality was assessed using scoring rules. Although predictions and respective scoring rules were also generated for the training data, this was just to compare how the different models behaved for the training and testing data, to help identify model overfitting and compare how consistent models were. However, the model predictive capability was assessed just with the predictions of the testing data. The scores calculated were the following: a) the Dawid-Sebastiani score (S\_DS), which measures the concentration of predictive distribution and coverage probabilities (Dawid and Sebastiani, 1999), and b) the logarithmic score (S\_LOG), which is a strictly proper scoring rule, depending on the full probabilistic distribution and

accounting for the skewness in the data (Good, 1952). The two used scoring rules are negatively oriented, which means the lower the score, the better the predictive model (Held and Meyer, 2019). Functions were created to calculate each one of the scores, the respective standard deviation (SD, which gives the dispersion of the data from the mean) and standard error (SE, which gives the precision of the mean). The mean difference between scores (keeping one model as reference) and the respective P-value for those differences were also estimated.

After comparing the models' predictions, Models 1 and 2 were run again with the full dataset, to predict the expected posterior density for *I. ricinus* nymphs and the associated uncertainty. The predictions were made over the regular grid of triangles (mesh) covering the domain of the Scottish mainland. The uncertainty map was based on SD values.

As discussed in Chapter 2, the coverage of data points over the covariates used in the predictions affects the predicted distribution and the predicted uncertainty. In addition, the random field can explain the same spatial effect as some covariates (making them irrelevant for predictive purposes) and can account for variability not accounted for by the covariates. Therefore, in this analysis it was necessary to assess a combination of different covariates in the presence of the random field, to create a better spatial model for tick distribution (when compared with Model 2). Using the training dataset, three best random field models (Model 3, Model 4 and Model 5) resulting from different combination of covariates were selected based on a forward approach and lower values of the DIC and Watanabe-Akaike information criteria (WAIC), as measures of goodness of model fit, and the CPO score as a measure for the predictive quality of the model (Gelman, Hwang and Vehtari, 2013). Predictions for the training and testing datasets for these three models were generated using 1,000 Monte Carlo samples. Scores were computed to help select the most suitable random field model. Finally, the best predictive model was used to assess the effect of the two random effects (drag and site)

and the ZIP distribution on the predicted uncertainty by comparing the prediction with these components against the prediction without these components.

The following functions were used for the predictions in the inlabru code (provided by Professor Finn Lindgren specifically for this work):

**Specification of the zero-inflated Poisson (Type 1) model:**

$$\theta \sim p(\theta)$$

Where  $\theta$  is a vector of parameters

$$x|\theta \sim N(0, Q(\theta)^{-1})$$

$x$  is the covariates used in the model ( $x$  given  $\theta$ ), and the covariates are normally distributed with mean equal to 0; and variance  $Q(\theta)^{-1}$

$$\eta = Ax$$

$\eta$  is the linear predictor (defined in page 78)

$$\lambda = \exp(\eta)$$

$\lambda$  is the exponent function of the linear predictor, that predicts the average nymph abundance

$$y|x, \theta \sim ZIP_1(\lambda, p_0)$$

$y$  is nymph abundance, which depends on the covariates;  $p_0$  is the probability of excess zeros in the model

The probability function for the ZIP<sub>1</sub> is

$$P_{ZIP_1}(y = K|\lambda, p_0) = p_0 1_{[k=0]} + (1 - p_0)P_{Poisson}(y = k|\lambda)$$

$K$  is the observed value of abundance in each drag

Moments:

The conditional expectation (mean) and variance of a ZIP<sub>1</sub> model are:

(This is the process for obtaining the formula of the mean and the variance of the zero-inflated Poisson 1 process)

$$\begin{aligned}
E_{ZIP_1(\lambda, p_0)}(y) &= (1 - p_0)E_{Poisson(\lambda)}(y) = (1 - p_0)\lambda \\
Var_{ZIP_1(\lambda, p_0)}(y) &= E\left(Var_{ZIP_1|1_{[y=0]}}(y|1_{[y=0]}, \lambda, p_0) \middle| \lambda, p_0\right) \\
&\quad + Var\left(E_{ZIP_1|1_{[y=0]}}(y|1_{[y=0]}, \lambda, p_0) \middle| \lambda, p_0\right) \\
&= E\left((1 - 1_{[y=0]})\lambda \middle| \lambda, p_0\right) + Var\left((1 - 1_{[y=0]})\lambda \middle| \lambda, p_0\right) \\
&= (1 - p_0)\lambda + p_0(1 - p_0)\lambda^2 = (1 - p_0)\lambda(1 + p_0\lambda)
\end{aligned}$$

The inlabru predict function generates posterior Monte Carlo samples

$$\left\{(\lambda^{[i]}, p_0^{[i]}), i = 1, \dots, N\right\}$$

Where  $i$  is the number of values (in this work  $N$  was equal to 1,000 samples)

giving estimated means and variances

$E$  is the expectation function

$$\hat{E}_{(\lambda, p_0)}[E(y|\lambda, p_0)] = \frac{1}{N} \sum_{i=1}^N E_{ZIP_1(\lambda^{[i]}, p_0^{[i]})}(y)$$

$$\widehat{Var}_{(\lambda, p_0)}[E(y|\lambda, p_0)] = \frac{1}{N-1} \sum_{i=1}^N \left( E_{ZIP_1(\lambda^{[i]}, p_0^{[i]})}(y) - \hat{E}_{(\lambda, p_0)}[E(y|\lambda, p_0)] \right)^2$$

$$\hat{E}_{(\lambda, p_0)}[Var(y|\lambda, p_0)] = \frac{1}{N} \sum_{i=1}^N Var_{ZIP_1(\lambda^{[i]}, p_0^{[i]})}(y)$$

The integrated predictive expectation and variance Monte Carlo estimates are

$$\hat{E}_F = \hat{E}_{(\lambda, p_0)}[E(y|\lambda, p_0)]$$

$$\widehat{Var}_F = \hat{E}_{(\lambda, p_0)}[Var(y|\lambda, p_0)] + \widehat{Var}_{(\lambda, p_0)}[E(y|\lambda, p_0)]$$

For estimating the scoring rules:

1 – Dawid-Sebastiani: the Dawid-Sebastiani score for a prediction  $F$  and observation  $y$  is

$$S_{DS}(F, y) = \frac{(y - \mu_F)^2}{\sigma_F^2} + \log(\sigma_F^2), \text{ where in this case } \mu_F = \hat{E}_F \text{ and } \sigma_F^2 = \widehat{Var}_F$$

Where  $\mu$  is the mean of the predicted values and  $\sigma^2$  is the variance of the predicted values

2 – Logarithmic score: the Logarithmic likelihood-score is

$$S_{LOG}(F, y) = -\log [P_{Y \sim F}(Y = y)]$$

We estimate  $P_{Y \sim F}(Y = y)$  with a Monte Carlo estimator, as before,

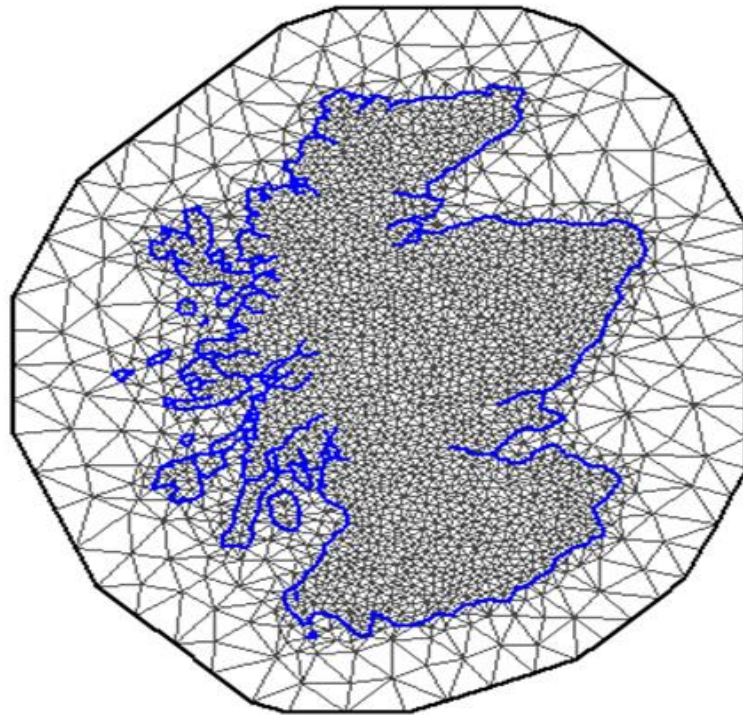
The predicted function

$$\hat{P}_{Y \sim F}(Y = y) = \hat{P}_{(\lambda, p_0)}[P_{Y \sim ZIP_1(\lambda, p_0)}(Y = y)] = \frac{1}{N} \sum_{i=1}^N P_{Y \sim ZIP_1(\lambda^{[i]}, p_0^{[i]})}(Y = y)$$

### 3.3. Results

#### 3.3.1. The constrained refined Delaunay triangulation: mesh

Figure 3.2 shows the spatial dependency structure used in the models with the random field. This refined mesh is composed by small triangles with almost the same dimensions in the inner domain, where the predictions are important, and bigger triangles in the outer extension (after the boundaries), to decrease the boundary effect.



**Figure 3.2:** The two-dimensional constrained refined Delaunay triangulation (mesh) used in this work to define the spatial dependency structure.

The blue line represents the boundaries of mainland Scotland. The small triangles are in the inner domain and the larger triangles are in the outer extension.

#### 3.3.2. Assessment of model predictive quality: comparison between Models 1 and 2

Table 3.1 describes the covariates used in Models 1 and 2, the presence or not of the random field, and the values of DIC, WAIC and CPO, when the models were run with the training data. Although Model 2 presented slightly



higher values of DIC and WAIC than Model 1, adding the random field resulted in a better value of predictive performance than Model 1 as depicted by the CPO (Table 3.1).

**Table 3.1:** Description of Models 1 and 2.

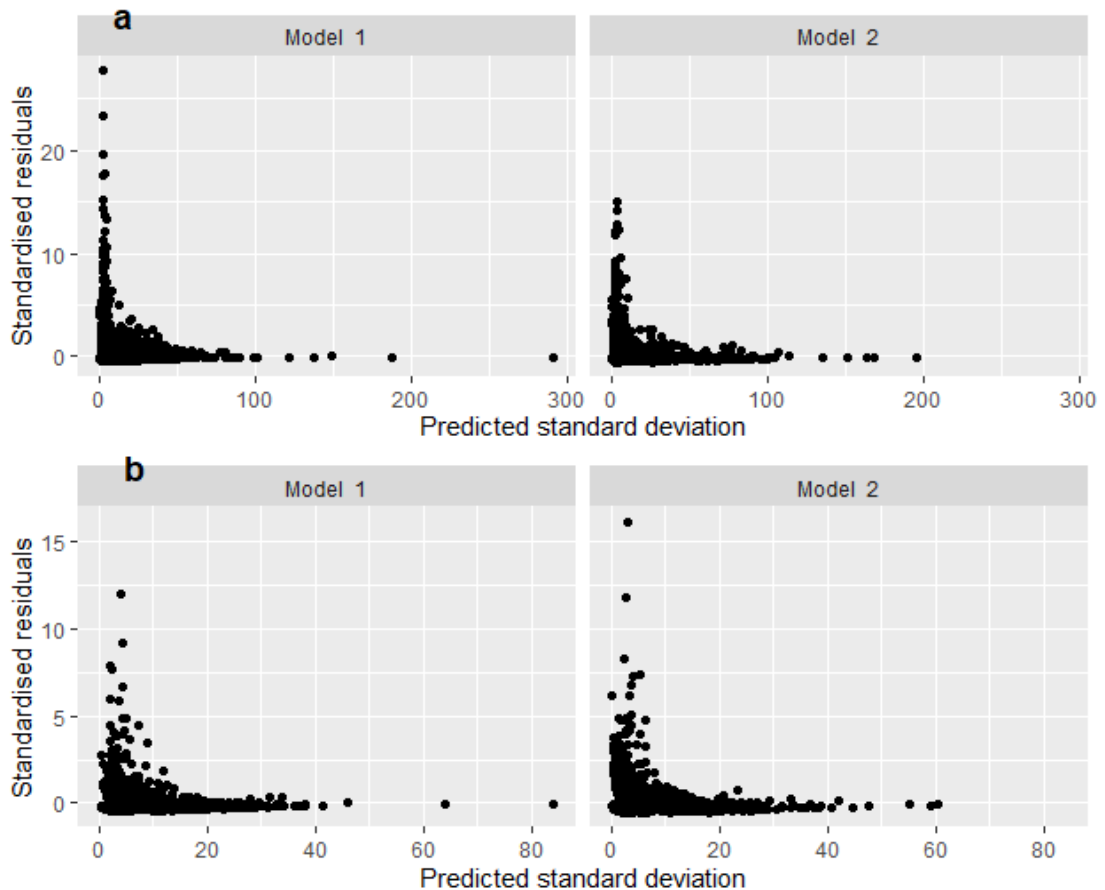
Models 1 and 2 training set covariates (fixed effects), presence or not of the random field, and respective Bayesian assessment criteria, DIC, WAIC and CPO.

Model	Fixed effects	Presence of random field?	DIC	WAIC	CPO
Model 1	LST July	No	23,341	22,978	16,300
	Number of days of frost in September				
	% cover of deciduous woodland				
	% cover of coniferous woodland				
	Roe deer presence				
	Latitude				
	Longitude				
Model 2	LST July	Yes	23,358	23,051	14,743
	Number of days of frost in September				
	% cover of deciduous woodland				
	% cover of coniferous woodland				
	Roe deer presence				
	Latitude				
	Longitude				

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; LST, land surface temperature; WAIC, Watanabe-Akaike information criteria

Exploratory graphs were generated to show the distribution of the standardised residuals of Models 1 and 2 for a) the training and b) testing datasets (Figure 3.3). The standardised residual is the residual (the difference between the observed and the predicted values) divided by its SD (y-axis). Figure 3.3 analyses the standardised residuals from Model 1 and Model 2, in order to compare variability between the two models. Figure 3.3 shows that variability is higher in the training dataset than in the testing dataset. This is due to the higher proportion of data in the training dataset (80% of the sample), which resulted in a larger SD in counts of nymphs per drag, visit and site. This may

explain why the SD of predicted values for the training dataset is larger than the SD of predicted values of the testing dataset. The effect of the large SD in the training dataset also reflects on the scores, which are slightly higher for the training than for the testing dataset. In addition, the training and testing datasets were skewed and have outliers. These characteristics were important for selecting the most appropriate scoring rule to inform the decision regarding the most suitable predictive model.



**Figure 3.3:** (a) Model standardised residuals (y-axis) for the training dataset for Models 1 and 2; (b) model standardised residuals (y-axis) for the testing dataset for Models 1 and 2.

The x-axis shows the predicted SD. Different scales were used because the interest is to compare the testing datasets in both models separately.

For both the training and testing datasets, under  $S_{DS}$  and  $S_{LOG}$  scores, Model 1 has lower predictive capability than Model 2, the model with the random field (Table 3.2). In this table, the SD of the score is a measure of the

dispersion of the data from the mean, being very susceptible to outliers, whereas the SE gives the precision of the mean. However, both results of the SD and the SE were not completely informative about the differences between the scores of the different models. Therefore a next step was to assess differences between mean scores of the different models, and test if the differences between means were statistically significant or not.

**Table 3.2:** S\_DS and S\_LOG scores for Models 1 and 2.

Score mean and respective SD and SE of the predictions for Models 1 and 2, of the counts of nymphs per drag in mainland Scotland, for the training and testing datasets.

Subset	Model	S_DS			S_LOG		
		mean	SD	SE	mean	SD	SE
Training, n=8510	Model 1	2.221	6.95	0.075	1.821	1.76	0.019
	Model 2	1.624	3.24	0.035	1.671	1.56	0.017
Testing, n=2101	Model 1	1.921	2.39	0.052	1.574	1.47	0.032
	Model 2	1.619	3.77	0.082	1.562	1.45	0.032

Abbreviations: SD, standard deviation; S\_DS, Dawid-Sebastiani score; SE, standard error; S\_LOG, logarithmic score

Table 3.3 shows the results of the difference between scores (the score mean) and respective P-values when comparing Model 1 and Model 2, considering Model 1 as the reference model (i.e. test if the difference between score means is statistically significant or not). Results should be interpreted by considering the score results of Model 1 in the first place. Therefore if differences between scores were positive, Model 1 was worse than Model 2. If the differences were negative, Model 1 was a better model than Model 2. Model 1 was consistently inferior to Model 2 for the training and testing datasets. For the testing dataset, differences were statistically significant for S\_DS (Table 3.3).

**Table 3.3:** Differences in the score mean, and respective P-value, between Model 1 and Model 2, for the training and testing datasets.

The significance level is less than 0.05. If the P-value is close to 1, means that Model 1 is significantly worse than the one it is being compared to (Model 2). If the P-value is close to 0, Model 1 is significantly better than Model 2.

Model	Subset	Mean S_DS	P-value	Mean S_LOG	P-value
Model 2	Training	0.598	1	0.150	1
	Testing	0.303	1	0.012	0.938

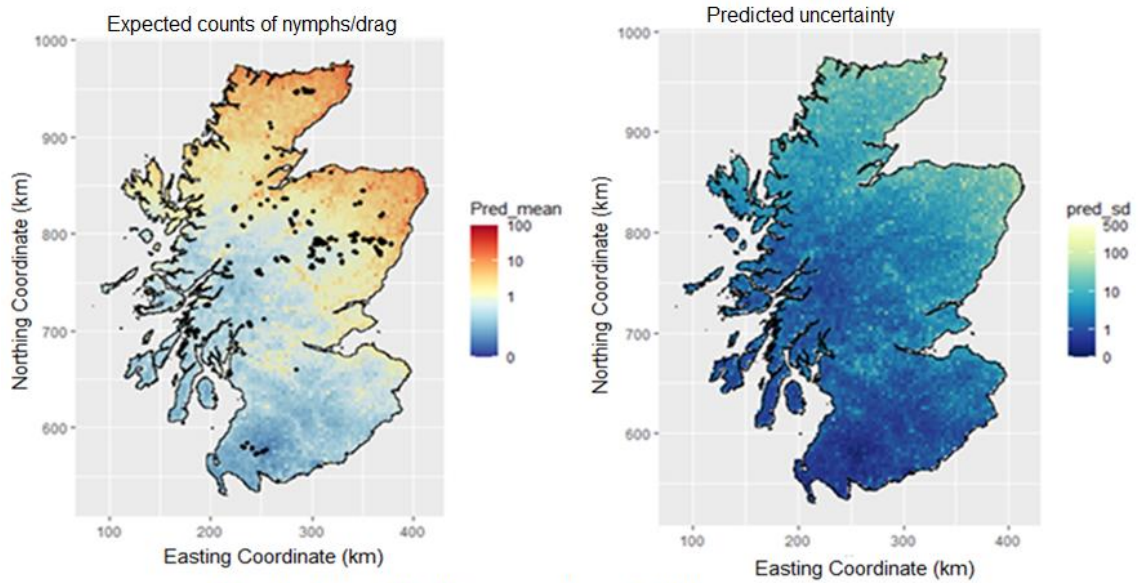
Abbreviations: S\_DS, Dawid-Sebastiani score; S\_LOG, logarithmic score

### **3.3.2.1. Predicted posterior tick density and predicted uncertainty**

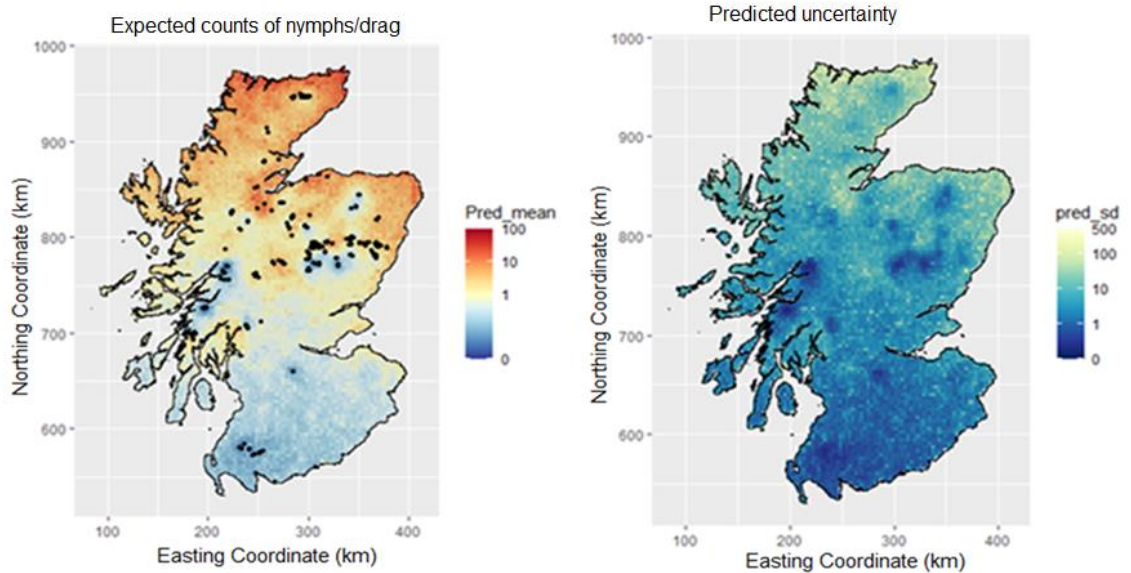
To assess the effect of the introduction of random field in model predictions and in the predicted uncertainty, Model 1 and Model 2 were re-run with the full dataset. Figure 3.4 compares Model 1 and Model 2, showing the expected tick density (the predicted counts of nymphs per drag) and the respective uncertainty in the prediction.

Predictions from Model 1 show evidence of increasing trend of predicted tick density from the south west to the north east coast of Scotland. Model 2, a model with the same covariates as Model 1 but with the additional effect of the random field, was able to identify other areas as low and high tick density. The uncertainty maps also show differences between the two models. Whereas the uncertainty of Model 1 followed the same trend as the predicted posterior tick density, the map of predicted uncertainty for Model 2 identified some areas of high predicted tick density with lower uncertainty, and some areas of low predicted tick density as high uncertainty. Additionally, in the map of predicted uncertainty for Model 2, areas with sites (where questing tick surveys were carried out) have lower predicted uncertainty.

### Model 1 without the random field



### Model 2 with the random field



**Figure 3.4:** Prediction of the posterior density surface of counts of *I. ricinus* nymphs per drag using the grid of triangulations covering the domain and respective uncertainty (the SD) for Models 1 and 2. The black points are the sites where questing tick surveys were carried out.

#### 3.3.3. New predictive models with the random field

Table 3.4 describes the covariates used in Models 3, 4 and 5, the three new selected models with the random field, and the respective values of DIC, WAIC

and CPO, when the models were run with the training data. According to CPO values, Model 3 has better predictive performance amongst the three models with random field (Table 3.4).

**Table 3.4:** Description of Models 3, 4 and 5. Model covariates (fixed effects) and respective Bayesian assessment criteria, DIC, WAIC and CPO, for the training dataset from the three best models with the random field.

Model	Fixed effects	DIC	WAIC	CPO
Model 3	Elevation	23,324	23,017	14,719
	Rain in July			
	% cover of deciduous woodland			
	% cover of coniferous woodland			
	Red deer presence			
	Latitude			
Longitude				
Model 4	Elevation	23,230	22,830	16,025
	Rain in July			
	% cover of deciduous woodland			
	Red deer presence			
	Latitude			
	Longitude			
Model 5	Elevation	23,360	23,050	14,817
	% cover of deciduous woodland			
	Red deer presence			
	Latitude			
	Longitude			

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; WAIC, Watanabe-Akaike information criteria

The model predictive quality was assessed for the three new models with the random field (Table 3.5). Although for the training dataset, Model 4 appears to be a better model (as it has a better fit to the data), the predictive quality for new data was worse, particularly when compared with Model 3 (Models 3 and 4 only differ in the presence or not of the covariate coniferous forest). For the testing dataset, Model 3 showed a better predictive performance when compared with the other two models. Considering Model 3 as the reference model (Table 3.6), Model 3 has better predictive performance than Models 4 and 5 for the testing dataset. Significant differences were found between

Model 3 and Model 5 for the two scores, and between Model 3 and Model 4 for S\_LOG (the strictly proper scoring rule).

**Table 3.5:** S\_DS and S\_LOG scores for Models 3, 4 and 5.

Score mean and respective SD and SE of the predictions for the three best models with the random field, of the counts of nymphs per drag in mainland Scotland, for the training and testing datasets.

Subset	Model	S_DS			S_LOG		
		mean	SD	SE	mean	SD	SE
Training, n=8510	Model 3	1.603	3.92	0.042	1.669	1.56	0.017
	Model 4	1.492	4.34	0.047	1.647	1.55	0.017
	Model 5	1.596	4.70	0.051	1.672	1.55	0.017
Testing, n=2101	Model 3	1.560	2.84	0.062	1.556	1.43	0.031
	Model 4	1.602	4.24	0.093	1.581	1.46	0.032
	Model 5	1.624	4.16	0.091	1.560	1.44	0.031

Abbreviations: SD, standard deviation; S\_DS, Dawid-Sebastiani score; SE, standard error; S\_LOG, logarithmic score

**Table 3.6:** Differences in the score mean, and respective P-value, between Models 2, 4 and 5 and Model 3 (used as reference model), for the training and testing datasets.

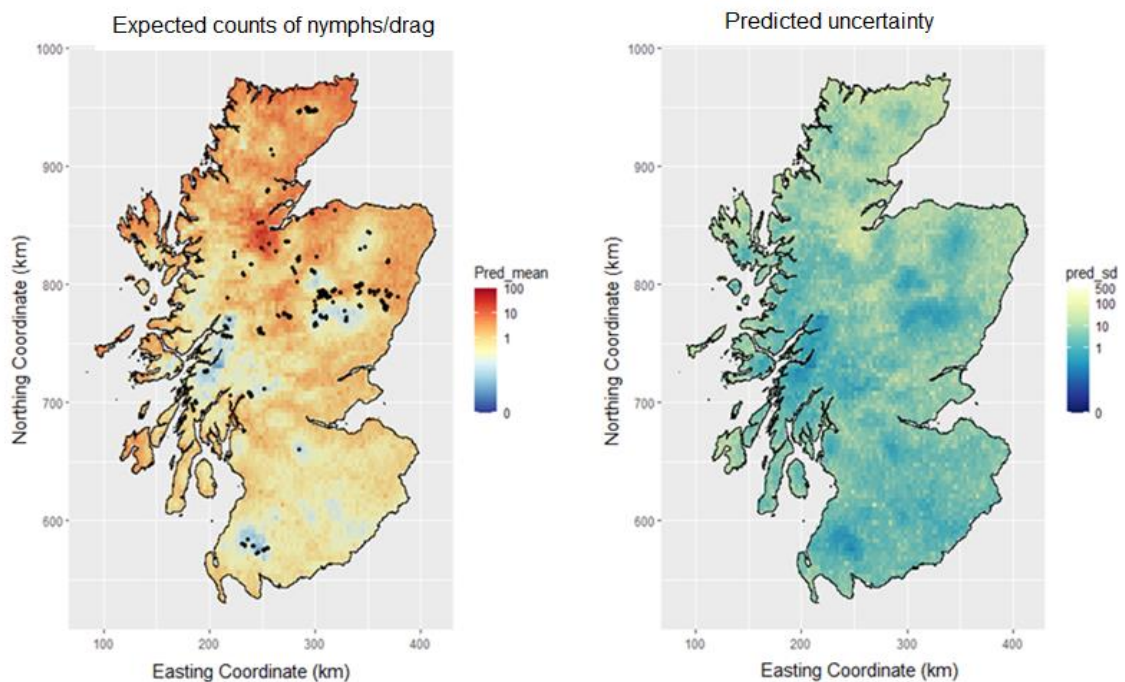
The significance level is less than 0.05. If the P-value is close to 1, the reference model is significantly worse than the one it is being compared to. If the P-value is close to 0, the reference model is significantly better than the model it is being compared to.

Model	Subset	Mean S_DS	P-value	Mean S_LOG	P-value
Model 2	Training	-0.021	0.204	-0.002	0.160
Model 4		0.111	1	0.022	1
Model 5		0.007	0.613	-0.003	0.014
Model 2	Testing	-0.058	0.063	-0.006	0.041
Model 4		-0.042	0.172	-0.025	0
Model 5		-0.064	0.027	-0.004	0.006

Abbreviations: S\_DS, Dawid-Sebastiani score; S\_LOG, logarithmic score

Comparing Model 3 with Model 2, Table 3.2 and Table 3.5 indicate that Model 3 has lower S\_DS and S\_LOG scores than Model 2. When checking for significant differences in the mean score, no significant differences were found between Models 2 and 3 for the training dataset. Regarding the testing dataset,

Model 3 performs better than Model 2 at S\_DS and S\_LOG (with significant statistical differences at S\_LOG) (Table 3.6).



**Figure 3.5:** Prediction of the posterior density surface of counts of *I. ricinus* nymphs per drag using the grid of triangulations covering the domain and uncertainty (the SD) for Model 3. The black points are the sites where questing tick surveys were carried out.

For Model 3, the model with the random field and better predictive quality, the map of posterior density surface depicted areas of low, medium and high counts of nymphs per drag over mainland Scotland, and the existence of a markedly increasing trend from the west to the east coast of Scotland is not evident (Figure 3.5). Some areas with no survey sites were identified as areas with high predicted posterior nymph density, and some areas where surveys were conducted were identified as areas with low predicted posterior nymph density. The predicted uncertainty differentiated areas with low and high uncertainty over mainland Scotland, and areas where surveys were conducted presented high to medium to low uncertainty. The same is observed for areas with no survey sites. As Model 3 was considered an improvement on Model 2, its outputs for the fixed effects (the covariates) and the hyperparameters are



presented in Table 3.7. The posterior estimates accounted for the negative effects of elevation (posterior mean = -0.0022), the amount of rain in July (posterior mean = -0.0251) and longitude (posterior mean = -0.3749), and the positive effects of the proportion of forest that is deciduous (posterior mean = 0.0044) and coniferous (posterior mean = 0.3607), the probability of red deer presence (posterior mean = 0.0105) and latitude (posterior mean = 0.6702). Posterior estimates also included the effects of the hyperparameters, the random field, the two random effects (drag and site) and the ZIP distribution.

**Table 3.7:** Results of Model 3. Posterior mean, SD, 2.5% and 97.5% quantiles and estimates (logarithm) of fixed effects and hyperparameters.

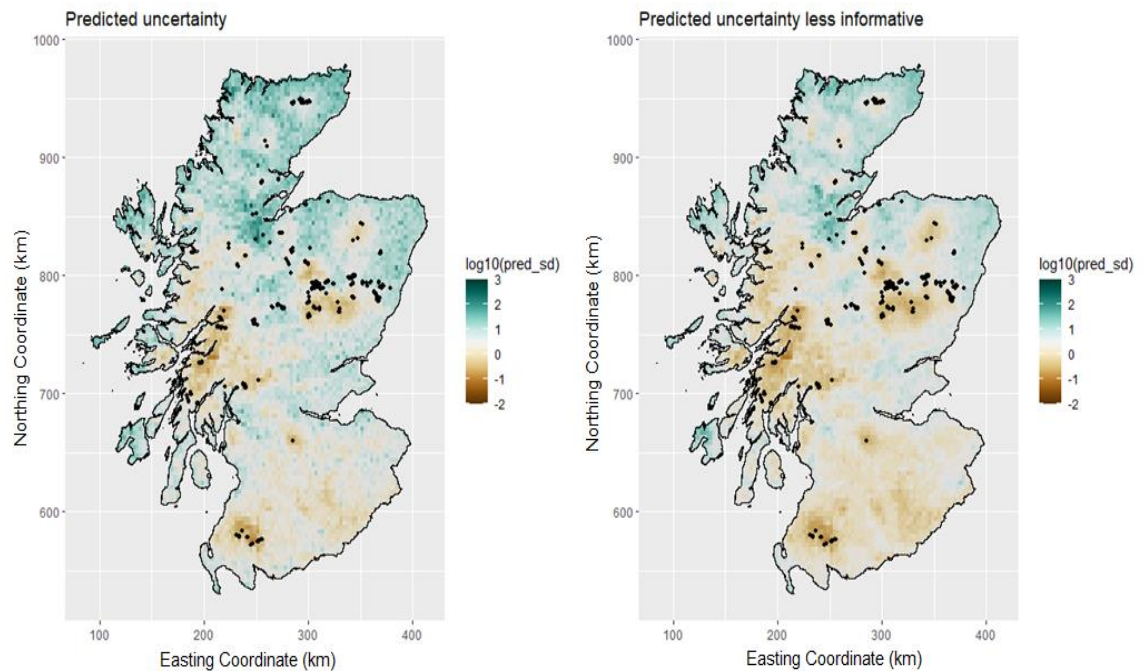
<b>Fixed effects</b>	<b>Posterior mean</b>	<b>SD</b>	<b>2.5% quantile</b>	<b>97.5% quantile</b>
Intercept	-1.5253	0.2830	-2.0820	-0.9710
Elevation	-0.0022	0.0009	-0.0038	-0.0005
Rain in July	-0.0251	0.0077	-0.0404	-0.0100
% cover of deciduous woodland	0.0044	0.0116	-0.0186	0.0270
% cover of coniferous woodland	0.3607	0.2755	-0.1807	0.9006
Red deer presence	0.0105	0.0058	-0.0008	0.0218
Latitude	0.6702	0.3588	-0.0324	1.3758
Longitude	-0.3749	0.3014	-0.9668	0.2162
<b>Model hyperparameters</b>	<b>Mean</b>	<b>SD</b>	<b>2.5% quantile</b>	<b>97.5% quantile</b>
zero-probability parameter for ZIP <sub>1</sub>	0.1367	0.0004	0.1359	0.1375
Range random field	22.8367	0.1140	22.6135	23.0618
SD for random field	1.3811	0.0079	1.3655	1.3967
Precision for site	0.8292	0.0042	0.8210	0.8374
Precision for drag	2.0962	0.0101	2.0764	2.1162

Abbreviation: SD, standard deviation; ZIP<sub>1</sub>, zero-inflated Poisson type 1

### 3.3.4. Assessment of the extra variation due to the drag and site effects, and the ZIP process in model predictions

Figure 3.6 compares the predicted uncertainty with and without the variability due to drag and site random effects and the ZIP process for Model 3. The map on the left (the predicted uncertainty which accounts for the extra sources of variation) has higher spatial uncertainty (more green and dark green areas) because it captures all sources of uncertainty, whereas the map on the right side has lower predicted uncertainty (more areas coloured light green, brown

and dark brown), not depicting the uncertainty present in some areas, and is therefore less informative. This is particularly obvious in some of the areas where no questing tick surveys were carried out.



**Figure 3.6:** Comparison between the predicted uncertainty from Model 3 with and without the variations due to the random effects and the ZIP process. Results are log scaled ( $\log_{10}$ ). The black points are the sites where questing tick surveys were carried out.

### 3.4. Discussion

Tick distribution data used for predictive mapping are frequently characterised by a heterogeneous distribution of survey sites over the domain, variability in the number of drags per site, and lack of data coverage for the covariate range in the domain. These characteristics negatively affect the reliability of the predictions outside the core survey areas and the ability to adequately capture uncertainty in predictions. This issue is common when predicting tick distribution, and the most applied statistical techniques are not equipped to capture all the variability in the data, resulting in incorrect or incomplete results. The objective of the present study was to improve mapping of the predicted distribution of *I. ricinus* nymphs in mainland Scotland and the predicted

uncertainty, using tick data with the above characteristics. For this purpose, the most extensive and up-to-date scientifically collected (using standard technique) tick distribution dataset for mainland Scotland was used.

To achieve the objective, a sophisticated approach using `inlabru` package was developed, which allowed the implementation of the SPDE approach to estimate the underlying spatial stochastic process and account for spatial dependency, and INLA for hierarchical Bayesian inference. It was hypothesised that the introduction of the random field component would improve the predictions of the posterior tick density and the uncertainty, and that adding the random effects of drag and site, and the ZIP process in the predictions, would improve the predicted uncertainty, resulting in a more informative uncertainty. In general, results indicated that models with the random field have a better predictive performance than models without this component; the predicted density of nymphs per drag and predicted uncertainty were more informative when the random field was added. Finally, using the same model, without adding the extra variation due to the effects of drag, site and the ZIP process in the predictions, the resultant uncertainty was less informative than estimates of uncertainty which accounted for these extra sources of variation.

#### **3.4.1. The accuracy and robustness of model validation and selection**

In this study, to measure the statistical consistency between the predictions and the observations, model predictions were validated based on scoring rules. These are summary metrics (penalties) which measure the accuracy of probabilistic predictions, by assigning a numerical score based on the predicted distribution and on the observation that materialises (Gneiting and Raftery, 2007; Ehm and Gneiting, 2012).

As demonstrated by other authors (such as Held and Meyer, 2019), different scoring rules applied to the same data can be inconsistent, therefore reaching different conclusions on the predictions of different models. This implies that

researchers should be careful when considering the scoring rule that is tailored to their specific situation (Armstrong, 2001; Merkle and Steyvers, 2013). For this study it was decided to apply two strictly proper scoring rules, S\_DS and S\_LOG. Dawid-Sebastiani score is considered a strictly proper scoring rule in the presence of a Gaussian distribution. However, although S\_DS is a good metric for general predictive model choice, if the distribution is not fully determined by the first two moments, as in this study, then this score may not be able to detect miscalibration (Held, Rufibach and Balabdaoui, 2010). The logarithmic score was the most relevant metric used since it is the unique strictly proper scoring rule (Gelman, Hwang and Vehtari, 2013), depending on the full distribution and being adequate to use in case of a skewed distribution with outliers. Using these scoring rules to assess the similarities and differences between models, allowed the demonstration that Model 2 (with the random field) resulted in better predictions over the domain than Model 1. For selecting the most suitable predictive model, the analysis of S\_LOG mean, SD and SE, as well as the differences between models, were decisive in choosing Model 3 (with the random field) as the most suitable predictive model of nymph counts per drag over mainland Scotland.

For this study, the squared error and the absolute error scores (Hyndman and Koehler, 2006; Held and Meyer, 2019) were also calculated. However, after preliminary analysis they were found not appropriate to use with the data as their results were inconsistent, and so they were excluded. Although both scores are frequently used to evaluate model predictive capability (Olden, Lawler and Poff, 2008; Gneiting, 2011; Ducheyne *et al.*, 2013; Kiewra, Kryza and Szymanowski, 2014; Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017; Held and Meyer, 2019), the absolute error score is not a proper scoring rule, as it is not very sensitive to outliers and fails to punish larger errors in the predictions. The squared error score can be excessively affected by the presence of outliers and penalises larger errors than smaller errors (e.g. if a method fails just one time but the error is large, the assessment will be affected negatively). Particularly for the testing dataset, there were some

inconsistencies in results when compared with the metrics used that are more appropriate to the data, therefore questioning the reliability and appropriateness of the squared error and absolute error scores to use in this study. As seen in Figure 3.3, the distribution of the standardised residuals was skewed with outliers, and neither the squared error score nor the absolute error score (which only depends on the mean and median of the distribution, respectively) were able to capture the variability present.

### **3.4.2. The value in adding the spatial random field to improve model predictions and the predicted uncertainty**

As shown in Figure 3.4 and Figure 3.5, the spatial pattern captured by the SPDE played an important role in estimating the posterior density of nymphs per drag and the predicted uncertainty. This study demonstrated that adding the random field improved the model prediction quality for mainland Scotland. For both Models 2 and 3, but particularly obvious in Model 3, the predictive map of the posterior density of nymphs per drag differentiates areas with high and low predictive density over the domain, without a marked increasing trend from the west to the east coast (where more data were collected), and predictions for areas without questing tick survey sites were improved through the creation of the neighbourhood structure.

To confirm the value of the random field in improving the predicted nymph density and uncertainty, besides adding the random field to Model 1 (to give Model 2), it was important to construct a new model that has the random field and a new set of covariates. This step was very important since the random field accounts not only for variation that is not explained by the covariates, but it also substitutes the effect of some covariates if they explain the same spatial structure and variation. Therefore the combination of rain in July, elevation, probability of presence of red deer, proportion of forest that is deciduous and coniferous, and latitude and longitude, resulted in a more suitable model than when using the covariates in Models 1 and 2 that were not selected in combination with the random field (LST in July, frost in September, probability

of presence of roe deer and proportion of forest that is deciduous and coniferous, latitude and longitude). The random field allowed modelling spatially of autocorrelated random effects, and model patterns in residuals that cannot be explained by the covariates. Additionally, by modelling spatial autocorrelation, models with the SPDE do not have the risk of drawing biased inference on the effects of the covariates due to unmodelled correlations (Yuan *et al.*, 2016).

When using the random field, knowledge and rigour when creating the mesh and defining the SPDE model are required, and it is advised to follow authors' information and guidelines (Lindgren, Rue and Lindström, 2011; Lindgren and Rue, 2015; Krainski *et al.*, 2017). This is important because the properties of the triangulation will affect model accuracy. Coarse meshes, where mesh vertices are too far apart, will not capture the dependency between values of the random field at adjacent mesh points. Typically, the triangulation is defined to maximise the minimum internal triangle (the Delaunay triangulation), which helps to ensure a smooth transition between small and larger triangles, and the total required number of triangles is minimised to fulfil the size and shape constraints (Lindgren, Rue and Lindström, 2011; Lindgren and Rue, 2015; Krainski *et al.*, 2017). There are also problems if the maximum edge of the triangles is longer than the estimated practical range, breaking the assumption of the SPDE approach and resulting in a wrong approximation of the Matérn covariance function (Fuglstad and Beguin, 2018). In this study, the mesh was created using the interactive app 'meshbuilder', a tool which provided assistance to develop the mesh, assess its quality, and constantly reshaped the mesh until a mesh with the triangles as regular as possible in size and shape is obtained. To decrease the variance near the boundary (control for 'boundary effects') but at the same time decrease the computational time due to the extension where no data were available, the model domain was extended far from the boundaries, with triangles with a coarser resolution. Conversely, the inner triangles had a fine resolution to enable a higher level of

detail of the analysis. However, this fine resolution inside the domain was subject to available computational resources.

### **3.4.3. The value of adding the random effects of drag and site, and the ZIP process, in improving model uncertainty**

It was hypothesised that adding the extra variation effects due to the drag and site, and the ZIP process in the predictive process, would improve the predicted uncertainty. In the dataset used, drag and site represented a high source of variability, and not accounting for these two sources of variability in the predictions would result in underestimates of the predicted uncertainty. As demonstrated in Figure 3.6, the map of uncertainty is more informative when it accounts for the extra variation effects of drag and site, and the ZIP process. Frequently, when predicting species distribution, researchers adopt models for linear and additive responses, in which the response variable increases by a constant value and the point estimate does not change. Additionally, because problems related to uncertainty are difficult to deal with, they are often ignored and there is no interest in studying the uncertainty of the predictions (Elith and Leathwick, 2009). However, in cases of non-linearity in the response variable, and particularly when interested in predicting the uncertainty, it becomes essential to add these extra variation effects as sources of variation that need to be accounted for.

### **3.4.4. Inlabru vs INLA/SPDE**

This complex model procedure was developed in the new R package *inlabru*. *Inlabru* builds on the *INLA* package, being tailored for spatial and spatial-temporal analyses (Bakka *et al.*, 2018; Bachl *et al.*, 2019). As demonstrated with this work, *inlabru* allows sophisticated spatial (and spatial-temporal) models with an easier syntax and shorter code (e.g. the code for the *inla* stack in *INLA* required to define the SPDE model is hidden in *inlabru* (Bakka *et al.*, 2018)). Additionally, different from *INLA*, *inlabru* includes the possible option to predict any function of any subset of the components of the model specification (Bakka *et al.*, 2018; Bachl *et al.*, 2019) which also besides

returning the data, adds several columns with relevant information on the predictions such as the mean, the SD, the variance, the quantiles and the density.

### **3.5. Conclusions**

The objective of this study was achieved statistically by applying a robust methodology, namely the introduction of a Gaussian Markov random field. This enabled the model to capture the existence of spatial dependency among neighbouring partitions and account for spatial variation that was not explained by the covariates. In addition, the use of a neighbourhood spatial structure improved the predictions for areas with no survey sites. The introduction of the extra variation effects of drag, site and the ZIP process in predictions captured all the sources of variation and improved the uncertainty map, making it more informative. The most suitable model with the random field component resulted in the most complete and informative predictive maps of tick posterior density and uncertainty. Accounting for the spatial autocorrelation was a necessary procedure because the underlying mechanisms that dictate the spatial distribution of questing nymphs are complex. Using proper and strictly proper scoring rules was valuable in the process of decision-making because these scoring rules allow measurement of the quality of predictions and ranking of competing models. In particular the logarithmic score allowed for careful and honest assessments, based on a fair comparison of different probabilistic predictive models and encouraging truthful report (Garthwaite, Kadane and O'Hagan, 2005; Ehm and Gneiting, 2012; Merkle and Steyvers, 2013; Held and Meyer, 2019). Additionally, this study provided an opportunity to test inlabru R package with real and complex ecological data, constituting one of the first studies to use the inlabru approach.

This is the first time that this approach has been used for tick distribution modelling. This study presents novel approaches applied to tick data, which add value in improving the spatial predictive distribution of vectors. Predictive maps with vector distribution are often used by public health agencies and



other stakeholders involved in the control and prevention of VBDs, to make decisions about public health policy. Therefore, it is essential to get accurate distribution maps and to understand the uncertainty in the predictions. Sophisticated model techniques, such as the one presented here, proved to be valuable in dealing with all sources of data variability, and in improving predicted uncertainty.

## **Chapter 4: Linking human tick bite risk with tick abundance in the environment: a survey of orienteers in Scotland**

### **4.1. Introduction**

The benefits for human health and wellbeing of outdoor recreation are well recognised (Herrington and Brussoni, 2015; Murray *et al.*, 2017; Levinger *et al.*, 2018). However, outdoor recreation increases our exposure to common vectors of diseases. Current worldwide public health concerns are dengue, chikungunya, Zika and malaria, which are carried by mosquitoes (Medlock *et al.*, 2012; Musso, Cao-Lormeau and Gubler, 2015; WHO, 2016; Zanotto and Leite, 2018); Leishmaniasis, which is spread by sandflies (Horrillo *et al.*, 2019); and LB and TBE, which are transmitted by ticks (ECDC, 2019b; PHE, 2019; Tulloch *et al.*, 2019).

For public health decision-making, it is important to know the risk of a specific hazard. At the simplest level 'risk' is defined as the probability for a certain hazard to occur combined with the exposure to that hazard, the 'hazard times exposure' (Braks *et al.*, 2016). Considering TBDs, it is critical to assess the risk of a tick bite, since a tick bite is required for the transmission of pathogenic agents to occur (Porter *et al.*, 2019). The risk of a tick bite is the product of the hazard, generally considered to be the density of infected ticks (the environmental component of the risk (Bouchard *et al.*, 2018)) and the human exposure or the likelihood of human contact with ticks (Ginsberg, 1993; Eisen and Eisen, 2016), such as the number of people entering into a tick habitat (Sprong *et al.*, 2012; Hall *et al.*, 2017). In a specific environment, the density of infected ticks can be high, but if it is an area infrequently visited by humans, the risk of tick bites occurring in this area is low (Ginsberg, 1993; Eisen and Eisen, 2016).

Data measuring risk directly are rare because it is difficult to accurately determine the rate that people acquire tick bites in a particular environment. There are few published studies on tick bite and tick encounter (including tick bites and ticks crawling on body and/or clothes) rates on humans. Those that have been reported are based on CS projects or volunteer submission of data (Faulde *et al.*, 2014; Keukeleire *et al.*, 2015; Hall *et al.*, 2017), and report the number of tick bites per person per time of exposure (approximate time of exposure). However, these studies are based on passive submission of tick bites, and do not include the submission of reports where people have not observed ticks (which are essential to accurately infer tick bite rate in the population, and its seasonal and spatial variation), or account for the time and/or distance of human exposure.

In addition, human exposure is difficult to assess, because exposure also encompasses other factors that influence the likelihood of an encounter with a tick, such as degree of contact a person has with vegetation, but these factors are poorly understood and have not been quantified. In the absence of data on risk, or exposure, many studies instead use hazard, such as the density of infected questing nymphs (such as Pepin *et al.*, 2012; Li *et al.*, 2016; Vourc'h *et al.*, 2016; Takken *et al.*, 2017), or proxies for hazard such as the abundance or presence-absence of questing ticks (such as Schwarz *et al.*, 2009; Swart *et al.*, 2014; Boehnke *et al.*, 2015; Brugger *et al.*, 2016; Bouchard *et al.*, 2018; Kjær *et al.*, 2019a, 2019b), parameters that are usually measured during questing tick surveys.

It is not clear how well these measures of the hazard, particularly questing ticks in the environment, which is the parameter often measured, reflect the actual risk of tick bite to an individual present in that area. Some studies have compared the abundance of questing ticks (from blanket dragging or flagging) with the frequency that people picked up ticks in the same area. The methodologies included quantifying the number of ticks found on leggings or long cotton trousers and protective clothing used by researchers after walking

through vegetation (per time or per distance walked) (Walker *et al.*, 2001; Faulde and Robbins, 2008; Dobson, Taylor and Randolph, 2011). However, measuring ticks on clothing is not an accurate indicator of tick bite risk, since it does not quantify the number of actual tick bites on people. Therefore, there is a need to measure actual tick bites after exposure, as only this parameter is a true measure of tick bite risk that incorporates both the hazard (ticks in the environment) and human exposure.

Accordingly, it is necessary to identify the relationship between human tick bite risk and results from questing tick surveys and quantify tick bite rate of people exposed to ticks, including reports where people did not receive tick bites. These objectives are essential to confirm the value of questing tick surveys in identifying high risk areas, and to identify risk factors for tick bites, which could help inform risk reduction strategies. The results of this study will be fundamental to link tick ecology to human risk of tick bites and can inform public health messaging to communicate risk and prevention strategies aiming to decrease the risk of tick bites and TBDs.

This study aims to address these objectives by collecting data from orienteers. Orienteering is, alongside other outdoor activities such as gardening and hiking (Mulder *et al.*, 2013), yard work, mowing, walking with or without pets (Porter *et al.*, 2019), scouting (Keukeleire *et al.*, 2015), forestry work (Rath *et al.*, 1996; Santino *et al.*, 2004; Cisak *et al.*, 2005; Kaya *et al.*, 2008), hunting (Hjetland *et al.*, 2013) and military activities (Faulde *et al.*, 2014), considered to be a high risk activity for TBDs due to the high exposure to tick bites (Fahrer *et al.*, 1991; Zhioua *et al.*, 1998). Orienteers are frequently exposed to ticks because they run through rough vegetation to find their way as quickly as possible via a series of checkpoints ('control points', each of which is marked with an orange and white kite), using a map and a compass (SOA, 2020). Orienteers often track their route through the landscape, providing information on time and route of exposure. However, although considered to be at high risk of tick bites, the tick bite rate has not been reported for orienteers.

The overarching aim of this study was to assess whether questing tick surveys reflect the probability of human tick bites. The study objectives were:

1. To assess the relationship between relative abundance of questing ticks, measured through questing tick surveys, and tick bite rate of exposed orienteers. This led to two sub-objectives:
  - 1.1. To identify the most suitable denominator for human exposure (as number of people, number of people per time of exposure, and/or per distance of exposure).
  - 1.2. To assess the relationship between tick encounter rate (defined as the rate people encounter ticks crawling on their body or clothes or/and have a tick bite) with tick bite rate and relative abundance of questing ticks.
2. To identify environmental risk factors for tick bite risk in humans. This objective tested several hypotheses: that temperature on the day, rain conditions on the day, season, land cover type, not running on paths, and start time of the running, would affect the human risk of tick bites.
3. To estimate the mean incidence of tick bites and tick encounters of orienteers.

## **4.2. Materials and Methods**

For this study, it was essential to have accurate data on tick bite and tick encounter rates in humans, for which there is currently minimal information in the existing literature. This dataset needed to include: a) information on when tick bites and tick encounters have occurred, and importantly, not occurred; b) time of actual exposure; and c) distance travelled (both as possible denominators for tick bites and tick encounters). Therefore, the study objectives were addressed by sharing with orienteers via a website, a questionnaire to collect data on their encounters with ticks (including zeros), the number of people exposed, the time spent orienteering and the route they covered. Questing tick surveys were conducted in the same geographical area and approximate time period.

#### **4.2.1. Development and testing of the questionnaire and website for data collection**

With the objective of estimating tick bite and tick encounter rates, a questionnaire to report tick records was created in a server and shared with orienteers, who reported information on their tick bites and tick encounters through a website. A website was selected rather than an app since preparatory discussions with orienteers indicated that they generally do not carry their smartphones when running, and they often use global positioning system (GPS) watches to track their route. The questionnaire was designed to collect the following information in each report (please see Appendix C: Supplementary materials for Chapter 4 (S1), which shows the screenshots of the website with the questionnaire):

- a) Event identification, so it was possible to group data per event and account for event variability. This information was collected automatically after login.
- b) The number of ticks crawling and/or attached to the body after exposure. This question was accompanied by a picture identifying the three active stages of *I. ricinus* ticks. Due to small size and the high clustering nature of larvae, orienteers were asked to report adult ticks and nymphs but not larvae (Randolph and Steele, 1985; Berglund *et al.*, 1995; Sormunen *et al.*, 2016).
- c) The number of people exposed (because orienteers can run as an individual or as a group) and the duration of the activity (in hours), as denominators of the two rates (Faulde *et al.*, 2014; Keukeleire *et al.*, 2015; Hall *et al.*, 2017).
- d) The location and the trajectory of the route, to identify the area of exposure, calculate the distance travelled (e.g. possible denominator for both rates, in km) and calculate the proportion of land cover per report (e.g. to identify land cover types associated with human exposure to ticks). Orienteers had two options in the website: to add points manually indicating their trajectory on a map (where it was possible to

zoom-in, zoom-out and pan) or to upload a GPS exchange format (.GPX) file with their route.

e) The date of activity, to account for seasonal effect. In here, participants chose a date from a calendar.

f) The start time of activity, to account for any potential effect of the time of day in the reporting rates of tick bites and tick encounters. This information was inputted manually.

g) The type of paths orienteers mostly used during the activity, since the degree of contact with vegetation can influence the likelihood of a tick bite or tick encounter. This information was provided through answering a multiple-choice question with the following categories of surfaces mainly run on: no paths; small paths; or large paths/roads.

Due to the nature of the required data and to decrease the potential for errors, the questionnaire was composed of closed questions. The questionnaire was introduced by a brief title and explanation of the purpose and value of the study (Thrusfield, 2005) and by the consent form (Appendix C: Supplementary materials for Chapter 4 (S2)). The questionnaire was designed to collect information per single run/walk, per individual or group, and per event. Before the development of the website, to visualise the sequence and structure of the future system, wireframes with the system design were created. The website was designed by the author and coded by a member of the Scotland's Rural College (SRUC) staff. The website was developed using R software (R Core Team, 2019a) and the packages shiny (Chang *et al.*, 2020), leaflet (Cheng *et al.*, 2019) and shinyjs (Attali, 2020).

The questionnaire and the website were tested, validated and piloted by the developers and work colleagues. The data provided through the website were sent to and kept in an internal database, created for the purpose, on a server at SRUC. Access to the server was limited to the PhD student and database developer and manager (SRUC staff). SRUC adheres to the requirements described in ISO/IEC 27001:2013 (Information technology - Security

techniques - Information security management systems - Requirements), and regularly undergoes internal security audits. Data on the server database were extracted and saved as comma delimited values, to then be analysed via other software. The study protocol was approved by the Human (Research) Ethical Review Committee (HERC) at the Royal (Dick) School of Veterinary Studies (RDSVS), University of Edinburgh, including research, recruitment, mitigation risk, participant consent, legal and codes of conduct, rights of human subjects and data management (HERC\_ 224\_18).

#### **4.2.2. Engagement with the orienteering community and selection of events**

A first email was sent to the Scottish Orienteering Association (SOA), and a face-to-face meeting with the events manager was arranged. Once SOA agreed to participate, individual orienteering events were identified with their assistance, and local event organisers approached. A total of 11 orienteering events at world, national, regional and local levels were attended in Scotland between August 2018 to September 2019, only covering the months of tick activity (March to October). Dissemination of the project to the SOA community was done through the SOA website and newsletter. All participants of the events attended were informed of the researcher's presence on the day through the local organisation and host team Facebook, SOA website and/or SOA newsletter.

#### **4.2.3. Data collection at orienteering events**

In an orienteering event, the competition courses generally open at 10am or 10:30am and close at 2pm, and orienteers can participate as individuals (for competition) or as groups (generally families or beginners). Orienteering events have different courses available, which vary in terms of navigational and fitness difficulty (by order of difficulty from easy to difficult: white < yellow < orange < light green < short green < green < blue < brown). In an event, orienteers select their course according to their capabilities. White courses are easy to navigate and have an average of 1 to 2 km (linear distance), green



courses are technically challenging and corresponding to an average of 4 to 5 km (linear distance), and brown courses are technically and fitness challenging, corresponding to an average of 10 to 11 km (linear distance) (SOA, 2020).

At each event, a 'tick tent' was provided with information about the project and an area to check for ticks (private cubicle, mirror and lights) (Figure 4.1). Orienteering participants who had finished competing were asked to participate and if they agreed, they were asked to check for ticks and to report into the website the number of ticks found (including zeros), either via a tablet at the time, or by visiting the website afterwards. They were given a log-in and a password specific to the event, and no personal information was collected. Participants confirmed consent when logging in to the website. In addition to the number of ticks found, participants provided the information referred in Section 4.2.1. The temperature, the rain conditions and the number of competitors in each event were recorded on the event day. The conditions of rain were classified as an ordinal variable (0, no rain and dry day; 1, showers or light rain; 2, continuous rain and wet vegetation).



**Figure 4.1:** The 'tick tent' at two orienteering events. The 'tick tent' included a 'cabin' to check for ticks which had a full body mirror. To help checking for ticks, a light was provided if required.

#### **4.2.4. Questing tick surveys: data collection**

All the orienteering events attended in this study occurred in areas that were predominantly forested/wooded (i.e. deciduous, coniferous and mixed forests) but also included open habitats with grassland, moorland and/or pastures. Before each event, the event organisers provided a map of the event area, including the control points which the runners would visit. The maps showed the different types of vegetation cover.

Questing tick surveys were conducted using the blanket drag technique in three sites across each one of the event areas (for 11 events, i.e. a total of 33 sites), with each site chosen to represent a distinct habitat. The definition of habitat was based on the standard habitat (land cover) category definition (according to EEA, 1995), and the three most contrasting habitats covered by the majority of the orienteering control points were selected. The main habitats in the areas of the orienteering events attended were:

- a) deciduous or broad-leaved forest or woodland (“vegetation formation composed principally of trees, including shrub and bush understoreys, where broad-leaved species predominate”);
- b) coniferous forest or woodland (“vegetation formation composed principally of trees, including shrub and bush understoreys, where coniferous species predominate”);
- c) mixed forest or woodland (“vegetation formation composed principally of trees, including shrub and bush understoreys, where neither broad-leaved nor coniferous species predominate”);
- d) moorland (“vegetation with low and closed cover, dominated by bushes, shrubs and herbaceous plants (heather, briars, broom, gorse, laburnum, etc.)”);
- e) natural grassland (“low productivity grassland; often situated in areas of rough, uneven ground; frequently includes rocky areas, briars and heathland”);

f) pastures (“dense grass cover, of floral composition, dominated by graminaceae, not under a rotation system; mainly for grazing, but the fodder may be harvested mechanically; includes areas with hedges”).

Questing tick surveys were conducted one or two days before the event (except for one event when this occurred four days before due to logistical constraints), with the day of sampling chosen to ensure that the weather conditions were as close as possible to the event day. However, questing tick surveys were not done during heavy rain as this adversely affects the effectiveness of the blanket drag technique (Walker *et al.*, 2001; James *et al.*, 2012). Surveys occurred between 10am and 4pm, and questing ticks were sampled using the standard technique of dragging a white wool blanket of 1 m<sup>2</sup> along the ground for 10 m. Vegetation height and density were measured three times per drag (at 0 m, 5 m and 10 m) (Gilbert, 2010; James *et al.*, 2012). Questing ticks that were touched by the blanket attached to it and when the blanket was turned over, the ticks were counted and collected live in a plastic vial containing a small piece of humid paper. This procedure was done to aid tick survival until freezing at -80 degrees Celsius (in case they are needed for pathogen diagnostics in a future study). In each of the three sites of each orienteering event, 25 drags were performed (i.e. 75 drags per event area). On wet days, multiple blankets were carried to allow regular switching if vegetation was wet.

Ticks were not identified to species level, because all previous tick studies using blanket drags in Scotland that have identified ticks to species level have found 100% of ticks collected from vegetation in Scotland to be *I. ricinus* (James *et al.*, 2012; Millins *et al.*, 2015; Hall *et al.*, 2017; Gandy, 2020). Tick life stage (adult male, adult female or nymph) was recorded. Nymphs are responsible for much higher number of bites on people than the other stages (Robertson, Gray and Stewart, 2000; Hugli *et al.*, 2009; Briciu *et al.*, 2011; James *et al.*, 2012; Faulde *et al.*, 2014; Li *et al.*, 2016; Cull *et al.*, 2019), because their small size makes them less noticed and they are approximately

ten times more abundant than adults (Gilbert *et al.*, 2017). Larvae were not collected because of the much greater degree of aggregation in larvae distribution on the vegetation when compared with nymphs and adults (Randolph and Steele, 1985; Berglund *et al.*, 1995; Sormunen *et al.*, 2016); therefore the count data distribution is zero-inflated and over-dispersed to such a large degree that the assumptions of statistical tests may be invalid (Ruiz-Fons and Gilbert, 2010). The latitude and longitude of the start point of each sampling site was recorded using GPS. Temperature and conditions of rain (using the same categories described above) were recorded on the day.

#### **4.2.5. Questing tick surveys: data management**

##### ***4.2.5.1. Calculating the proportion of land cover at sampling sites and event***

In order to determine the relationship between questing tick abundance and tick bite and tick encounter rates, it was necessary to first assess whether the field surveys performed were a good representation of the habitat where orienteers actually ran. To quantify the proportion of land cover types at the sampling sites (and to mirror the analyses of land cover in orienteers' trajectories described below in Section 4.2.6), the 'buffer' tool in ArcGIS version 10.7.1 (ESRI, 2019) was used to create circular areas of 250 m of radius around the geographic coordinates of the start location point of each site. A radius of 250 m (500 m diameter) was chosen to make sure that the area of survey was well represented. Finally, the proportion of each land cover type (deciduous, coniferous and mixed forest, moorland, grassland and pastures; from Corine land cover 2018 (EEA, 2018)), was extracted to each buffer.

##### ***4.2.5.2. Calculating the mean of questing ticks at sampling sites and event***

The mean of questing ticks (nymphs and adult ticks) and questing nymphs per drag (10 m<sup>2</sup>) was calculated for each site of collection and for each event area (11 areas). As volunteers were asked to report both nymphs and adult ticks,

the mean of questing ticks (nymphs and adult ticks) was used in the statistical analyses (section 4.2.7).

#### **4.2.6. Orienteering data: management and analysis**

##### **4.2.6.1. *Calculating the distance run***

R software (version 3.6.2 (R Core Team, 2019b)) was used to estimate the distance run by each individual or group per report. This was done using the R packages *dplyr*, *purrr* and particularly *sf* (Pebesma *et al.*, 2020), to create a geo-tibble (a modern re-imagining of a data frame with geographic data) (Wickham and Grolemund, 2017) and then the distance in km for each row (report) in the tibble was calculated. After this process, it was identified that 25.3% (86/340) of the reports marked just one point identifying the area of the competition (so the distance was zero). When documenting their trajectories on the website, orienteers were asked to mark one trajectory point per km. Therefore, it was assumed that orienteers that just marked one point only competed within that area of 1 km<sup>2</sup>, and these reports were assigned a distance of 1 km. In some reports, the distance was estimated as higher than 20 km (5.3%, 18 of the 340 records), which was unlikely to be correct in the orienteering competitions attended. These reports were further investigated using the raw tracking data, and their running distance was recalculated.

##### **4.2.6.2. *Calculating the proportion of land cover from tracking data***

In order to estimate the proportion of land cover type covered by orienteers during the competition, two approaches were used based on the distance per report. For reports with 1 km distance (25.3% of total reports), a 1 km diameter circle area from the starting point was created. This was performed using the 'buffer' tool in ArcGIS (version 10.7.1 (ESRI, 2019)). For the rest of the reports, the points of the tracking data per report were converted into line features by unifying points with the same identifier (the timestamp when each individual report was submitted), using the tool 'points to line'. Subsequently, a buffer of 5 m around each one of the lines was created, to mimic the possible contact between people and the vegetation and allow for some uncertainty in the exact

trajectory. The proportion of each one of the land cover types (deciduous, coniferous and mixed forest, moorland, grassland and pastures (EEA, 2018)) in the selected buffers was then estimated using the ArcGIS tool 'tabulate areas' (version 10.7.1 (ESRI, 2019)).

#### **4.2.7. Statistical analysis**

As a descriptive analysis preceding each statistical model, the correlation between response and explanatory variables was assessed either using Pearson or Spearman correlation tests. The selection between Pearson or Spearman tests was performed after Shapiro test to check the normality of the distribution of each variable. Linearity between the two variables and the existence of outliers also influenced the choice of method. Correlation between variables was assessed at event level (i.e. as average of each variable per event) to account for the variability and clustering at the event level. However, since report level analyses account for individual variation and so give more information on data variability, all the statistical models were performed at report level. All analyses and plots were performed in R software (version 3.6.2 (R Core Team, 2019b)).

##### ***4.2.7.1. Assessment of the relationship between questing tick surveys and human tick bite risk***

##### **Comparing habitat and weather conditions between questing tick surveys and orienteering event days**

In the first instance, as a method to check that the assessment of the relationship between questing tick surveys and tick bite risk was robust, the average proportion of land cover covered by orienteers in each event was compared with the average proportion of land cover covered during questing tick surveys in each event area. Conditions of temperature and rain between days of ecological surveys and event were also compared.

## **Selection of the appropriate denominator for tick bite and tick encounter rates**

Before assessing the relationship between results from questing tick surveys and tick bite risk, the number of tick bites submitted per report (and tick bites per person per report) was modelled as a function of the time of exposure (hours) and distance travelled (km), to identify the most appropriate relevant parameter. The same approach was conducted for the number of tick encounters submitted per report and tick encounters per person per report. Due to the zero inflation and over-dispersion of the response variables, the selected model was fitted with a negative binomial distribution. The negative binomial distribution fitted the data better, than a Poisson, ZIP or zero-inflated negative binomial distribution (the selection was based in the values of DIC, WAIC and CPO in a model just with the intercept term). When modelling the number of tick bites/tick encounters per person per report, to standardise the response variables as rates, the logarithm of the number of people per report was considered as offset. To account for the variability introduced by the event level, event was considered as a random effect. Models were developed under a Bayesian framework using INLA R package (Rue *et al.*, 2013). Models were selected based on Bayesian criterion of model goodness of fit, the DIC and the WAIC and model predicted capability, specifically the CPO score, with models with the lowest values preferred.

Model specification for the negative binomial regression is given as follows:

$$\text{Log}(TB_i) \sim \alpha + \beta_1 \text{Time} + \varepsilon + f(\text{event}_i)$$

Where  $TB_i$  is the mean number of tick bites (or tick encounters) in each event  $i$ ;  $\alpha$  is the intercept;  $\beta_1$  is the measure of Time (or distance) effect and  $f()$  denote the random effects due to event. When modelling the number of tick bites (or tick encounters) per person per report, the logarithm of the number of people per report was added as an offset ( $\varepsilon$ ).

In order to test the relationship between questing tick surveys and tick bite rate, and to identify the most relevant denominator for the rate (the one that better agreed with results from questing tick surveys), a univariable model was developed with questing ticks (sum of nymphs and adult ticks) included as a fixed effect, and different response variables relating to human exposure: a) tick bites per person; b) tick bites per person per time orienteering; and c) tick bites per person per distance run. The same models were conducted using tick encounter response variables (per person, per person per hour, and per person per distance run). The selected model was fitted with a negative binomial distribution using the same approach described previously; considering as response variables the number of tick bites or tick encounters per report, and with an offset to standardise the response variables as rates (i.e. the logarithm of the number of people per report, the number of people multiplied by the time spent orienteering per report, the number of people multiplied by the distance run per report). In these models, the environmental variables were not included because they were considered likely to drive questing ticks, and so tick exposure and risk, and because the study design ensured that weather, season and habitat were the same for both questing ticks and reported ticks. It is important to determine how relevant tick bite and tick encounter rates are to questing ticks (or vice versa), irrespective of weather, climate, habitat, or other variables.

### **The effect of the variable ‘group of orienteers’**

In the exploratory analyses, it was noted that tick encounter rate was influenced by the variable ‘group of orienteers’. This ‘group of orienteers’ variable had four categories (1 to 4), representing the events grouped by club or organisation (in one situation two neighbouring clubs regularly participating in the events of each other, hence were analysed as one group; due to the dimension, type and other characteristics, the two world events were grouped together). To further test the value of measuring tick encounters compared to tick bites as equally good measures of risk, it was therefore necessary to test whether tick bite and tick encounter rates (with the previously selected suitable



denominator) were influenced by 'group of orienteers', while statistically controlling for the abundance of questing ticks in the environment since this might reflect differences in reporting behaviour or engagement level.

To test this question, tick bite and tick encounter rates were analysed in a model with two fixed effects, questing tick abundance and 'group of orienteers'. The model approach used was the one previously described: a) the response variable was the number of tick bites/tick encounters per report; b) the offset was the denominator for the rate (per person, per person per hour, or per person per distance); c) event was considered as random effect; d) the model was fitted with a negative binomial distribution. To demonstrate the influence of 'group of orienteers', it was also necessary to characterise the relative abundance of questing ticks per drag per event in the four different groups.

Model specification (as explained in page 114):

$$\text{Log}(TB_i) \sim \alpha + \beta_1 \text{QuestingTicks} + \beta_2 \text{Group of orienteers} + \varepsilon + f(\text{event}_i)$$

#### **4.2.7.2. Determination of the environmental risk factors for tick bite risk to humans**

From the previous analyses, the most suitable response variable (i.e. tick bite/tick encounters per person, or per person per hour, or per person per distance) was selected. The correlation between environmental variables and the selected response variable was assessed at event level through correlation plots, and their relationship evaluated at report level, using the same model as previously described (fitted with a negative binomial distribution, with the corresponding offset and event as random effect).

Regarding the fixed effects included in the models, if the previous models showed that 'group of orienteers' was important, then this variable was included. The other variables included in the models were variables collected during the events which are known to affect questing tick activity: the temperature and rain on the event day; variables related with the seasonality

such as season categories: spring (March to May), summer (June to August) and autumn (September to October) or month; and habitat type (proportion of land cover covered around the track per report that was deciduous, coniferous and mixed forest, moorland, grassland and pastures). Other variables collected at report level which can affect reporting rates such as start time in each report and type of paths used were also included. Tick abundance was not included in the models because the aim was to see if other risk factors could predict tick bite/tick encounter rates, and because the environmental variables used likely affect tick activity.

Before constructing the models, a correlation matrix allowed the identification of highly correlated variables (correlation higher than 0.6), and variables correlated with each other were excluded, where the one with the strongest biological effect was preferred. First, each explanatory variable was tested on its own in a sequence of univariable models, and only significant variables or with low DIC, WAIC and CPO score were included at the multivariable stage. Backward stepwise procedure was performed to exclude variables from the multivariable models, and select the model with the lowest DIC, WAIC and CPO score.

#### ***4.2.7.3. Estimation of the mean incidence of tick bites and tick encounters of orienteers***

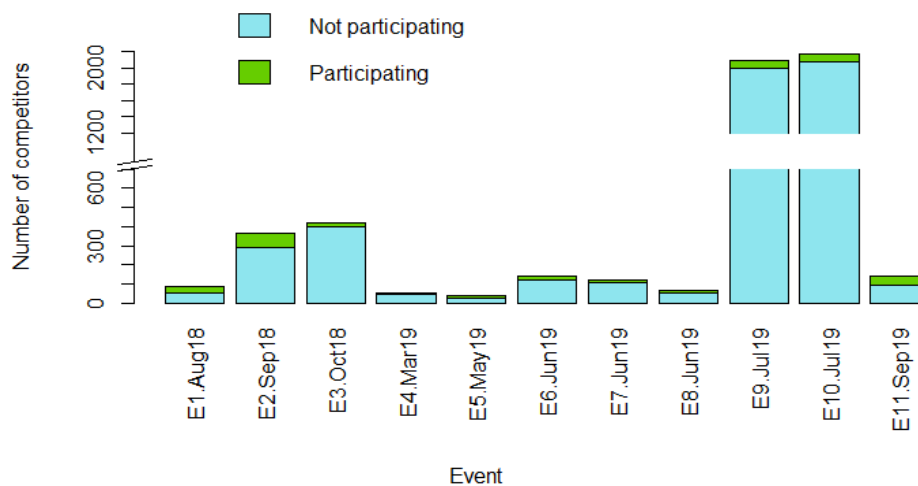
The mean incidence of tick bites and tick encounters per 1,000 person-hours (participant-hours) and the number of person-hours required for one tick bite were estimated for the 11 events, with the respective 95% confidence intervals.

### **4.3. Results**

#### **4.3.1. Descriptive analysis of data collected at orienteering events**

The 11 orienteering events attended had a median of 137 competitors (range 37 – 2,170 competitors) and in each event, a median of 20 competitors participated in the study (eight to 98 participants) (Figure 4.2). In the 11 events

attended, a total of 441 orienteers shared information on their contact with ticks, submitting 340 reports. Whilst most reports were from individual participants (81.8%), some participated as groups (18.2% of reports). The average start time of participants in the 11 events was at 11:50am, ranging from 8am to 6pm. Participants reported an average duration of running of 1 hour and 30 minutes (median of 1 hour, range of 0.5 to 4 hours), and the mean distance run was 4.2 km (median of 3.3 km, range of 1 to 20 km). Most participants reported predominantly not running on paths (85.9%), whilst a smaller number reported predominantly running on small paths (11.2%) or large paths (2.9%).

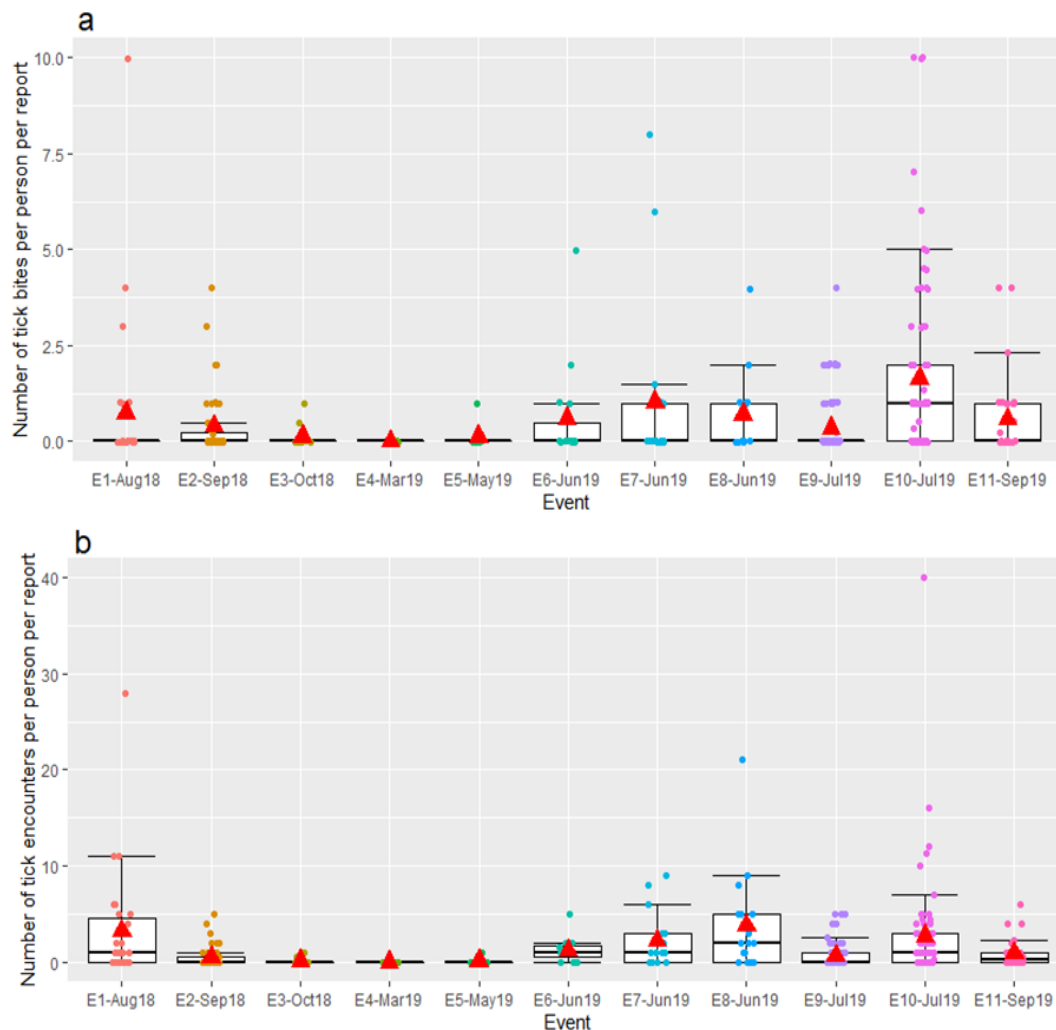


**Figure 4.2:** Distribution of the number of event competitors who did and did not participate in the study.

Event 1 (August18), event 2 (September18), event 3 (October18), event 4 (March19), event 5 (May19), event 6 (June19), event 7 (June19), event 8 (June19), event 9 (July19), event 10 (July19), event 11 (September19). This plot has a break in the y-axis because of the high number of competitors in July (events 9 and 10 were at world level) when compared with the other events.

From the total of 340 reports, 113 (33.2%) reported at least one tick bite and 162 (47.6%) reported at least one tick encounter. In total, 285 tick bites and 595 tick encounters were reported. More tick bites and tick encounters (per person) were reported in events carried out in the summer, particularly June

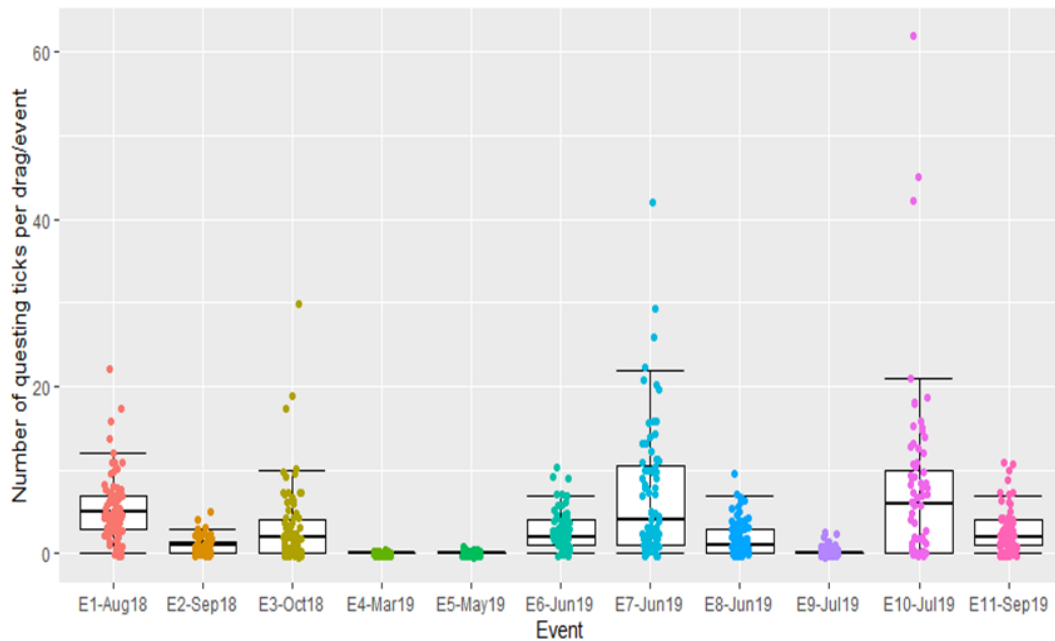
and July, visible when comparing the mean of observations (red triangles) (Figure 4.3).



**Figure 4.3:** Distribution of the number of tick bites and tick encounters, per person per report, in each event.

The distribution of the number of (a) tick bites and (b) tick encounters per person is shown as boxplots and the reports (with respective value of tick bites/tick encounters per person) are the dots, with different colour per event. The mean of tick bites and tick encounters per person per report is marked as a red triangle. The number of tick bites per person has a maximum of 10 per report, and the number of tick encounters has a maximum of 40 per report.

### 4.3.2. Descriptive analysis of the tick data collected during questing tick surveys



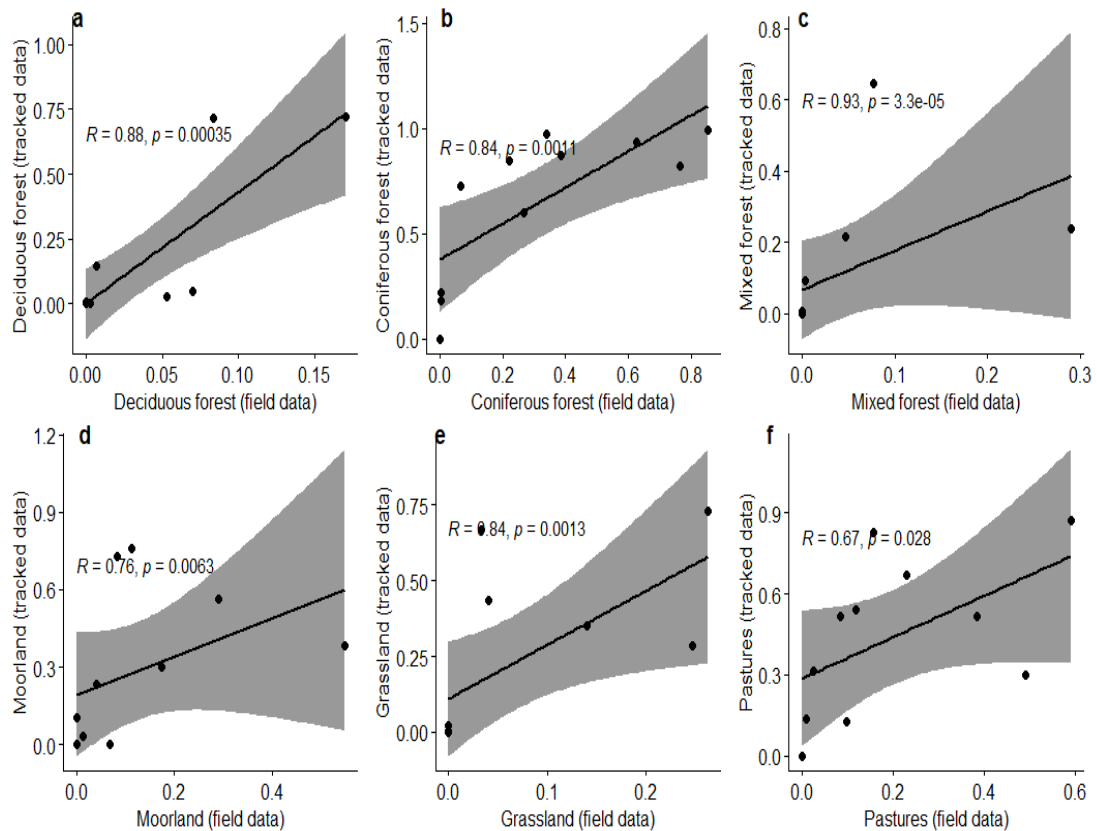
**Figure 4.4:** Distribution of the number of questing ticks per blanket drag (10m<sup>2</sup>) per event area.

The distribution of the number of ticks per blanket drag per event area (average of the three sites) is shown as boxplots. The counts of questing ticks per blanket drag per event (dots) have different colour per event.

From the questing tick surveys conducted before each event, a total of 2,379 ticks were counted from 825 blanket drag transects, and an average of 2.9 ticks were counted per blanket drag (median of 1, and range of 0 to 62 ticks per blanket drag). From the 825 drags, 92% of the total ticks counted were nymphs and only 8% were adult ticks. There was a wide range in the average ticks per drag counted per event area, with no ticks counted in the event in early March, and the highest average of 7.4 ticks per drag was collected in an event in July (Figure 4.4). Considering the three sites of collection per event area, the highest number of ticks and nymphs was always collected in a woodland site (45.4% of the 11 events in deciduous woodland, 27.3% in mixed and 27.3% in coniferous woodlands).

### 4.3.3. Relationship between questing tick surveys and tick bite risk

#### Comparing habitat and weather conditions between questing tick surveys and orienteering event days



**Figure 4.5:** Correlation coefficient (as R) and respective P-value, between the average proportion of land cover types around the tracks of orienteers and the average proportion of land cover types at the sites of questing tick surveys, in each of the 11 event areas.

The figure shows the proportion of the areas covered that is (a) deciduous forest; (b) coniferous forest; (c) mixed forest; (d) moorland; (e) grassland; (f) pastures. The black dots indicate each orienteering event, the dark line is the regression line and the grey shading represents the 95% confidence intervals of the regression line. In panels (b) and (d) the limit of the confidence interval above one (y-axis) is due to an artefact of the method used. None of the values of proportion of land cover used go above one.

The sites of questing tick surveys were a good representation of the habitats where orienteers actually ran. Figure 4.5 shows the coefficients and respective 95% confidence intervals of the strong and positive correlations between the average proportion of vegetation (land cover types) around the track of

orienteers, and the average proportion of vegetation (land cover types) on the sites of questing tick surveys, at the 11 event areas.

The temperature and humidity at the time of questing tick surveys were also compared to the event day. The differences in temperature between the period of questing tick surveys and the event day varied from zero to six degrees Celsius, with a median of one degree of variation. Only one event had substantial differences in weather when compared with the day of surveying questing ticks in the environment (from dry and sunny to rainy and cold conditions, with a drop of six degrees Celsius in the temperature).

### **Selection of the appropriate denominator for tick bite and tick encounter rates**

At report level, results demonstrate a positive and significant relationship between the number of tick bites (and tick bites per person per report) and the number of hours of exposure. The relationship between the distance travelled and the number of tick bites (and tick bites per person per report) was not significant and models presented a worse fit (Table 4.1). When performing the same analysis for the number of tick encounters (and tick encounters per person per report), similar results were found, with hours of exposure presenting a better relationship (Table 4.1).

**Table 4.1:** Results from four negative binomial models for the number of tick bites and tick encounters.

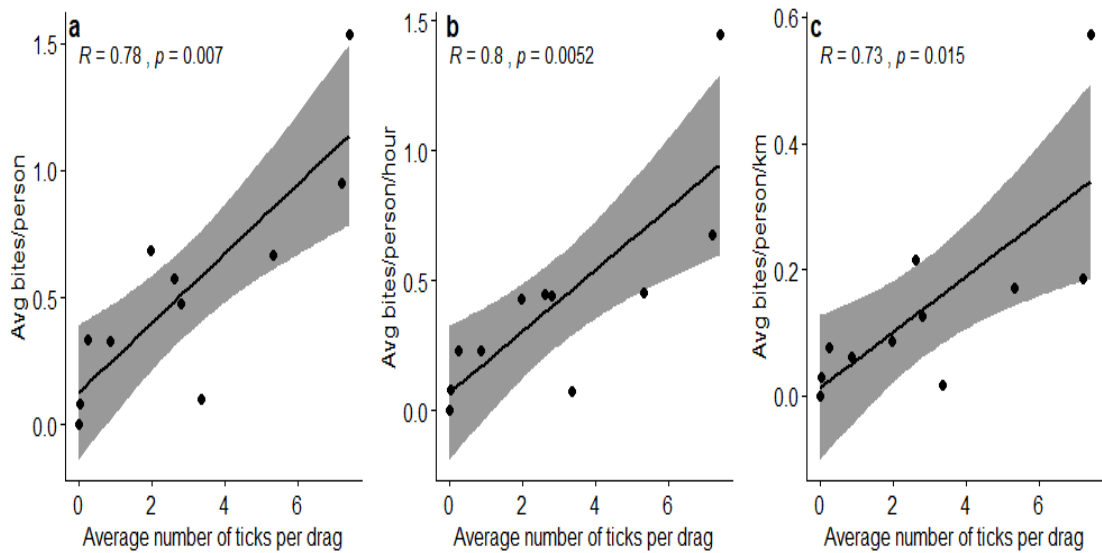
The estimates of the fixed effects are presented in the logarithm form. The number of tick bites/tick encounters per report, and tick bites/tick encounters per person per report are the response variables, and the time of exposure (number of hours) and the distance travelled (km), are the fixed effects. The offset is the number of people reporting.

Response variable	Fixed effects	Posterior mean	2.5% quantile	97.5% quantile	DIC	WAIC	CPO
Number of tick bites per report	Intercept	-0.93	-1.63	-0.29	790.8	792.4	396.5
	Time	0.34	0.01	0.70			
	Intercept	-0.61	-1.20	-0.08	793.6	795.1	397.8
	Distance	0.03	-0.03	0.10			
Number of tick bites per person report	Intercept	-1.21	-1.94	-0.53	790.7	792.5	396.5
	Time	0.41	0.07	0.78			
	Intercept	-0.90	-1.52	-0.34	792.9	794.4	397.5
	Distance	0.06	-0.007	0.12			
Number of tick encounters per report	Intercept	-0.37	-1.22	0.41	1087.4	1089.8	545.2
	Time	0.34	0.04	0.66			
	Intercept	-0.11	-0.87	0.56	1089.2	1090.9	545.8
	Distance	0.05	-0.001	0.10			
Number of tick encounters per person per report	Intercept	-0.66	-1.56	0.16	1,080.6	1,082.0	541.3
	Time	0.42	0.11	0.74			
	Intercept	-0.39	-1.19	0.34	1,082.0	1,082.5	541.5
	Distance	0.06	0.01	0.11			

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; WAIC, Watanabe-Akaike information criteria

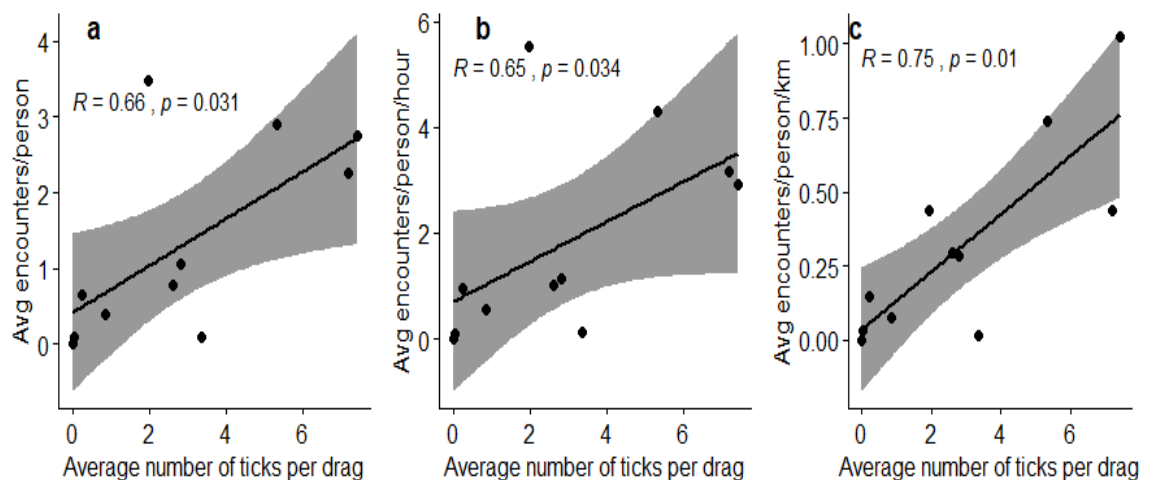
The correlation between the average number of ticks per blanket drag in each event area and tick bite rate is higher when the rate considers as the denominator the number of people reporting multiplied by the time (hours) spent orienteering (Figure 4.6). Regarding tick encounter rate, the correlation is strongest when considering the number of people multiplied by the distance run (Figure 4.7). Therefore, the descriptive analysis at event level suggests that the most relevant human exposure parameter for tick bites is the number of tick bites per person per hour of exposure, and for tick encounters is the number of tick encounters per person per distance (km) travelled.





**Figure 4.6:** Correlation coefficient (as R) and respective P-value, between the average number of ticks collected per drag ( $10\text{m}^2$ ) per event area (three sites), and tick bite rate.

The response reporting rates are: (a) tick bites per person; (b) tick bites per person per hour; and (c) tick bites per person per distance (km). The black dots are each orienteering event, the dark line is the regression line and the grey shading represents the 95% confidence intervals of the regression line.



**Figure 4.7:** Correlation coefficient (as R) and respective P-value between the average number of ticks collected per drag ( $10\text{m}^2$ ) per event area (three sites), and tick encounter rate.

The response reporting rates are: (a) tick encounters per person; (b) tick encounters per person per hour; and (c) tick encounters per person per distance (km). The black dots are each orienteering event, the dark line is the regression line and the grey shading represents the 95% confidence intervals of the regression line.

Univariable negative binomial models indicate a positive and significant relationship between questing ticks and all the reporting rates of tick bites and tick encounters. However, models present a better fit and predictive power when the denominator is (for both cases) the number of people multiplied by the hours spent orienteering (Table 4.2).

**Table 4.2:** Results from six univariable negative binomial models for tick bite and tick encounter rates.

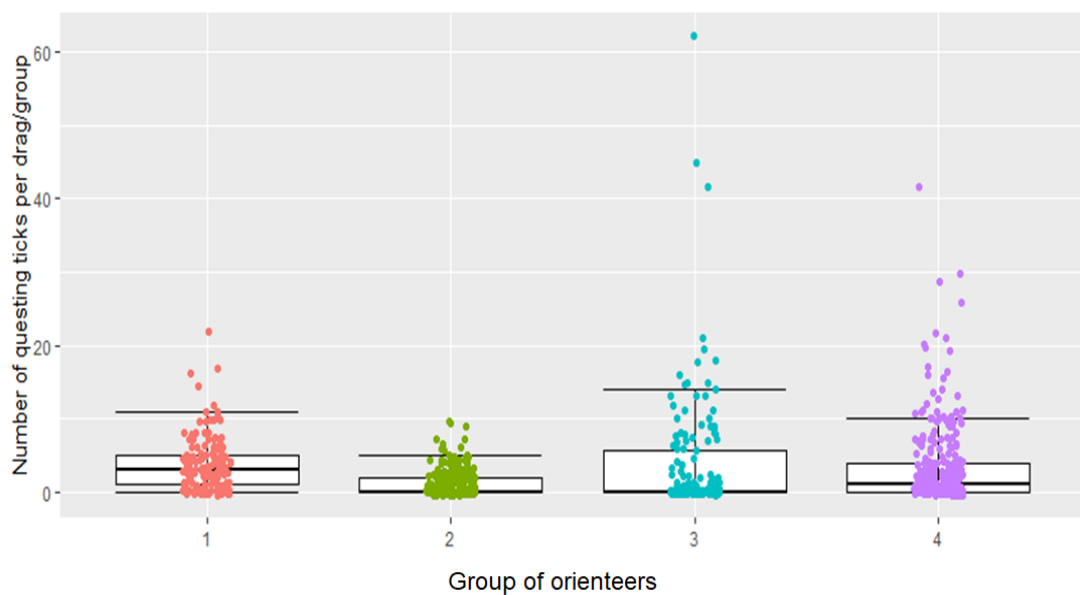
The estimates of the fixed effects are presented in the logarithm form. The response variables are the number of tick bites/encounters per person; per person per time (hour); and per person per distance (km). The explanatory variable is the average of questing ticks per blanket drag/per event.

Response variable	Fixed effects	Posterior mean	2.5% quantile	97.5% quantile	DIC	WAIC	CPO
Number of tick bites per person per report	Intercept	-1.27	-1.79	-0.79	790.6	791.6	395.9
	Questing ticks	0.22	0.10	0.33			
Number of tick bites per person per hour per report	Intercept	-1.60	-2.15	-1.08	785.3	786.4	393.4
	Questing ticks	0.24	0.11	0.36			
Number of tick bites per person per km per report	Intercept	-2.49	-3.32	-1.74	831.1	832.4	416.3
	Questing ticks	0.25	0.05	0.43			
Number of tick encounters per person per report	Intercept	-0.93	-1.99	-0.08	1,085.7	1,087.4	543.9
	Questing ticks	0.29	0.08	0.54			
Number of tick encounters per person per hour per report	Intercept	-1.16	-2.14	-0.39	1,081.0	1,082.5	541.3
	Questing ticks	0.30	0.10	0.53			
Number of tick encounters per person per km per report	Intercept	-1.98	-3.11	-1.09	1,136.6	1,138.2	569.3
	Questing ticks	0.29	0.06	0.55			

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; WAIC, Watanabe-Akaike information criteria

### The effect of 'group of orienteers'

Figure 4.8 characterises the variation of questing ticks between categories of the variable 'group of orienteers'. More questing ticks were counted in events in group 3 (average of  $3.8 \pm 8.3$  ticks per drag, range of 0 to 62, first quartile of 0 and third quartile of 6), than from events in group 1 (average of  $3.6 \pm 3.8$  ticks per drag, range of 0 to 22, first quartile of 1 and third quartile of 5), group 4 (average of  $3.3 \pm 5.4$  ticks per drag, range of 0 to 42, first quartile of 0 and third quartile of 4) and less from events from group 2 (average of  $1.2 \pm 1.9$  ticks per drag, range of 0 to 10, first quartile of 0 and third quartile of 2).



**Figure 4.8:** Distribution of the number of questing ticks per blanket drag (10m<sup>2</sup>) per 'group of orienteers'. Each group includes two or more orienteering events. The counts of questing ticks per blanket drag per group (dots) have different colour per group. The distribution of the number of questing ticks per blanket drag is shown as boxplots.

The effect of the variable 'group of orienteers' in the relationship between questing ticks and tick bite and tick encounter rates was assessed considering the rates per person per hour. According to the multivariable models with questing ticks per blanket drag and 'group of orienteers' as fixed effects, results indicated that the variable 'group of orienteers' did not influence tick bite rate,

not just because it was not significant but also because it decreased model goodness of fit and predictive power (Tables 4.2 and 4.3). In contrast, results demonstrate that the 'group of orienteers' influenced tick encounter rate (Table 4.3). Results indicate that group 1 reported more tick encounters than the other groups, with significant differences between groups 1 and 4.

**Table 4.3:** Results from the multivariable negative binomial models for tick bite and tick encounter rates (per person per hour).

The estimates of the fixed effects are presented in the logarithm form. The explanatory variables are the average of questing ticks per blanket drag per event area and 'group of orienteers'.

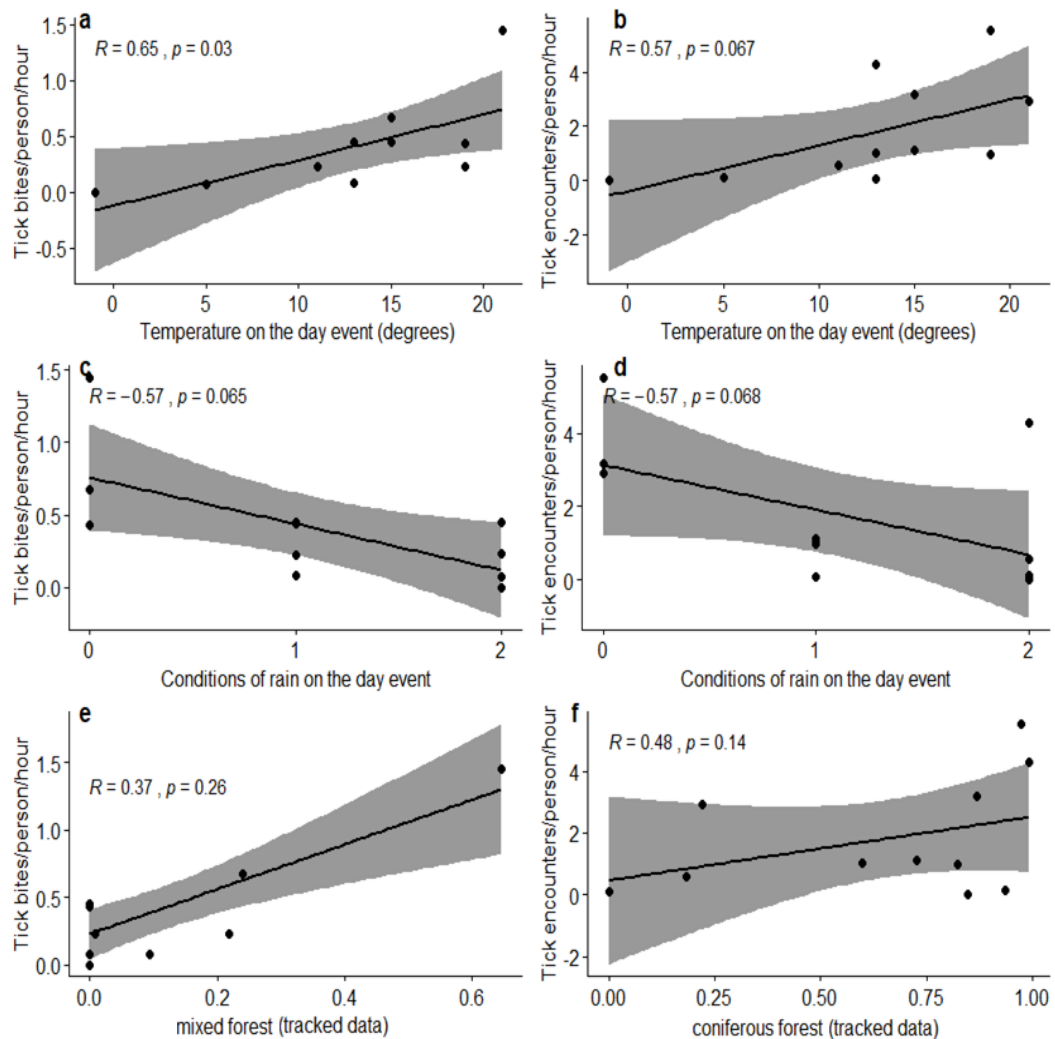
Response variable	Fixed effects	Posterior mean	2.5% quantile	97.5% quantile	DIC	WAIC	CPO
Number of tick bites per person per hour per report	Intercept	-1.73	-2.80	-0.68	787.4	789.6	395.0
	Questing ticks	0.25	0.09	0.41			
	Group of orienteers	G2 versus G1: 0.22	-1.05	1.43			
		G3 versus G1: 0.33	-0.85	1.46			
		G4 versus G1: -0.20	-1.36	0.89			
Number of tick encounters per person per hour per report	Intercept	-0.31	-1.86	-1.05	1,080.7	1,082.5	541.5
	Questing ticks	0.29	0.09	0.55			
	Group of orienteers	G2 versus G1: -0.88	-2.57	0.77			
		G3 versus G1: -0.68	-2.35	0.95			
		G4 versus G1: -1.42	-3.01	-0.03			

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; G, group of orienteers; WAIC, Watanabe-Akaike information criteria

#### 4.3.4. Environmental risk factors for human tick bite risk

For subsequent analyses, the selected response variables are tick bite and tick encounter rates per person per hour. Tick bite rate correlated positively with the temperature on the event day (correlation coefficient = 0.65,  $p = 0.03$ ), and negatively with the rain conditions on the event day (although the correlation was not significant, correlation coefficient = -0.57,  $p = 0.065$ ). Although with a positive and negative trend respectively, it was not possible to

demonstrate a significant correlation between temperature ( $p = 0.067$ ) and rain conditions on the event day ( $p = 0.068$ ) with tick encounter rate (Figure 4.9). None of the land cover variables showed a significant correlation at event level with the two rates. For tick bite rate, the better correlation was with the proportion of mixed forest around the track ran by orienteers (correlation coefficient = 0.37,  $p = 0.26$ ), and the strongest correlation with tick encounter rate was with the proportion of coniferous forest around the track ran by orienteers (correlation coefficient = 0.48,  $p = 0.14$ ). In both cases, the correlation was not significant, but the nature of the correlation was positive (Figure 4.9). At report level, there was multicollinearity between temperature and rain conditions on the event day, and between these two covariates and season (categorical variable).



**Figure 4.9:** Correlation coefficient (as R) and respective P-value between tick bite and tick encounter rates (per person per hour) and temperature, rain and proportion of forest covered.

The environmental variables for tick bite rate (per person per hour) are: (a) the temperature on the event day, (c) the conditions of rain on the event day, and (e) the proportion of mixed forest covered by orienteers. For tick encounter rate (per person per hour) are: (b) the temperature on the event day, (d) the conditions of rain on the event day, and (f) the proportion of coniferous forest covered by orienteers. The black dots are each orienteering event, the dark line is the regression line and the grey shading represents the 95% confidence intervals of the regression line.

Table 4.4 shows the results from the multivariable negative binomial models to identify risk factors for tick bite and tick encounter rates. The variable 'group of orienteers' was included in the multivariable model for environmental risk factors of tick encounters, since this variable affects tick encounter rate. Tick

bite and tick encounter rates vary negatively with the start time and the proportion of pastures around the track ran by orienteers, and positively with the temperature on the event day. Models for both rates have the same environmental covariates, except that the model for tick encounter rate included the variable 'group of orienteers'. Group 1 shows a trend of reporting more tick encounters when compared with the other three groups, and the differences are significant when compared with group 3.

**Table 4.4:** Results from the final multivariable negative binomial models for tick bite and tick encounter rates (per person per hour) and the studied environmental covariates.

The estimates of the fixed effects are presented in the logarithm form.

Response variable	Fixed effects	Posterior mean	2.5% quantile	97.5% quantile	DIC	WAIC	CPO
Tick bite rate	Intercept	-0.265	-2.906	2.336	780.7	782.8	391.6
	Start time	-0.226	-0.422	-0.031			
	Proportion of pastures	-0.848	-1.579	-0.134			
	Temperature	0.159	0.073	0.253			
Tick encounter rate	Intercept	-1.639	-5.076	1.467	1,071.2	1,072.7	536.6
	Start time	-0.184	-0.354	-0.013			
	Proportion of pastures	-0.804	-1.456	-0.166			
	Temperature	0.305	0.161	0.478			
	Group of orienteers	G2 versus G1: -0.554	-1.894	0.856			
		G3 versus G1: -1.744	-3.270	-0.340			
		G4 versus G1: -0.087	-1.378	1.424			

Abbreviations: CPO, conditional predictive ordinate; DIC, deviance information criteria; G, group of orienteers; WAIC, Watanabe-Akaike information criteria

The variable rain conditions on the event day showed a negative and significant effect on both rates. Nonetheless as it was correlated with temperature, it was later excluded, as temperature on the event day had the strongest effect and resulted in a better model fit. The same happened with season which showed, for both rates, a positive effect of the summer category when comparing with spring and autumn categories. However it was excluded from the final models because it was correlated with temperature on the event

day. Besides the negative effect of the proportion of pastures around the track run by orienteers, a significant effect for any of the other land cover types, including the proportion of mixed or coniferous forest around the track of orienteers, was not demonstrated.

#### **4.3.5. Mean incidence of tick bites and tick encounters among orienteers**

The mean number of tick bites reported varied from 0 to 1.45 (average of 0.41) per person per hour per orienteering event, and 1 bite was reported for every 4 person-hours across all events (95% confidence interval of 1-7). The mean bite rate was 409 tick bites per 1,000 person-hours of activity (95% confidence interval of 140-678). The mean number of tick encounters reported varied from 0 to 5.5 (average of 1.80) per person per hour per orienteering event, and 1 tick encounter was reported for every 3 person-hours across all events (95% confidence interval of 0.1-7). The mean tick encounter rate was 981 human-tick encounters per 1,000 person-hours (95% confidence interval of 347-1,616).

#### **4.4. Discussion**

The aim of this study was to establish the relationship between results from questing tick surveys and human tick bite risk, as this analysis is currently lacking in the research literature. To achieve this aim, the relationship between questing ticks and tick bite rate in orienteers was assessed, including defining the most appropriate denominator for tick bite rate; testing if tick encounter rate was also a good measure of tick risk to humans or whether actual tick bite is needed to accurately reflect the risk to people, and finally identifying the environmental factors for tick bite risk to humans. To estimate tick bite and tick encounter rates, this study tested a new methodology which included the report of zeros (where people were exposed but do not have a tick bite or tick encounter), time of exposure (hours) and tracking of participants. Finally, the mean incidence of tick bites and tick encounters among orienteers was estimated.



#### **4.4.1. Relationship between questing tick surveys and the actual tick bite risk to humans**

Results from questing tick surveys are frequently used to infer tick risk to humans. However, the relationship between results from questing tick surveys such as tick relative abundance in the environment, and tick bite risk to humans has not been previously directly assessed. This study indicates that, even with the uncertainty related with the blanket drag efficiency (Ruiz-Fons and Gilbert, 2010; Kjellander *et al.*, 2021) and related to volunteer submission, questing tick abundance is strongly correlated with tick bite rate, confirming the value of questing tick surveys in identifying areas for human risk to tick bites.

In this study, several measures were taken to control for other factors that might influence the relationship between results from questing tick surveys and tick bite rate: conducting questing tick surveys in the same areas as the competitors ran; sampling one to two days beforehand and attempting to keep weather conditions as similar as possible; conducting surveys during the same time of day as the events (i.e. between 10am to 4pm); and more than one blanket available for dragging (total of three blankets) to switch in case of wet vegetation. Analyses confirmed that this approach was successful in reducing potential differences in conditions between days of questing tick surveys and event. Results showed a positive and strong correlation between the land cover type in the sites where questing tick surveys were carried out, and in areas ran by orienteers, with correlations between 0.67 and 0.93 (Figure 4.5).

Differences in weather conditions and time of day of sampling influence tick microhabitat temperature and relative humidity, and therefore tick questing (Randolph *et al.*, 2002; Schulz, Mahling and Pfister, 2014; Tomkins *et al.*, 2014). However in this study, just one event day (in October 2018) showed marked differences in conditions of rain and temperature, when compared with the day of sampling (Figure 4.3 and Figure 4.4). In this case, questing tick surveys were carried out four days before the event due to logistical constraints, and the temperature unexpectedly decreased by six degrees

Celsius and the conditions changed from dry to very wet. This may have decreased the tick bite rate on the event day, and few competitors participated (Figure 4.2). In case of wet vegetation, as humidity affects blanket drag efficiency (Walker *et al.*, 2001; James *et al.*, 2012), using more than one blanket helped to reduce the effects of a wet blanket in some situations.

In this study, orienteers were asked to just report nymphs and/or adult ticks attached and/or crawling. Larvae were excluded because their report could lead to errors in larvae numbers, and larvae are thought to present negligible risk of LB transmission (Gray *et al.*, 1998; Hubálek and Halouzka, 1998). Larvae tend to be clustered in the environment and one tends to either find many or none (Randolph and Steele, 1985; Berglund *et al.*, 1995; Sormunen *et al.*, 2016), and their small size makes them difficult to detect (which would increase the number of false negatives). Another reason to exclude larvae was that larvae quest close to the ground and their numbers could be subject to bias when dragging (Gilbert, 2010; Dobson, Taylor and Randolph, 2011).

At the same time, this study confirms that the number of tick bites reported (per person) was better explained by time (hours) than by distance (km) of exposure (Table 4.1). This is corroborated by previous published research, where the number of hours spent in the yard showed a positive effect on the likelihood of a tick encounter (Mead *et al.*, 2018). In addition, when compared with other denominators for tick bite rate, the number of tick bites per person per hour showed a better correlation and relationship with the relative questing tick abundance (Figure 4.6 and Table 4.2).

This study also assessed whether tick encounter rate (which considers ticks crawling and attached) provided an equally useful measure of human tick bite risk as tick bite rate. Although crawling ticks may be detected and removed prior to attachment and so do not directly contribute to the risk of pathogen transmission (Eisen and Eisen, 2016), some studies account for them when reporting tick encounters. Correlation between questing tick abundance and

the number of tick encounters per person per hour was positive but not as strong (correlation coefficient of 0.65, Figure 4.7) when compared with the correlation between questing tick abundance and tick bite rate (correlation coefficient of 0.80, Figure 4.6). These results were perhaps surprising because it might be expected that tick encounter rate (which considers both crawling and attached ticks) would better reflect questing tick abundance in an area compared to tick bite rate, which is likely to be influenced by people's behaviour.

Results from the multivariable model with questing tick abundance and 'group of orienteers' indicated that, in contrast to tick bite rate, tick encounter rate was influenced by the variable 'group of orienteers' (Table 4.3). Previous studies indicated that data from volunteer participation often contain biases and considerable variation. Differences in volunteers' engagement (which influence volunteer effort) and in behaviour are some of the causes for the high variability in volunteer data (Bird *et al.*, 2014; Johnston *et al.*, 2017; Palmer *et al.*, 2017). Therefore, because in this study each report is related to an event and not to an individual volunteer (orienteer), and event was already considered as a random effect in the models, grouping events per club (and other similar characteristics) could indicate a trend for reporting more or less ticks due to engagement or other conditions at the time. Group behaviour could also determine whether orienteers were more likely to report at the time (more crawling ticks) or later (fewer crawling ticks and perhaps more tick bites). Results indicate that tick bite rate is a more robust and consistent measure of tick exposure and risk to humans than tick encounter rate, which is subject to bias, error and has higher variability. In general, orienteers are 'tick aware'. Crawling ticks, particularly when in high number, tend to be brushed away and not accounted. Therefore, there is a certain amount of variability in the data that can mean that not all participants reported all the ticks crawling on their body or clothing.

#### **4.4.2. Environmental risk factors for tick bite risk to humans**

Previous studies on risk factors for tick bites and tick encounters on humans analysed differences in risk between age groups, activities, habitats, time of exposure, educational level and pet ownership (Bartosik *et al.*, 2011; Hjetland *et al.*, 2013; Mulder *et al.*, 2013; Mead *et al.*, 2018; Jore *et al.*, 2020). Although it is recognised that the risk of a tick bite is shaped by environmental factors (Keukeleire *et al.*, 2015), there is still little understanding of the environmental factors for tick bite risk to humans. This study demonstrated that both tick bite and tick encounter rates (per person per hour) among orienteers decreased when the activity started later in the day, and with an increase in the proportion of pastures around the track ran by orienteers, and increased with higher temperature on the event day. Conditions of rain on the event day showed a significant negative influence on both rates, and significantly more tick bites and tick encounters were reported in summer than in autumn.

Temperature and relative humidity are critical for tick activity (Randolph *et al.*, 2002; Tomkins *et al.*, 2014). In Scotland, ticks start questing when the weekly average maximum temperature is between six to seven degrees Celsius (Gilbert, Aungier and Tomkins, 2014), and warmer temperatures are associated with an increase in tick questing. In general, high temperatures associated with low relative humidity have a negative impact on the questing of ticks due to the risks of desiccation (Schwarz *et al.*, 2009). However in Scotland, summer temperatures are not high, with an average high temperature of 18 degrees Celsius and an average low temperature of 11 degrees Celsius, and during the year the average monthly quantity of rainfall varies from 55 millimetres in spring and 70 millimetres in autumn (Scottish Government, 2020). Temperature can increase the likelihood of a tick bite (and tick encounters) by increasing the human exposure to ticks, through its influence on tick activity or by favouring outdoor activities by humans and the use of light clothing (Keukeleire *et al.*, 2015). Although relative humidity is critical for tick activity, excess rain can negatively affect the relative abundance of questing ticks (James *et al.*, 2012; Ribeiro *et al.*, 2019). Previous studies

have also identified a positive and negative association between tick bite rate and, respectively, the temperature and an excess of rain (Robertson, Gray and Stewart, 2000; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017).

Tick activity is seasonal, and it was previously demonstrated that the number of tick bites on orienteers varied with season of tick activity (Zhioua *et al.*, 1998). In this study, more tick bites and tick encounters were reported per person in events occurring in the summer (Figure 4.3), and the results from the multivariable model confirmed that summer had higher tick bite and tick encounter rates (per person per hour). This agrees with results of previous studies identifying the summer as the season with the highest number of tick bites reported, with peaks in June or July every year (Berglund *et al.*, 1995; Stjernberg and Berglund, 2002; Bartosik *et al.*, 2011; Faulde *et al.*, 2014; Cull *et al.*, 2019). The higher number of tick bites in summer in orienteers can be explained by the fact that most events in this study occurred after the peak of questing activity of *I. ricinus* in the environment (spring to earlier summer) (Gray, 1991), and that in summer, orienteers are likely to wear light clothing. Although seasonal patterns were found, not enough events were attended over a long time period to get a detailed picture of the variation in risk over the year.

Habitat type influences tick survival and abundance, and determines the chance of human-tick contact (Tack *et al.*, 2012; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017). Results identified that open habitats covered with pastures were associated with lower tick bite and tick encounter rates (Table 4.4). These results agree with knowledge that questing tick abundance is lower in open habitats than in forest areas (Lindström and Jaenson, 2003; Dobson, Taylor and Randolph, 2011; Pfäffle *et al.*, 2013) and that areas with a high proportion of arable land and pastures are associated with lower number of tick bites or crawling ticks on cotton trousers (Walker *et al.*, 2001; Keukeleire *et al.*, 2015).

It is known that deciduous, mixed and also coniferous forests support higher tick abundance (Ruiz-Fons and Gilbert, 2010; Tack *et al.*, 2012; Schulz, Mahling and Pfister, 2014) and previous research identified forest habitats as a high risk for tick bites (Robertson, Gray and Stewart, 2000; Bartosik *et al.*, 2011; Mulder *et al.*, 2013; Keukeleire *et al.*, 2015; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017). However in this study, the results of the correlation analyses and the relationship between the proportion of forest covered around the track ran by orienteers and tick bite and tick encounter rates was not significant. When comparing questing tick abundance between sites in each event, more ticks were always collected in the deciduous, mixed or coniferous forest sites (45.4% of the 11 events in deciduous woodland, 27.3% in mixed and 27.3% in coniferous woodlands), and not in the sites with moorland, grassland or pastures. The strong and positive correlation between results from questing tick surveys and tick bite rate potentially indicate that orienteers were exposed to ticks mainly when running through deciduous, mixed or coniferous forest. However because events occurred in areas that were predominantly covered by forest, most being tick-suitable areas, differences between the proportion of forest area covered by orienteers were not enough to explain the risk to tick bites. Since orienteers reported their tick bites and tick encounters after their run, it was not possible to know in which habitat they had picked up any ticks.

Orienteers that started their activity earlier in the day reported higher tick bite and tick encounter rates, when compared with later starts. This could be related with the time of day, as it is known that this influences tick activity as the temperature fluctuates during the day. However, in the events attended, orienteers ran mainly between 10am and 2pm (with an average start time of 11:50am). This relatively limited time period may not have allowed meaningful inference about the effect of the time of the day on tick bite and tick encounter rates. It is plausible that the first competitors to run through an area could have been exposed to questing ticks, removing them from the area by the time later competitors ran through. This is an interesting finding that could have

implications for human tick bite risk in other activities, but further research would be needed to confirm this.

The only behavioural determinant considered in the analyses, apart from 'group of orienteers' (which was previously discussed) was the type of path used by orienteers. It was expected that using clear footpaths would decrease tick bite and tick encounter rates (Walker *et al.*, 2001). However in this study, no effect was detected. A possible reason for this could be that most orienteers ran mainly off paths, or because even those competitors that predominantly ran on paths still did some off-road running. It could be that because orienteers often run off-road, other outdoor activities would be more appropriate to quantify the effect of the type of paths on tick bite and tick encounter rates.

#### **4.4.3. Tick bite and tick encounter rates among orienteers**

Although orienteers are frequently exposed to tick bites (Fahrer *et al.*, 1991; Zhioua *et al.*, 1998), tick bite rate has never been estimated for orienteers. This study found that orienteers are exposed to a high tick bite rate, with 1 bite for every 4 person-hours across all events, and a mean incidence of 409 tick bites per 1,000 person-hours. Previous studies have estimated the tick bite rate for other groups: a) 568 soldiers frequenting an outdoor training in Germany, with a reported mean incidence rate of 2.3 bites/1,000 person-days (Faulde *et al.*, 2014); b) 931 scouts attending summer camps in Belgium, with a reported mean incidence rate of 22.8 tick bites/1,000 person-days (Keukeleire *et al.*, 2015); c) 624 mountain marathon runners in the Highlands of Scotland in a two day competition, with a reported mean incidence rate of 677 bites/1,000 competitor-days and 1 bite every 35 competitor-hours (Hall *et al.*, 2017); and d) 235 persons living in a tick-endemic area in Sweden, with a reported mean incidence of 0.14 tick bites/10 hours outdoors (Stjernberg and Berglund, 2002). The tick bite rate of orienteers identified in this study is markedly higher than what has been previously reported amongst other comparable groups. A direct comparison with all the studies previously referenced above is difficult because the rates have been calculated in different ways.

In this study, a novel methodology to estimate tick bite rate was applied which contributed towards decreasing the possibility of errors and bias in measuring tick bite rate, by accounting for positive and negative tick encounters, considering the real time of exposure to ticks when performing a specific activity and tracking participants. To decrease the bias and errors of estimating tick bite rate based on reports of the same people for the same area and time period, in this study an effort was made to attend orienteering events held in different areas and months, therefore being subject to differences in weather and local characteristics.

The present study accounted not just for positive tick encounters, but also for negative ones. Orienteers were engaged before and during the study and were encouraged to report after checking for ticks (a tent was provided to check for ticks which also helped publicise the study). The importance of reporting even when participants did not find ticks was highlighted at every opportunity. Information was provided via the project website, during attendance at orienteer events in the 'tick tent', through SOA website and newsletter, and clubs' Facebook pages. However, there is still a risk that the data submitted is biased towards reporting when people saw ticks, hence underestimating the negative reports and overestimating the tick bite rate.

Most studies with volunteer data on tick bites or tick encounters account for only positive reports (Mulder *et al.*, 2013; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017; Cull *et al.*, 2018, 2019; Garcia-Martí *et al.*, 2018). However this limits the information available to a subset of the population that had tick bites, and therefore it cannot be used to fully assess tick bite rate and how this rate changes over time or space (Pearce and Boyce, 2006; Abdullah *et al.*, 2016; Cull *et al.*, 2019). Studies based on passive submission of tick bites also suffer from underreporting, because tick bites can be missed (Berglund *et al.*, 1995) due to difficulties in detecting feeding nymphs due to their small size, detection efficiency and individual sensitivity to tick bites (Eisen and Eisen, 2016). It is also common in studies involving reporting of tick bites to assume that all



people exposed are participating in the study (such as Faulde *et al.*, 2014; Hall *et al.*, 2017). Therefore people who do not report tick bites (such as people that do not want to participate) or if some tick reports are missing, are assumed to be negative, hence leading to underestimates of the tick bite rate. However, when a study depends on volunteer engagement, it is difficult to get a high rate of participation. This was demonstrated by Keukeleire *et al.*, (2015) and in the present study (Figure 4.2) where, although there was a large amount of effort put in to achieve a high response rate from the competitors, the mean response rate from the 11 events was only 19% (range between four and 36%).

Besides accounting for positive and negative tick bites and tick encounters, this study considered the real time of exposure to ticks and tracked participants' route. Most of the previous studies collected data on a daily basis rather than considering the hours of exposure during a specific activity, and so included as the denominator for tick bite rate, periods of the day where participants did not have contact with ticks, therefore underestimating the tick bite rate. Tracking participants' movement allowed us to establish a firm spatial link between a tick bite and the environment in which it took place.

There are other reasons that could help explain the differences in the tick bite rate reported between studies, namely differences in questing tick abundance between areas of different studies, and the higher tick bite rate for orienteers when compared with other outdoor activities. From the studies referenced above, only Hall *et al.*, (2017) conducted questing tick surveys in the environment, reporting a total of 291 adult ticks and nymphs in the three sites of collection, which was not substantially different from the average of ticks collected in the eleven event areas in this study, 216 adult ticks and nymphs. These results therefore do not explain the lower tick bite rate reported by Hall *et al.*, (2017).

Human behaviour, and the adoption or not of preventive measures (Robertson, Gray and Stewart, 2000; Bartosik *et al.*, 2011; Hjetland *et al.*, 2013; Marcu *et al.*, 2013; Mead *et al.*, 2018; Cull *et al.*, 2019) could help explain differences found in the tick bite rate between studies. Orienteers are exposed to high tick bite and tick encounter rates because they normally run through rough vegetation and generally wear short pants and sport shirts (Fahrer *et al.*, 1991) which means that their skin is exposed to tick bites. However, mountain marathon runners (Hall *et al.*, 2017) also used shorts and short-sleeved tops, exposing their skin to ticks when brushing on vegetation. In contrast, military personal (Faulde *et al.*, 2014) used tick bite preventive measures such as battle-dress uniforms and arthropod-repellent skin cream.

#### **4.4.4. Contributions and study limitations**

This study has confirmed the value of data from questing tick surveys in defining risk areas for human tick bite risk, and the definition of a rate of tick bites which accounts for human exposure and risk. To achieve this, it was necessary to engage orienteers to report positive and negative tick bites and tick encounters, to account for the real time of exposure when orienteering, and to track participants. Success was achieved by creating a website specifically for this purpose with an easy and informative interface, and by encouraging orienteers both before and during the study to report their activity even if they did not observe tick bites or tick encounters. Encouraging orienteers to check for ticks brought not just scientific but also public health value to this study.

However, this study had some limitations. One limitation was that tick reports were not confirmed by specialists, which can be common in studies with data from volunteer reporting (such as Mulder *et al.*, 2013; Keukeleire *et al.*, 2015; Mead *et al.*, 2018; Eisen and Eisen, 2021). Data from tick bites based on volunteer submissions are more reliable if an entomologist is responsible for tick identification (Eisen and Eisen, 2016), since not everyone knows how to identify a tick (Mowbray, Amlôt and Rubin, 2014). However, as LB is a problem

in Scotland (Mavin, Watson and Evans, 2015) and orienteers are well informed about ticks (through SOA advice, through our publications in the SOA website and newsletter, and through the project website which contained a photo with *I. ricinus* ticks in the three active stages), it is reasonable to assume a good level of confidence in the results. Associated with this limitation, the tick species responsible for the tick encounters with orienteers is not known. Although questing ticks collected in the environment were not speciated, *I. ricinus* is the tick species commonly found in the vegetation in Scotland (James *et al.*, 2012; Millins *et al.*, 2015; Hall *et al.*, 2017; Gandy, 2020), and the species commonly found biting humans in Europe (Stjernberg and Berglund, 2002; Hugli *et al.*, 2009; Wilhelmsson *et al.*, 2013; Faulde *et al.*, 2014), UK (Robertson, Gray and Stewart, 2000; Jameson and Medlock, 2011; Cull *et al.*, 2019) and specifically in Scotland (Hall *et al.*, 2017).

Although neither the questing ticks collected in the vegetation nor the ticks collected from participants during orienteering events were tested for human pathogens such as *B. burgdorferi* (*s.l.*), the presence of infection in the ticks by Borrelia and other pathogens is expected (James *et al.*, 2012; Millins *et al.*, 2015; Hall *et al.*, 2017). A refinement of this study would be to quantify the frequency of participant exposure to pathogens such as *B. burgdorferi* (*s.l.*), *B. miyamotoi*, *Babesia venatorum* and TBEV. For future research, the better we understand where and how people encounter questing infected ticks (particularly nymphs), the better we can design prevention and control strategies.

#### **4.5. Conclusions**

This study demonstrated that results from questing tick surveys are related with the actual tick bite risk to humans, confirming the value of questing tick surveys in identifying risk areas for ticks. The number of tick bites per person per hour (tick bite rate) was identified as the most valuable measure to infer tick risk to humans. Tick encounter rate was also calculated and showed interesting results. However, because tick encounter rate has more potential

for errors and bias and has a weaker relationship with results from questing tick surveys, tick bite rate is a more accurate measure. Results of this study using orienteers in Scotland confirmed that data from volunteer observations can be used to estimate tick bite rate, including reporting of zero values, accounting for the time (hours) covered and tracking participants. Orienteers are exposed to high tick bite and tick encounter rates. Both rates vary with environmental conditions such as weather and habitat type, and therefore identification of events likely to have higher rates of tick bite could allow for targeted measures to reduce tick bite risk.

## **Chapter 5: Use of a novel citizen science approach to assess human risk of tick bites.**

### **5.1 Introduction**

Ticks are the primary vectors of a wide range of pathogenic agents responsible for human diseases which have serious public health impacts. In Europe, *I. ricinus* is the most abundant and widespread tick species, and the main vector of zoonotic tick-borne pathogens (Gray, 1998; Estrada-Peña *et al.*, 2013; Oechslin *et al.*, 2017), the most prevalent and important of which are the complex of bacteria which cause LB (van den Wijngaard *et al.*, 2015; Stone, Tourand and Brissette, 2017; Azagi *et al.*, 2020).

To directly measure the risk of tick encounters and tick bites to humans, it is necessary to have information of where and when people are being exposed to ticks, an interaction that depends on human activity and tick ecology (Finch *et al.*, 2014; Eisen and Eisen, 2016, 2021; Garcia-Martí *et al.*, 2018; Porter *et al.*, 2019). Citizen science projects present a potentially valuable approach to tick data collection at scales and resolutions impossible to cover using the traditional survey methods (Garcia-Martí *et al.*, 2018; Nieto *et al.*, 2018; Porter *et al.*, 2019), and to gather data on human exposure to ticks and tick bites (Eisen and Eisen, 2021). It also provides opportunities for engaging with the public about TBD risk and prevention strategies (Lewis *et al.*, 2018; Sakamoto, 2018).

Data from CS projects where participants submit tick sightings (of human or animal encounters with ticks) have been used to create predictive maps of tick distribution (Koffi *et al.*, 2012; Atkinson *et al.*, 2014; Soucy *et al.*, 2018; Ribeiro *et al.*, 2019). Presence-only data from CS projects have also been combined with presence data from questing tick surveys in large composite datasets to predict tick distribution at larger scales (Estrada-Peña, Venzal and Acedo, 2006; Feria-Arroyo *et al.*, 2014; Alfredsson *et al.*, 2017; Alkiske, Peterson and

Samy, 2017; ECDC, 2019a; Ribeiro *et al.*, 2019). However, the major potential of CS is that this approach can provide valuable information on human-tick encounters, over large and fine-grained geographic areas and times (Garcia-Martí *et al.*, 2018; Nieto *et al.*, 2018; Porter *et al.*, 2019).

However, CS can present analytical challenges due to the non-structured, opportunistic nature of the data collected. The number of reports from citizen scientists can vary with the popularity of an area, local population density, degree of media attention and volunteer motivation, which leads to bias in CS datasets (Bird *et al.*, 2014; Langley, Messina and Moore, 2017; Zhang and Zhu, 2018). The most important limitations of CS are that studies usually lack absence data and information on the degree of human effort and of exposure. The lack of absences limit the type of scientific questions that can be asked and the predictive power of inference (i.e. predictive spatial and/or temporal models) (Pearce and Boyce, 2006; Bird *et al.*, 2014). Information on volunteer effort (i.e. how much a volunteer contributes to the project) can be included in analysis to decrease associated bias (e.g. Palmer *et al.*, 2017), but is rarely done. Variations in effort lead to spatial bias, with data from certain areas being constantly over or underestimated (Laaksonen *et al.*, 2017; Nieto *et al.*, 2018) and because effort may change over time, it can be difficult to distinguish seasonal patterns in human-tick encounters from those due to changes in effort (Bird *et al.*, 2014; Langley, Messina and Moore, 2017). These challenges were also identified by Ribeiro *et al.*, (2019) (Chapter 2), who demonstrated dissimilarities in the predicted distribution pattern of *I. ricinus* between such passive surveillance data and data from questing tick surveys in the environment, suggesting the potential for future CS data to consider the report of negative encounters (absence data) and to account for volunteer effort.

Whilst there are some studies on tick bites that also account for the number of people and time of exposure as denominators for the number of tick bites, they are based on passive submission of tick encounters (such as Faulde *et al.*, 2014; Hall *et al.*, 2017). Without accurate denominator information (number of

people exposed and/or time of exposure and/or distance travelled), it is difficult to accurately assess tick bite rate, which is necessary for understanding risk, as highlighted in Chapter 4. Therefore, although CS shows much promise for collection of data to measure tick bite rate and risk, there remains a key question to understand if the potential biases of CS can be overcome to generate high quality data, in particular by accounting for the report of zeros and by collecting data on exposure such as number of people, duration of activity and activity route.

In Chapter 4 it was confirmed that data from questing tick surveys closely reflect the tick bite and tick encounter rates in orienteers, when conducted in the same place and time. This demonstrated that, under ideal data gathering conditions, questing tick surveys are close indicators of true tick bite risk. However, this study used orienteers as volunteers (who often run off paths, accurately record their time and routes, and were motivated to report both positive and negative tick encounters), and researchers invested considerable resources conducting questing tick surveys at the same place and time as the orienteering events. As a further step, it is therefore important to test the relationship between questing tick surveys and tick bite and tick encounter rates within the general context of a CS project, with a wider range of outdoor recreation activities, human behaviour and environmental conditions. This analysis is important not only to validate CS data, but also to identify the components of tick bite rate when it is not possible to control for the environmental factors that affect questing tick activity and behavioural factors that affect human activity.

Therefore, the main aims of this study were to assess the potential of CS data collection to generate robust measures of tick bite rate, by using a novel approach where volunteers reported positive and negative tick encounters, as well as the time spent doing outdoor activities, and the routes taken. It would thus be possible to examine the relationship between questing tick abundance with CS data on human-tick encounters.

The specific objectives were:

1. To measure tick bite and tick encounter rates from CS and compare them with questing tick abundance in the same area, using a predictive map developed with data from questing tick surveys.
2. To identify risk factors for human tick bite and tick encounter rates.
3. To compare the relative contribution of human behavioural risk factors and questing tick abundance in explaining the variation of tick bite and tick encounter rates.

This study directly compared questing tick surveys with data on human-tick encounters in a realistic scenario when there is no control on the influence from the surrounding environment. The results of this study will improve our understanding of the drivers of tick bite risk and identify additional data which should be collected in CS projects to improve the usefulness of CS datasets.

## **5.2. Materials and Methods**

To accomplish the objectives, data on tick bite and tick encounter rates (per person per hour) from volunteers participating in a CS project, were compared with predictions of questing tick abundance derived from a statistical model. To be able to undertake this comparison, two environmentally contrasting areas in Scotland were identified, where both CS and questing tick surveys were carried out. The two areas, Cairngorms and Lochaber, are in the Highlands, the region in Scotland with highest prevalence of LB (Mavin, Watson and Evans, 2015). Risk factors for tick bite and tick encounter rates were assessed using information gathered from the CS project (type of activity, type of paths used, month or season of activity, distance, start time, year, and area).

### **5.2.1. Engagement with volunteers**

A priority and, at same time, a challenge of this project was to collect data each time volunteers carried out outdoor activities in order to generate data on when they did not encounter ticks as well as when they did. Therefore, volunteers



were recruited in two ways, via outdoor recreational and forestry organisations, and from the general public.

Outdoor recreational and forestry organisations in the two study areas were approached to participate. For those organisations which engaged with the project, face-to-face meetings with the groups were arranged for preliminary discussions, and to identify the best tools to record data, followed by workshops to demonstrate the data collection tools. Posters were distributed to create awareness about the project and engage other volunteers. A major effort was undertaken to recruit from outdoor recreational organisations that were willing to report regularly, to try to reduce the risk of reporting bias towards positive reports.

Members of the general public were also able to register to report data on their outdoor activities. Most of the volunteers came across the project via engagement with outdoor recreational organisations that were participating, or by the engagement with the orienteering community (Chapter 4). However, some users approached the project independently (for example after conferences, University presentations, or by word of mouth). Although engagement was only promoted in the two study areas, reports could be submitted from anywhere in Scotland.

### **5.2.2. The citizen science project 'TickApp'**

A CS project called 'TickApp' was developed and fully implemented in Scotland between May and November 2018 and March and November 2019. The project was closed during the winter months (December to February), when ticks are less active, since there was concern that volunteers may be less motivated to report routinely if few or no tick encounters occurred.

The questionnaire that was previously developed and used to collect tick encounter data from orienteers (Chapter 4) was adopted in this study. The questionnaire used was described fully in Chapter 4, Section 4.2.1. Briefly, the

information requested included: user identification (ID); date of activity; start time of activity; the number of ticks crawling and/or tick bites; the number of people involved in the same activity at the same time (individual or group); the type of path used; the type of activity and the route taken. As for Chapter 4, both the number of tick bites and the number of tick encounters (bites plus crawling ticks on the body or clothes) were collected. For this project, an ID was set up for each individual, family or group of volunteers who wished to participate (i.e. in a group, the group leader reported one activity performed in the same time and space by N people). The meetings with the outdoor recreational and forestry organisations who volunteered to participate helped to identify the categories of activities to be included in the drop-box menu of the questionnaire: orienteering, forestry work, running, walking, biking, camping, climbing and canoeing/kayaking, and other. If the user selected 'other' activity, another free-text box appeared so the user could type the specific activity.

The same questionnaire was shared with volunteers via an app and a website. Although there has been a recent rise in the popularity of apps for smartphones, some volunteers expressed a preference for a website-based questionnaire, because when people are working outdoors they do not always take their smartphones, they often want to maintain smartphone battery for safety reasons, and people often find tick bites one or two days after the activity. The website was described in Chapter 4 (Section 4.2.1). When using the website, interested participants had to contact the project team (at [tick.project@sruc.ac.uk](mailto:tick.project@sruc.ac.uk)) to request a username and password to be able to participate in the project.

The app was designed by the author and coded by a freelancer, and later refined by a member of SRUC staff. The app was designed for Android (as many outdoor institutions used Android phones). The user identification for using the app was the Google email of each user; this was the only personal information collected. When using the app information regarding the date, start

time, period of activity and the route taken was automatically recorded by the app. After the initial log in, users were asked to start the activity and, if affirmative, the app started recording the time and GPS location every minute (Appendix D: Supplementary materials for Chapter 5 (S1), with information on the software used to develop the app, and app screenshots).

When logging in to the website, users could access the consent form with the terms and conditions of use (Appendix C: Supplementary materials for Chapter 4 (S2)). If using the app, users were required to confirm consent via a statement provided at the first log in (Appendix D: Supplementary materials for Chapter 5 (S2)). In both the app and website, volunteers were given the option to send photos of the ticks seen to the project team (tick.project@sruc.ac.uk) as a way to help data validation.

The app and website were initially tested by the developers and work colleagues and, at the final stage, also piloted by project volunteers (staff of outdoor recreational organisations). To improve the quality of the data recorded: a) workshops were performed before the beginning of the study with some volunteers to increase engagement, ensure that users were comfortable using the website or app, and ensure the level of information provided via the website and app were appropriate; b) all the data were checked on submission for errors and inconsistencies. For the analysis, participants were anonymised and represented by a code only. The study protocol also received ethics approval, RDSVS HERC\_224\_18 (see Chapter 4, Section 4.2.1).

#### **5.2.2.1. Citizen science data management**

Citizen science data submitted by volunteers, using the app and the website, were stored in the internal SRUC database (as comma delimited values files). All files were opened in R by running a script to process all submissions, using the R packages tidyverse (Wickham, 2020) and sf (Pebesma *et al.*, 2020). One report sent in by volunteers contained information related to one activity performed in a single day per volunteer (individual or group). The first

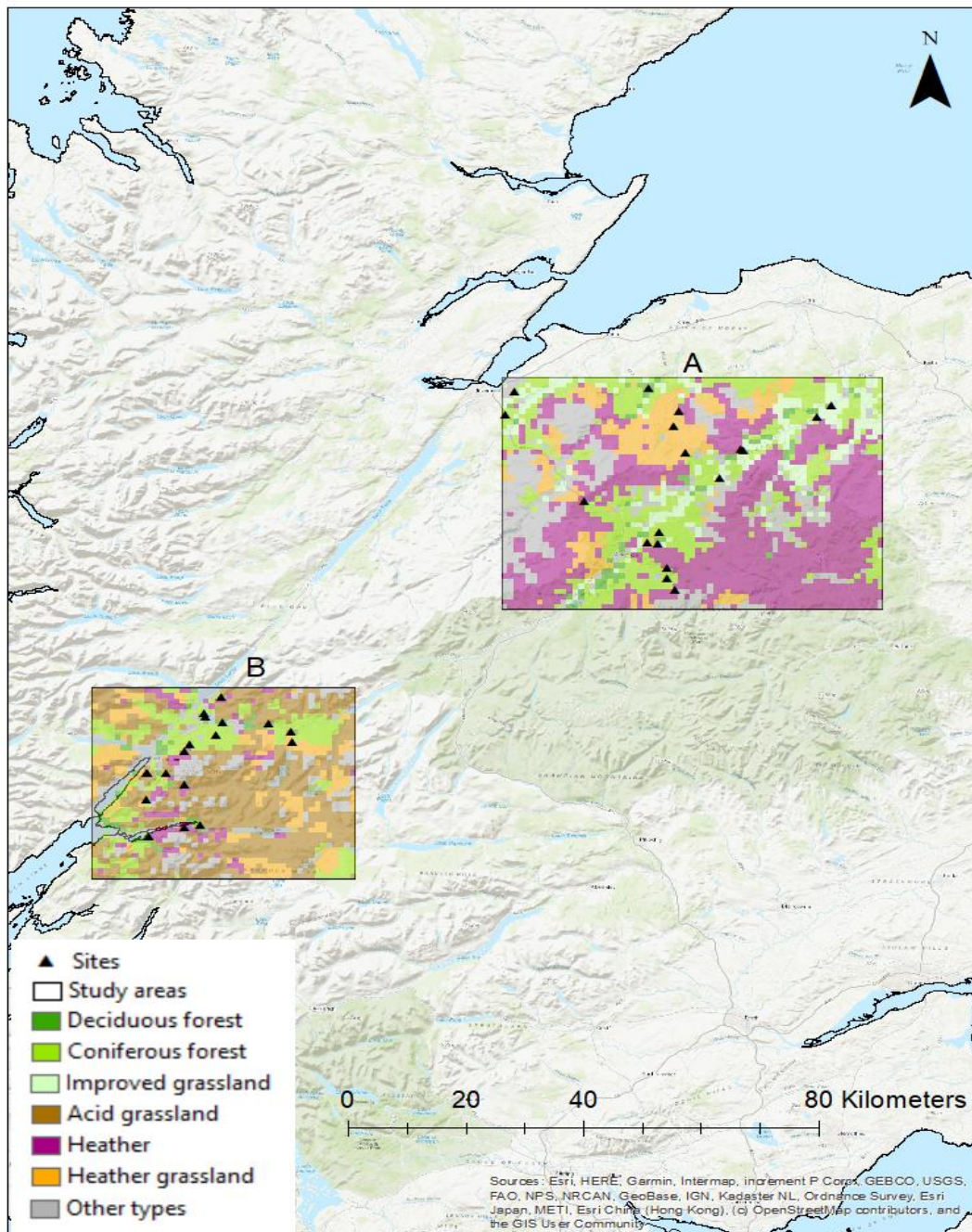
step was to define the latitude and longitude for each report, and to extract the files with missing location references. After this, the distance covered in each report was calculated using geosphere package for R, specifically the function 'distGeo()/1,000', which calculates the distance between points (and considers that the points are on a sphere and not on a flat surface (Hijmans, 2019; Hijmans, Williams and Vennes, 2019)).

The data were cleaned to remove a small number of reports from the testing and pilot phase before the project started fully in May 2018, and from the winter months when the project was not active. Afterwards all the points were transformed into a shapefile of spatial points, and a shapefile of Scotland's boundaries was used to clip the points, just keeping the points which corresponded to activities carried out in Scotland. This step was done using the R packages sp (Pebesma and Bivand, 2005), rgeos (Bivand, Rundel, *et al.*, 2020) and rgdal (Bivand, Keitt, *et al.*, 2020). The same process was repeated to extract the CS data reported in the two study areas, using the corresponding shapefiles of each area (see next Section 5.2.3 for the definition of the two study areas and Figure 5.1).

For each report, the tick bite and tick encounter rates were calculated as the number of tick bites or tick encounters divided by the number of people included in each report multiplied by the time spent doing the activity. Tick bite rate is directly related to the potential risk of TBDs. In addition, in Chapter 4, a stronger correlation was found between data from questing tick surveys and tick bite rate, than with tick encounter rate. However, as tick encounters (crawling and attached ticks) happen more frequently, accounting for tick encounters might help keeping people more engaged to the project. In addition, accounting for tick encounters might give more power to the statistical analyses.

### **5.2.3. Definition of the two study areas**

To compare human-tick encounter data with data from questing tick surveys, two independent study areas were selected. These two areas were selected to have contrasting environments within the Scottish context, and to increase the applicability and relevance of these research findings to a wider range of areas. Additionally, these areas are often used for outdoor recreational activities and are accessible from Inverness (where the PhD student was based). The areas chosen were northern Cairngorms (about 63 km south-east of Inverness) and Lochaber (about 98 km south-west of Inverness on the west coast). The west coast has a milder climate with more rain, whilst the Cairngorms has higher altitude, being colder and drier (Met Office, 2016). Polygons were defined in ArcGIS 10.2.2 (ESRI, 2012); the north Cairngorms polygon included part of Inverness, Nairn and Speyside and covered an area of 2,911 km<sup>2</sup>, and the Lochaber polygon covered part of Glen Gloy, Fort William and Kinlochleven with an area of 1,667 km<sup>2</sup> (Figure 5.1).



**Figure 5.1:** The two study areas where questing tick surveys were carried out and volunteers reported to the CS project; ‘A’ the Cairngorms, and ‘B’ Lochaber.

The black triangles represent the sites where questing tick surveys were carried out. In each area, a layer with the dominant target land cover types used to stratify the sites is overlaid. A layer with the topography of Scotland was used as a base map.

### **5.2.3.1. Identification of habitats for questing tick surveys in the two study areas**

Questing tick surveys were carried out in selected habitats in the two study areas, to provide data for modelling and predicting questing tick abundance in each area. The sampling sites were stratified by the main land cover types to be representative of each area and where people go, and to cover the range of variables that affect tick abundance. Since elevation can be an influential source of variation of questing tick abundance (Gilbert, 2010), within each area, sites for questing tick surveys were selected under 500 m of elevation.

For the stratification, the land cover data from land cover map 2015 was used (Rowland *et al.*, 2017b, 2017a). Using ArcGIS version 10.2.2 (ESRI, 2012), a layer with the dominant target land cover types was overlaid and extracted in the polygons of the two study areas. Then, the number of pixels of each land cover class in the two polygons was counted. In each area, the five most highly represented land cover types (with higher proportion of pixels) were selected. In the Cairngorms area, the dominant land cover types were classified as heather, coniferous forest, deciduous forest, heather grassland and improved grassland. In Lochaber area, the dominant land cover types were the same as for Cairngorms except that the grassland is classified as acid instead of improved (CEH, 2017). Questing tick surveys were carried out in 36 sites (18 in each area), distributed through the five dominant land cover types (Figure 5.1). In each one of the five land cover types, three to five sites were identified for questing tick surveys based on: a) being an area frequently visited by people; b) being accessible to make surveys practicable; c) permission for access being granted by the landowner or manager. Questing tick surveys were carried out during 2019.

Since tick activity varies seasonally, three sites in each area were visited every month between March to November. The remaining 15 sites were visited three times during that period: March/May, June/August and September/November.

### **5.2.3.2. Data collection during questing tick surveys**

Questing tick surveys were conducted between 9am and 4pm using the standard technique of dragging a white blanket of 1 m<sup>2</sup> across the ground vegetation area for 10 m (Figure 5.2) (Gilbert, 2010; James *et al.*, 2012). In each site and visit, 25 drags were performed. Tick life stage (adult male, adult female or nymph) was recorded and nymphs and adult ticks were collected in a numbered plastic vial (Figure 5.2). Ticks were kept in the laboratory freezer at -80 degrees Celsius. Ticks were not identified to species level, however previous studies in Scotland have confirmed 100% of the ticks collected during questing tick surveys to be *I. ricinus* (James *et al.*, 2012; Millins *et al.*, 2015; Hall *et al.*, 2017). Larval numbers were not recorded because of the much greater degree of aggregation in larvae distribution, and because volunteers were not asked to report encounters with larvae. Ground vegetation height and density was recorded using a sward stick at the beginning, middle and end of each blanket drag (Figure 5.2), because vegetation height and density can affect the efficiency of the blanket drag technique (Ruiz-Fons and Gilbert, 2010). The latitude and longitude of the start point of each site was recorded using GPS. Temperature and conditions of rain (using the same categories described in Chapter 4, Section 4.2.3) were recorded on the day (Figure 5.2).





**Figure 5.2:** Practices during questing tick surveys.

Panels (a) to (f) identify some of the steps during questing tick surveys; (a) first notes before dragging (i.e. habitat type, geographic coordinates of the initial location, time, temperature and rain conditions); (b) and (c) dragging and measuring vegetation height and density in two distinct tick habitats; (d) inspection of the blanket; (e) identification of ticks; (f) placement of the ticks to a plastic tub with damp paper.

#### **5.2.4. Statistical analysis**

Although the CS project received records of human encounters with ticks from all over Scotland, to address the specific aims of this study, only the data reported in the two study areas was used. Analyses were performed using R software (version 3.6.2 (R Core Team, 2019b)).

##### **5.2.4.1. Descriptive analysis**

Descriptive analyses of the CS data were conducted, including the risk factors for tick bite and tick encounter rates (specifically activity type, type of path, seasonality and month), and of the questing tick data that was then used to develop predictive models and maps of questing tick abundance. The Kruskal-Wallis test was used to test for differences between mean tick bite and tick encounter rates (regarding the type of activity and path, and season). In order to identify months with high rates, tick bite and tick encounter rates were plotted per month. The descriptive analysis of questing ticks also includes the variation of questing tick abundance per month. Here, only data from sites where questing tick surveys were carried out every month were used. Descriptive analysis was performed using R packages, tidyverse (Wickham, 2019), dplyr (Wickham, François, *et al.*, 2020), ggplot2 (Wickham, Chang, *et al.*, 2020) and inlabru (Bachl *et al.*, 2019).

##### **5.2.4.2. Relationship between predicted questing tick abundance and human tick bite and tick encounter rates in the Cairngorms and Lochaber**

###### **Mapping tick bite and tick encounter rates in the Cairngorms and Lochaber**

ArcGIS version 10.7.1 (ESRI, 2019) was used to create a grid of pixels with the same extent and projection of the shapefile with the boundaries of the two study areas. A grid was created for each area with 1 km<sup>2</sup> resolution per pixel, due to the availability of covariates for modelling questing tick abundance at the same scale, and because the questionnaire in the website also indicated that volunteers should draw their route using points separated by 1 km.

The CS data reported in the two study areas were uploaded into ArcGIS and transformed into a shapefile of points (representing the trajectories of volunteers) with the same projection of the two grids. A 'spatial join' added to each pixel of each grid the points that intersected (the pixels were defined as the target attributes and the spatial points as the join attributes). Due to the presence of duplicate numbers of reports in each pixel, the output table was exported and analysed in R to clean all the duplicate information in each target pixel. Finally, using that information, the number of people carrying out activities, tick bite and tick encounter rates per pixel, were calculated for each area. Results were plotted with the corresponding grid in ArcGIS. For mapping purposes, classes of data were defined by quantile due to the skewed distribution of both rates.

## Predictive model and map of questing tick abundance in the Cairngorms and Lochaber

### Georeferenced environmental data and variable selection

**Table 5.1:** Georeferenced environmental variables used in the study.

Variables		Resolution
ERA5 variables	Monthly precipitation data (2014-2018)	rescaled to 1 km <sup>2</sup>
MODIS variables	Frost days by month calculated from the MOD11A1 (2014-2018)	rescaled to 1 km <sup>2</sup>
	Synoptic monthly day LST calculated from the MOD11A2 (2014-2018)	rescaled to 1 km <sup>2</sup>
	Synoptic monthly NDVI calculated from the MOD13C1 (2014-2018)	rescaled to 1 km <sup>2</sup>
	Evapotranspiration (provides information for water resource management in the land surface such as soil water status (NASA, 2016) by month calculated from the MOD16A2 (2014-2018)	rescaled to 1 km <sup>2</sup>
	Elevation	rescaled to 1 km <sup>2</sup>
Host related variables	Index of presence of roe (Alexander <i>et al.</i> , 2014)	0.008333 * 0.008333 degrees (~1 km <sup>2</sup> )
	Index of presence of red deer (Wint <i>et al.</i> , 2014)	0.008333 * 0.008333 degrees (~1 km <sup>2</sup> )
Land cover variables	Proportion of land cover that was coniferous, deciduous, heather, heather grassland, improved grassland, and acid grassland (Rowland <i>et al.</i> , 2017b)	1 km <sup>2</sup>

In order to develop a predictive map of tick abundance in the two study areas, ecologically relevant climatic, topographic, land cover and host-related variables for tick distribution were selected. These variables were available as GIS-based raster maps (Table 5.1). All variables had a final spatial resolution of 1 km<sup>2</sup> per pixel. For data extraction compatibility and modelling purposes, all variables were converted to the same extent and projection as the shapefiles of the two study areas.

Environmental data were extracted for each of the sites of questing tick surveys using the tool 'extract multiple values at points' from ArcGIS version 10.7.1 (ESRI, 2019). Before model implementation, a correlation analysis was performed to exclude variables highly correlated to each other (correlation coefficient higher than 0.6), where the one with less biological relevance was dropped.

#### Statistical models and predictive maps for questing tick abundance in the Cairngorms and Lochaber

Preliminary analysis of the CS data in the Cairngorms and Lochaber indicated that there were not enough human-tick encounter records submitted each month to map tick bite and tick encounter rates in space and time. Therefore, to assess the relationship between questing tick abundance and tick bite and tick encounter rates, spatial models of the average of questing tick abundance were developed.

To model and predict questing tick abundance, the average number of ticks (adult ticks and nymphs) collected per drag per visit at each site, was used as the response variable. Each model (one for Cairngorms, one for Lochaber) included site as a random effect to account for between site variation, and was fitted as a Gaussian process using the INLA package in R (Rue *et al.*, 2013). Before multivariable models, univariable analysis was conducted to select the most relevant covariates as model candidates. Models were evaluated using the DIC and WAIC, as measures for goodness of fit and a parameter from the

cross-validation leave-one-out, namely the logarithmic CPO as a measure for the predictive quality of the model (Gelman, Hwang and Vehtari, 2013). A backward stepwise procedure was used to select the models with the best fit and predictive power, based on the lowest values of DIC, WAIC and CPO, amongst competing models with various covariate combinations. If two covariates interacted in the model, they were removed in stepwise. However, if the model goodness of fit and/or predictive power decreased, and if the interaction was biologically plausible, they were added in the model as an interaction.

Model specification:

$$\mu_i \sim \alpha + \beta_1 \text{Deciduous} + \beta_2 \text{NDVI} + \beta_3 \text{LST:Rain} + \dots + \beta_k \text{Lat:Long} + f(\text{site}_i)$$

Where  $\mu_i$  is the average of tick abundance per site  $i$  per visit (the average of the 25 drags in each visit);  $\alpha$  is the intercept;  $\beta_k$  are the measure of covariate effects; and  $f()$  denote the random effects due to site.

The model posterior means were used to produce the predictive maps of questing tick abundance in the two study areas, and the difference between the 97.5% and 2.5% quantiles of the predicted values were used to create uncertainty maps. The resolution of all maps was 1 km<sup>2</sup> per pixel. Maps were plotted using ggplot2 (Wickham, Chang, *et al.*, 2020) and inlabru (Bachl *et al.*, 2019). In order to validate the quality of the predictive maps, Spearman correlation test was used to assess the correlation between observed and predicted values.

### **Correlation between predicted tick abundance from questing tick surveys and tick bite and tick encounter rates in Cairngorms and Lochaber**

Spearman test was used to assess the correlation between tick bite and tick encounter rates and the predicted questing tick abundance in each 1 km<sup>2</sup> pixel where volunteers carried out activities. Values of both variables were

transformed to  $\log_{10}(y+1)$ . The correlation findings were corroborated with results from univariable models fitted as Gaussian distribution, where each one of the rates was modelled as function of the tick abundance in each area.

#### **5.2.4.3. Identification of risk factors for tick bite and tick encounter rates**

Two statistical models were developed in order to identify risk factors for human tick bite and tick encounter rates. Additionally, the relative contribution of those factors for explaining the variation of both rates was quantified. To account for the excess of zeros, the selected models were fitted with a negative binomial distribution, using as the response variable the number of tick bites or tick encounters submitted per report. The logarithmic of the number of people (included in each report) multiplied by the time spent in the activity was included as an offset (to standardise the number of tick bites and tick encounters as a rate), and the volunteer ID was set as a random effect (to account for variation due to volunteer effort, as the number of times each volunteer participated). The models were developed using INLA package in R.

Model specification (the model is very similar to the models developed in Chapter 4):

$$\text{Log}(TB_i) \sim \alpha + \beta_1 \text{QuestingTicks} + \beta_2 \text{Activity} + \dots \beta_k \text{Paths} + \varepsilon + f(\text{volunteerID}_i)$$

Where  $TB_i$  is the mean number of tick bites (or tick encounters) submitted per report per volunteer  $i$ ;  $\alpha$  is the intercept;  $\beta_k$  is the measure of covariate effects;  $\varepsilon$  is the offset and  $f()$  denote the random effects due to volunteer ID.

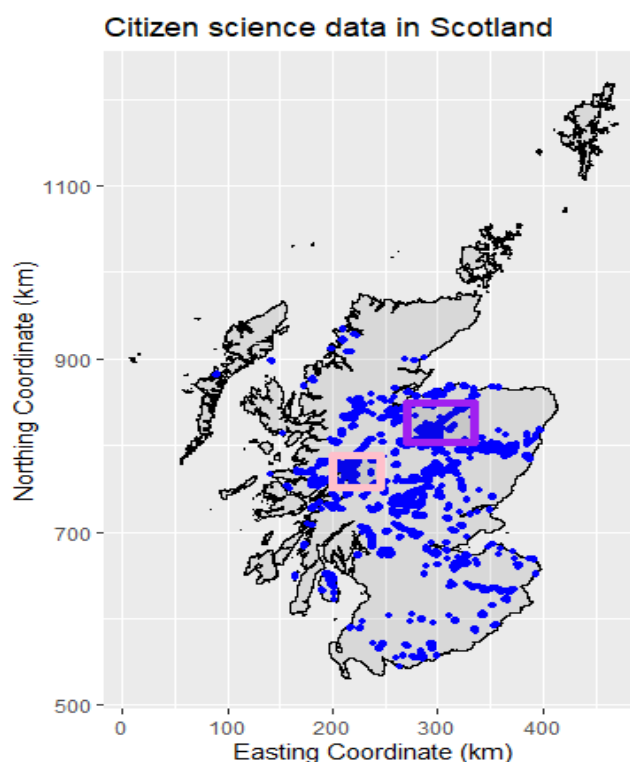
The covariates (fixed effects) included in the analysis were: the type of activity performed (biking, camping, canoeing/kayaking, climbing, orienteering, running, walking, other), the type of path mostly used (no paths, small paths or large paths), the month of activity (March to November) or instead, the season (spring from March to May, summer from June to August, and autumn from September to November), the start time of the activity, the distance

covered, the year (2018 or 2019), the area (Lochaber or Cairngorms, to account for intrinsic differences between them), and the predicted questing tick abundance in the area. It is known that these covariates are related (i.e. certain activities such as orienteering are often carried out off paths, and certain months/seasons of the year have more risk of tick encounters than others). However, to quantify the relative contribution of each covariate in tick bite and tick encounter rates, all these covariates entered in the analysis, and if suitable, in the multivariable model.

Univariable analysis was used to identify the most relevant set of covariates for the multivariable models, based on DIC, WAIC and CPO. The two best models were selected following the approach previously described in Section 5.2.4.2. After selecting the best model for tick bite and tick encounter rates, the contribution of each covariate to explain the likelihood of tick bites and tick encounters in the selected models was assessed using the pseudo- $R^2$  calculated with McFadden's approximation (the higher the value, more variance is explained).

## 5.3. Results

### 5.3.1. Descriptive analysis of the CS data

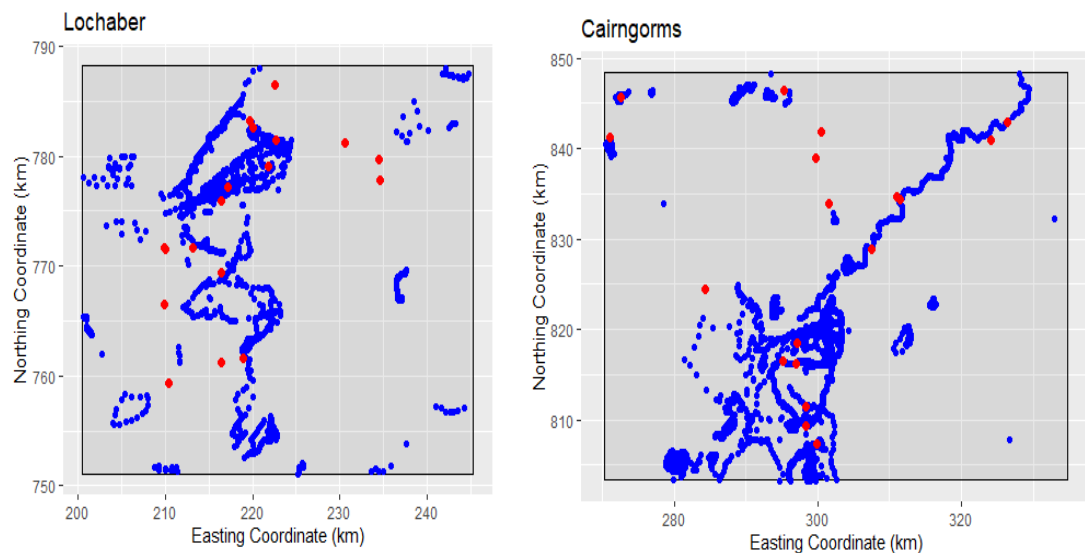


**Figure 5.3:** Citizen science data submitted to the project 'TickApp' in Scotland, between May and November 2018 and March and November 2019. The blue dots (total of 143,820 points) are the trajectories of volunteers. The two boxes denote the study areas in Cairngorms (purple) and Lochaber (pink). Data was submitted using the website and the app.

The 'TickApp' CS project ran between May and November 2018 and March and November 2019. However, in Lochaber, data was received between June and October 2018, and March and October 2019. Volunteers submitted reports mostly using the website (99.6% of reports) rather than the app. Sixty-five volunteers submitted 1,914 reports from all over Scotland (Figure 5.3). A total of 231 reports were submitted in the Cairngorms and 118 in Lochaber. There was a better spatial overlay between volunteers' reports and the sites where



questing tick surveys were carried out in the Cairngorms than in Lochaber (Figure 5.4). From this point forward, all the analysis will focus on the reports submitted in the two study areas, since only these were used to answer the scientific questions. Descriptive analysis for the reports submitted in Scotland can be seen in Appendix D: Supplementary materials for Chapter 5 (S3).



**Figure 5.4:** Citizen science data submitted to the project ‘TickApp’ from the two study areas, between May and November 2018 and March and November 2019.

The blue dots are the trajectories of volunteers. Lochaber plot has 2,782 points and Cairngorms has 42,685 points. The red dots in each area are the locations of the sites where questing tick surveys were carried out.

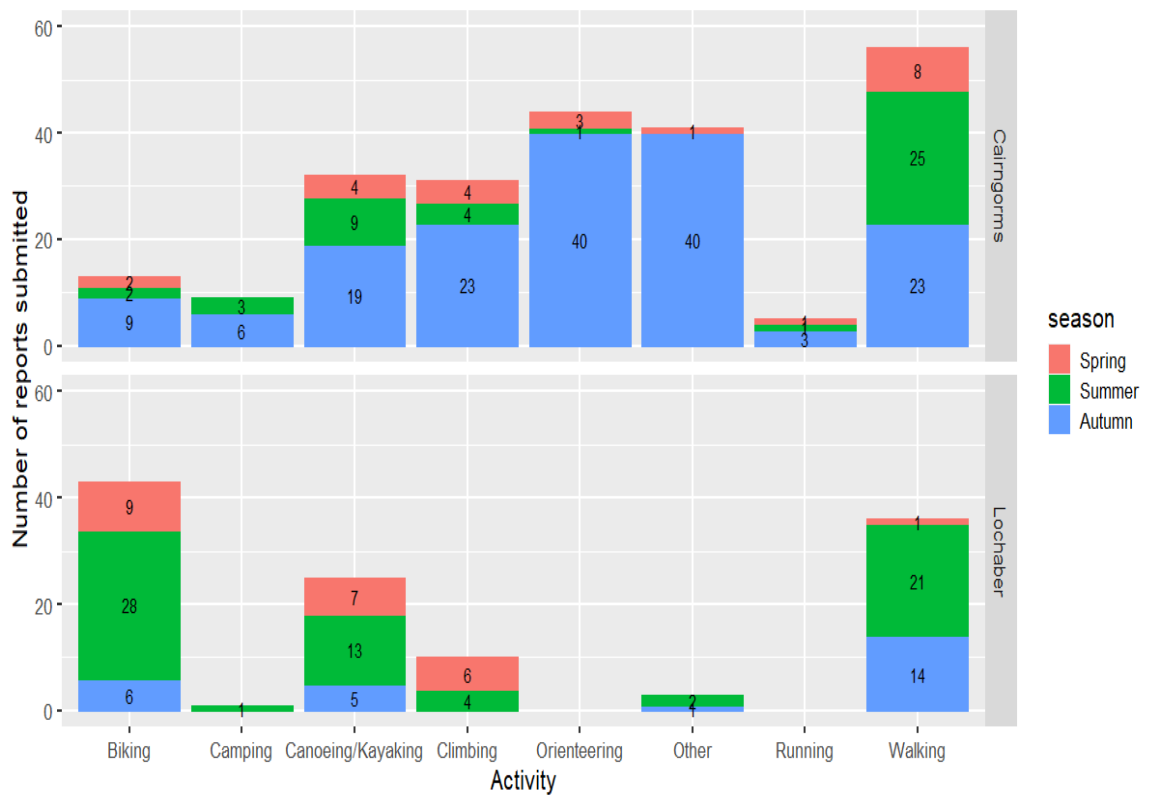
Summary report characteristics are shown in Table 5.2. Lochaber had less engagement than Cairngorms, with less volunteers participating in the project, less reports submitted and low tick bite and tick encounter rates. The two outdoor recreational organisations reported more often than all the other volunteers. There were no reports from forestry organisations in the two study areas.

**Table 5.2:** Summary statistics of the number of reports submitted to the CS project in the Cairngorms and Lochaber. The table includes information regarding the number of tick bites, tick encounters, and tick bite and tick encounter rates.

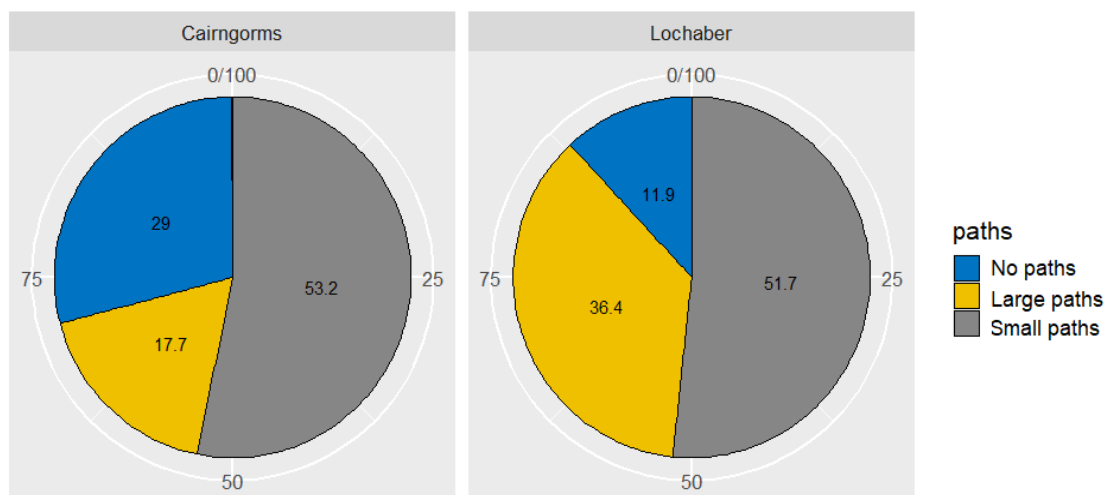
	<b>Cairngorms</b>	<b>Lochaber</b>
Number of reports	231	118
Number (and percentage) of reports with at least one tick bite	20 (8.7%)	8 (6.8%)
Number (and percentage) of reports with at least one tick encounter	26 (11.3%)	9 (7.6%)
Number of volunteers	21	9
Total number of people included*	2,088	722
Number (and percentage) of volunteers who were outdoor recreational organisations	2 (9.5%)	2 (22.2%)
Number (and percentage) of reports from outdoor recreational organisations	170 (73.6%)	78 (66.1%)
Number and percentage of volunteers who reported one time only	7 (33.3%)	4 (44.4%)
Total number of tick bites	51	13
Total number of tick encounters	117	25
Tick bite rate (per person per hour): mean, median and range	0.083; 0; 0 - 7	0.018; 0; 0 - 1
Tick encounter rate (per person per hour): mean, median and range	0.268; 0; 0 - 33	0.028; 0; 0 - 2

\*Since volunteers could report as individual or as a group of N people; this value has duplicate information regarding each specific volunteer who reported more than one time.

In Cairngorms, most reports were from walking and orienteering. Volunteers in Cairngorms also often reported 'other' category, which included predominantly activities organised by one of the outdoor recreation groups for children and teenagers. Additional activities included in 'other' category and reported less frequently (less than one percent of the total activities reported) included berry picking, bird watching and/or monitoring of birds/raptors, work in maintenance and control, gardening and golf. In Lochaber, most reports were from biking and walking, and no reports were submitted for orienteering or running (Figure 5.5). In both areas, most activities were performed on small paths (Figure 5.6).



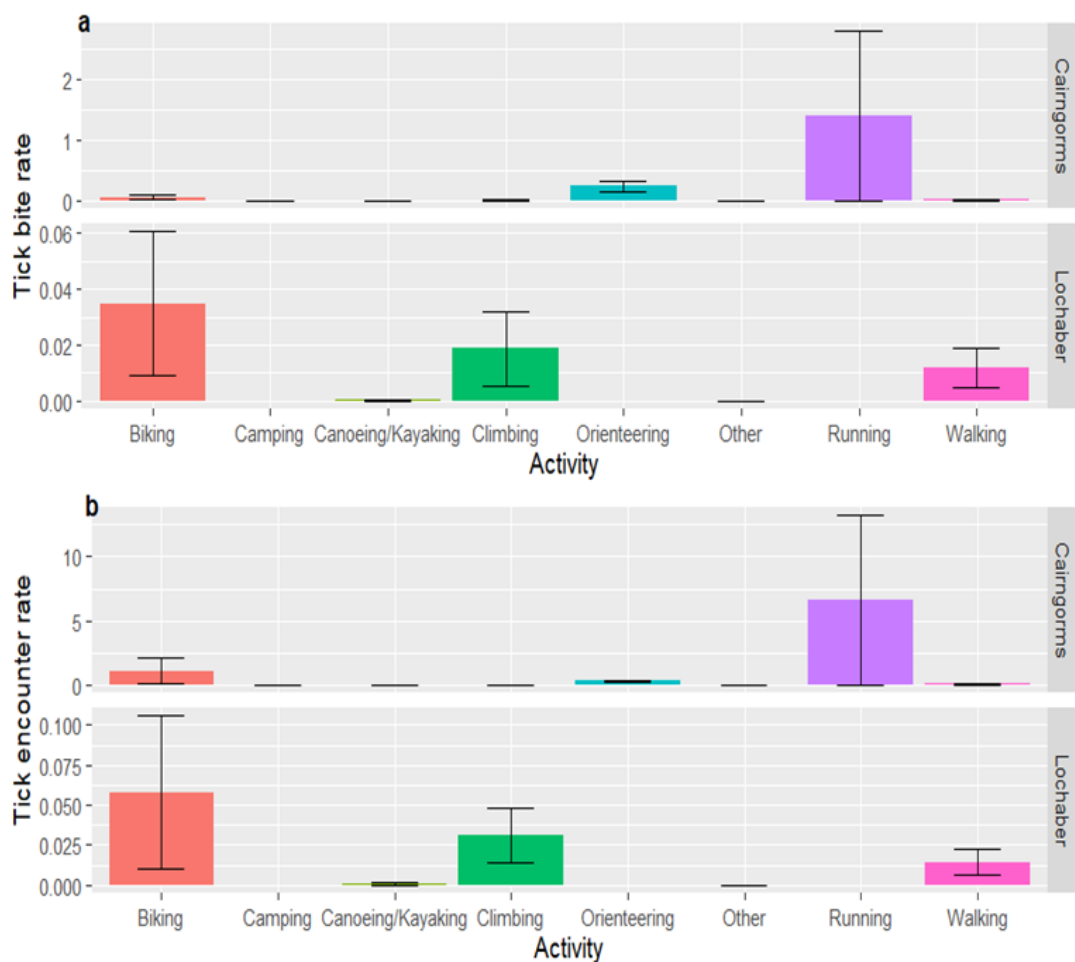
**Figure 5.5:** Frequency and seasonal distribution of the reports submitted by activities. Reports submitted between May and November 2018, and March and November 2019, in the Cairngorms (top panel) and Lochaber (bottom panel).



**Figure 5.6:** Percentage of reports submitted by type of path used. Reports submitted between May and November 2018 and March and November 2019, in the Cairngorms and Lochaber.

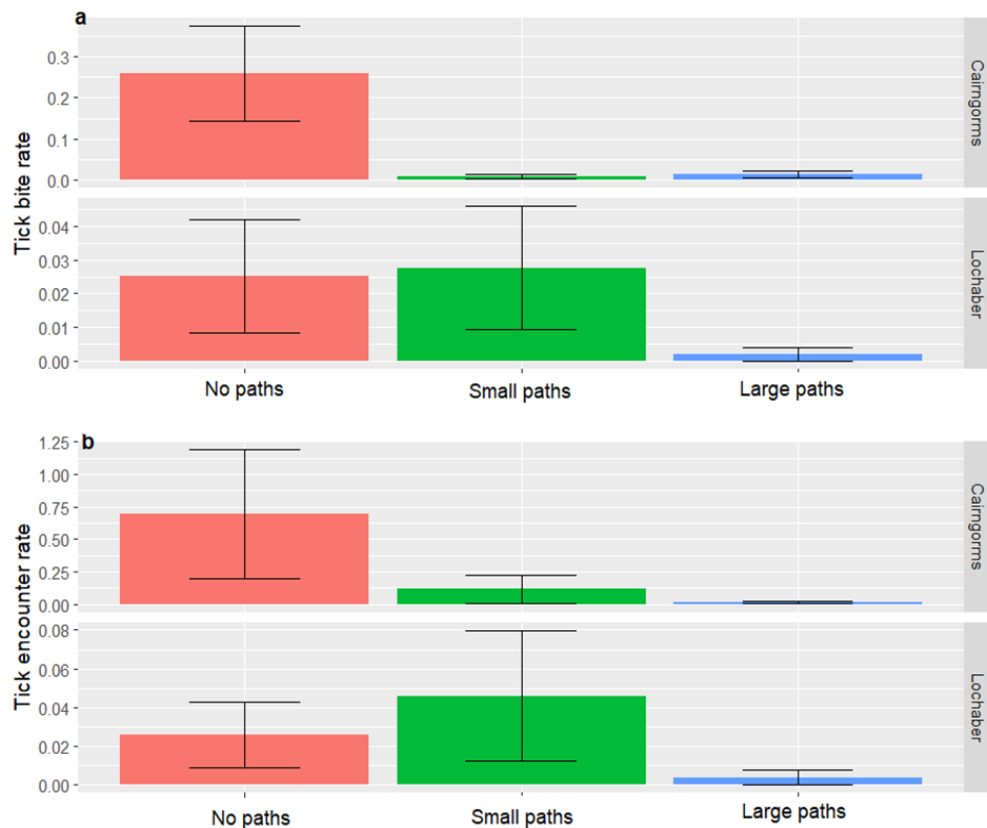
## Descriptive analysis of the risk factors for tick bite and tick encounter rates

Kruskal-Wallis test showed that, in the Cairngorms, both tick bite ( $p < 0.001$ ) and tick encounter rates ( $p = 0.0014$ ) varied with activity type. Running, orienteering and biking had the highest rates of tick bite and tick encounter (Figure 5.7). In Lochaber, the activities with highest rates were biking, climbing and walking. The range of values for tick bite and tick encounter rates were low in Lochaber when compared with the rates in the Cairngorms (Figure 5.7), and Kruskal-Wallis test did not identify significant statistical differences between activities in Lochaber ( $p > 0.05$ ).



**Figure 5.7:** Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per type of activity, in the Cairngorms and Lochaber.

In Cairngorms, tick bite ( $p < 0.001$ ) and tick encounter rates ( $p < 0.001$ ) varied with the type of path used. Not using paths was the category with highest rates, compared with using small and large paths (Figure 5.8). For Lochaber, using small paths had the highest rates (Figure 5.8), and statistical differences in tick bite ( $p = 0.016$ ) and tick encounter rates ( $p = 0.018$ ) were found between not using paths and using large paths.

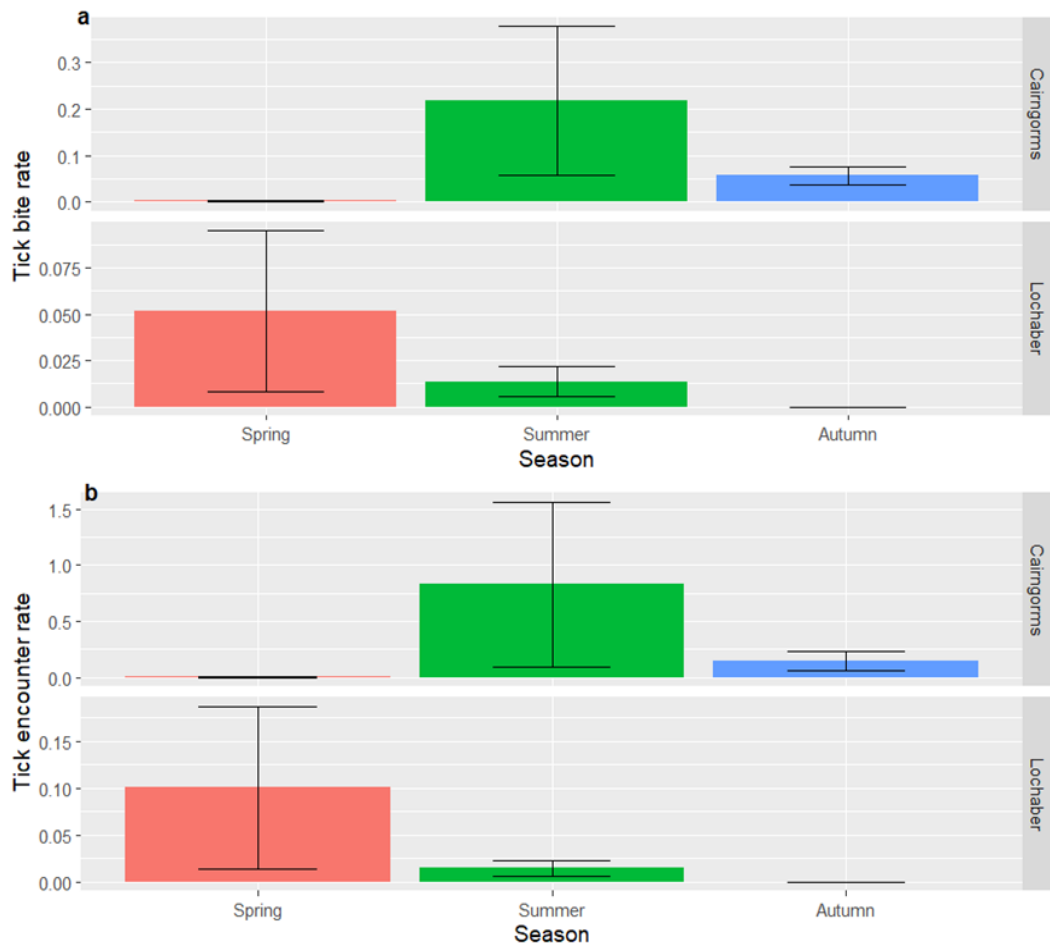


**Figure 5.8:** Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per type of path, in the Cairngorms and Lochaber.

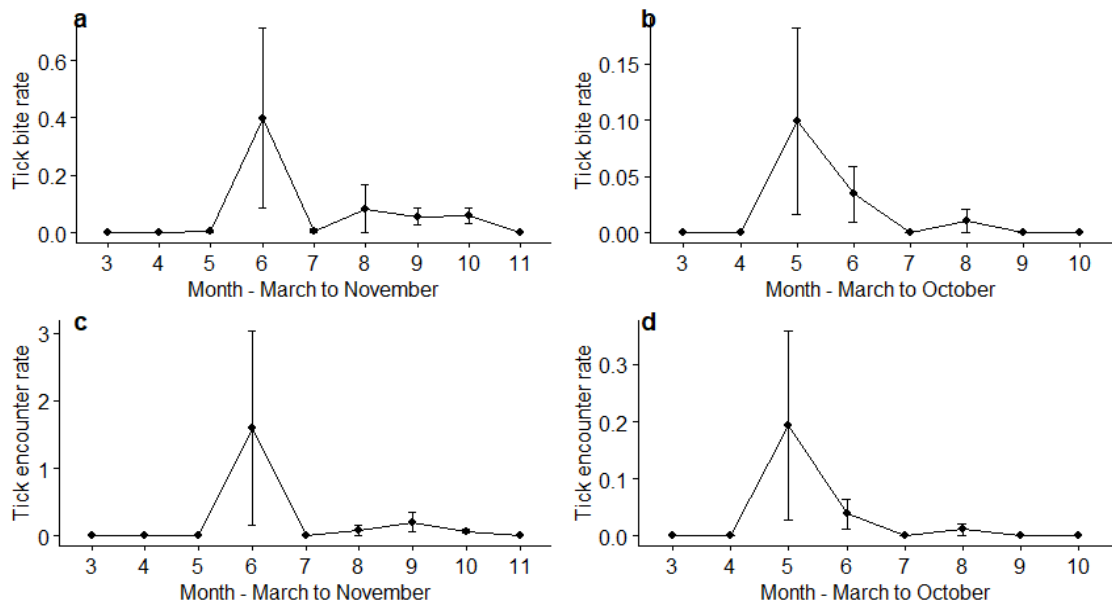
For both rates, the scale of the y-axis is higher for the Cairngorms than Lochaber.

Cairngorms had the highest tick bite and tick encounter rates in the summer, whereas in Lochaber highest rates were reported in the spring (Figure 5.9). Kruskal-Wallis test only identified significant statistical differences in the tick encounter rate ( $p = 0.03$ ) reported between spring and autumn in Lochaber. In

Cairngorms, tick bite and tick encounter rates peaked in June. Tick bite rate peaked again in late summer and in autumn, and tick encounter rate just in early autumn. In Lochaber, both tick bite and tick encounter rates peaked in May, and then presented a small peak in late summer. No peak in autumn was observed (Figure 5.10).



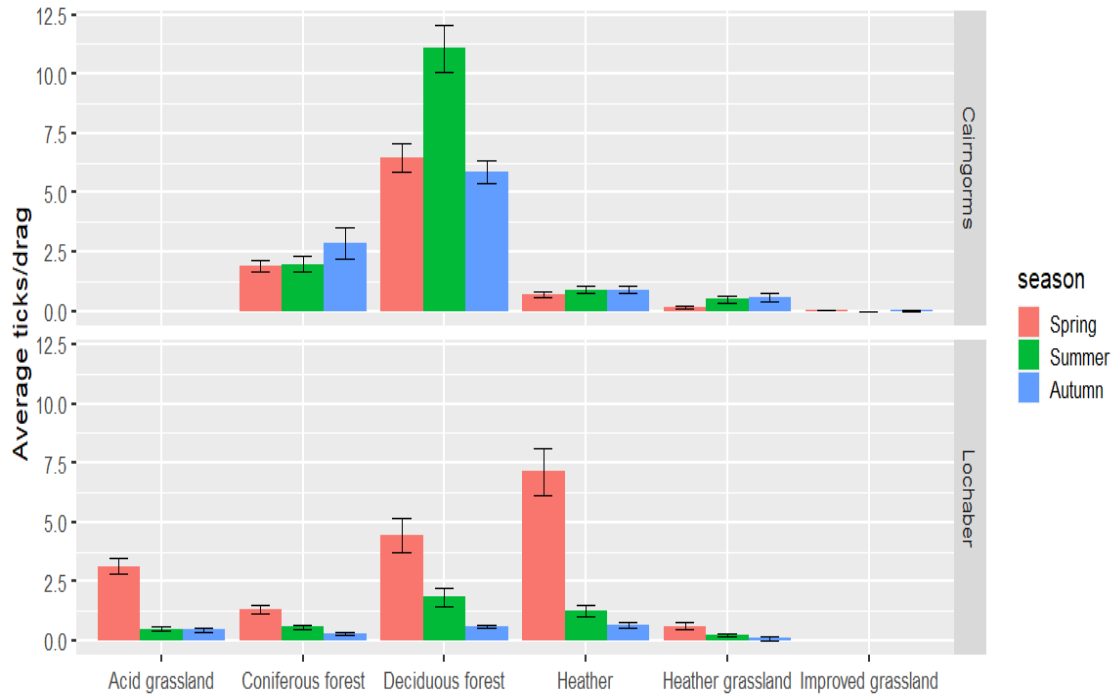
**Figure 5.9:** Mean (with SE bars) of (a) tick bite rate, and (b) tick encounter rate (per person per hour), per season, in the Cairngorms and Lochaber. For both rates, the scale of the y-axis is higher for the Cairngorms than Lochaber.



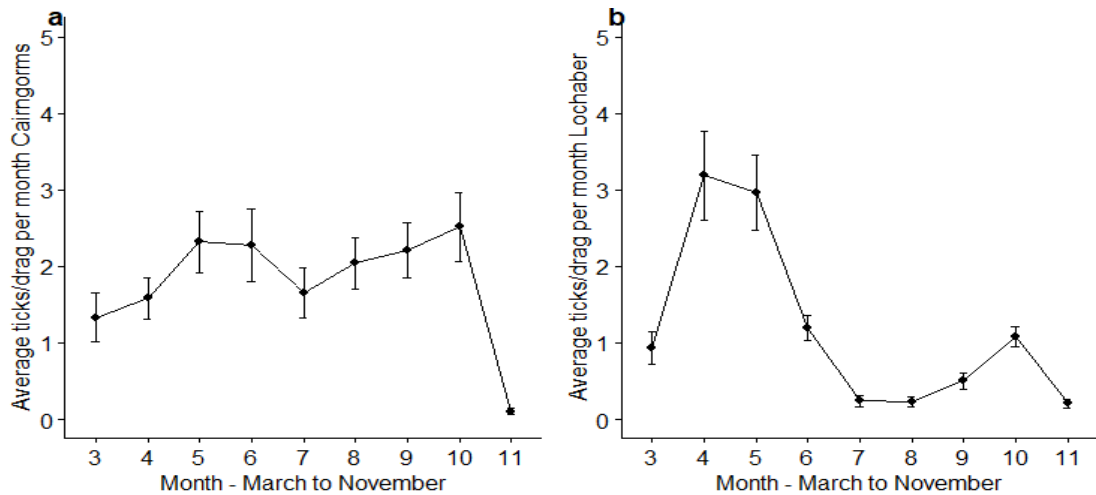
**Figure 5.10:** Mean (with SE bars) of tick bite and tick encounter rates (per person per hour) between March to October/or November using combined data from 2018 and 2019. Tick bite rate in (a) Cairngorms and (b) Lochaber, and tick encounter rate in (c) Cairngorms and (d) Lochaber.

### Descriptive analysis of questing tick data

In the Cairngorms, 4,461 questing ticks (nymphs and adult ticks) were counted. The average number of questing ticks per drag per visit per site was 2.48 (minimum of 0 and maximum of 24.2). In Lochaber, 2,650 questing ticks were counted, with an average number of questing ticks per drag per visit per site of 1.47 (minimum of 0 and maximum of 14.4). In the Cairngorms, more questing ticks were collected at the sites characterised by deciduous forest, and less ticks at the sites of improved grassland. In Lochaber, the differences between land cover types were not so distinctive as in the Cairngorms, but more questing ticks were collected in sites characterised by moorland (heather) and by deciduous forest (Figure 5.11, and Appendix D: Supplementary materials for Chapter 5 (S4)). In Lochaber, more ticks were collected earlier in the year (in spring), decreasing in summer and in autumn. In Cairngorms, more ticks were collected in the summer (Figure 5.11). In 2019, questing ticks peaked first in May in Cairngorms and in April in Lochaber, and then, again in October (Figure 5.12).



**Figure 5.11:** Average number (with SE bars) of questing ticks (nymphs and adult ticks) collected per 10m<sup>2</sup>, per land cover type and season, from questing tick surveys carried out in 2019, in the Cairngorms and Lochaber.



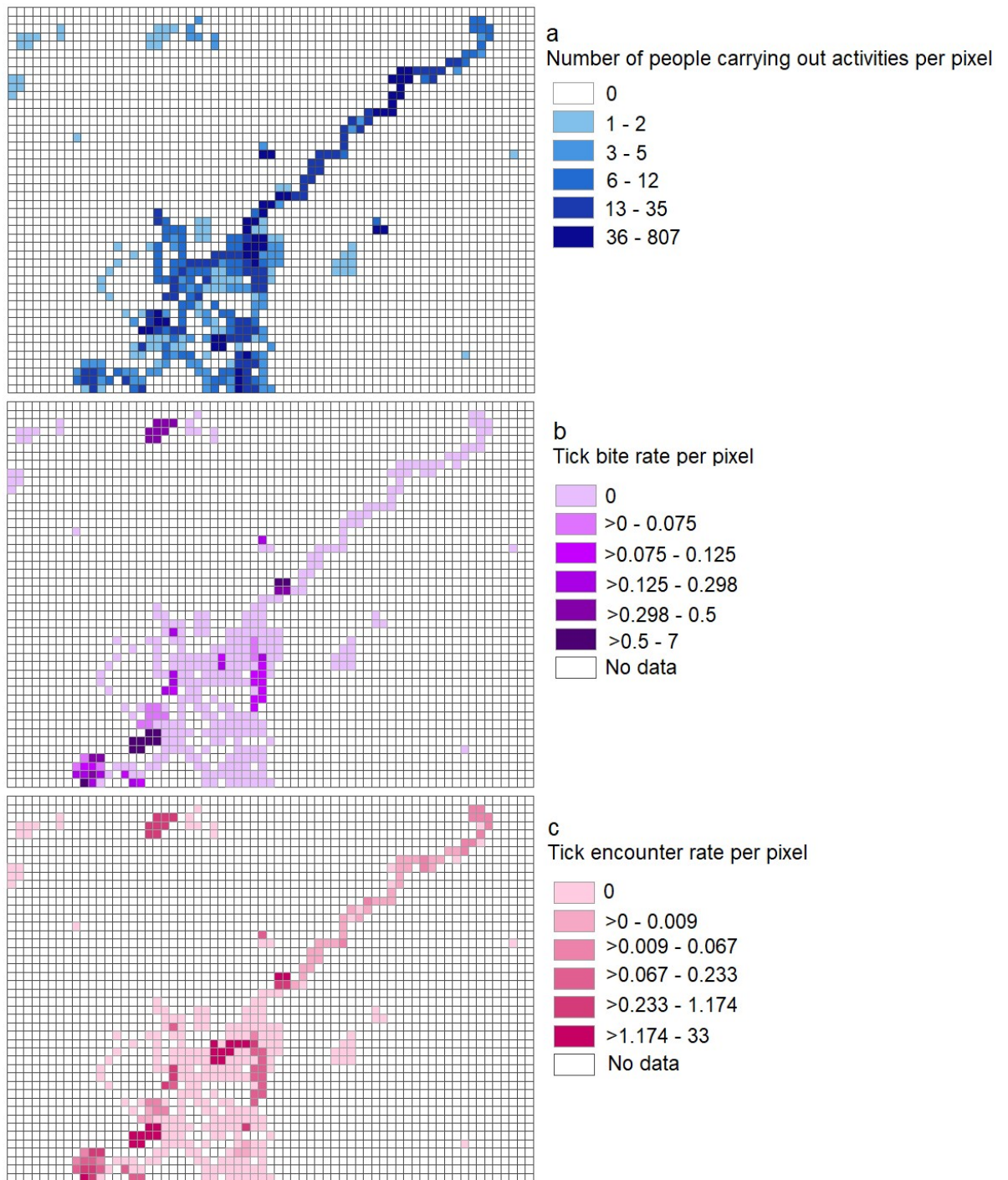
**Figure 5.12:** Average number per drag (with SE bars) of questing ticks (nymphs and adult ticks) collected per 10m<sup>2</sup> in each month, in (a) the Cairngorms and (b) Lochaber. Only data collected in the sites where questing tick surveys were carried out every month were used.



### **5.3.2. Relationship between predicted questing tick abundance and human tick bite and tick encounter rates in the Cairngorms and Lochaber**

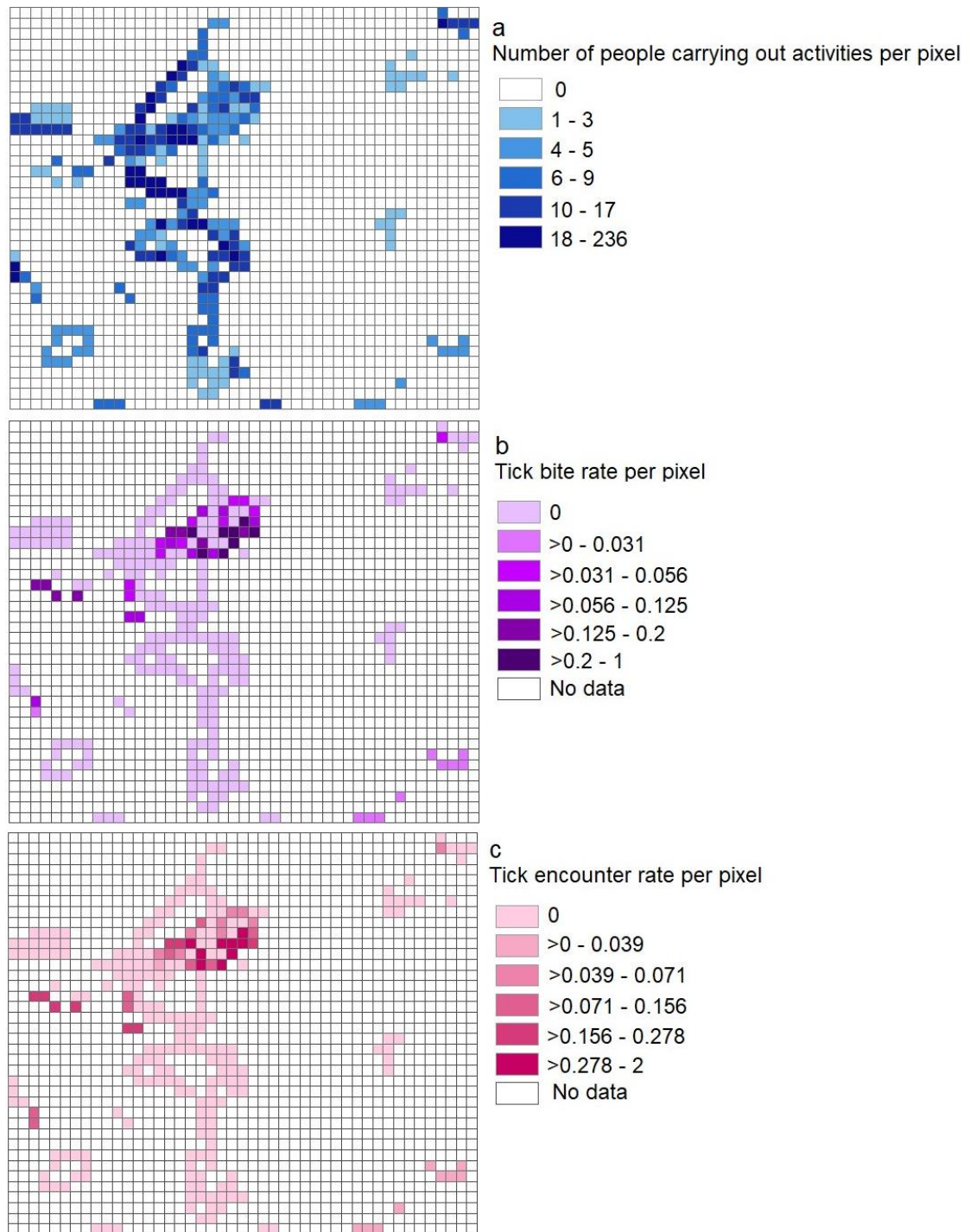
#### **Mapping tick bite and tick encounter rates in the Cairngorms and Lochaber**

Pixels coloured blue are those where volunteers reported data. Both Cairngorms and Lochaber were not uniformly covered, and a patchy spatial distribution of records of activities is visible. Lower tick bite and tick encounter rates were found in Lochaber when compared to the Cairngorms (Figures 5.13 and 5.14).



**Figure 5.13:** (a) Number of people (volunteers and other members of the groups) participating in the project in the Cairngorms, between May to November 2018 and March to November 2019 per 1 km<sup>2</sup> pixel; (b) tick bite and (c) tick encounter rates (per person per hour) in each 1 km<sup>2</sup> pixel.

Due to the presence of a skewed distribution of the three variables, six classes of values were created based on the sextiles.



**Figure 5.14:** (a) Number of people (volunteers and other members of the groups) participating in the project in Lochaber, between May to November 2018 and March to November 2019 per 1 km<sup>2</sup> pixel; (b) tick bite and (c) tick encounter rates (per person per hour) in each 1 km<sup>2</sup> pixel.

Due to the presence of a skewed distribution of the three variables, six classes of values were created based on the sextiles.

## Predictive model and map of questing tick abundance in the Cairngorms and Lochaber

**Table 5.3:** Final model outputs for questing tick abundance in the Cairngorms and Lochaber, for parameterising respective predictive maps. Estimates for the posterior mean, SD, 2.5% and 97.5% quantiles are presented.

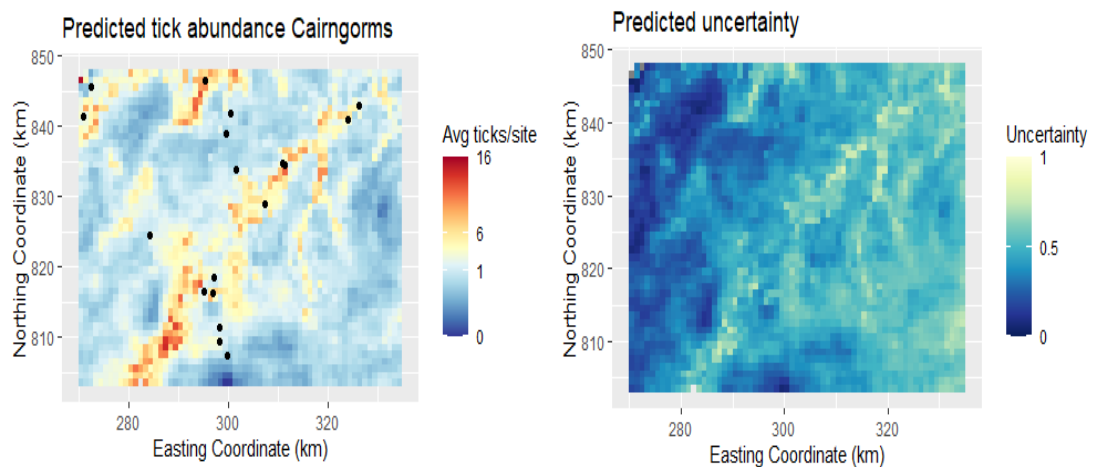
Area	Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Cairngorms	Intercept	-6.246	21.899	-49.889	37.223
	% cover of deciduous woodland	0.173	0.084	0.006	0.342
	NDVI (average 2014-2018)	0.330	0.230	-0.127	0.788
	Interaction term between LST and rain (both average 2014-2018)	0.089	0.077	-0.063	0.242
	Interaction term between latitude and longitude	-0.046	0.079	-0.204	0.112
Lochaber	Intercept	-5.773	8.695	-22.826	11.445
	% cover of deciduous woodland	0.071	0.034	0.002	0.138
	% cover of heather grassland	0.029	0.020	-0.012	0.069
	Rain (average 2014-2018)	-0.131	0.340	-0.802	0.538
	Interaction term between frost (average 2014-2018) and elevation (mean/1km <sup>2</sup> )	-0.001	0.001	-0.004	0.002
	Interaction term between LST and evapotranspiration (both average 2014-2018)	0.002	0.003	-0.003	0.008

Abbreviations: SD, standard deviation; LST, land surface temperature; NDVI, normalised difference vegetation index. The final model for Cairngorms presents a DIC of 328.96, a WAIC of 339.45 and a CPO of 173.70. The final model for Lochaber presents a DIC of 355.26, a WAIC of 364.01 and a CPO of 183.36.

In the Cairngorms, questing tick abundance increased in deciduous forest areas, with an increase of vegetation biomass (NDVI), and with an increase in LST and rainfall (a proxy for humidity). In Lochaber, areas of deciduous forests, heather grassland, and an increase in LST and evapotranspiration showed a positive effect on questing tick abundance. An increase in rainfall and in frost

(which is greater at high altitudes) decreased questing tick abundance (Table 5.3).

A map of predicted questing tick abundance was parameterised using these model outputs. In the Cairngorms, the uncertainty map identifies a crescent of higher uncertainty in predicted questing tick abundance from the west to the east, which correlates partially with increased predicted tick abundance (Figure 5.15). The distribution pattern of the tick bite and tick encounter rates, particularly the tick encounter rate (Figure 5.13), have some similarities with the distribution pattern of the predicted questing tick abundance in Cairngorms (Figure 5.15).

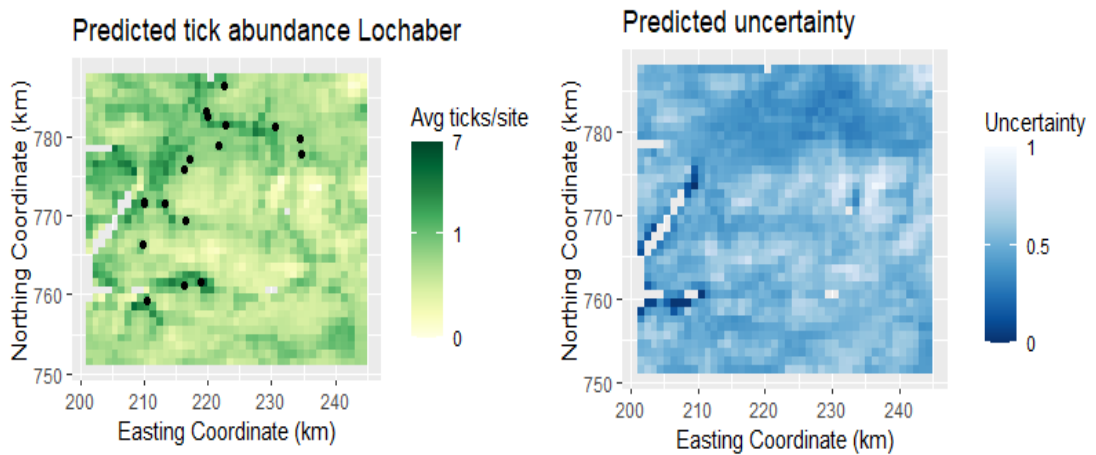


**Figure 5.15:** Predictive map of tick abundance and uncertainty map for the Cairngorms.

The uncertainty maps were calculated from the range of 95% confidence intervals of predicted values and rescaled to a 0-1 scale. Darker areas of blue have lower uncertainty, which means more confidence in the predictions for that area.

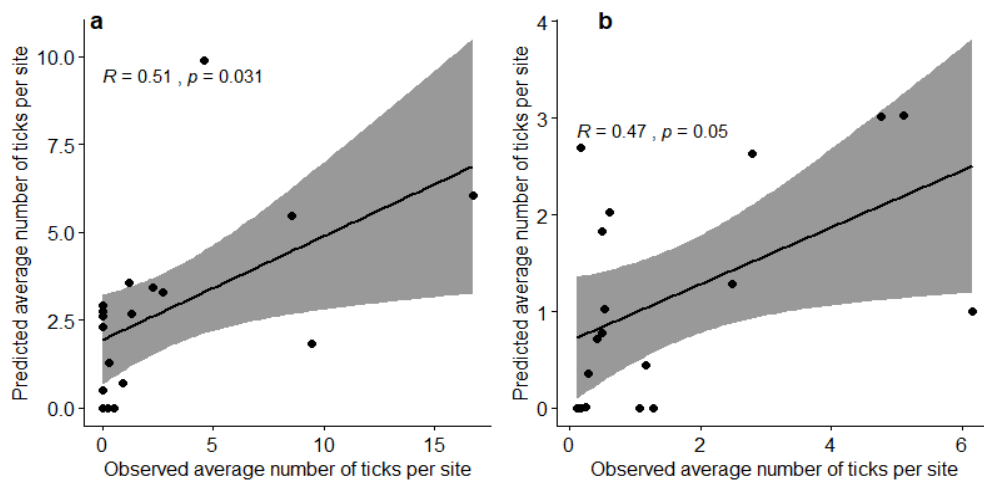
The predicted questing tick abundance for Lochaber is shown in Figure 5.16. The uncertainty map in general identifies the centre and east as areas with high uncertainty (where no questing tick surveys were carried out) (Figure 5.16). When comparing the mapped tick bite and tick encounter rates for Lochaber (Figure 5.14) with the predicted questing tick abundance (Figure

5.16), it is visible that some pixels with high rates correspond to areas with high predicted questing tick abundance.



**Figure 5.16:** Predictive map of tick abundance and uncertainty map for Lochaber.

The uncertainty maps were calculated from the range of 95% confidence intervals of predicted values and rescaled to a 0-1 scale. Darker areas of blue have lower uncertainty, which means more confidence in the predictions for that area.

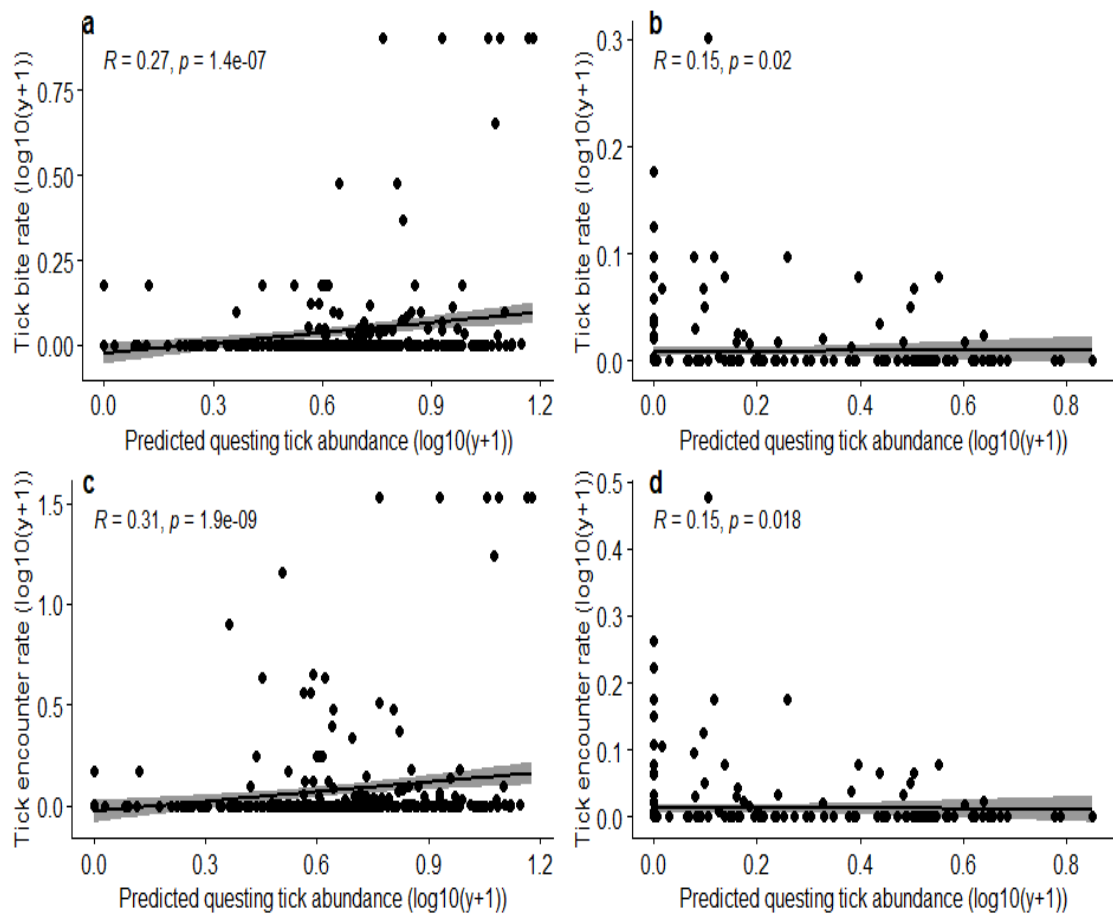


**Figure 5.17:** Correlation coefficient ( $R$ ) and significance of the correlation between the observed number of questing ticks collected per  $10\text{m}^2$ , visit and site, and the predicted questing tick abundance at (a) Cairngorms and (b) Lochaber.

The dots are the sites of questing tick surveys and the grey shades represent the 95% confidence intervals.

Both for the Cairngorms and Lochaber areas, there was a moderate positive correlation between the observed and the predicted questing tick abundance. For Cairngorms, the correlation was significant ( $P < 0.05$ ). For Lochaber, the correlation was significant at  $P = 0.05$  (Figure 5.17).

**Correlation between predicted tick abundance from questing tick surveys and tick bite and tick encounter rates in the Cairngorms and Lochaber**



**Figure 5.18:** Correlation coefficient ( $R$ ) and significance of the correlation between the predicted questing tick abundance and: (a) tick bite and (c) tick encounter rates, in the Cairngorms; and (b) tick bite and (d) tick encounter rates, in Lochaber.

The dots represent the correlation at each 1km<sup>2</sup> pixel, and the grey shadows represent the 95% confidence intervals. Values of both variables were transformed into log<sub>10</sub>(y+1).

In both the Cairngorms and Lochaber, there was a significant positive but weak correlation between predicted questing tick abundance and tick bite and tick encounter rates (Figure 5.18). When using the same information to undertake univariable models fitted with a Gaussian distribution to explain tick bite and tick encounter rates as a function of questing tick abundance, the models confirmed a positive and significant relationship with both rates and questing tick abundance in the Cairngorms. However, for Lochaber, the same positive relationship was not significant (Appendix D: Supplementary materials for Chapter 5 (S5)).

### **5.3.3. Identification of risk factors for tick bite and tick encounter rates**

Tick bite rate was influenced by the type of activity, type of path, month, area and questing tick abundance in the environment. Walking and canoeing/kayaking were activities associated with a low tick bite rate. Although not significant in the multivariable model, orienteering and running were associated with a higher tick bite rate than other activities. Not using paths when carrying out activities was identified as a risky behaviour for high tick bite rate. An increase in tick bite rate was also influenced by the month of the year, with high risk from May to October. Finally, the Cairngorms also presented higher tick bite rates than Lochaber, which may be associated with questing tick abundance in the area (Table 5.4). In the multivariable model, the type of activity (pseudo- $R^2 = 0.035$ ) explained more variation in tick bite rate than the type of path (pseudo- $R^2 = 0.029$ ), questing tick abundance (pseudo- $R^2 = 0.017$ ), area (pseudo- $R^2 = 0.01$ ) and month (pseudo- $R^2 = -0.019$ ).

Tick encounter rate increased with an increase in questing tick abundance and was higher in the Cairngorms than in Lochaber. Carrying out activities off paths was also identified as a risky behaviour for an increase in tick encounter rate. Finally, high tick encounter rate was observed in spring and in summer (Table 5.5). Tick encounter rate was better explained by questing tick abundance (pseudo- $R^2 = 0.006$ ) and type of paths (pseudo- $R^2 = 0.002$ ), than area (pseudo- $R^2 = -0.013$ ) and season (pseudo- $R^2 = -0.029$ ).



**Table 5.4:** Results of the final model to identify risk factors for tick bite rate. Volunteer ID was included as random effect to account for volunteer effort. Estimates (logarithm) of the posterior mean, SD, 2.5% and 97.5% quantiles are presented.

Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Intercept	-13.993	5.968	-27.884	-5.018
Camping	-9.297	10.543	-34.080	5.992
Canoeing/kayaking	-6.142	1.607	-9.553	-3.234
Climbing	-0.924	0.906	-2.737	0.824
Orienteering	0.526	0.988	-1.392	2.495
Other	-7.985	10.829	-33.498	7.572
Running	0.677	2.062	-3.453	4.685
Walking	-2.739	0.978	-4.716	-0.871
Small paths	-2.760	0.791	-4.414	-1.298
Large paths	-1.599	0.736	-3.094	-0.199
March	-2.554	18.042	-45.111	23.226
May	13.449	5.968	4.432	27.323
June	13.263	5.952	4.284	27.108
July	10.678	5.962	1.689	24.549
August	14.683	6.007	5.593	28.643
September	10.362	5.944	1.401	24.192
October	11.060	5.934	2.121	24.871
November	-2.924	18.498	-46.564	23.485
Interaction term between area Cairngorms and tick abundance	0.190	0.096	0.016	0.395
Interaction term between area Lochaber and tick abundance	0.312	0.348	-0.323	1.043

Abbreviation: SD, standard deviation. The final model presents a DIC of 367.88, a WAIC of 422.13 and a CPO of 645.74. Biking, not using paths, and April, are the baseline categories regarding the type of activity, type of path and month.

**Table 5.5:** Results of the final model to identify risk factors for tick encounter rate.

Volunteer ID was included as random effect to account for volunteer effort. Estimates (logarithm) of the posterior mean, SD, 2.5% and 97.5% quantiles are presented.

Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Intercept	-2.441	0.688	-3.746	-1.035
Tick abundance	0.179	0.087	0.008	0.352
Lochaber area	-0.465	0.790	-2.108	1.001
Small paths	-1.720	0.617	-2.948	-0.517
Large paths	-2.847	0.698	-4.237	-1.493
Spring	2.122	0.808	0.611	3.797
Summer	0.944	0.649	-0.304	2.249

Abbreviation: SD, standard deviation. The final model presents a DIC of 380.72, a WAIC of 392.11 and a CPO of 270.94. Not using paths and Autumn, are the baseline categories regarding the type of paths and season.

## 5.4. Discussion

The key objectives of this study were to assess the quality of a novel CS data collection approach to generate robust measures of tick bite and tick encounter rates; to compare tick bite and tick encounter rates generated from CS with questing tick abundance in the same area; to identify risk factors for tick bite and tick encounter rates and finally, to compare the relative contribution of human behavioural risk factors and questing tick abundance in explaining the variation of tick bite and tick encounter rates.

To achieve these objectives, a CS project was implemented in Scotland, where volunteers reported the number of ticks encountered (including zeros) after carrying out outdoor activities, the time spent (hours) and their route. To be able to compare tick bite and tick encounter rates from CS with predictions of questing tick abundance, two study areas were defined, where volunteers reported their activities and questing tick surveys were carried out. This study indicates that, when it is not possible to control for the influence of the environment, questing tick abundance has a positive weak correlation with tick bite and tick encounter rates, and that tick bite and tick encounter rates result

from the interaction between questing tick abundance, environmental factors that affect tick activity, and factors related to human behaviour and activity.

#### **5.4.1. Relationship between predicted questing tick abundance and human tick bite and tick encounter rates**

When environmental and human behaviour variables are controlled for, data from questing tick surveys are a good representation of tick bite and encounter rates, and therefore tick risk (Chapter 4). However, in this study, results in the Cairngorms and Lochaber areas showed a positive weak correlation between questing tick abundance and tick bite and encounter rates (Figure 5.18). A similar result was obtained in previous research conducted in Canada, where predictions from nymph density estimates (using data from active tick surveillance, which included from blanket dragging) were found to be weakly correlated ( $R=0.33$ ,  $P<0.001$ ) with ticks collected from humans (from passive surveillance) (Ripoche *et al.*, 2018).

In this study, this weak correlation was likely due to two main reasons: a) the presence of spatial and temporal differences between locations used by volunteers and site locations for questing tick surveys; and b) the effect on tick bite and tick encounter rates of factors related to tick activity and to human activity and behaviour. Tick bite and tick encounter rates from CS were validated using predictions of questing tick abundance, because both the CS project and the questing tick surveys were overlapping in the same year, hence it was difficult to get observed tick data for the locations people used. In addition, to correctly validate CS data, it was necessary to collect accurate and representative data from questing tick surveys of the two study areas, and therefore a stratified sample of sites was previously defined. However, this spatial and temporal dissimilarity between locations used by volunteers and sites of questing tick surveys (Figure 5.4) contributed to a decrease in the strength of the correlation previously observed (Chapter 4). Predicted questing tick abundance and tick bite and tick encounter rates correlated more strongly in the Cairngorms than in Lochaber, which may have been due to the better

spatial overlay between sites of questing tick surveys and volunteer reports in the Cairngorms (Figure 5.4).

This study permitted the quantification of the relative contribution to tick bite and tick encounter rates, of tick abundance and factors directly related to tick and human activity. This information has not previously been reported in the literature. This study indicated that tick bite rate varies with the type of activity, type of path, tick abundance in the environment, area and month (Table 5.4). Therefore, this confirms that the likelihood of tick bites depends on both tick abundance and activity, and human behaviour, which has previously been indicated but not quantified (Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017). In addition factors related to human behaviour, specifically type of activity and of path, were more important in explaining the variation of tick bite rate than tick abundance and other intrinsic characteristics of the area (such as environment, since both study areas represented two contrasting environments), and month of the year (which can relate to both tick and human activity). Previous research suggested that human-related factors (namely the number of warm and dry days per season, proximity to forests and recreational areas) were more relevant to model tick bites, than climatic and land cover variables (Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017). Therefore, questing tick abundance in an area does not necessary translate into tick bites, because human behaviour affects human exposure to ticks and the chances of getting a tick bite (Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017).

This study identified that questing tick abundance in the environment and the type of paths used contributed the most to explaining tick encounter rate than season and area (environment). A previous study using scouts as volunteers also indicated that both hazard and exposure-related variables significantly contributed to the frequency of scout-tick contact (Keukeleire *et al.*, 2015). Previously in Chapter 4, it was demonstrated that, when controlling for environmental and human behavioural factors, tick bite rate correlated better with questing tick abundance than tick encounter rate, which was subject to

high variability. However, this study indicated that tick bite rate is more affected by human behaviour and activity than tick encounter rate. This therefore explains the better correlation between questing tick abundance and tick encounter rate. These results suggest that some activities, with their inherent characteristics, make people more vulnerable to tick bites (such as type of clothes worn (such as pointed out in Faulde *et al.*, 2014; Hall *et al.*, 2017), or the fact of interacting more with the environment when practicing the activity). Because crawling ticks do not directly contribute to the risk of pathogen transmission (Eisen and Eisen, 2016), future studies could just analyse tick bite rate. However, the report of crawling ticks has its benefits. In large scale CS projects, tick encounter rate can be used to validate CS data with questing tick surveys, as it presents a better correlation with questing tick abundance; may contribute to better volunteer engagement (i.e. the importance of spotting a tick) and perhaps to a higher engagement of reporting 'zeros'; and it can also contribute to an earlier check for ticks, hence decreasing the attachment rate and potential of infection.

However, this study had some limitations and uncertainty, which also contributed to the weak strength of correlation between questing tick abundance and tick bite and tick encounter rates. These are related to the variability or bias in the CS dataset and the latent uncertainty in the predictions of questing tick abundance.

### **Citizen science data**

In order to generate accurate tick bite and tick encounter rates, and hence to better understand risk, this study presents a novel CS approach: volunteers were engaged to report both positive and negative tick encounters, the number of people involved in the same activity and the actual time spent in the activity. In addition, volunteer effort was considered in order to decrease associated bias. This information allowed the calculation of the number of tick bites and tick encounters per person per hour, as only these rates reflect actual human exposure and risk. Although there are other studies which presented measures

of tick bite rate (such as Faulde *et al.*, 2014; Keukeleire *et al.*, 2015; Hall *et al.*, 2017), they did not include the report of negative tick encounters, did not account for the specific time of exposure, and were related to the practice of a specific activity. Currently this is the only study reported to be using this approach to measure tick bite and tick encounter rates.

In order to increase the quality of the CS data reported, in this study additional measures were taken to: a) engage with volunteers to regularly report outdoor activities with and without tick encounters, to try to reduce bias towards positive reports; b) train the volunteers responsible for the highest number of reports (the outdoor recreational organisations), as volunteer training was identified as an important measure to decrease errors and bias (Palmer *et al.*, 2017; Tyson *et al.*, 2018); c) publicise the project before and during its implementation and communicate results with volunteers (i.e. first results were communicated to the volunteers via a website (Predicting Lyme Disease Risk | Animal Health Projects and Resources (sruc.ac.uk)), as these steps are important for project dissemination and volunteer engagement (Guilbaud and Guilbaud, 2017; Palmer *et al.*, 2017; van den Wijngaard *et al.*, 2017; Sakamoto, 2018; Tyson *et al.*, 2018); d) develop an app and a website for easy reporting, which also contributes to increase the quality of the information provided and volunteer engagement (August *et al.*, 2015; Hines and Sibbald, 2015; Guilbaud and Guilbaud, 2017; Palmer *et al.*, 2017; Eritja *et al.*, 2019); and e) track volunteers' routes, to decrease errors associated with lack of spatial precision for the location of the tick encounter (Laaksonen *et al.*, 2017; Salkeld *et al.*, 2019; Eisen and Eisen, 2021).

One weakness and at the same time, a strength of this data, is the patchy spatial distribution of CS reports in the Cairngorms and in Lochaber (Figure 5.4), which was perhaps influenced by the places where volunteers often went. Therefore, in one aspect, it is assumed that a certain degree of spatial bias is present in both CS datasets, associated with places more or less attractive, places often used to carry out outdoor recreational activities, or due to spatial

differences in volunteer motivation and effort (Palmer *et al.*, 2017; Zhang and Zhu, 2018). However, if the project was rolled out more widely and for longer, the spatial distribution of the CS reports would give information on the human usage of each area, as well as tick bite and tick encounter rates, enabling the mapping of human risk for tick bites. This is certainly helpful for identifying high risk areas, for targeting prevention, management, and information campaigns.

One of the limitations of this CS project is related to the lack of data representativeness, which increases the presence of spatial and temporal bias (van den Wijngaard *et al.*, 2017; Tyson *et al.*, 2018). Data representativeness is generally ensured by keeping volunteers engaged and motivated. In this study, although the same level of effort was made to engage with volunteers in the two areas, Lochaber had less engagement than the Cairngorms, and therefore less volunteers contributed reports (Table 5.2). This gives less certainty in the rates estimated for Lochaber (Figure 5.14). However in the Cairngorms, because more volunteers participated in the project, more people covered the area, which increased the confidence in the tick bite and tick encounter rates reported (Figure 5.13). This variation likely also contributed to the weaker correlation between tick bite and tick encounter rates with questing tick abundance in Lochaber than in the Cairngorms.

Another limitation of this study was that tick reports from CS were not confirmed by specialists, which increases the uncertainty with the CS data (Laaksonen *et al.*, 2017; Palmer *et al.*, 2017; Garcia-Martí *et al.*, 2018; Hamer, Curtis-Robles and Hamer, 2018; Nieto *et al.*, 2018; Tyson *et al.*, 2018; Switters and Osimo, 2019). Although volunteers were informed that they could send photos of their tick encounters to the project team, no photos were received to help with data validation. Confidence in the reports submitted increases because people in Scotland are normally well informed about ticks and TBDs, as LB is a problem in Scotland (Mavin, Watson and Evans, 2015); training was offered to the volunteers contributing the most reports (the outdoor recreational organisations); when using the website, individual volunteers needed to

contact the project to participate, and so in the first place they knew about the subject; and both the website and app had a photo with *I. ricinus* ticks in the three active stages. In addition, associated with this limitation, the tick species responsible for the tick encounters with volunteers is not known. Therefore, results cannot be generalised for *I. ricinus*. However in the UK, particularly in Scotland, *I. ricinus* was confirmed as the tick species that most often bites humans (Robertson, Gray and Stewart, 2000; Jameson and Medlock, 2011; Hall *et al.*, 2017; Cull *et al.*, 2019), and it is also the species most often collected from the vegetation during questing tick surveys (e.g. Hall *et al.*, 2017; Gandy, 2020). Finally, in this study, the presence of false negative reports cannot be ruled out, such as people that report a negative tick encounter but found a tick bite the next day. However, volunteers have occasionally emailed the project team to correct for additional tick submissions.

### **Predictive maps of questing tick abundance**

In this study, effort was made to guarantee the quality of the questing tick data collected. To ensure that the two distinct study areas were well represented in terms of habitats and where people could go for outdoor recreational activities, sites for questing tick surveys were stratified by land cover type. Additionally, because tick activity is seasonal (Randolph *et al.*, 2002), in some sites questing tick surveys were carried out every month, and in the remaining sites, three times per year (one timepoint each season).

Cairngorms and Lochaber have contrasting environments. Therefore, differences in the covariates that influence questing tick abundance, and in the predictive patterns of tick abundance, were expected. However for both areas, the results of the predictive models (Table 5.3) are explained by the knowledge of questing tick ecology in Scotland. In general, this study found high questing tick abundance in deciduous forests, in areas with a high NDVI (an index of green vegetation and soil moisture), when evapotranspiration increased, and in warmer temperatures. In contrast, frost and elevation were found to be



associated with low tick abundance. The positive effect of deciduous forest and warm temperature, and the negative effect of frost conditions and elevation in questing tick abundance, are in line with previous studies carried out in Scotland (Gilbert, 2010; Braga, 2012; James *et al.*, 2012; Ribeiro *et al.*, 2019). NDVI and evapotranspiration have been extensively used to predict questing tick presence and abundance (Estrada-Peña, Estrada-Sánchez and Estrada-Sánchez, 2015; Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017; ECDC, 2019a; Ribeiro *et al.*, 2019). Evapotranspiration, a measure of the water extracted from land surfaces due to evaporation and transpiration, and used sometimes as a proxy of water saturation deficit (such as Ruiz-Fons *et al.*, 2012) has been identified as one of the most important features to explain tick dynamics (Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017).

In this study, rain showed an opposite effect in questing tick abundance in the two study areas. Whereas in the Cairngorms, high questing tick abundance correlated with warm temperature and high humidity, in Lochaber, which is a very humid area located in the west coast of Scotland (Met Office, 2016), rain showed a negative effect in questing tick abundance. More questing ticks are predicted in areas with warmer climate and higher humidity or higher rainfall (Ruiz-Fons *et al.*, 2012). However, the negative effect of rainfall was also found in previous research conducted in Scotland (Braga, 2012; James *et al.*, 2012; Ribeiro *et al.*, 2019). The findings of this study can be explained by the increased gradient of rainfall from the east to the west coast of Scotland. Therefore, the wettest conditions of Lochaber may be too wet for *I. ricinus* to quest. This is the first study to identify different relationships between rain and questing tick abundance between the east and west coast of Scotland.

In addition in Lochaber, sites characterised by heather had the highest questing tick abundance in spring (Figure 5.11, Appendix D: Supplementary materials for Chapter 5 (S4)). Hence, it was expected that the final model captured a positive effect of heather for questing tick abundance. Perhaps due to the high variation in questing tick abundance over the year in heather, the

final predictive model for Lochaber did not include heather, but identified instead the effect of heather grassland, which had small but stable values of questing tick abundance over the year.

Latent uncertainty in the predictive maps (Figure 5.15 and Figure 5.16) arises from the variability present in the questing tick data and in the covariates used (Ribeiro *et al.*, 2019). Uncertainty was calculated as the difference between 97.5% and 2.5% quantiles of the posterior mean, meaning that uncertainty correlated with questing tick abundance. The limitations of this measure were highlighted in Chapter 2. Cairngorms area accounted for almost 63% of the total questing ticks collected and showed high variability in tick counts between sites. High variability in questing tick counts between seasons was observed in Lochaber (Figure 5.11, Appendix D: Supplementary materials for Chapter 5 (S4)). Another cause of high uncertainty in the predictions is related to lack of data coverage of the covariates used in the predictions. Although adding extra sites for questing tick surveys would have given better model fit, helping to identify other important covariates for questing tick abundance, and to decrease uncertainty related to the seasonal variation in questing ticks, the study design was limited by logistic constraints.

Besides the importance of the predicted uncertainty in the two study areas, the analysis of Lochaber area presented a smaller correlation between observed and predicted questing tick abundance, when compared to the Cairngorms (Figure 5.17). Therefore, this result added extra uncertainty, which influenced the differences found in the correlation between predicted questing tick abundance and tick bite and tick encounter rates in Lochaber and the Cairngorms.

#### **5.4.2. Risk factors for tick bite and tick encounter rates**

Published studies concerning risk factors for tick bites and/or tick encounters are generally based on passive tick submission and do not include information of where people have not observed ticks, or are questionnaire-based studies

implemented in a specific sample of the population (Bartosik *et al.*, 2011; Hjetland *et al.*, 2013; Mulder *et al.*, 2013; Garcia-Martí, Zurita-Milla, Swart, *et al.*, 2017; Mead *et al.*, 2018; Jore *et al.*, 2020). These studies have identified some habitats (such as forests), environmental conditions (warm and dry days), some age groups, owning a pet, and certain activities (running and walking in forests, paddling, rowing, hiking and hunting) with high risk for tick bites. Risk factors for tick bite and tick encounter rates are poorly known, since only a few studies which were focused on specific activities (i.e. military training, scouting and marathon runner (fell running)) have calculated these rates (Faulde *et al.*, 2014; Keukeleire *et al.*, 2015; Hall *et al.*, 2017). Just one of these studies related tick bite rate with landscape, and identified that being in complex and fragmented landscapes close to a forest patch resulted in high tick bite rate (Keukeleire *et al.*, 2015).

The novel CS approach used in this study considered both positive and negative tick encounters, the specific time of exposure, the number of people exposed in the same outdoor activity, and which tracked volunteers, has facilitated new analyses on risk factors for tick bite and tick encounter rates. As previously discussed, this approach also permitted quantification of the relative contribution for the risk, of factors directly related to human behaviour (such as type of activity and type of path, distance), factors related to tick activity (questing tick abundance in the area), and factors related to both (such as month or season).

In this study, canoeing/kayaking and walking were associated with low tick bite rates compared to orienteering, running and biking, which had high tick bite rate. Orienteering and running (i.e. fell running, running in forests) have been previously identified as activities with high exposure to tick bites (including in Chapter 4 and (Fahrer *et al.*, 1991; Zhioua *et al.*, 1998; Hall *et al.*, 2017; Jore *et al.*, 2020)). Whereas in other studies walking has been associated with high exposure to tick bites (Porter *et al.*, 2019; Jore *et al.*, 2020), the descriptive analysis indicated that walking was one of the activities more often reported,

which therefore increases the confidence in the results obtained. One value of this study is that activity type and tick abundance were both included in the final model. Therefore, the high rate is not just because people were conducting these activities in habitats with high questing tick abundance. Previous studies did not separate out the relative contribution of the type of activity from simply being in areas where tick exposure was more likely. Results also indicated that being off paths when performing activities increased the risk of tick bites. This is corroborated by previous research which indicated that using clear footpaths reduces human exposure to ticks (Walker *et al.*, 2001).

High tick bite rate is expected from May to October, particularly in May, June and August (the months with the highest model coefficients). These results can be explained by questing tick activity in the two areas. In 2019, tick bite rate peaked in May in Lochaber, and in June in the Cairngorms, respectively one month after the peak of questing ticks in the two areas. Additionally in both areas, a late peak in questing ticks was also observed in October (Figure 5.12). These findings are consistent with previous studies from the UK, that indicate that tick bites are more frequently reported from May to August (Hansford *et al.*, 2017; Cull *et al.*, 2018). There are no previous studies on tick bite seasonality in Scotland.

Regarding tick encounters, the results of this study indicated that the rate that people encounter ticks in the environment increases with high questing tick abundance, by not using paths, in spring and summer than in autumn, and was higher in the Cairngorms than in Lochaber. Not using clear footpaths was identified as an important risk factor for both tick bite and encounter rate. Season affects not only questing tick activity (Hancock, Brackley and Palmer, 2011; Cayol *et al.*, 2017) but also human behaviour and exposure to ticks (Cull *et al.*, 2019). In spring and summer, with the increase in temperature and number of hours of daylight, people are more likely to spend time outdoors, hence being more exposed to ticks.

CS data from the Cairngorms indicated higher tick bite and tick encounter rates than Lochaber. This is consistent with tick abundance data. However, in Lochaber, activities associated with low tick bite rate such as canoeing/kayaking, climbing and even walking were more frequently reported, and activities were more commonly conducted on paths which also contributed to the low rates of tick bite and tick encounter in Lochaber.

#### **5.4.3. Implications of this work and future research priorities**

The major scientific improvements arising from this study resulted from the application of a new CS approach. Accounting for negative encounters, for volunteer effort and using denominator data provided more confidence in the accuracy of tick bite and tick encounter rates and allowed a direct comparison with questing tick abundance. In addition, this approach enabled the identification of the relative contribution of questing tick abundance and of human activity and behaviour to tick bite and tick encounter rates, which was still missing from the literature. Therefore, when parameterising models and developing predictive maps of tick bite and/or tick encounter risk for public health decision-making, researchers should include information on human activity and behaviour, in addition to information on factors that affect tick activity and abundance (such as weather, climate, land cover and hosts). Results of this study identified risky activities and behaviours which could help targeted communication and prevention strategies. Future work related to this research will use all CS data gathered in Scotland to identify risk factors for tick bite and tick encounter rates, including risky habitats, and to develop predictive models and maps of risk to humans.

Results of this study bring not only scientific value, but they also have public health relevance. The full CS project received reports from the whole of Scotland (and even data from England and US). Engagement with the project is likely to have raised awareness about ticks and TBDs to people who may be exposed when carrying out outdoor activities. To report activities with and without tick encounters, volunteers needed to check for crawling and attached

ticks, which may have contributed to earlier removal of tick bites, hence decreasing the risk of infection. The educational and public health relevance of CS, and the contribution of volunteer engagement in decreasing the human-vector contact rate, hence the risk of infection, has been widely recognised for tick and other vectors of diseases (vonHedemann *et al.*, 2016; Palmer *et al.*, 2017; Sakamoto, 2018; Tyson *et al.*, 2018; Cull *et al.*, 2019; Switters and Osimo, 2019; Asingizwe *et al.*, 2020).

The approach described in this study generated accurate data on tick bite and tick encounter rates, which are essential to measure tick risk in humans. A wider application of this approach is recommended to improve data collection through CS, but some challenges need to be overcome. The major challenge of this CS approach was the very intensive work to try to keep volunteers engaged enough to report routinely, and even then, there was still low engagement noticed particularly in Lochaber.

In CS projects it is common to observe an initial phase of volunteer enthusiasm followed by a decrease in effort (Palmer *et al.*, 2017; Lewis *et al.*, 2018). However, difficulties with engagement increase when volunteers need additional effort to contribute to the project, such as when volunteers are asked to use a standard and repetitive methodology for data collection (Seifert *et al.*, 2016; Lewis *et al.*, 2018), or as seen in this study to report frequently and include negative tick encounters. This study adopted certain measures, which were previously identified as valuable strategies to increase volunteer engagement, such as recruiting volunteers to help in the initial stage of design process, training volunteers, sharing results with volunteers, giving free learning material (such as leaflets) and using a website and an app for data collection (August *et al.*, 2015; Sakamoto, 2018; Tyson *et al.*, 2018; Asingizwe *et al.*, 2020). However, there are other strategies that could be applied to increase long term engagement with this CS approach. In US, volunteers of 'The Tick App' project, besides reporting positive tick encounters, need to complete daily logs with information of the activities performed and if they

found any ticks or not. In order to achieve a wider representation of the daily logs, researchers only ask for 15 daily logs and reward volunteers with badges (Fernandez *et al.*, 2019). Other strategies could be considered to increase engagement, namely: a) to provide a predictive map of tick activity (also adopted by (Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017; Garcia-Martí *et al.*, 2018)); b) to increase gamification features (such as publishing volunteer league tables, offering rewards and avatars) (Bowser *et al.*, 2013; Vianna *et al.*, 2014; August *et al.*, 2015; Fernandez *et al.*, 2019); and c) to offer monetary incentives.

Therefore in future, for a wider implementation of this CS approach, strategies such as gamification and the rewarding of volunteers who put more effort into data collection, and presenting volunteers with a robust predictive map of tick abundance or activity to help in their decision-making, should be considered. Although monetary incentives might result in fraud for monetary gain, it was found that offering incentives to complete regular surveys did not encourage false responses when the same reward was offered regardless of the answer (Bell *et al.*, 2016). Bell *et al.*, (2016) also found that frequent task repetition with shorter recall periods (such as every week in contrast to monthly) resulted in more comprehensive reporting and improved data quality. Therefore, to decrease the effort in reporting negatives, volunteers could be randomly allocated to report activities just one time every week. Future research is needed to compare the accuracy of risk measures calculated with and without the negative encounters, to then measure the effect of bias towards positive reports and the importance of this approach.

To increase the application of this CS approach, it is recommended to develop a collaboration between researchers, government and public health agencies. This may facilitate the communication of the public health message and the scientific value of collecting data when no tick encounters occurred. Although this study showed the difficulties in collecting data on negative tick encounters, this is not impossible. For example, the new COVID Symptom Study app asks

health citizens to report their symptoms daily. This information helps researchers (i.e. predicting infection, gain understanding of the different symptoms, to study risk factors) and the app triggers an invitation to book a test if volunteers report symptoms that might be caused by COVID-19 (Kings College London, 2020; Wise, 2020; ZOE Global Limited & King's College London, 2020). One valuable study in the field of social sciences would be to identify successful and ineffective strategies of volunteer enrolment in this type of approach.

## **5.5. Conclusions**

In summary, this study demonstrated that the components of tick bite and tick encounter rates are questing tick abundance in the environment, and factors related to tick and human activity. Human activity explains more variation in tick bite rate than environmental factors, which explains the positive but weak correlation between tick bite rate and questing tick abundance. Therefore, to accurately predict tick bite risk to humans, it is necessary to include factors related to human activity and behaviour. This study constitutes a gateway for improving modelling methodologies to predict tick bite risk to humans, to help the prevention and control of TBDs. Additionally, this is the first study which identified accurate measures of tick bite and tick encounter rates for different outdoor activities carried out in Scotland, which can help targeted information, prevention and management measures.

Results of this study confirmed that volunteers can provide valuable data to estimate tick bite and tick encounter rates, including reporting of zero values, accounting for the time covered and tracking participants. This new methodology can be adopted by other CS projects to improve the quality of CS data. However, although CS is a worthwhile venture, many volunteers are needed to secure representative data.



## Chapter 6: Final discussion

Policy makers often require accurate maps depicting *I. ricinus* distribution and risk of tick bites to target control and prevention measures regarding TBDs. However, the development of accurate maps of tick distribution and risk of tick bites is not a trivial task. Researchers often lack robust long-term and geographically extensive tick distribution data, and information on human exposure to tick bites, which is necessary for measuring risk. Most of the maps published use data on questing ticks collected during surveys in the environment. However, it is not clear how data from questing tick surveys relate to the actual human exposure and risk. Citizen science projects have the potential to provide valuable data on human exposure to tick bites and on tick distribution but are often based on the passive submission of reports which by their nature do not include tick absence, thus limiting their usefulness.

The overarching aim of this study was therefore to assess methodologies to improve public health decision-making through distribution mapping of ticks and tick bite risk for Scotland. Accordingly, Chapter 1 reviewed the current methodologies and types of data for mapping tick distribution and assessing human risk of tick bites and identified the current knowledge gaps. Chapter 2 compared the quality and robustness for predictive mapping of the three types of tick data most often used for predictive mapping. Chapter 3 explored the application of a sophisticated statistical approach to improve the quality of predictions of the distribution of *I. ricinus* in Scotland, including the associated uncertainty. Chapter 4 assessed whether questing tick surveys reflect human-tick encounter risk, in a study conducted using orienteers as volunteer citizen scientists. Finally, Chapter 5 tested the feasibility of a new CS approach to assess human risk of tick encounters and investigated the environmental and human components which explain tick bite and tick encounter rates.

## 6.1. Connecting the dots and interpreting results

**Chapter 2:** In order to accomplish the overall aim of this PhD, it was essential to first assess the quality for predictive mapping of the three available datasets on *I. ricinus* distribution in Scotland. The issue of using imperfect data is almost ubiquitous in research that involves estimating population size or distribution of any organism, as is the challenge of trying to compare, and make decisions based on, predictive maps in the absence of gold standards. The tick datasets available for Scotland constitute exemplars of the main types of data used for distribution mapping of a wide range of organisms, hence the findings of this study have wider application. The results indicated that: 1) whereas tick data from questing tick surveys generated detailed predictive maps at local scale, at national level predictions were affected by poor data coverage; 2) data from public submissions (e.g. from passive surveillance, CS projects, museums and data curators) resulted in the predictive map with low spatial detail and higher uncertainty, and 3) the combined dataset produced the most appropriate map for national scale decision-making in Scotland.

Although data from questing tick surveys are considered to be the gold standard, the quality for predictive mapping of these datasets depends, among other characteristics, on the stratification of survey sites and on the data coverage, both geographically and over the covariate range. For predictions at country level, in the case of a heterogeneous coverage of sites for questing tick surveys, researchers can adopt the strategy described in ECDC (2019a) as used in Dataset 3, to create combined datasets. This strategy improves the coverage both geographically and of the range of the covariates used, hence improving the accuracy of predictions. Finally, it was not surprising that the dataset from public submissions generated the predictive map with low spatial detail and high uncertainty. Although the results of this study were affected by the small number of datapoints, it illustrates the challenge that data from public submissions provide, as they are usually based on passive and opportunistic submission of tick records. This brings more variability and uncertainty to predictive models. The comparison between the predictions and associated

uncertainty resultant from the three datasets illustrated the potential to improve the quality of data from public submissions, such as incorporating information on when people did not find ticks and on volunteer effort.

These results helped to identify the value and limitations associated with each tick dataset, to then pinpoint subsequent steps required for improving predictive mapping of tick distribution in Scotland and internationally. These initial results suggested two new steps for further development in this PhD: 1) to investigate statistical methods to improve the predictive distribution of *I. ricinus* in Scotland, including the predicted uncertainty (Chapter 3); and 2) to investigate ways to improve tick data from public submissions (Chapters 4 and 5).

**Chapter 3:** This chapter used the tick dataset collected during questing tick surveys (used in Chapter 2) to explore whether statistical approaches could improve predictive maps. This dataset represents the most robust scientifically collected dataset available for Scotland (as it was collected using a standard technique), but the data coverage is patchy, and so there was a strong interest in improving the predictions for areas where no surveys were carried out. In this work I have applied the SPDE approach to incorporate spatial correlation in the model. Although this approach has shown advantages for mapping other disease vectors (Kifle, Hens and Faes, 2017; Myer, Campbell and Johnston, 2017; Stanton *et al.*, 2018), this method has never been used to map tick distribution. The SPDE approach was selected after a careful review of the literature of the available statistical methods capable of dealing with complex hierarchical data and several sources of variability. The matrix of triangulations created over the domain of the predictions has the potential to capture the existence of spatial dependency among neighbouring partitions, to account for spatial variability that was not explained by the covariates, and therefore to improve the predictions for areas with no data points.

The analyses indicated that models with the random field have a better predictive performance than models without this component. The predicted density of nymphs per drag and the predicted uncertainty were more informative when the random field was added, particularly for areas where no surveys were undertaken. Finally, adding the variation effects of the drag, site and ZIP process in the predictions contributed to capturing these extra sources of uncertainty, making the uncertainty more informative. This study highlighted that when researchers use data from questing tick surveys where survey site and drag represent high sources of variation, these effects need to be accounted for in making predictions, otherwise the uncertainty measure is underestimated.

**Chapter 4:** As discussed previously, it is also important to assess whether data from questing tick surveys accurately reflect human-tick encounter risk. To achieve this, questing tick surveys were carried out whilst collecting contemporaneous data on tick encounters from orienteers running in the same areas at (almost) the same time. Results showed a strong and positive correlation between tick abundance from questing tick surveys and tick bite and tick encounter rates on orienteers. This indicated that outputs from questing tick surveys can be accurate proxies of the actual tick bite and tick encounter risk to humans (with orienteers as a 'gold standard' test case), confirming the value of questing tick surveys in identifying risk areas. In addition, results demonstrated that the number of tick bites reported per person was better explained by length of time (hours) than by distance (km) of exposure.

Although a positive relationship between questing tick abundance and the two rates was expected, this is the first study that directly defines a relationship between data from questing tick surveys with tick bite and tick encounter rates (as far as I am aware). Usually, using risk assessment terminology, tick abundance in the environment is considered to be the hazard, which is then multiplied by human exposure to give the tick bite risk. However, as illustrated

in Figure 6.1, there are factors that affect tick activity and human exposure that also impact risk, which are not fully considered by the terms of hazard and exposure. In this study, these factors (such as weather and type of activity) were kept constant between questing tick surveys and orienteering event days, which contributed to the strong positive relationship between questing tick abundance and the two rates. Using a similar approach, Gilbert *et al.*, (2017) found strong correlations between questing tick density on pastures and counts of ticks biting lambs on the same pasture, demonstrating that cloth lure transects are a good proxy of risk to livestock of tick exposure and TBDs. Importantly for studies in humans, Chapter 4 included a measure of effort for estimating tick bite and tick encounter rates: the number of orienteers running in a group or as an individual per length of time of exposure (hour). These outputs were therefore more comparable with the outputs from questing tick surveys, which also include effort (e.g. tick counts per 10 m<sup>2</sup>), in contrast to the passive CS data used in Chapter 2 where no effort was measured.

**Chapter 5:** The final results chapter describes a novel CS approach for gathering data on human-tick encounters which included collection of denominator data (number of people exposed and time spent during the recorded outdoor activity), combined with information on when people did and, importantly, when they did not encounter ticks. These human-tick encounter data were compared with questing tick survey data. This data collection approach was validated in the previous chapter using orienteers and a targeted questing tick survey approach as a 'gold standard' case study to test the correlation between questing tick surveys and orienteer tick encounters. Chapter 5 applies this general approach to a broader setting of wider outdoor user groups and comparing their tick encounter rates with surveys on questing ticks conducted in the general area but not necessarily at the same time nor exactly the same place. Thus, Chapter 5 represents the sort of data that is easier to collect through websites and apps for wider scale CS projects. The approach described in Chapter 5 is therefore considered to have wider applicability than the intense and targeted approach used to create a gold

standard in Chapter 4. In order to address the objectives of this study, tick bite and tick encounter rates calculated from volunteer reports were compared with predictions of questing tick abundance in two study areas, using models based on data collected from questing tick surveys. This study showed a positive weak correlation between predicted questing tick abundance and tick bite and encounter rates. This weak correlation was likely due to two main reasons: 1) questing tick surveys were conducted in the same area but not exactly in the same locations used by volunteers, hence predictions of tick abundance were used to evaluate the correlation with tick bite and tick encounter rates from CS; and 2) results indicated that tick bite and tick encounter rates depend both on questing tick abundance, and on factors related to human activity and behaviour. Although conducted with a different objective, the study developed by Ripoche *et al.*, (2018) also showed that predictions from nymph density estimates (using data from active tick surveillance including from blanket dragging) were weakly correlated with ticks collected from humans (from passive submissions).

This study permitted the quantification of the relative contribution of tick abundance and factors related to tick and human activity to tick bite and tick encounter rates, which had not previously been assessed. Results indicated that factors relating to human behaviour, such as activity type, explained more variation in tick bite rate than tick abundance in the environment. Hence, the inherent characteristics of some activities make people more vulnerable to tick bites, such as the type of clothes worn, or interacting more with the environment (small scale interactions with tick habitats). These results are supported by the research conducted by Garcia-Martí *et al.*, (2017), which indicated that human-related factors (namely the number of warm and dry days and the proximity to forests) were more relevant to model tick bites, than climatic and land cover variables. In Chapters 4 and 5, although both tick bite and tick encounter rates were measured, future research should use tick bite rate in order to assess human risk where possible. This is because, whilst tick encounters happen more frequently and data can be collected in order to

engage citizens and/or to give more power to the statistical analyses, tick bite rate is a more robust measure (as it is not as variable as tick encounter rate), gives more information about the real risk (such as the identification of risky activities) and is more directly related to the potential risk of TBDs.

## **6.2. Main contributions and implications: what gaps have the results filled?**

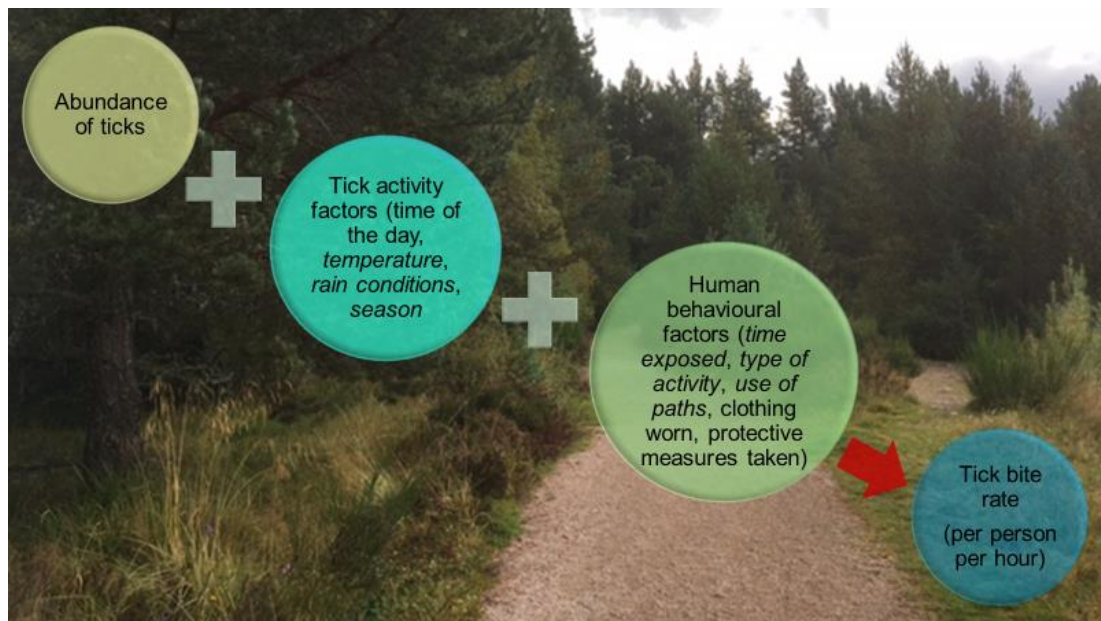
### **6.2.1. Scientific contributions**

One scientific contribution of this PhD was the generation of robust predictive models and maps of tick distribution (*I. ricinus*) in mainland Scotland, and in specific areas including Aberdeenshire, the Cairngorms and Lochaber (Chapters 2, 3 and 5). The only published map of *I. ricinus* distribution in Scotland is based on a mechanistic model (Li *et al.*, 2016); two other maps have been developed but are unpublished beyond postgraduate theses (Braga, 2012; Worton, 2016). Therefore, this thesis represents the most complete effort to accurately map *I. ricinus* distribution in Scotland. New covariates that affect *I. ricinus* distribution have been identified, a detailed spatial pattern for *I. ricinus* distribution in Scotland has been described (at country and local levels), and the uncertainty associated with the predictions has been measured. The sophisticated methodology used in Chapter 3 proved valuable in dealing with all sources of data variability, in improving the predicted uncertainty, and the predictions of *I. ricinus* for areas where no questing tick surveys have been carried out. It is likely that the map generated in Chapter 3 represents the best predictive map of *I. ricinus* distribution in Scotland, based on the data currently available. This was the first time the SPDE approach has been applied to tick distribution data, and the results demonstrated its value for a wide application in tick distribution modelling and mapping. Finally, this study contributed as an opportunity to test inlabru R package with real and complex ecological data. The dataset used in this study was used specifically to improve some features of the internal code of inlabru.

In this study, a novel CS approach to estimate tick bite and tick encounter rates was applied, which contributed towards decreasing the possibility of errors and bias in measuring these rates (Chapters 4 and 5). This approach includes accounting for negative and positive tick encounters, number of people and hours of exposure, and tracking volunteers. The submission of positive and negative tick encounters adds value to the rates estimated, as only this gives information on the likelihood of human contact with ticks, hence creating real estimates of human exposure and risk. In addition, this is the first study that directly assesses how outputs from questing tick surveys compare with tick bite and tick encounter rates (Chapters 4 and 5). This was fundamental to evaluate the link between tick ecology and the human risk of tick bites and tick encounters, and to quantify the components making up the risk of tick bites (Figure 6.1). This enabled conclusions to be reached that: 1) outputs from questing tick surveys are accurate proxies of human tick bite rate as long as factors that affect tick activity and human activity and behaviour are controlled (Chapter 4); 2) however, when it is impossible to measure questing tick abundance at the same time and place of human exposure, and when other factors also vary, tick bite and tick encounter rates depend not only on questing tick abundance in the environment, but on factors that affect tick activity and human behaviour (Chapter 5), and so these additional factors need to be recorded.

Tick occurrence or abundance therefore cannot be used alone to infer the human risk of tick encounters and tick bites. In order to predict and map tick bite risk, this study supports the future application of this type of CS approach. Predictions should include information on human behaviour factors (e.g. activity type and type of paths used), in addition to information on factors that affect tick activity and abundance (such as weather, climate, land cover and hosts) that can either be easily collected by the participant or using available spatial data sources (Figure 6.1 and Table 6.1). A measure of volunteer effort is essential to account for associated bias.





**Figure 6.1:** The components of tick bite rate, according to the studies carried out in Chapters 4 and 5. The main factors that affect tick activity and human behaviour are listed. I have highlighted (i.e. italic) those included in the studies carried out in Chapters 4 and 5.

### 6.2.2. Contributions for policy decision-making

The studies presented in this thesis are of direct relevance for policy decision-making in Scotland and also have strong application for other areas. For decision-making in Scotland, policy makers can adopt the spatial predictive map and associated predicted uncertainty generated in Chapter 3, since this study has demonstrated its quality. The study conducted in Chapter 4 is the first study to present an accurate measure of mean tick bite rate per 1,000 person-hours for orienteers, a group with high exposure to tick bites. In addition, this study identified that tick bite rate among orienteers varied with habitat composition and weather conditions. Therefore, these results can inform public health messaging to communicate risk and prevention strategies aiming to decrease the risk of tick bites in orienteers. This study can also help the identification of events likely to have higher rates of tick bite, which could allow for targeted measures to reduce tick bite risk (such as encouraging orienteers to wear clothing that covers the arms and legs, to apply arthropod-repellent skin cream, to check for ticks as soon as possible, and to have at the

event area a tent for tick checks). The CS project implemented in the whole of Scotland (Chapter 5) generated measures of tick bite rate for different outdoor activities and behaviours (such as the type of paths used in outdoor activity). This information can help target prevention and management measures for certain higher risk activities (e.g. orienteering, running and biking), and some human behaviours (i.e. not using clear footpaths). Citizens can identify which activities and/or behaviours may increase the risk of tick bites to better protect themselves.

The work presented in this thesis forms a good exemplar of the challenges of identifying an appropriate tick risk map for Scotland and has value in identifying lessons relevant to other studies. This thesis highlights how, as good scientific practice, tick data need be collected in specific ways to address a specific objective. The scenarios where different approaches are most appropriate are outlined in Table 6.1. Finally, the studies in Chapters 2 and 3 highlight that uncertainty needs to be communicated alongside predictive maps, to provide information on the reliability of predictions and help the process of decision-making. Knowledge of the predicted uncertainty can help policy makers select the most accurate predictive map, as well as indicate the necessity for further data, or identify areas that need further research, or areas where control and preventive measures need to be prioritised.

**Table 6.1:** Appropriate tick data to use to predict and map tick distribution and/or human risk of tick bites at local, regional or country levels.

Objective	Tick data	Observations
Predict and map tick distribution at local or regional levels	Data from questing tick surveys	Sites for questing tick surveys should be stratified for key covariates (for example vegetation cover) within the local area and/or region, and questing tick surveys should be conducted regularly within the year(s) (such as Boehnke <i>et al.</i> , 2015; Vourc'h <i>et al.</i> , 2016).
Predict and map tick distribution at country level	Data from questing tick surveys	Sites for questing tick surveys should be stratified for key covariates within the country, and questing tick surveys should be conducted regularly within the year(s) (such as Kjær <i>et al.</i> , 2019b). However, if resources are insufficient to generate good coverage, and other tick data are available, a combined dataset can be used instead (such as ECDC, 2019a).
Measure tick bite risk for one specific activity, including assessing differences in tick bite risk for different seasons and habitats	Calculate tick bite rate from data gathered through a CS project using volunteers of that activity	Citizen science projects can be targeted to specific activities to generate data on tick bite rates. In addition, data from questing tick surveys could also be used. For this it is necessary to sample areas frequently used for the activity, and questing tick surveys should be carried out every season during an approximate time of when the activity is generally carried out.
Predict and map tick bite risk at local, regional or country levels, including assessment of differences in tick bite risk for different activities, seasons, and exposure factors.	Calculate tick bite rate from data gathered through CS project on human tick bites	For spatial or spatial-temporal predictions of tick bite risk it will be appropriate to use information on factors that affect human activity and behaviour, and on factors that affect tick activity and abundance.

### 6.2.3. Public health contributions

The major public health contribution of the work developed in this PhD stem from the work carried out with volunteers (Chapters 4 and 5). These volunteers were orienteers and other citizens who often carried out outdoor activities in groups (e.g. work-related, friends or families) or as individuals. The reporting of activities both with and without tick encounters likely encouraged

participants to conduct checks for ticks and removal of crawling and attached ticks, decreasing the risk of infection after a tick bite. Engaging with the orienteering community through participation in orienteering events, writing articles for the orienteering newsletter, and promotion of tick research on orienteering websites has contributed to raising awareness for ticks and TBDs. During the attendance at orienteering events, orienteers often approached the 'tick tent' with queries and to express their perceptions on the risks involved. The full CS project received reports from the whole of Scotland. Engagement with the project is likely to have raised awareness about ticks and TBDs to people who may be exposed when carrying out outdoor activities. The results of this research will be disseminated to project volunteers.

### **6.3. Study limitations and future research**

One general limitation common to most of these studies is that spatial maps are usually static, whereas both tick activity and risk are dynamic, varying with factors such as time of day and influence of weather conditions. One possible method to make more dynamic maps is to develop spatial-temporal models and maps using weather and satellite daily data, to map daily forecasts of questing tick activity (such as Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017). Another possibility is to mask the base spatial map with weekly or daily layers of cumulative land surface temperature, based on the knowledge that ticks start to quest after a specific temperature threshold. This is particularly important before and after the winter break period of tick activity, which is perhaps more uncertain due to climate changes. However, there are two challenges associated with the environmental data with daily temporal resolution: 1) most of the daily layers have pixels with missing values, and 2) the lowest spatial resolution of currently available daily temperature data is 1 km<sup>2</sup>.

In addition, since risk is often related with decisions made at a very detailed scale (e.g. walking on a path or not), maps of tick distribution and tick bite risk could be rescaled at the finest resolution possible (e.g. using the fine resolution

of the vegetation type layers available). The resolution of maps needs to be adapted according to the objective of the predictions, and the intended audience. For country level decision-making, a resolution of 1 km<sup>2</sup> may often be enough to identify areas where interventions are most needed. However, for local level predictions aiming to help individual decision-making and to implement measures in small areas such as small patches of heavily used woodland, 1 km<sup>2</sup> may be too coarse, and 100 m<sup>2</sup> resolution (or less) may be necessary. Hence, predictive maps for vector distribution and risk will likely benefit from the availability of covariates with fine spatial resolution. The resolution and appearance of a map is very important because it can help to engage the users of the maps (e.g. policy makers, citizens), whilst also raising awareness about risk.

Although this work has presented novel statistical methods to model and map tick distribution, future research in this field can explore ways to decrease the uncertainty of the predictions associated with the data quality. This can be done by adding *a priori* information of habitat preferences into the model structure (Conn, Johnson and Boveng, 2015), or as a mask of unsuitable habitats and climatic conditions for the vector of interest. For example in the report from ECDC (2019a), a mask of unsuitable habitats for *I. ricinus* was overlaid on the final predictive distribution map, to highlight locations where ticks were predicted as present, but the scientific knowledge of experts stated the opposite. Another statistical methodology that has never been used on tick studies is to develop a complex hierarchical joint model which combines tick data from questing tick surveys with tick data from CS projects (such as developed by. This joint model examines two response variables, tick abundance and tick bite rate (for example), and the pattern of the spatial distribution of tick abundance will consider the information on tick bites (and vice-versa). Outputs of this model could be useful to improve maps of tick distribution and of human risk of tick bites. A similar approach was developed by Jones-Todd *et al.*, (2018) to predict the spatial distribution of three species

of birds in UK, and by Nelli, Ferguson and Matthiopoulos (2020) in modelling malaria in West Africa.

In this thesis, the importance of presenting predictive maps of tick distribution alongside the associated uncertainty was highlighted. However, the uncertainty generated was found to be correlated with the posterior mean of the predictions and did not include uncertainty associated with extrapolation outside the covariate range. This approach to measure uncertainty has been implemented in other studies with vectors (such as Garcia-Martí, Zurita-Milla, van Vliet, *et al.*, 2017; Lord *et al.*, 2018). Chapter 3 generated a more informative uncertainty, but the uncertainty estimated was still related with the posterior mean of the model predictions. I recommend that future studies should consider the development of statistical techniques to accurately quantify the uncertainty of the predictions. These techniques could be used to assess where predictions for certain areas are not reliable due to the lack of data, or to include the uncertainty associated with extrapolation outside the covariate range. There are some approaches already in use that could be investigated. The Mobility-Oriented Parity metric (Owens *et al.*, 2013) highlights areas in the maps where extrapolation (i.e. values outside of the range of environments in the reference region/time) occurs. This method was applied by Alkische *et al.*, (2017) to assess the uncertainty associated with extrapolation of the influences of climate change on the geographic distribution of *I. ricinus* in Europe.

It is highly likely that some of the ticks collected during this PhD would be infected with *Borrelia burgdorferi* (s.l.), the LB pathogens, and other microorganisms. From the study presented in Chapter 4, ticks are available both from questing tick surveys and from the orienteers themselves. A refinement of this study would be to quantify the prevalence of pathogenic agents causing human disease in ticks gathered from orienteers and from the vegetation. This would enable estimation of an infected tick bite rate (from the events where ticks were collected from orienteers), and to calculate the risk of

tick-borne pathogens in areas where orienteering activity is often carried out by clubs. From the study developed in Chapter 4, and from three orienteering events where ticks were specifically collected from orienteers (tick bites and crawling ticks), a total of 145 ticks have already been analysed by the Scottish Lyme Disease and Tick-borne infections Reference Laboratory (Inverness). Preliminary results showed that 2.8% of the ticks (4/145) were positive for *B. burgdorferi* (s.l.) and 13.8% (20/145) were positive for *Anaplasma phagocytophilum*. However there was insufficient funding and time to allow further testing to be undertaken, and so these results have not been included in this thesis. From Chapter 5, the knowledge of the prevalence of *B. burgdorferi* (s.l.) (for example) in my survey sites in the Cairngorms and Lochaber, would provide information about the environmental hazard of infection in these areas, which could help target management and preventive measures. The prevalence of *B. burgdorferi* (s.l.) in ticks is very variable due to the complex nature of tick ecology and the relationship with multiple host species. Therefore, more data are needed to understand the mechanisms driving risk between the different areas I studied.

One limitation associated with the CS project presented in Chapter 5 is the challenge of volunteer engagement, and its potential impact on data representativeness. Due to this, data from CS were not used to make spatial predictions of tick bite and tick encounter rates in the Cairngorms and in Lochaber. Hence, the quality of this novel CS approach for predictive mapping of tick bite and tick encounter rates was not assessed. Another refinement of this study would be to model and map tick bite and tick encounter rates using the CS data submitted from all over Scotland. The predictive model will add information on other covariates, such as climate and habitat type. This is fundamental to identify risk factors for the Scottish context. Although CS data collection is inherently biased towards areas which are more frequently visited, this can also provide a benefit in identifying high use areas that are also at high risk of tick bites, if enough data are collected. Therefore, these results would

constitute a valuable tool to be used by policy makers, in planning and improving tick and disease control strategies in Scotland.

The CS project implemented in this PhD highlighted that volunteers can provide valuable data to estimate tick bite rate, including reporting of zero values, accounting for the time spent, volunteer effort and tracking of participants. This new methodology can be adopted by other CS projects to improve the quality of CS data. However, future research is needed to compare the accuracy of risk measures calculated with and without these recorded negative encounters. This would enable quantification of the effect of bias towards positive reports, and the importance of this approach. In addition, more research is needed to study ways to increase people's engagement with this novel CS approach. This could be done by identifying efficient 'gaming' strategies (such as rewarding volunteers with badges, identification of volunteers who contribute most reports), or other strategies to increase the likelihood that volunteers consistently report all data, even when no ticks are observed (for example volunteers could be randomly allocated to report activities just one time every week or every month). Last but not least, this study opens the doors to research in the field of social science to address how to better communicate tick and TBD risk to policy makers, managers and citizens in the form of risk maps (including the respective uncertainty) as well as improving understanding of the relationship between people's perceptions about tick bite risks and their use of the outdoors.

In summary, the work developed in this thesis contains unique data on tick bite rate from CS and constitutes a gateway to improve our understanding of drivers of people's risk of tick bites. In addition, this work presents new methodologies to improve predictive mapping of ticks and tick bite risk to help decrease the number of tick bites and the risk of TBDs. Understanding the distribution of ticks and tick bite risk is only likely to become more important with issues such as climate change increasing both tick distribution and the



subsequent increased likelihood of interactions between vectors and humans for disease transmission.

## References

- Abdullah, S., Helps, C., Tasker, S., Newbury, H. and Wall, R. (2016) 'Ticks infesting domestic dogs in the UK: a large-scale surveillance programme', *Parasites & Vectors*, 9(391), pp. 1–9. doi: 10.1186/s13071-016-1673-4.
- Aguilar-Vega, C., Fernández-Carrión, E., Lucientes, J. and Sánchez-Vizcaíno, J. M. (2020) 'A model for the assessment of bluetongue virus serotype 1 persistence in Spain', *PLoS ONE*, 15(4), pp. 1–22. doi: 10.1371/journal.pone.0232534.
- Albon, S., McLeod, J., Potts, J., Brewer, M., Irvine, J., Towers, M., Elston, D., Fraser, D. and Irvine, R. (2017) *Estimating national trends and regional differences in red deer density on open-hill ground in Scotland: identifying the causes of change and consequences for upland habitats, Scottish Natural Heritage Commissioned , report no. 981*. Available at: [https://www.nature.scot/sites/default/files/2017-11/Publication 2017 - SNH Commissioned Report 981 - Estimating national trends and regional differences in red deer.pdf](https://www.nature.scot/sites/default/files/2017-11/Publication%202017%20-%20SNH%20Commissioned%20Report%20981%20-%20Estimating%20national%20trends%20and%20regional%20differences%20in%20red%20deer.pdf) (Accessed: 18 October 2018).
- Alexander, N. S., Morley, D., Medlock, J., Searle, K. and Wint, W. (2014) 'A First Attempt at Modelling Roe Deer (*Capreolus capreolus*) Distributions Over Europe', *Open Health Data*, 2(1), pp. 1–5. doi: 10.5334/ohd.ah.
- Alfredsson, M., Olafsson, E., Eydal, M., Unnsteinsdottir, E. R., Hansford, K. M., Wint, W., Alexander, N. and Medlock, J. M. (2017) 'Surveillance of *Ixodes ricinus* ticks (Acari: Ixodidae) in Iceland', *Parasites & Vectors*, 10(466), pp. 1–11. doi: 10.1186/s13071-017-2375-2.
- Alkische, A. A., Peterson, A. T. and Samy, A. M. (2017) 'Climate change influences on the potential geographic distribution of the disease vector tick *Ixodes ricinus*', *PLoS ONE*, 12(12), pp. 1–14. doi: 10.1371/journal.pone.0189092.
- Anderson, J. F. and Magnarelli, L. A. (2008) 'Biology of Ticks', *Infectious Disease Clinics of North America*, 22(2), pp. 195–215. doi: 10.1016/j.idc.2007.12.006.
- Andriopoulos, P., Economopoulou, A., Spanakos, G. and Assimakopoulos, G. (2013) 'A local outbreak of autochthonous *Plasmodium vivax* malaria in Laconia, Greece—a re-emerging infection in the southern borders of Europe?', *International Journal of Infectious Diseases*, 17(2), pp. 125–128. doi: 10.1016/j.ijid.2012.09.009.
- Antonise-Kamp, L., Beaujean, D. J. M. A., Crutzen, R., van Steenberg, J. E. and Ruwaard, D. (2017) 'Prevention of tick bites: An evaluation of a smartphone app', *BMC Infectious Diseases*, 17(744), pp. 1–15. doi: 10.1186/s12879-017-2836-4.
- Araújo, M. B. and New, M. (2007) 'Ensemble forecasting of species distributions.', *Trends in Ecology and Evolution*, 22(1), pp. 42–47. doi: 10.1016/j.tree.2006.09.010.

- Armstrong, J. S. (2001) 'Evaluating Forecasting Methods, Chapter 14', in J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Norwell: Kluwer Academic Publishers. doi: 10.1007/978-0-306-47630-3\_20.
- Asingizwe, D., Murindahabi, M. M., Koenraad, C. J. M., Poortvliet, P. M., van Vliet, A. J. H., Ingabire, C. M., Hakizimana, E., Mutesa, L., Takken, W. and Leeuwis, C. (2019) 'Co-Designing a Citizen Science Program for Malaria Control in Rwanda', *Sustainability*, 11(24), pp. 1–17. doi: 10.3390/su11247012.
- Asingizwe, D., Poortvliet, P. M., van Vliet, A. J. H., Koenraad, C. J. M., Ingabire, C. M., Mutesa, L. and Leeuwis, C. (2020) 'What do people benefit from a citizen science programme? Evidence from a Rwandan citizen science programme on malaria control', *Malaria Journal*, 19(283), pp. 1–13. doi: 10.1186/s12936-020-03349-8.
- Atkinson, S. F., Sarkar, S., Aviña, A., Schuermann, J. A. and Williamson, P. (2014) 'A determination of the spatial concordance between Lyme disease incidence and habitat probability of its primary vector *Ixodes scapularis* (black-legged tick)', *Geospatial Health*, 9(1), pp. 203–212. doi: 10.4081/gh.2014.17.
- Attali, D. (2020) 'shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds.' R package version 2.0.0.
- August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T. and Jepson, P. (2015) 'Emerging technologies for biological recording', *Biological Journal of the Linnean Society*, 115, pp. 731–749. doi: 10.1111/bij.12534.
- Azagi, T., Hoornstra, D., Kremer, K., Hovius, J. W. R. and Sprong, H. (2020) 'Evaluation of disease causality of rare ixodes ricinus-borne infections in Europe', *Pathogens*, 9(150), pp. 1–22. doi: 10.3390/pathogens9020150.
- Bachl, F. E., Lindgren, F., Borchers, D. L. and Illian, J. B. (2019) 'inlabru: an R package for Bayesian spatial modelling from ecological survey data', *Methods in Ecology and Evolution*, 10(6), pp. 760–766. doi: 10.1111/2041-210X.13168.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. and Lindgren, F. (2018) 'Spatial modeling with R-INLA: A review', *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), pp. 1–33. doi: 10.1002/wics.1443.
- Barbet-Massin, M., Jiguet, F., Albert, C. H. and Thuiller, W. (2012) 'Selecting pseudo-absences for species distribution models: How, where and how many?', *Methods in Ecology and Evolution*, 3(2), pp. 327–338. doi: 10.1111/j.2041-210X.2011.00172.x.
- Barry, S. and Elith, J. (2006) 'Error and uncertainty in habitat models', *Journal of Applied Ecology*, 43(3), pp. 413–423. doi: 10.1111/j.1365-2664.2006.01136.x.
- Bartosik, K., Sitarz, M., Szymańska, J. and Buczek, A. (2011) 'Tick bites on humans in the agricultural and recreational areas in south-eastern Poland.', *Annals of Agricultural and Environmental Medicine*, 18, pp. 151–157. doi: 10.1080/085640032000089762.

- Bartumeus, F., Oltra, A. and Palmer, J. R. B. (2018) 'Citizen Science: A Gateway for Innovation in Disease-Carrying Mosquito Management?', *Trends in Parasitology*, 34(9), pp. 727–729. doi: 10.1016/j.pt.2018.04.010.
- Beerling, D. J., Huntley, B. and Bailey, J. P. (1995) 'Climate and the distribution of *Fallopia japonica* : use of an introduced species to test the predictive capacity of response surfaces', *Journal of Vegetation Science*, 6(2), pp. 269–282. doi: 10.2307/3236222.
- Beguín, J., Martino, S., Rue, H. and Cumming, S. G. (2012) 'Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation', *Methods in Ecology and Evolution*, 3(5), pp. 921–929. doi: 10.1111/j.2041-210X.2012.00211.x.
- Bell, A. R., Ward, P. S., Killilea, M. E. and Tamal, E. H. (2016) 'Real-Time Social Data Collection in Rural Bangladesh via a "Microtasks for Micropayments" Platform on Android Smartphones', *PLoS ONE*, 11(11), pp. 1–14. doi: 10.1371/journal.pone.0165924.
- Berglund, J., Eitrem, R., Ornstein, K., Lindberg, A., Ringnér, A., Elmrud, H., Carlsson, M., Runehegen, A., Svanborg, C. and Norrby, R. (1995) 'An epidemiologic study of lyme disease in southern sweden', *New England Journal of Medicine*, 333(20), pp. 1319–1324. doi: 10.1056/NEJM199511163332004.
- Bern, C., Kjos, S., Yabsley, M. J. and Montgomery, S. P. (2011) 'Trypanosoma cruzi and chagas' disease in the United States', *Clinical Microbiology Reviews*, 24(4), pp. 655–681. doi: 10.1128/CMR.00005-11.
- Bernardo, J. M. and Smith, A. F. M. (2000) *Bayesian Theory*. 1st edn. Chichester, West Sussex: John Wiley & Sons Ltd.
- Beugnet, F., Chalvet-Monfray, K. and Loukos, H. (2009) 'FleaTickRisk: A meteorological model developed to monitor and predict the activity and density of three tick species and the cat flea in Europe', *Geospatial Health*, 4(1), pp. 97–113. doi: 10.4081/gh.2009.213.
- Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N. and Frusher, S. (2014) 'Statistical solutions for error and bias in global citizen science datasets', *Biological Conservation*, 173(2014), pp. 144–154. doi: 10.1016/j.biocon.2013.07.037.
- Bisanzio, D., Giacobini, M., Bertolotti, L., Mosca, A., Balbo, L., Kitron, U. and Vazquez-Prokopec, G. M. (2011) 'Spatio-temporal patterns of distribution of West Nile virus vectors in eastern Piedmont Region, Italy', *Parasites & Vectors*, 4(230), pp. 1–11. doi: 10.1186/1756-3305-4-230.
- Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Rouault, E., Warmerdam, F., Ooms, J. and Rundel, C. (2020) 'rgdal: Bindings for the "Geospatial" Data Abstraction Library.' R package version 1.5-18.
- Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K. O., Giraudoux, P., Davis, M. and Santilli, S. (2020) 'rgeos: Interface to

- Geometry Engine - Open Source ('GEOS').' R package version 0.5-5.
- Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013) 'Spatial and spatio-temporal models with R-INLA.', *Spatial and spatio-temporal epidemiology*, 4(2013), pp. 33–49. doi: 10.1016/j.sste.2012.12.001.
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-Qing, D., Clark, N. E., O'Connor, K. and Mace, G. M. (2010) 'Distorted views of biodiversity: Spatial and temporal bias in species occurrence data', *PLoS Biology*, 8(6), pp. 1–11. doi: 10.1371/journal.pbio.1000385.
- Boeckmann, M. and Joyner, T. A. (2014) 'Old health risks in new places? An ecological niche model for I. ricinus tick distribution in Europe under a changing climate', *Health and Place*, 30(2014), pp. 70–77. doi: 10.1016/j.healthplace.2014.08.004.
- Boehnke, D., Brugger, K., Pfäffle, M., Sebastian, P., Norra, S., Petney, T., Oehme, R., Littwin, N., Lebl, K., Raith, J., Walter, M., Gebhardt, R. and Rubel, F. (2015) 'Estimating Ixodes ricinus densities on the landscape scale', *International Journal of Health Geographics*, 14(23), pp. 1–12. doi: 10.1186/s12942-015-0015-7.
- Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J. and Wilderman, C. C. (2009) *Public Participation in Scientific Research: defining the field and assessing its potential for science education., A CAISE Inquiry Group Report*. Washington, D.C. Available at: [www.caise.insci.org](http://www.caise.insci.org) (Accessed: 19 April 2018).
- Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J. and Parrish, J. K. (2014) 'Citizen science: Next steps for citizen science', *Science*, 343(6178), pp. 1436–1437. doi: 10.1126/science.1251554.
- Bouchard, C., Aenishaenslin, C., Rees, E. E., Koffi, J. K., Pelcat, Y., Ripoche, M., Milord, F., Lindsay, L. R., Ogden, N. H. and Leighton, P. A. (2018) 'Integrated social-behavioral and ecological risk maps to prioritize local public health responses to lyme disease', *Environmental Health Perspectives*, 126(4), pp. 1–13. doi: 10.1289/EHP1943.
- Bowser, A., Hansen, D., He, Y., Boston, C., Reid, M., Gunnell, L. and Preece, J. (2013) 'Using Gamification to Inspire New Citizen Science Volunteers', in *Gamification '13: Proceedings of the First International Conference on Gameful Design, Research, and Applications*. New York, NY, USA: Association for Computing Machinery, pp. 18–25. doi: 10.1145/2583008.2583011.
- Braga, J. (2012) *Predicting current and future tick abundance across Scotland*. Master thesis. University of Aberdeen.
- Braks, M., Mulder, A., Swart, A. and Wint, W. (2016) 'Grasping risk mapping', in Braks, M., Wieren, van S., Takken, W., and Sprong, H. (eds) *Ecology and prevention of Lyme borreliosis*. Wageningen Academic Publishers, pp. 351–371. doi: 10.3920/978-90-8686-838-4.
- Breiman, L. (2001) 'Random forests', *Machine Learning*, 45, pp. 5–32. doi: 10.1201/9780367816377-11.
- Briciu, V. T., Titilincu, A., Țățulescu, D. F., Cârstina, D., Lefkaditis, M. and Mihalca, A. D. (2011) 'First survey on hard ticks (Ixodidae) collected from humans in Romania: Possible risks for tick-borne diseases',

- Experimental and Applied Acarology*, 54(2), pp. 199–204. doi: 10.1007/s10493-010-9418-0.
- Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011) *Handbook of Markov chain Monte Carlo*. 1st edn. Edited by S. Brooks, A. Gelman, G. Jones, and X. Meng. CRC Press, Taylor & Francis Group.
- Brownstein, J. S., Holford, T. R. and Fish, D. (2003) 'A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States', *Environmental Health Perspectives*, 111(9), pp. 1152–1157. doi: 10.1289/ehp.6052.
- Brugger, K., Boehnke, D., Petney, T., Dobler, G., Pfeffer, M., Silaghi, C., Schaub, G. A., Pinior, B., Dautel, H., Kahl, O., Pfister, K., Süss, J. and Rubel, F. (2016) 'A density map of the tick-borne encephalitis and Lyme borreliosis vector *Ixodes ricinus* (Acari: Ixodidae) for Germany', *Journal of Medical Entomology*, 53(6), pp. 1292–1302. doi: 10.1093/jme/tjw116.
- Burgdorfer, W., Barbour, A., Hayes, S., Péter, O. and Aeschlimann, A. (1983) 'Erythema chronicum migrans -a tickborne spirochetosis.', *Acta tropica*, 40(1), pp. 79–83.
- Cadenas, F. M., Rais, O., Jouda, F., Douet, V., Humair, P., Moret, J. and Gern, L. (2007) 'Phenology of *Ixodes ricinus* and Infection with *Borrelia burgdorferi* sensu lato Along a North- and South-Facing Altitudinal Gradient on Chaumont Mountain, Switzerland', *Journal of Medical Entomology*, 44(4), pp. 683–693. doi: 10.1093/jmedent/44.4.683.
- Cairns, V., Wallenhorst, C., Rietbrock, S. and Martinez, C. (2019) 'Incidence of Lyme disease in the UK: A population-based cohort study', *BMJ Open*, 9(8), pp. 1–8. doi: 10.1136/bmjopen-2018-025916.
- Calisher, C. H. (1988) 'Antigenic classification and taxonomy of flaviviruses (family Flaviviridae): emphasizing a universal system for the taxonomy of viruses causing tick-borne encephalitis.', *Acta virologica*, 32(5), pp. 469–478.
- Callcutt, K., Croft, S. and Smith, G. C. (2018) 'Predicting population trends using citizen science data: do subsampling methods produce reliable estimates for mammals?', *European Journal of Wildlife Research*, 64(28), pp. 1–6. doi: 10.1007/s10344-018-1189-7.
- Cameletti, M., Lindgren, F., Simpson, D. and Rue, H. (2012) 'Spatio-temporal modeling of particulate matter concentration through the SPDE approach', *AStA Advances in Statistical Analysis*, 97(2), pp. 109–131. doi: 10.1007/s10182-012-0196-3.
- Carvalho, B. M., Rangel, E. F., Ready, P. D. and Vale, M. M. (2015) 'Ecological niche modelling predicts southward expansion of *Lutzomyia* (*Nyssomyia*) *flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania* (*Leishmania*) *amazonensis* in South America, under climate change', *PLoS ONE*, 10(11), pp. 1–21. doi: 10.1371/journal.pone.0143282.
- Cayol, C., Koskela, E., Mappes, T., Siukkola, A. and Kallio, E. R. (2017) 'Temporal dynamics of the tick *Ixodes ricinus* in northern Europe: epidemiological implications', *Parasites & Vectors*, 10(166), pp. 1–11.

- doi: 10.1186/s13071-017-2112-x.
- CEH (2017) *Land Cover Map 2015 (Dataset documentation)*. Centre for Ecology & Hydrology. doi: 10.5285/bb15e200-9349-403c-bda9-b430093807c7.
- Chalghaf, B., Chemkhi, J., Mayala, B., Harrabi, M., Benie, G. B., Michael, E. and Ben Salah, A. (2018) 'Ecological niche modeling predicting the potential distribution of *Leishmania* vectors in the Mediterranean basin: Impact of climate change', *Parasites & Vectors*, 11(461), pp. 1–9. doi: 10.1186/s13071-018-3019-x.
- Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A. and Turak, E. (2017) 'Contribution of citizen science towards international biodiversity monitoring', *Biological Conservation*, 213(2017), pp. 280–294. doi: 10.1016/j.biocon.2016.09.004.
- Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2020) 'shiny: Web Application Framework for R.' R package version 1.5.0.
- Chauhan, G., McClure, J., Hekman, J., Marsh, P. W., Bailey, J. A., Daniels, R. F., Genreux, D. P. and Karlsson, E. K. (2020) 'Combining Citizen Science and Genomics to Investigate Tick, Pathogen, and Commensal Microbiome at Single-Tick Resolution', *Frontiers in Genetics*, 10(1322), pp. 1–12. doi: 10.3389/fgene.2019.01322.
- Cheng, J., Karambelkar, B., Xie, Y., Wickham, H., Russell, K., Johnson, K., Schloerke, B. and Agafonkin, V. (2019) 'leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet.' R package version 2.0.3.
- Chiaravalloti-Neto, F., Da Silva, R. A., Zini, N., Da Silva, G. C. D., Da Silva, N. S., Parra, M. C. P., Dibo, M. R., Estofolete, C. F., Fávares, E. A., Dutra, K. R., Mota, M. T. O., Guimarães, G. F., Terzian, A. C. B., Blangiardo, M. and Nogueira, M. L. (2019) 'Seroprevalence for dengue virus in a hyperendemic area and associated socioeconomic and demographic factors using a cross-sectional design and a geostatistical approach, state of São Paulo, Brazil', *BMC Infectious Diseases*, 19(441), pp. 1–14. doi: 10.1186/s12879-019-4074-4.
- Cisak, E., Chmielewska-Badora, J., Zwolinski, J., Wojcik-Fatla, A., Polak, J. and Dutkiewicz, J. (2005) 'Risk of tick-borne bacterial diseases among workers of Roztocze National Park (south-eastern Poland)', *Annals of Agricultural and Environmental Medicine*, 12(1), pp. 127–132. doi: 10.1097/00124743-200208000-00004.
- Ciss, M., Biteye, B., Fall, A. G., Fall, M., Gahn, M. C. B., Leroux, L. and Apolloni, A. (2019) 'Ecological niche modelling to estimate the distribution of *Culicoides*, potential vectors of bluetongue virus in Senegal', *BMC Ecology*, 19(45), pp. 1–12. doi: 10.1186/s12898-019-0261-9.
- Clutton-Brock, T., Coulson, T. and Milner, J. (2004) 'Red deer stocks in the Highlands of Scotland', *Nature*, 429(6989), pp. 11–12. doi: 10.1038/nature02616.
- Conn, P. B., Johnson, D. S. and Boveng, P. L. (2015) 'On extrapolating past the range of observed data when making statistical predictions in ecology', *PLoS ONE*, 10(10), pp. 1–16. doi:

10.1371/journal.pone.0141416.

- Cooper, C. (2017) *Citizen Science: how ordinary people are changing the face of discovery*. 1st edn. London: Duckworth Publishers.
- Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J. and Waller, D. M. (2011) 'Assessing citizen science data quality: An invasive species case study', *Conservation Letters*, 4(6), pp. 433–442. doi: 10.1111/j.1755-263X.2011.00196.x.
- Cuéllar, A. C., Jung Kjær, L., Baum, A., Stockmarr, A., Skovgard, H., Nielsen, S. A., Andersson, M. G., Lindström, A., Chirico, J., Lühken, R., Steinke, S., Kiel, E., Gethmann, J., Conraths, F. J., Larska, M., Smreczak, M., Orłowska, A., Hamnes, I., Sviland, S., Hopp, P., Brugger, K., Rubel, F., Balenghien, T., Garros, C., Rakotoarivony, I., Allène, X., Lhoir, J., Chavernac, D., Delécolle, J. C., Mathieu, B., Delécolle, D., Setier-Rio, M. L., Venail, R., Scheid, B., Chueca, M. Á. M., Barceló, C., Lucientes, J., Estrada, R., Mathis, A., Tack, W. and Bødker, R. (2018) 'Monthly variation in the probability of presence of adult Culicoides populations in nine European countries and the implications for targeted surveillance', *Parasites & Vectors*, 11(1), pp. 1–19. doi: 10.1186/s13071-018-3182-0.
- Cull, B., Pietzsch, M. E., Gillingham, E. L., Mcginley, L., Medlock, J. M. and Hansford, K. M. (2019) 'Seasonality and anatomical location of human tick bites in the United Kingdom', *Zoonoses and Public Health*, 00, pp. 1–10. doi: 10.1111/zph.12659.
- Cull, B., Pietzsch, M. E., Hansford, K. M., Gillingham, E. L. and Medlock, J. M. (2018) 'Surveillance of British ticks: An overview of species records, host associations, and new records of Ixodes ricinus distribution', *Ticks and Tick-borne Diseases*, 9(3), pp. 605–614. doi: 10.1016/j.ttbdis.2018.01.011.
- Curtis-Robles, R., Hamer, S. A., Lane, S., Levy, M. Z. and Hamer, G. L. (2018) 'Bionomics and spatial distribution of triatomine vectors of trypanosoma cruzi in Texas and other Southern States, USA', *American Journal of Tropical Medicine and Hygiene*, 98(1), pp. 113–121. doi: 10.4269/ajtmh.17-0526.
- Curtis-Robles, R., Wozniak, E. J., Auckland, L. D., Hamer, G. L. and Hamer, S. A. (2015) 'Combining Public Health Education and Disease Ecology Research: Using Citizen Science to Assess Chagas Disease Entomological Risk in Texas', *PLoS Neglected Tropical Diseases*, 9(12), pp. 1–12. doi: 10.1371/journal.pntd.0004235.
- van Dam, A. P., Kuiper, H., Vos, K., Widjojokusumo, A., de Jongh, B. M., Spanjaard, L., Ramselaar, A. C. P., Kramer, M. D. and Dankert, J. (1993) 'Different Genospecies of Borrelia burgdorferi Are Associated with Distinct Clinical Manifestations of Lyme Borreliosis', *Clinical Infectious Diseases*, 17(4), pp. 708–717. doi: 10.1093/clinids/17.4.708.
- Daniel, M., Danielová, V., Kříž, B., Jirsa, A. and Nožička, J. (2003) 'Shift of the tick Ixodes ricinus and tick-borne encephalitis to higher altitudes in Central Europe', *European Journal of Clinical Microbiology and Infectious Diseases*, 22(5), pp. 327–328. doi: 10.1007/s10096-003-0918-2.



- Danielsen, F., Jensen, P. M., Burgess, N. D., Altamirano, R., Alviola, P. A., Andrianandrasana, H., Brashares, J. S., Burton, A. C., Coronado, I., Corpuz, N., Enghoff, M., Fjeldså, J., Funder, M., Holt, S., Hübertz, H., Jensen, A. E., Lewis, R., Massao, J., Mendoza, M. M., Ngaga, Y., Pipper, C. B., Poulsen, M. K., Rueda, R. M., Sam, M. K., Skielboe, T., Sørensen, M. and Young, R. (2014) 'A Multicountry Assessment of Tropical Resource Monitoring by Local Communities', *BioScience*, 64(3), pp. 236–251. doi: 10.1093/biosci/biu001.
- Dawid, A. P. and Sebastiani, P. (1999) 'Coherent dispersion criteria for optimal experimental design', *Annals of Statistics*, 27(1), pp. 65–81. doi: 10.1214/aos/1018031101.
- Demicheli, V., Debalini, M. and Rivetti, A. (2009) 'Vaccines for preventing tick-borne encephalitis (Review)', *Cochrane Database of Systematic Reviews*, (1), pp. 1–38. doi: 10.1002/14651858.CD000977.pub2.www.cochranelibrary.com.
- Diarra, M., Fall, M., Fall, A. G., Diop, A., Lancelot, R., Seck, M. T., Rakotoarivony, I., Allène, X., Bouyer, J. and Guis, H. (2018) 'Spatial distribution modelling of Culicoides (Diptera: Ceratopogonidae) biting midges, potential vectors of African horse sickness and bluetongue viruses in Senegal', *Parasites & Vectors*, 11(341), pp. 1–15. doi: 10.1186/s13071-018-2920-7.
- Dickinson, E. R., Millins, C. and Biek, R. (2020) 'Sampling scale and season influence the observed relationship between the density of deer and questing Ixodes ricinus nymphs', *Parasites & Vectors*, 13(493), pp. 1–11. doi: 10.1186/s13071-020-04369-8.
- Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T. and Purcell, K. (2012) 'The current state of citizen science as a tool for ecological research and public engagement', *Frontiers in Ecology and the Environment*, 10(6), pp. 291–297. doi: 10.1890/110236.
- Dickinson, J. L., Zuckerberg, B. and Bonter, D. N. (2010) 'Citizen science as an ecological research tool: Challenges and benefits', *Annual Review of Ecology, Evolution, and Systematics*, 41, pp. 149–172. doi: 10.1146/annurev-ecolsys-102209-144636.
- Diuk-Wasser, M. A., Hoen, A. G., Cislo, P., Brinkerhoff, R., Hamer, S. A., Rowland, M., Cortinas, R., Vourc'h, G., Melton, F., Hickling, G. J., Tsao, J. I., Bunikis, J., Barbour, A. G., Kitron, U., Piesman, J. and Fish, D. (2012) 'Human risk of infection with Borrelia burgdorferi, the lyme disease agent, in Eastern United States', *American Journal of Tropical Medicine and Hygiene*, 86(2), pp. 320–327. doi: 10.4269/ajtmh.2012.11-0395.
- Diuk-Wasser, M. A., Vourc'h, G., Cislo, P., Hoen, A. G., Melton, F., Hamer, S. A., Rowland, M., Cortinas, R., Hickling, G. J., Tsao, J. I., Barbour, A. G., Kitron, U., Piesman, J. and Fish, D. (2010) 'Field and climate-based model for predicting the density of host-seeking nymphal Ixodes scapularis, an important vector of tick-borne disease agents in the eastern United States', *Global Ecology and Biogeography*, 19(4), pp. 504–514. doi: 10.1111/j.1466-8238.2010.00526.x.

- Diza, E., Papa, A., Vezyri, E., Tsounis, S., Milonas, I. and Antoniadis, A. (2004) 'Borrelia valaisiana in cerebrospinal fluid', *Emerging Infectious Diseases*, 10(9), pp. 1692–1693. doi: 10.3201/eid1009.030439.
- Dobson, A. D. M., Finnie, T. J. R. and Randolph, S. E. (2011) 'A modified matrix model to describe the seasonal population ecology of the European tick *Ixodes ricinus*', *Journal of Applied Ecology*, 48(4), pp. 1017–1028. doi: 10.1111/j.1365-2664.2011.02003.x.
- Dobson, A. D. M. and Randolph, S. E. (2011) 'Modelling the effects of recent changes in climate, host density and acaricide treatments on population dynamics of *Ixodes ricinus* in the UK', *Journal of Applied Ecology*, 48(4), pp. 1029–1037. doi: 10.1111/j.1365-2664.2011.02004.x.
- Dobson, A. D. M., Taylor, J. L. and Randolph, S. E. (2011) 'Tick (*Ixodes ricinus*) abundance and seasonality at recreational sites in the UK: Hazards in relation to fine-scale habitat types revealed by complementary sampling methods', *Ticks and Tick-borne Diseases*, 2(2), pp. 67–74. doi: 10.1016/j.ttbdis.2011.03.002.
- Domşa, C., Mihalca, A. D. and Sándor, A. D. (2018) 'Modeling the distribution of *Ixodes ricinus* in Romania', *North-Western Journal of Zoology*, 14(1), pp. 25–29.
- Ducheyne, E., Chueca, M. A. M., Lucientes, J., Calvete, C., Estrada, R., Boender, G. J., Goossens, E., De Clercq, E. M. and Hendrickx, G. (2013) 'Abundance modelling of invasive and indigenous Culicoides species in Spain', *Geospatial Health*, 8(1), pp. 241–254. doi: 10.4081/gh.2013.70.
- Ducheyne, E., Tran Minh, N. N., Haddad, N., Bryssinckx, W., Buliva, E., Simard, F., Malik, M. R., Charlier, J., De Waele, V., Mahmoud, O., Mukhtar, M., Bouattour, A., Hussain, A., Hendrickx, G. and Roiz, D. (2018) 'Current and future distribution of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in WHO Eastern Mediterranean Region', *International Journal of Health Geographics*, 17(4), pp. 1–13. doi: 10.1186/s12942-018-0125-0.
- Dumler, J. S., Barbet, A. F., Bekker, C. P. J., Dasch, G. A., Palmer, G. H., Ray, S. C., Rikihisa, Y. and Rurangirwa, F. R. (2001) 'Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of Ehrlichia with Anaplasma, Cowdria with Ehrlichia and Ehrlichia with Neorickettsia, descriptions of six new species combi', *International Journal of Systematic and Evolutionary Microbiology*, 51, pp. 2145–2165. doi: 10.1099/00207713-51-6-2145.
- ECDC (2012) *Second expert consultation on TBD with emphasis on Lyme borreliosis and tick-borne encephalitis*. Stockholm. Available at: [http://www.esamid.org/research\\_projects/study\\_groups/esgbor/](http://www.esamid.org/research_projects/study_groups/esgbor/) (Accessed: 10 January 2019).
- ECDC (2019a) *A spatial modelling method for vector surveillance*. Stockholm. doi: 10.2900/633757.
- ECDC (2019b) *Tick-borne encephalitis. Annual Epidemiological Report for 2017*. Stockholm. Available at:

- <https://www.ecdc.europa.eu/en/publications-data/tick-borne-encephalitis-annual-epidemiological-report-2017> (Accessed: 10 July 2020).
- ECDC and EFSA (2014) *VectorNet: A European network for sharing data on the geographic distribution of arthropod vectors, transmitting human and animal disease agents.*, Stockholm. Available at: <https://vectornet.ecdc.europa.eu/> (Accessed: 10 January 2018).
- ECDC and EFSA (2018) *The importance of vector abundance and seasonality. Results from an expert consultation.* Stockholm. doi: 10.2900/37171.
- EEA (1995) *CORINE land cover.* Copenhagen. Available at: <https://www.eea.europa.eu/publications/COR0-landcover> (Accessed: 12 March 2020).
- EEA (2014) *CORINE land cover 2006 raster data.* Available at: <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3> (Accessed: 9 November 2019).
- EEA (2018) *CORINE land cover 2018 raster data.* Available at: <https://land.copernicus.eu/pan-european/corine-land-cover> (Accessed: 11 June 2020).
- Ehm, W. and Gneiting, T. (2012) 'Local proper scoring rules of order two', *Annals of Statistics*, 40(1), pp. 609–637. doi: 10.1214/12-AOS973.
- Ehrmann, S., Liira, J., Gärtner, S., Hansen, K., Brunet, J., Cousins, S. A. O., Deconchat, M., Decocq, G., Frenne, P., Smedt, P., Diekmann, M., Gallet-Moron, E., Kolb, A., Lenoir, J., Lindgren, J., Naaf, T., Paal, T., Valdés, A., Verheyen, K., Wulf, M. and Scherer-Lorenzen, M. (2017) 'Environmental drivers of Ixodes ricinus abundance in forest fragments of rural European landscapes', *BMC Ecology*, 17(31), pp. 1–14. doi: 10.1186/s12898-017-0141-0.
- Eidsvik, J., Finley, A. O., Banerjee, S. and Rue, H. (2012) 'Approximate Bayesian inference for large spatial datasets using predictive process models', *Computational Statistics and Data Analysis*, 56(6), pp. 1362–1380. doi: 10.1016/j.csda.2011.10.022.
- Eisen, L. and Eisen, R. J. (2016) 'Critical Evaluation of the Linkage Between Tick-Based Risk Measures and the Occurrence of Lyme Disease Cases', *Journal of Medical Entomology*, 53(5), pp. 1050–1062. doi: 10.1093/jme/tjw092.
- Eisen, L. and Eisen, R. J. (2021) 'Benefits and Drawbacks of Citizen Science to Complement Traditional Data Gathering Approaches for Medically Important Hard Ticks (Acari: Ixodidae) in the United States', *Journal of Medical Entomology*, 58(1), pp. 1–9. doi: 10.1093/jme/tjaa165.
- Eisen, L., Eisen, R. J. and Lane, R. S. (2006) 'Geographical distribution patterns and habitat suitability models for presence of host-seeking ixodid ticks in dense woodlands of Mendocino County, California', *Journal of Medical Entomology*, 43(2), pp. 415–427. doi: 10.1603/0022-2585(2006)043[0415:GDPAHS]2.0.CO;2.
- Eisen, R. J., Lane, R. S., Fritz, C. L. and Eisen, L. (2006) 'Spatial patterns of lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure', *American Journal of Tropical*

- Medicine and Hygiene*, 75(4), pp. 669–676. doi: 10.4269/ajtmh.2006.75.669.
- Elith, J. and Leathwick, J. R. (2009) 'Species Distribution Models: Ecological Explanation and Prediction Across Space and Time', *Annual Review of Ecology, Evolution, and Systematics*, 40, pp. 677–697. doi: 10.1146/annurev.ecolsys.110308.120159.
- Elith, J., Leathwick, J. R. and Hastie, T. (2008) 'A working guide to boosted regression trees', *Journal of Animal Ecology*, 77(4), pp. 802–813. doi: 10.1111/j.1365-2656.2008.01390.x.
- Ellis, J., Oyston, P. C. F., Green, M. and Titball, R. W. (2002) 'Tularemia', *Clinical Microbiology Reviews*, 15(4), pp. 631–646. doi: 10.1128/CMR.15.4.631-646.2002.
- Ellwood, E. R., Primack, R. B. and Talmadge, M. L. (2010) 'Effects of Climate Change on Spring Arrival Times of Birds in Thoreau's Concord From 1851 to 2007.', *The Condor: Ornithological Applications*, 112(4), pp. 754–762. doi: 10.1525/cond.2010.100006.
- Embling, C. B., Walters, A. E. M. and Dolman, S. J. (2015) 'How much effort is enough? The power of citizen science to monitor trends in coastal cetacean species', *Global Ecology and Conservation*, 3(2015), pp. 867–877. doi: 10.1016/j.gecco.2015.04.003.
- Eritja, R., Ruiz-Arrondo, I., Delacour-Estrella, S., Schaffner, F., Álvarez-Chachero, J., Bengoa, M., Puig, M. Á., Melero-Alcibar, R., Oltra, A. and Bartumeus, F. (2019) 'First detection of *Aedes japonicus* in Spain: An unexpected finding triggered by citizen science', *Parasites & Vectors*, 12(53), pp. 1–9. doi: 10.1186/s13071-019-3317-y.
- ESRI (2012) 'ArcGIS Desktop: Release 10.2.2'. Redlands, CA: Environmental Systems Research Institute.
- ESRI (2019) 'ArcGIS Desktop: Release 10.7.1'. Redlands, CA: Environmental Systems Research Institute.
- Estrada-Peña, A. (2001) 'Distribution, Abundance, and Habitat Preferences of *Ixodes ricinus* (Acari: Ixodidae) in Northern Spain', *Journal of Medical Entomology*, 38(3), pp. 361–370. doi: 10.1603/0022-2585-38.3.361.
- Estrada-Peña, A., Alexander, N. and Wint, G. R. W. (2016) 'Perspectives on modelling the distribution of ticks for large areas: So far so good?', *Parasites & Vectors*, 9(179), pp. 1–10. doi: 10.1186/s13071-016-1474-9.
- Estrada-Peña, A., Estrada-Sánchez, A. and Estrada-Sánchez, D. (2015) 'Methodological caveats in the environmental modelling and projections of climate niche for ticks, with examples for *Ixodes ricinus* (Ixodidae)', *Veterinary Parasitology*, 208(2015), pp. 14–25. doi: 10.1016/j.vetpar.2014.12.016.
- Estrada-Peña, A., Farkas, R., Jaenson, T. G. T., Koenen, F., Madder, M., Pascucci, I., Salman, M., Tarres-Call, J. and Jongejan, F. (2013) 'Association of environmental traits with the geographic ranges of ticks (Acari: Ixodidae) of medical and veterinary importance in the western Palearctic. A digital data set', *Experimental and Applied Acarology*, 59, pp. 351–366. doi: 10.1007/s10493-012-9600-7.

- Estrada-Peña, A. and Venzal, J. M. (2006) 'Changes in habitat suitability for the tick *Ixodes ricinus* (Acari: Ixodidae) in Europe (1900-1999)', *EcoHealth*, 3(3), pp. 154–162. doi: 10.1007/s10393-006-0036-6.
- Estrada-Peña, A., Venzal, J. M. and Acedo, C. S. (2006) 'The tick *Ixodes ricinus*: distribution and climate preferences in the western Palaearctic.', *Medical and Veterinary Entomology*, 20(2), pp. 189–197. doi: 10.1111/j.1365-2915.2006.00622.x.
- Fahrer, H., Linden, S. M. Van Der, Sauvain, M.-J., Gern, L., Zhioua, E. and Aeschlimann, A. (1991) 'The Prevalence and Incidence of Clinical and Asymptomatic Lyme Borreliosis in a Population at Risk', *Oxford Journals*, 163(2), pp. 305–310. doi: 10.1093/infdis/163.2.305.
- Faulde, M. K. and Robbins, R. G. (2008) 'Tick infestation risk and *Borrelia burgdorferi* s.l. infection-induced increase in host-finding efficacy of female *Ixodes ricinus* under natural conditions', *Experimental and Applied Acarology*, 44(2), pp. 137–145. doi: 10.1007/s10493-008-9131-4.
- Faulde, M. K., Rutenfranz, M., Hepke, J., Rogge, M., Görner, A. and Keth, A. (2014) 'Human tick infestation pattern, tick-bite rate, and associated *Borrelia burgdorferi* s.l. infection risk during occupational tick exposure at the Seedorf military training area, northwestern Germany', *Ticks and Tick-borne Diseases*, 5(5), pp. 594–599. doi: 10.1016/j.ttbdis.2014.04.009.
- Fehr, J. S., Bloemberg, G. V., Ritter, C., Hombach, M., Lüscher, T. F., Weber, R. and Keller, P. M. (2010) 'Septicemia caused by tick-borne bacterial pathogen candidate *Neoehrlichia mikurensis*.', *Emerging Infectious Diseases*, 16(7), pp. 1127–1129. doi: 10.3201/eid1607.091907.
- Feria-Arroyo, T. P., Castro-Arellano, I., Gordillo-Perez, G., Cavazos, A. L., Vargas-Sandoval, M., Grover, A., Torres, J., Medina, R. F., De León, A. A. P. and Esteve-Gassent, M. D. (2014) 'Implications of climate change on the distribution of the tick vector *Ixodes scapularis* and risk for Lyme disease in the Texas-Mexico transboundary region', *Parasites & Vectors*, 7(199), pp. 1–16. doi: 10.1186/1756-3305-7-199.
- Fernandez, M. P., Bron, G. M., Kache, P. A., Larson, S. R., Maus, A., Gustafson, D., Tsao, J. I., Bartholomay, L. C., Paskewitz, S. M. and Diuk-Wasser, M. A. (2019) 'Usability and feasibility of a smartphone app to assess human behavioral factors associated with tick exposure (the Tick app): Quantitative and qualitative study', *JMIR mHealth and uHealth*, 7(10), pp. 1–20. doi: 10.2196/14769.
- Ferro e Silva, A. M., Sobral-Souza, T., Vancine, M. H., Muylaert, R. L., de Abreu, A. P., Pelloso, S. M., de Barros Carvalho, M. D., de Andrade, L., Ribeiro, M. C. and Toledo, M. J. de O. (2018) 'Spatial prediction of risk areas for vector transmission of *Trypanosoma cruzi* in the State of Paraná, southern Brazil', *PLoS Neglected Tropical Diseases*, 12(10), pp. 1–17. doi: 10.1371/journal.pntd.0006907.
- Finch, C., Al-Damluji, M. S., Krause, P. J., Niccolai, L., Steeves, T., O'Keefe, C. F. and Diuk-Wasser, M. A. (2014) 'Integrated assessment of behavioral and environmental risk factors for Lyme disease infection on

- Block Island, Rhode Island', *PLoS ONE*, 9(1), pp. 1–8. doi: 10.1371/journal.pone.0084758.
- Franck, M., Ghozzi, R., Pajaud, J., Lawson-Hogban, N. E., Mas, M., Lacout, A. and Perronne, C. (2020) 'Borrelia miyamotoi: 43 Cases Diagnosed in France by Real-Time PCR in Patients With Persistent Polymorphic Signs and Symptoms', *Frontiers in Medicine*, 7(55), pp. 1–10. doi: 10.3389/fmed.2020.00055.
- Friedman, J. (2001) 'Greedy function approximation: a gradient boosting machine.', *Annals of Statistics*, 29(5), pp. 1–39. doi: 10.1214/aos/1013203451.
- Fuglstad, G. A. and Beguin, J. (2018) 'Environmental mapping using Bayesian spatial modelling (INLA/SPDE): A reply to Huang et al. (2017)', *Science of the Total Environment*, 624(2018), pp. 596–598. doi: 10.1016/j.scitotenv.2017.12.067.
- Fuglstad, G. A., Simpson, D., Lindgren, F. and Rue, H. (2019) 'Constructing Priors that Penalize the Complexity of Gaussian Random Fields', *Journal of the American Statistical Association*, 114(525), pp. 445–452. doi: 10.1080/01621459.2017.1415907.
- Fukunaga, M., Takahashi, Y., Tsuruta, Y., Matsushita, O., Ralph, D., McClelland, M. and Nakao, M. (1995) 'Genetic and phenotypic analysis of *Borrelia miyamotoi* sp. nov., isolated from the ixodid tick *Ixodes persulcatus*, the vector for Lyme disease in Japan.', *International Journal of Systematic Bacteriology*, 45(4), pp. 804–810. doi: 10.1099/00207713-45-4-804.
- Gallo, T. and Waitt, D. (2011) 'Creating a Successful Citizen Science Model to Detect and Report Invasive Species', *BioScience*, 61(6), pp. 459–465. doi: 10.1525/bio.2011.61.6.8.
- Gallup, J. L. and Sachs, J. D. (2001) 'The economic burden of malaria', *American Journal of Tropical Medicine and Hygiene*, 64(1–2), pp. 85–96. doi: 10.4269/ajtmh.2001.64.85.
- Gandy, S. L. (2020) *The impacts of host community composition on Lyme disease risk in Scotland*. PhD thesis. University of Glasgow.
- Garcia-Martí, I. (2019) *Mapping tick dynamics and tick bite risk using data-driven approaches and volunteered observations*. PhD thesis. University of Twente. Available at: [https://library.itc.utwente.nl/papers\\_2019/phd/garciamarti.pdf](https://library.itc.utwente.nl/papers_2019/phd/garciamarti.pdf) (Accessed: 20 March 2020).
- Garcia-Martí, I., Zurita-Milla, R., Harms, M. G. and Swart, A. (2018) 'Using volunteered observations to map human exposure to ticks', *Nature Scientific Reports*, 8(15435), pp. 1–10. doi: 10.1038/s41598-018-33900-2.
- Garcia-Martí, I., Zurita-Milla, R., Swart, A., van den Wijngaard, K. C., van Vliet, A. J. H., Bennema, S. and Harms, M. (2017) 'Identifying Environmental and Human Factors Associated With Tick Bites using Volunteered Reports and Frequent Pattern Mining', *Transactions in GIS*, 21(2), pp. 277–299. doi: 10.1111/tgis.12211.
- Garcia-Martí, I., Zurita-Milla, R., van Vliet, A. J. H. and Takken, W. (2017) 'Modelling and mapping tick dynamics using volunteered

- observations', *International Journal of Health Geographics*, 16(41), pp. 1–15. doi: 10.1186/s12942-017-0114-8.
- Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005) 'Statistical methods for eliciting probability distributions', *Journal of the American Statistical Association*, 100(470), pp. 680–701. doi: 10.1198/016214505000000105.
- Gasmi, S., Ogden, N. H., Leighton, P. A., Lindsay, L. R. and Thivierge, K. (2016) 'Analysis of the human population bitten by Ixodes scapularis ticks in Quebec, Canada: Increasing risk of Lyme disease', *Ticks and Tick-borne Diseases*, 7(6), pp. 1075–1081. doi: 10.1016/j.ttbdis.2016.09.006.
- Gassner, F., van Vliet, A. J. H., Burgers, S. L. G. E., Jacobs, F., Verbaarschot, P., Hovius, E. K. E., Mulder, S., Verhulst, N. O., van Overbeek, L. S. and Takken, W. (2011) 'Geographic and Temporal Variations in Population Dynamics of Ixodes ricinus and Associated Borrelia Infections in The Netherlands', *Vector-Borne and Zoonotic Diseases*, 11(5), pp. 523–532. doi: 10.1089/vbz.2010.0026.
- Gehring, H., Schacht, E., Maylaender, N., Zeman, E., Kaysser, P., Oehme, R., Pluta, S. and Splettstoesser, W. D. (2012) 'Presence of an emerging subclone of Francisella tularensis holarctica in Ixodes ricinus ticks from south-western Germany', *Ticks and Tick-borne Diseases*, 4(1–2), pp. 93–100. doi: 10.1016/j.ttbdis.2012.09.001.
- Gelfand, A. E., Silander, J. A., Wu, S., Latimer, A., Lewis, P. O., Rebelo, A. G. and Holder, M. (2006) 'Explaining species distribution patterns through hierarchical modeling', *Bayesian Analysis*, 1(1), pp. 41–92. doi: 10.1214/06-BA102.
- Gelman, A., Hwang, J. and Vehtari, A. (2013) 'Understanding predictive information criteria for Bayesian models', *Statistics and Computing*, 24(6), pp. 997–1016. doi: 10.1007/s11222-013-9416-2.
- Gern, L., Cadenas, F. M. and Burri, C. (2008) 'Influence of some climatic factors on Ixodes ricinus ticks studied along altitudinal gradients in two geographic regions in Switzerland', *International Journal of Medical Microbiology*, 298(1), pp. 55–59. doi: 10.1016/j.ijmm.2008.01.005.
- Gern, L. and Rais, O. (1996) 'Efficient Transmission of Borrelia burgdorferi Between Cofeeding Ixodes ricinus Ticks (Acari: Ixodidae)', *Journal of Medical Entomology*, 33(1), pp. 189–192. doi: 10.1093/jmedent/33.1.189.
- Gilbert, L. (2010) 'Altitudinal patterns of tick and host abundance: A potential role for climate change in regulating tick-borne diseases?', *Oecologia*, 162, pp. 217–225. doi: 10.1007/s00442-009-1430-x.
- Gilbert, L. (2013) 'Can restoration of afforested peatland regulate pests and disease?', *Journal of Applied Ecology*, 50(5), pp. 1226–1233. doi: 10.1111/1365-2664.12141.
- Gilbert, L. (2016) 'Louping ill virus in the UK: a review of the hosts, transmission and ecological consequences of control', *Experimental and Applied Acarology*, 68(3), pp. 363–374. doi: 10.1007/s10493-015-9952-x.
- Gilbert, L., Aungier, J. and Tomkins, J. (2014) 'Climate of origin affects tick

- (*Ixodes ricinus*) host-seeking behavior in response to temperature: Implications for resilience to climate change?', *Ecology and Evolution*, 4(7), pp. 1186–1198. doi: 10.1002/ece3.1014.
- Gilbert, L., Brunner, K., Lande, U., Klingens, I. and Grøva, L. (2017) 'Environmental risk factors for *Ixodes ricinus* ticks and their infestation on lambs in a changing ecosystem : Implications for tick control and the impact of woodland encroachment on tick-borne disease in livestock', *Agriculture, Ecosystems and Environment*, 237(2017), pp. 265–273. doi: 10.1016/j.agee.2016.12.041.
- Gilbert, L., Maffey, G. L., Ramsay, S. L. and Hester, A. J. (2012) 'The effect of deer management on the abundance of *Ixodes ricinus* in Scotland', *Ecological Applications*, 22(2), pp. 658–667. doi: 10.1890/11-0458.1.
- Gillingham, E. L., Hall, J. L., Birtles, R. J., Bown, K. J., Medlock, J. M., Smith, R., Hansford, K. M., Warner, J. C., Dryden, M. and Pietzsch, M. E. (2020) 'Study of general practitioner consultations for tick bites at high, medium and low incidence areas for Lyme borreliosis in England and Wales.', *Zoonoses and Public Health*, 00, pp. 1–9. doi: 10.1111/zph.12694.
- Ginsberg, H. S. (1993) 'Transmission Risk of Lyme Disease and Implications for Tick Management', *American Journal of Epidemiology*, 138(1), pp. 65–73. doi: 10.1093/oxfordjournals.aje.a116778.
- Gneiting, T. (2011) 'Making and Evaluating Point Forecasts', *Journal of the American Statistical Association*, 106(494), pp. 746–762.
- Gneiting, T. and Raftery, A. E. (2007) 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association*, 102(477), pp. 359–378. doi: 10.1198/016214506000001437.
- Godana, A. A., Mwalili, S. M. and Orwa, G. O. (2019) 'Dynamic spatiotemporal modeling of the infected rate of visceral leishmaniasis in human in an endemic area of Amhara regional state, Ethiopia', *PLoS ONE*, 14(3), pp. 1–21. doi: 10.1371/journal.pone.0212934.
- Good, I. J. (1952) 'Rational Decisions', *Royal Statistical Society*, 14(1), pp. 107–114. doi: 10.1111/j.2517-6161.1952.tb00104.x.
- Goodchild, M. F. (2007) 'Citizens as sensors: The world of volunteered geography', *GeoJournal*, 69(4), pp. 211–221. doi: 10.1007/s10708-007-9111-y.
- Goodchild, M. F. and Li, L. (2012) 'Assuring the quality of volunteered geographic information', *Spatial Statistics*, 1(2012), pp. 110–120. doi: 10.1016/j.spasta.2012.03.002.
- Gray, A., Capewell, P., Loney, C., Katzer, F., Shiels, B. R. and Weir, W. (2019) 'Sheep as host species for zoonotic *Babesia venatorum*, United Kingdom', *Emerging Infectious Diseases*, 25(12), pp. 2257–2260. doi: 10.3201/eid2512.190459.
- Gray, J. S. (1991) 'The development and seasonal activity of the tick *Ixodes ricinus*: a vector of Lyme borreliosis.', *Review of Medical and Veterinary Entomology*, 79(6), pp. 323–333.
- Gray, J. S. (1998) 'The ecology of ticks transmitting Lyme borreliosis', *Experimental and Applied Acarology*, 22(1998), pp. 249–258. doi:



10.1023/A:1006070416135.

- Gray, J. S., Dautel, H., Estrada-Peña, A., Kahl, O. and Lindgren, E. (2009) 'Effects of Climate Change on Ticks and Tick-Borne Diseases in Europe', *Interdisciplinary Perspectives on Infectious Diseases*, 2009, pp. 1–12. doi: 10.1155/2009/593232.
- Gray, J. S., Kahl, O., Robertson, J. N., Daniel, M., Estrada-Peña, A., Gettinby, G., Jaenson, T. G. T., Jensen, P., Jongejan, F., Korenberg, E., Kurtenbach, K. and Zeman, P. (1998) 'Lyme borreliosis habitat assessment', *Zentralblatt fur Bakteriologie*, 287(3), pp. 211–228. doi: 10.1016/S0934-8840(98)80123-0.
- Gray, J. S. and Lohan, G. (1982) 'The development of a sampling method for the tick *Ixodes ricinus* and its use in a redwater fever area', *Annals Applied Biology*, 101, pp. 421–427. doi: 10.1111/j.1744-7348.1982.tb00842.x.
- Gritsun, T. S., Lashkevich, V. A. and Gould, E. A. (2003) 'Tick-Borne Encephalitis', *Antiviral Research*, 57(2003), pp. 129–146. doi: 10.1016/J.IDC.2008.03.013.
- Guerra, M., Walker, E., Jones, C., Paskewitz, S., Roberto Cortinas, M., Ashley Stancil, L. B., Bobo, M. and Kitron, U. (2002) 'Predicting the risk of Lyme disease: Habitat suitability for *Ixodes scapularis* in the north central United States', *Emerging Infectious Diseases*, 8(3), pp. 289–297. doi: 10.3201/eid0803.010166.
- Guilbaud, C. S. E. and Guilbaud, T. G. D. P. V. (2017) 'Mosquito Mapper: a phone application to map urban mosquitoes', *Scientific Phone Apps and Mobile Devices*, 3(6), pp. 1–6. doi: 10.1186/s41070-017-0018-9.
- Guisan, A. and Thuiller, W. (2005) 'Predicting species distribution: offering more than simple habitat models', *Ecology Letters*, 8(9), pp. 993–1009. doi: 10.1111/j.1461-0248.2005.00792.x.
- Haakenstad, A., Harle, A. C., Tsakalos, G., Micah, A. E., Tao, T., Anjomshoa, M., Cohen, J., Fullman, N., Hay, S. I., Mestrovic, T., Mohammed, S., Mousavi, S. M., Nixon, M. R., Pigott, D., Tran, K., Murray, C. J. L. and Dieleman, J. L. (2019) 'Tracking spending on malaria by source in 106 countries, 2000–16: an economic modelling study', *The Lancet Infectious Diseases*, 19(7), pp. 703–716. doi: 10.1016/S1473-3099(19)30165-3.
- Hadler, J. L., Patel, D., Nasci, R. S., Petersen, L. R., Hughes, J. M., Bradley, K., Etkind, P., Kan, L. and Engel, J. (2015) 'Assessment of arbovirus surveillance 13 years after introduction of West Nile virus, United States', *Emerging Infectious Diseases*, 21(7), pp. 1159–1166. doi: 10.3201/eid2107.140858.
- Hall, J. L., Alpers, K., Bown, K. J., Martin, S. J. and Birtles, R. J. (2017) 'Use of mass-participation outdoor events to assess human exposure to tickborne pathogens', *Emerging Infectious Diseases*, 23(3), pp. 463–467. doi: 10.3201/eid2303.161397.
- Hamer, S. A., Curtis-Robles, R. and Hamer, G. L. (2018) 'Contributions of citizen scientists to arthropod vector data in the age of digital epidemiology', *Current Opinion in Insect Science*, 28, pp. 98–104. doi: 10.1016/j.cois.2018.05.005.

- Hancock, P. A., Brackley, R. and Palmer, S. C. F. (2011) 'Modelling the effect of temperature variation on the seasonal dynamics of *Ixodes ricinus* tick populations', *International Journal for Parasitology*, 41(5), pp. 513–522. doi: 10.1016/j.ijpara.2010.12.012.
- Hanincová, K., Schäfer, S. M., Etti, S., Sewell, H. S., Taragelová, V., Ziak, D., Labuda, M. and Kurtenbach, K. (2003) 'Association of *Borrelia afzelii* with rodents in Europe.', *Parasitology*, 126(1), pp. 11–20. doi: 10.1017/s0031182002002548.
- Hanincová, K., Taragelova, V., Koci, J., Schafer, S. M., Hails, R., Ullmann, A. J., Piesman, J., Labuda, M. and Kurtenbach, K. (2003) 'Association of *Borrelia garinii* and *B. valaisiana* with Songbirds in Slovakia', *Applied and Environmental Microbiology*, 69(5), pp. 2825–2830. doi: 10.1128/AEM.69.5.2825.
- Hansford, K. M., Fonville, M., Gillingham, E. L., Coipan, E. C., Pietzsch, M. E., Krawczyk, A. I., Vaux, A. G. C., Cull, B., Sprong, H. and Medlock, J. M. (2017) 'Ticks and *Borrelia* in urban and peri-urban green space habitats in a city in southern England', *Ticks and Tick-borne Diseases*, 8(3), pp. 353–361. doi: 10.1016/j.ttbdis.2016.12.009.
- Hansford, K. M., Fonville, M., Jahfari, S., Sprong, H. and Medlock, J. M. (2015) '*Borrelia miyamotoi* in host-seeking *Ixodes ricinus* ticks in England', *Epidemiology and Infection*, 143(5), pp. 1079–1087. doi: 10.1017/S0950268814001691.
- Hansford, K. M., Pietzsch, M. E., Cull, B., Gillingham, E. L. and Medlock, J. M. (2018) 'Potential risk posed by the importation of ticks into the UK on animals: records from the Tick Surveillance Scheme', *Veterinary Record*, 182(107), pp. 1–7. doi: 10.1136/vr.104263.
- Harris, S. J., Massimino, S., Newson, S. E., Eaton, M. A., Marchant, J. H., Balmer, D. E., Noble, D. G., Gillings, S., Procter, D. and Pearce-Higgins, J. W. (2016) *The Breeding Bird Survey 2015, BTO Research Report 687*. British Trust for Ornithology, Thetford. Available at: <https://www.bto.org/sites/default/files/bbs-report-2015.pdf> (Accessed: 28 July 2020).
- Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. I., Shanks, G. D., Myers, M. F. and Snow, R. W. (2002) 'Climate change and the resurgence of malaria in the East African highlands', *Nature*, 415(6874), pp. 905–909. doi: 10.1038/415905a.
- Heath, A., Manolopoulou, I. and Baio, G. (2016) 'Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation', *Statistics in Medicine*, 35(23), pp. 4264–4280. doi: 10.1002/sim.6983.
- Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., DeAngelis, D., Dye, C., Eames, K. T. D., Edmunds, W. J., Frost, S. D. W., Funk, S., Hollingsworth, T. D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J. O., Metcalf, C. J. E., Mollison, D., Pellis, L., Pulliam, J. R. C., Roberts, M. G., Viboud, C., Arinaminpathy, N., Ball, F., Bogich, T., Gog, J., Grenfell, B., Lloyd, A. L., Mclean, A., O'Neill, P., Pearson, C., Riley, S., Tomba, G. S., Trapman, P. and Wood, J. (2015) 'Modeling infectious disease dynamics in the complex landscape of

- global health', *Science*, 347(6227), pp. 1216–1225. doi: 10.1126/science.aaa4339.
- Held, L. and Meyer, S. (2019) 'Forecasting Based on Surveillance Data', in Held, L., Hens, N., O'Neill, P., and Wallinga, J. (eds) *Handbook of Infectious Disease Data Analysis*. 1st edn. Chapman and Hall/CRC, pp. 1–19. doi: 10.1201/9781315222912-25.
- Held, L., Rufibach, K. and Balabdaoui, F. (2010) 'A Score Regression Approach to Assess Calibration of Continuous Probabilistic Predictions', *Biometrics*, 66(4), pp. 1295–1305. doi: 10.1111/j.1541-0420.2010.01406.x.
- Herrington, S. and Brussoni, M. (2015) 'Beyond Physical Activity: The Importance of Play and Nature-Based Play Spaces for Children's Health and Development', *Current Obesity Reports*, 4(4), pp. 477–483. doi: 10.1007/s13679-015-0179-2.
- Hijmans, R. J. (2019) *Introduction to the geosphere package (version 1.5-10)*, *R Journal*. Available at: <https://cran.r-project.org/web/packages/geosphere/vignettes/geosphere.pdf> (Accessed: 20 March 2020).
- Hijmans, R. J., van Etten, J., Sumner, M., Cheng, J., Baston, D., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Hingee, K., Karney, C., Mattiuzzi, M., Mosher, S., Nowosad, J., Pebesma, E., Lamigueiro, O. P., Racine, E. B., Rowlingson, B., Shortridge, A., Venables, B. and Wueest, R. (2020) 'raster: Geographic Data Analysis and Modeling.' R package version 3.4-5.
- Hijmans, R. J., Williams, E. and Vennes, C. (2019) 'geosphere: Spherical Trigonometry.' R package version 1.5-10.
- Hildebrandt, A., Gray, J. S. and Hunfeld, K. P. (2013) 'Human Babesiosis in Europe: What clinicians need to know', *Infection*, 41(6), pp. 1057–1072. doi: 10.1007/s15010-013-0526-8.
- Hines, D. and Sibbald, S. (2015) 'Citizen science: Exploring its application as a tool for prodromic surveillance of vector-borne disease', *Canada Communicable Disease Report*, 41(3), pp. 63–67. doi: 10.14745/ccdr.v41i03a04.
- Hirzel, A. H., Hausser, J., Chessel, D. and Perrin, N. (2002) 'Ecological-Niche Factor Analysis: How To Compute Habitat-Suitability Maps Without Absence Data?', *Ecology*, 83(7), pp. 2027–2036. doi: 10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2.
- Hirzel, A. H., Helfer, V. and Metral, F. (2001) 'Assessing habitat-suitability models with a virtual species', *Ecological Modelling*, 145(2–3), pp. 111–121. doi: 10.1016/S0304-3800(01)00396-9.
- Hjetland, R., Eliassen, K. E., Lindbæk, M., Nilsen, R. M., Grude, N. and Ulvestad, E. (2013) 'Tick bites in healthy adults from western Norway: Occurrence, risk factors, and outcomes', *Ticks and Tick-borne Diseases*, 4(4), pp. 304–310. doi: 10.1016/j.ttbdis.2013.02.003.
- Hoffmann, B., Scheuch, M., Höper, D., Jungblut, R., Holsteg, M. and Schirrmeyer, H. (2012) 'Novel Orthobunyavirus in Cattle, Europe, 2011', *Emerging Infectious Diseases*, 18(3), pp. 469–472. doi:

10.3201/eid1803.111905.

- Hofhuis, A., Harms, M., van den Wijngaard, C., Sprong, H. and van Pelt, W. (2015) 'Continuing increase of tick bites and Lyme disease between 1994 and 2009', *Ticks and Tick-borne Diseases*, 6(1), p. 69–74. doi: 10.1016/j.ttbdis.2014.09.006.
- Hofmeester, T. R., Sprong, H., Jansen, P. A., Prins, H. H. T. and Van Wieren, S. E. (2017) 'Deer presence rather than abundance determines the population density of the sheep tick, *Ixodes ricinus*, in Dutch forests', *Parasites & Vectors*, 10(433), pp. 1–8. doi: 10.1186/s13071-017-2370-7.
- Holding, M., Dowall, S. D., Medlock, J. M., Carter, D. P., Pullan, S. T., Lewis, J., Vipond, R., Rocchi, M. S., Baylis, M. and Hewson, R. (2020) 'Tick-borne encephalitis virus, United Kingdom', *Emerging Infectious Diseases*, 26(1), pp. 90–96. doi: 10.3201/eid2601.191085.
- Holding, M., Dowall, S. and Hewson, R. (2020) 'Detection of tick-borne encephalitis virus in the UK', *The Lancet*, 395(10222), p. 1. doi: 10.1016/S0140-6736(20)30040-4.
- Hönig, V., Švec, P., Marek, L., Mrkvička, T., Zubriková, D., Wittmann, M. V., Masař, O., Szturcová, D., Růžek, D., Pfister, K. and Grubhoffer, L. (2019) 'Model of risk of exposure to lyme borreliosis and tick-borne encephalitis virus-infected ticks in the border area of the Czech republic (South Bohemia) and Germany (lower bavaria and upper palatinate)', *International Journal of Environmental Research and Public Health*, 16(7), pp. 1–17. doi: 10.3390/ijerph16071173.
- Horrillo, L., Castro, A., Matía, B., Molina, L., García-Martínez, J., Jaqueti, J., García-Arata, I., Carrillo, E., Moreno, J., Ruiz-Giardin, J. M. and San Martín, J. (2019) 'Clinical aspects of visceral leishmaniasis caused by *L. infantum* in adults. Ten years of experience of the largest outbreak in Europe: what have we learned?', *Parasites & vectors*, 12(359), pp. 1–11. doi: 10.1186/s13071-019-3628-z.
- Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D. and Willis, S. G. (2014) 'Improving species distribution models: The value of data on abundance', *Methods in Ecology and Evolution*, 5(6), pp. 506–513. doi: 10.1111/2041-210X.12184.
- Hubálek, Z. and Halouzka, J. (1998) 'Prevalence rates of *Borrelia burgdorferi* sensu lato in host-seeking *Ixodes ricinus* ticks in Europe.', *Parasitology Research*, 84(3), pp. 167–72. doi: 10.1007/s004360050378.
- Hugli, D., Moret, J., Rais, O., Moosmann, Y., Erard, P., Malinverni, R. and Gern, L. (2009) 'Tick bites in a Lyme borreliosis highly endemic area in Switzerland', *International Journal of Medical Microbiology*, 299(2), pp. 155–160. doi: 10.1016/j.ijmm.2008.06.001.
- Hvidsten, D., Frafjord, K., Gray, J. S., Henningsson, A. J., Jenkins, A., Kristiansen, B. E., Lager, M., Rognerud, B., Slåtve, A. M., Stordal, F., Stuen, S. and Wilhelmsson, P. (2020) 'The distribution limit of the common tick, *Ixodes ricinus*, and some associated pathogens in north-western Europe', *Ticks and Tick-borne Diseases*, 11(4), pp. 1–10. doi: 10.1016/j.ttbdis.2020.101388.

- Hyndman, R. J. and Koehler, A. B. (2006) 'Another look at measures of forecast accuracy', *International Journal of Forecasting*, 22(4), pp. 679–688. doi: 10.1016/j.ijforecast.2006.03.001.
- Jado, I., Oteo, J. A., Aldámiz, M., Gil, H., Escudero, R., García-amil, C., Rodríguez-moreno, I. and Anda, P. (2007) 'Rickettsia monacensis and human disease, Spain', *Emerging Infectious Diseases*, 13(9), pp. 1405–1407. doi: 10.3201/eid1309.060186.
- Jaenson, T. G. T., Hjertqvist, M., Bergström, T. and Lundkvist, Å. (2012) 'Why is tick-borne encephalitis increasing? A review of the key factors causing the increasing incidence of human TBE in Sweden', *Parasites & Vectors*, 5(184), pp. 1–13. doi: 10.1186/1756-3305-5-184.
- Jaenson, T. G. T., Jaenson, D. G. E., Eisen, L., Petersson, E. and Lindgren, E. (2012) 'Changes in the geographical distribution and abundance of the tick Ixodes ricinus during the past 30 years in Sweden', *Parasites & Vectors*, 5(8), pp. 1–15. doi: 10.1186/1756-3305-5-8.
- Jaenson, T. G. T. and Tälleklint, L. (1992) 'Incompetence of Roe Deer as Reservoirs of the Lyme Borreliosis Spirochete', *Journal of Medical Entomology*, 29(5), pp. 813–817. doi: 10.1093/jmedent/29.5.813.
- James, M., Bowman, A., Forbes, K., Lewis, F., McLeod, J. and Gilbert, L. (2012) 'Environmental determinants of Ixodes ricinus ticks and the incidence of Borrelia burgdorferi sensu lato, the agent of Lyme borreliosis, in Scotland', *Parasitology*, 140(2), pp. 237–246. doi: 10.1017/S003118201200145X.
- James, M., Gilbert, L., Bowman, A. and Forbes, K. (2014) 'The Heterogeneity, Distribution, and Environmental Associations of Borrelia burgdorferi Sensu Lato, the Agent of Lyme Borreliosis, in Scotland', *Frontiers in Public Health*, 2(129), pp. 1–10. doi: 10.3389/fpubh.2014.00129.
- Jameson, L. J. and Medlock, J. M. (2011) 'Tick surveillance in Great Britain', *Vector-Borne and Zoonotic Diseases*, 11(4), pp. 403–412. doi: 10.1089/vbz.2010.0079.
- Jeffries, C. L., Mansfield, K. L., Phipps, L. P., Wakeley, P. R., Mearns, R., Schock, A., Bell, S., Breed, A. C., Fooks, A. R. and Johnson, N. (2014) 'Louping ill virus: An endemic tick-borne disease of Great Britain', *Journal of General Virology*, 95, pp. 1005–1014. doi: 10.1099/vir.0.062356-0.
- Johnson, T. L., Bjork, J. K. H., Neitzel, D. F., Dorr, F. M., Schiffman, E. K. and Eisen, R. J. (2016) 'Habitat Suitability Model for the Distribution of Ixodes scapularis (Acari: Ixodidae) in Minnesota', *Journal of Medical Entomology*, 53(3), pp. 598–606. doi: 10.1093/jme/tjw008.Habitat.
- Johnston, A., Fink, D., Hochachka, W. M. and Kelling, S. (2017) 'Estimates of observer expertise improve species distributions from citizen science data', *Methods in Ecology and Evolution*, 9, pp. 88–97. doi: 10.1111/2041-210X.12838.
- Jones-Todd, C. M., Swallow, B., Illian, J. B. and Toms, M. (2018) 'A spatiotemporal multispecies model of a semicontinuous response', *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 67(3), pp. 705–722. doi: 10.1111/rssc.12250.

- Jordan, R. C., Brooks, W. R., Howe, D. V. and Ehrenfeld, J. G. (2012) 'Evaluating the Performance of Volunteers in Mapping Invasive Plants in Public Conservation Lands', *Environmental Management*, 49(2), pp. 425–434. doi: 10.1007/s00267-011-9789-y.
- Jore, S., Vanwambeke, S. O., Slunge, D., Boman, A., Krogfelt, K. A., Jepsen, M. T. and Vold, L. (2020) 'Spatial tick bite exposure and associated risk factors in Scandinavia', *Infection Ecology and Epidemiology*, 10(1), pp. 1–12. doi: 10.1080/20008686.2020.1764693.
- Jore, S., Vanwambeke, S. O., Viljugrein, H., Isaksen, K., Kristoffersen, A. B., Woldehiwet, Z., Johansen, B., Brun, E., Brun-Hansen, H., Westermann, S., Larsen, I. L., Ytrehus, B. and Hofshagen, M. (2014) 'Climate and environmental change drives *Ixodes ricinus* geographical expansion at the northern range margin', *Parasites & Vectors*, 7(11), pp. 1–14. doi: 10.1186/1756-3305-7-11.
- Jore, S., Viljugrein, H., Hofshagen, M., Brun-Hansen, H., Kristoffersen, A. B., Nygård, K., Brun, E., Ottesen, P., Sævik, B. K. and Ytrehus, B. (2011) 'Multi-source analysis reveals latitudinal and altitudinal shifts in range of *Ixodes ricinus* at its northern distribution limit', *Parasites & Vectors*, 4(84), pp. 1–11. doi: 10.1186/1756-3305-4-84.
- Kahle, D. and Wickham, H. (2013) 'ggmap: Spatial visualization with ggplot2', *R Journal*, 5(1), pp. 144–161. doi: 10.32614/rj-2013-014.
- Kamal, M., Kenawy, M. A., Rady, M. H., Khaled, A. S. and Samy, A. M. (2018) 'Mapping the global potential distributions of two arboviral vectors *Aedes aegypti* and *Ae. Albopictus* under changing climate', *PLoS ONE*, 13(12), pp. 1–21. doi: 10.1371/journal.pone.0210122.
- Kaya, A. D., Parlak, A. H., Ozturk, C. E. and Behcet, M. (2008) 'Seroprevalence of *Borrelia burgdorferi* infection among forestry workers and farmers in Duzce, north-western Turkey', *New Microbiologica*, 31(2), pp. 203–209.
- Kelling, S., Johnston, A., Hochachka, W. M., Iloff, M., Fink, D., Gerbracht, J., Lagoze, C., La Sorte, F. A., Moore, T., Wiggins, A., Wong, W.-K., Wood, C. and Yu, J. (2015) 'Can Observation Skills of Citizen Scientists Be Estimated Using Species Accumulation Curves?', *PLOS ONE*, 10(10), pp. 1–20. doi: 10.1371/journal.pone.0139600.
- Keukeleire, M. De, Vanwambeke, S. O., Somassè, E., Kabamba, B., Luyasu, V., Robert, A., De Keukeleire, M., Vanwambeke, S. O., Somassè, E., Kabamba, B., Luyasu, V. and Robert, A. (2015) 'Scouts, forests, and ticks: Impact of landscapes on human-tick contacts', *Ticks and Tick-borne Diseases*, 6(5), pp. 636–644. doi: 10.1016/j.ttbdis.2015.05.008.
- Kiewra, D., Kryza, M. and Szymanowski, M. (2014) 'Influence of selected meteorological variables on the questing activity of *Ixodes ricinus* ticks in Lower Silesia, SW Poland', *Journal of Vector Ecology*, 39(1), pp. 138–145. doi: 10.1111/j.1948-7134.2014.12080.x.
- Kifle, Y. W., Hens, N. and Faes, C. (2017) 'Using additive and coupled spatiotemporal SPDE models: a flexible illustration for predicting occurrence of *Culicoides* species', *Spatial and Spatio-temporal Epidemiology*, 23, pp. 11–34. doi: 10.1016/j.sste.2017.07.003.
- Kilpatrick, A. M., Dobson, A. D. M., Levi, T., Salkeld, D. J., Sweit, A.,

- Ginsberg, H. S., Kjemtrup, A., Padgett, K. A., Jensen, P. M., Fish, D., Ogden, N. H. and Diuk-Wasser, M. A. (2017) 'Lyme disease ecology in a changing world: Consensus, uncertainty and critical gaps for improving control', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1722), pp. 1–15. doi: 10.1098/rstb.2016.0117.
- Kings College London (2020) *New research identifies those most at risk from 'long COVID'*. Available at: <https://www.kcl.ac.uk/news/study-identifies-those-most-risk-long-covid> (Accessed: 27 November 2020).
- Kirby, A. D., Smith, A. A., Benton, T. G. and Hudson, P. J. (2004) 'Rising burden of immature sheep ticks (*Ixodes ricinus*) on red grouse (*Lagopus lagopus scoticus*) chicks in the Scottish uplands', *Medical and Veterinary Entomology*, 18, pp. 67–70. doi: 10.1111/j.0269-283X.2004.0479.x.
- Kitron, U. (2000) 'Risk Maps: Transmission and Burden of Vector-Borne Diseases', *Parasitology Today*, 16(8), pp. 324–325. doi: 10.1016/S0169-4758(00)01708-7.
- Kjær, L. J., Soleng, A., Edgar, K. S., Lindstedt, H. E. H., Paulsen, K. M., Andreassen, Å. K., Korslund, L., Kjelland, V., Slettan, A., Stuen, S., Kjellander, P., Christensson, M., Teräsväinen, M., Baum, A., Klitgaard, K. and Bødker, R. (2019a) 'Predicting and mapping human risk of exposure to *Ixodes ricinus* nymphs using climatic and environmental data, Denmark, Norway and Sweden, 2016', *Eurosurveillance*, 24(9), pp. 1–11. doi: 10.2807/1560-7917.ES.2019.24.9.1800101.
- Kjær, L. J., Soleng, A., Edgar, K. S., Lindstedt, H. E. H., Paulsen, K. M., Andreassen, Å. K., Korslund, L., Kjelland, V., Slettan, A., Stuen, S., Kjellander, P., Christensson, M., Teräsväinen, M., Baum, A., Klitgaard, K. and Bødker, R. (2019b) 'Predicting the spatial abundance of *Ixodes ricinus* ticks in southern Scandinavia using environmental and climatic data', *Nature Scientific Reports*, 9(18144), pp. 1–16. doi: 10.1038/s41598-019-54496-1.
- Kjellander, P. L., Aronsson, M., Bergvall, U. A., Carrasco, J. L., Christensson, M., Lindgren, P. E., Åkesson, M. and Kjellander, P. (2021) 'Validating a common tick survey method: cloth-dragging and line transects', *Experimental and Applied Acarology*, 83(1), pp. 131–146. doi: 10.1007/s10493-020-00565-4.
- Koffi, J. K., Leighton, P. A., Pelcat, Y., Trudel, L., Lindsay, L. R., Milord, F. and Ogden, N. H. (2012) 'Passive surveillance for *I. scapularis* ticks: Enhanced analysis for early detection of emerging Lyme disease risk', *Journal of Medical Entomology*, 49(2), pp. 400–409. doi: 10.1603/ME11210.
- Kosmala, M., Wiggins, A., Swanson, A. and Simmons, B. (2016) 'Assessing data quality in citizen science', *Frontiers in Ecology and the Environment*, 14(10), pp. 551–560. doi: 10.1002/fee.1436.
- Kowalec, M., Szewczyk, T., Welc-Fałęciak, R., Siński, E., Karbowski, G. and Bajer, A. (2017) 'Ticks and the city - Are there any differences between city parks and natural forests in terms of tick abundance and prevalence of *spirochaetes*?', *Parasites & Vectors*, 10(573), pp. 1–19.

doi: 10.1186/s13071-017-2391-2.

- Kraemer, M. U. G., Reiner, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D., Perkins, T. A., Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., Van Bortel, W., Marsboom, C., Hendrickx, G., Schaffner, F., Moore, C. G., Nax, H. H., Bengtsson, L., Wetter, E., Tatem, A. J., Brownstein, J. S., Smith, D. L., Lambrechts, L., Cauchemez, S., Linard, C., Faria, N. R., Pybus, O. G., Scott, T. W., Liu, Q., Yu, H., Wint, G. R. W., Hay, S. I. and Golding, N. (2019) 'Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*', *Nature Microbiology*, 4(5), pp. 854–863. doi: 10.1038/s41564-019-0376-y.
- Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M., Barker, C. M., Moore, C. G., Carvalho, R. G., Coelho, G. E., Van Bortel, W., Hendrickx, G., Schaffner, F., Elyazar, I. R., Teng, H. J., Brady, O. J., Messina, J. P., Pigott, D. M., Scott, T. W., Smith, D. L., William Wint, G. R., Golding, N. and Hay, S. I. (2015) 'The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. Albopictus*', *eLife*, 4(e08347), pp. 1–18. doi: 10.7554/eLife.08347.
- Krainski, E. T., Lindgren, F., Simpson, D. and Rue, H. (2017) *The R-INLA tutorial on SPDE models, R-INLA tutorial series*. Available at: <https://folk.ntnu.no/fuglstad/Lund2016/Session6/spde-tutorial.pdf> (Accessed: 11 April 2018).
- Kreusch, T. M., Holding, M., Hewson, R., Harder, T., Medlock, J. M., Hansford, K. M., Dowall, S., Semper, A., Brooks, T., Walsh, A., Russell, K. and Wichmann, O. (2019) 'A probable case of tick-borne encephalitis (TBE) acquired in England, July 2019', *Eurosurveillance*, 24(47), pp. 1–5. doi: 10.2807/1560-7917.ES.2019.24.47.1900679.
- Laaksonen, M., Klemola, T., Feuth, E., Sormunen, J. J., Puisto, A., Mäkelä, S., Penttinen, R., Ruohomäki, K., Hänninen, J., Sääksjärvi, I. E., Vuorinen, I., Sprong, H., Hytönen, J. and Vesterinen, E. J. (2018) 'Tick-borne pathogens in Finland: Comparison of *Ixodes ricinus* and *I. persulcatus* in sympatric and parapatric areas', *Parasites & Vectors*, 11(556), pp. 1–13. doi: 10.1186/s13071-018-3131-y.
- Laaksonen, M., Sajanti, E., Sormunen, J. J., Penttinen, R., Hänninen, J., Ruohomäki, K., Sääksjärvi, I., Vesterinen, E. J., Vuorinen, I., Hytönen, J. and Klemola, T. (2017) 'Crowdsourcing-based nationwide tick collection reveals the distribution of *Ixodes ricinus* and *I. persulcatus* and associated pathogens in Finland', *Emerging Microbes and Infections*, 6(5), pp. 1–7. doi: 10.1038/emi.2017.17.
- Langley, S. A., Messina, J. P. and Moore, N. (2017) 'Using meta-quality to assess the utility of volunteered geographic information for science', *International Journal of Health Geographics*, 16(40), pp. 1–11. doi: 10.1186/s12942-017-0113-9.
- Latimer, A. M., Wu, S., Gelfand, A. E. and Silander, J. A. (2006) 'Building Statistical Models to Analyze Species Distributions', *Ecological Applications*, 16(1), pp. 33–50. doi: 10.1890/04-0609.
- Leathwick, J. R., Elith, J., Chadderton, W. L., Rowe, D. and Hastie, T. (2008)



- 'Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species', *Journal of Biogeography*, 35(8), pp. 1481–1497. doi: 10.1111/j.1365-2699.2008.01887.x.
- Lehmann, A., Overton, J. M. and Leathwick, J. R. (2002) 'GRASP: generalized regression analysis and spatial prediction', *Ecological Modelling*, 157(2–3), pp. 189–207. doi: 10.1016/S0304-3800(02)00195-3.
- Lernout, T., De Regge, N., Tersago, K., Fonville, M., Suin, V. and Sprong, H. (2019) 'Prevalence of pathogens in ticks collected from humans through citizen science in Belgium', *Parasites & Vectors*, 12(550), pp. 1–11. doi: 10.1186/s13071-019-3806-z.
- Levi, T., Keesing, F., Oggenfuss, K. and Ostfeld, R. S. (2015) 'Accelerated phenology of blacklegged ticks under climate warming', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(20130556), pp. 1–8. doi: 10.1098/rstb.2013.0556.
- Levinger, P., Sales, M., Polman, R., Haines, T., Dow, B., Biddle, S. J. H., Duque, G. and Hill, K. D. (2018) 'Outdoor physical activity for older people—the senior exercise park: Current research, challenges and future directions', *Health Promotion Journal of Australia*, 29(3), pp. 353–359. doi: 10.1002/hpja.60.
- Lewis, J., Boudreau, C. R., Patterson, J. W., Bradet-Legrís, J. and Lloyd, V. K. (2018) 'Citizen Science and Community Engagement in Tick Surveillance—A Canadian Case Study', *Healthcare*, 6(22), pp. 1–13. doi: 10.3390/healthcare6010022.
- Li, S., Gilbert, L., Harrison, P. A. and Rounsevell, M. D. A. (2016) 'Modelling the seasonality of Lyme disease risk and the potential impacts of a warming climate within the heterogeneous landscapes of Scotland', *Journal of The Royal Society Interface*, 13(20160140), pp. 1–9. doi: 10.1098/rsif.2016.0140.
- Lindgren, E. and Gustafson, R. (2001) 'Tick-borne encephalitis in Sweden and climate change', *The Lancet*, 358(9275), pp. 16–18. doi: 10.1016/S0140-6736(00)05250-8.
- Lindgren, E. and Jaenson, T. G. T. (2006) *Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures*, World Health Organization. Available at: [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0006/96819/E89522.pdf](http://www.euro.who.int/__data/assets/pdf_file/0006/96819/E89522.pdf) (Accessed: 13 April 2018).
- Lindgren, E., Talleklint, L. and Polfeldt, T. (2000) 'Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick *Ixodes ricinus*', *Environmental Health Perspectives*, 108(2), pp. 119–123. doi: 10.1289/ehp.00108119.
- Lindgren, F. and Rue, H. (2015) 'Bayesian spatial modelling with R-INLA', *Journal of Statistical Software*, 63(19), pp. 1–25. doi: 10.18637/jss.v063.i19.
- Lindgren, F., Rue, H. and Lindström, J. (2011) 'An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach', *Journal of the Royal Statistical*

- Society. Series B: Statistical Methodology*, 73(4), pp. 423–498. doi: 10.1111/j.1467-9868.2011.00777.x.
- Lindquist, L. and Vapalahti, O. (2008) 'Tick-borne encephalitis', *The Lancet*, 371(9627), pp. 1861–1871. doi: 10.1016/S0140-6736(08)60800-4.
- Lindström, A. and Jaenson, T. G. T. (2003) 'Distribution of the Common Tick, *Ixodes ricinus* (Acari: Ixodidae), in Different Vegetation Types in Southern Sweden', *Journal of Medical Entomology*, 40(4), pp. 375–378. doi: 10.1603/0022-2585-40.4.375.
- Ling, C. L., Joss, A. W. L., Davidson, M. M. and Ho-Yen, D. O. (2000) 'Identification of different *Borrelia burgdorferi* genomic groups from Scottish ticks', *Journal of Clinical Pathology - Molecular Pathology*, 53(2), pp. 94–98. doi: 10.1136/mp.53.2.94.
- Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R. C., Dobson, A. P., Hudson, P. J. and Grenfell, B. T. (2009) 'Epidemic Dynamics at the Interface, Human-animal', *Science*, 326(5958), pp. 1–13. doi: 10.1126/science.1177345.Epidemic.
- Lohr, B., Müller, I., Mai, M., Norris, D. E., Schöffski, O. and Hunfeld, K. (2015) 'Ticks and Tick-borne Diseases Epidemiology and cost of hospital care for Lyme borreliosis in Germany : Lessons from a health care utilization database analysis', *Ticks and Tick-borne Diseases*, 6(2015), pp. 56–62. doi: 10.1016/j.ttbdis.2014.09.004.
- Lord, J. S., Torr, S. J., Auty, H. K., Brock, P. M., Byamungu, M., Hargrove, J. W., Morrison, L. J., Mramba, F., Vale, G. A. and Stanton, M. C. (2018) 'Geostatistical models using remotely-sensed data predict savanna tsetse decline across the interface between protected and unprotected areas in Serengeti, Tanzania', *Journal of Applied Ecology*, 55(4), pp. 1997–2007. doi: 10.1111/1365-2664.13091.
- Lourenço, J. and Recker, M. (2014) 'The 2012 Madeira Dengue Outbreak: Epidemiological Determinants and Future Epidemic Potential', *PLoS Neglected Tropical Diseases*, 8(8), pp. 1–11. doi: 10.1371/journal.pntd.0003083.
- Lukan, M., Bullova, E. and Petko, B. (2010) 'Climate warming and tick-borne encephalitis, Slovakia', *Emerging Infectious Diseases*, 16(3), pp. 524–526. doi: 10.3201/eid1603.081364.
- Lwin, M. O., Jayasundar, K., Sheldenkar, A., Wijayamuni, R., Wimalaratne, P., Ernst, K. C. and Foo, S. (2017) 'Lessons From the Implementation of Mo-Buzz, a Mobile Pandemic Surveillance System for Dengue', *JMIR Public Health and Surveillance*, 3(4), pp. 1–9. doi: 10.2196/publichealth.7376.
- MacKechnie, C., Maskell, L., Norton, L. and Roy, D. (2011) 'The role of "Big Society" in monitoring the state of the natural environment', *Journal of Environmental Monitoring*, 13(10), pp. 2687–2691. doi: 10.1039/c1em10615e.
- MacLeod, J. (1936) '*Ixodes ricinus* in relation to its physical environment. An analysis of the ecological complexes controlling distribution and activities', *Parasitology*, 28(3), pp. 295–319. doi: 10.1017/S0031182000022502.
- Macleod, J. and Gordon, W. (1932) 'Studies in louping ill: II. Transmission by

- the sheep tick, *Ixodes ricinus* L.', *Journal of Comparative Pathology*, 45(3), pp. 240–256. doi: 10.1016/S0368-1742(32)80022-6.
- Macleod, J. and Gordon, W. (1933) 'Studies on tick-borne fever in sheep. I. Transmission by the tick *Ixodes ricinus* and description of the disease produced.', *Parasitology*, 25(2), pp. 273–283. doi: 10.1017/S0031182000019442.
- Maetzel, D., Maier, W. A. and Kampen, H. (2005) 'Borrelia burgdorferi infection prevalences in questing *Ixodes ricinus* ticks (Acari: Ixodidae) in urban and suburban Bonn, western Germany', *Parasitology Research*, 95, pp. 5–12. doi: 10.1007/s00436-004-1240-3.
- Maki, E. C. and Cohnstaedt, L. W. (2015) 'Crowdsourcing for large-scale mosquito (Diptera: Culicidae) sampling', *The Canadian Entomologist*, 147(1), pp. 118–123. doi: 10.4039/tce.2014.27.
- Marcu, A., Barnett, J., Uzzell, D., Vasileiou, K. and Connell, S. O. (2013) 'Experience of Lyme disease and preferences for precautions : a cross-sectional survey of UK patients', *BMC Public Health*, 13(481), pp. 1–7. doi: 10.1186/1471-2458-13-481.
- Massimino, D., Harris, S. J. and Gillings, S. (2018) 'Evaluating spatiotemporal trends in terrestrial mammal abundance using data collected during bird surveys', *Biological Conservation*, 226(2018), pp. 153–167. doi: 10.1016/j.biocon.2018.07.026.
- Materna, J., Daniel, M. and Danielová, V. (2005) 'Altitudinal distribution limit of the tick *Ixodes ricinus* shifted considerably towards higher altitudes in Central Europe: Results of three years monitoring in the Krkonoše Mts. (Czech Republic)', *Central European Journal of Public Health*, 13(1), pp. 24–28.
- Mavin, S., Watson, E. J. and Evans, R. (2015) 'Distribution and presentation of Lyme borreliosis in Scotland-analysis of data from a national testing laboratory', *The Journal of the Royal College of Physicians of Edinburgh*, 45(3), pp. 196–200. doi: 10.4997/JRCPE.2015.304.
- McCoy, G. W. and Chapin, C. W. (1912) 'Further observations on a plague-like disease of rodents with a preliminary note on the causative agent, bacterium tularensis', *Journal of Infectious Diseases*, 10(1), pp. 61–72. doi: 10.1093/infdis/10.1.61.
- Mead, P., Hinckley, A., Hook, S. and Beard, C. Ben (2015) 'TickNET - A Collaborative Public Health Approach to Tickborne Disease Surveillance and Research', *Emerging Infectious Diseases*, 21(9), pp. 1574–1577. doi: 10.3201/eid2109.150301.
- Mead, P., Hook, S., Niesobecki, S., Ray, J., Meek, J., Delorey, M., Prue, C. and Hinckley, A. (2018) 'Risk factors for tick exposure in suburban settings in the Northeastern United States', *Ticks and Tick-borne Diseases*, 9(2), pp. 319–324. doi: 10.1016/j.ttbdis.2017.11.006.
- Medlock, J. M., Hansford, K. M., Bormane, A., Derdakova, M., Estrada-peña, A., George, J., Golovljova, I., Jaenson, T. G. T., Jensen, J., Jensen, P. M., Kazimirova, M., Oteo, J. A., Papa, A., Pfister, K., Plantard, O. and Randolph, S. E. (2013) 'Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe', *Parasites & Vectors*, 6(1), pp. 1–11. doi: 10.1186/1756-3305-6-1.

- Medlock, J. M., Hansford, K. M., Schaffner, F., Versteirt, V., Hendrickx, G., Zeller, H. and Bortel, W. Van (2012) 'A review of the invasive mosquitoes in Europe: Ecology, public health risks, and control options', *Vector-Borne and Zoonotic Diseases*, 12(6), pp. 435–447. doi: 10.1089/vbz.2011.0814.
- Medlock, J. M., Hansford, K. M., Vaux, A. G. C., Cull, B., Gillingham, E. and Leach, S. (2018) 'Assessment of the Public Health Threats Posed by Vector-Borne Disease in the United Kingdom (UK)', *International Journal of Environmental Research and Public Health*, 15(2145), pp. 1–22. doi: 10.3390/ijerph15102145.
- Medlock, J. M. and Leach, S. A. (2015) 'Effect of climate change on vector-borne disease risk in the UK', *The Lancet Infectious Diseases*, 15(6), pp. 721–730. doi: 10.1016/S1473-3099(15)70091-5.
- Medlock, J. M., Pietzsch, M. E., Rice, N. V. P., Jones, L., Kerrod, E., Avenell, D., Los, S., Ratcliffe, N., Leach, S. and Butt, T. (2008) 'Investigation of Ecological and Environmental Determinants for the Presence of Questing *Ixodes ricinus* (Acari: Ixodidae) on Gower, South Wales', *Journal of Medical Entomology*, 45(2), pp. 314–325. doi: 10.1603/0022-2585(2008)45[314:IOEAED]2.0.CO;2.
- Mejlon, H. A. and Jaenson, T. G. T. (1997) 'Questing behaviour of *Ixodes ricinus* ticks (Acari: Ixodidae)', *Experimental and Applied Acarology*, 21(12), pp. 747–754. doi: 10.1023/A:1018421105231.
- Mengersen, K., Peterson, E. E., Clifford, S., Ye, N., Kim, J., Bednarz, T., Brown, R., James, A., Vercelloni, J., Pearse, A. R., Davis, J. and Hunter, V. (2017) 'Modelling imperfect presence data obtained by citizen science', *Environmetrics*, 28(5), pp. 1–29. doi: 10.1002/env.2446.
- Merkle, E. C. and Steyvers, M. (2013) 'Choosing a strictly proper scoring rule', *Decision Analysis*, 10(4), pp. 292–304. doi: 10.1287/deca.2013.0280.
- Met Office (2016) *UK regional climates*. Available at: <https://www.metoffice.gov.uk/research/climate/maps-and-data/regional-climates/index> (Accessed: 30 October 2020).
- Michael Thrusfield (2005) *Veterinary Epidemiology*. 3rd edn. Oxford: Blackwell Publishing Ltd.
- Miller-Rushing, A. J. and Primack, R. B. (2008) 'Global warming and flowering times in Thoreau's Concord: a community perspective.', *Ecology*, 89(2), pp. 332–341. doi: 10.1890/07-0068.1.
- Miller-Rushing, A. J., Primack, R. B. and Bonney, R. (2012) 'The history of public participation in ecological research', *Frontiers in Ecology and the Environment*, 10(6), pp. 285–290. doi: 10.1890/110278.
- Miller, J. (2010) 'Species Distribution Modeling', *Geography Compass*, 4(6), pp. 490–509. doi: 10.1111/j.1749-8198.2010.00351.x.
- Millins, C., Gilbert, L., Johnson, P., James, M., Kilbride, E., Birtles, R. and Biek, R. (2016) 'Heterogeneity in the abundance and distribution of *Ixodes ricinus* and *Borrelia burgdorferi* (sensu lato) in Scotland: implications for risk prediction', *Parasites & Vectors*, 9(595), pp. 1–12. doi: 10.1186/s13071-016-1875-9.

- Millins, C., Magierecka, A., Gilbert, L., Edoff, A., Brereton, A., Kilbride, E., Denwood, M., Birtles, R. and Bieka, R. (2015) 'An invasive mammal (the gray squirrel, *Sciurus carolinensis*) commonly hosts diverse and atypical genotypes of the zoonotic pathogen *Borrelia burgdorferi* Sensu lato', *Applied and Environmental Microbiology*, 81(13), pp. 4236–4245. doi: 10.1128/AEM.00109-15.
- Moffett, A., Shackelford, N. and Sarkar, S. (2007) 'Malaria in Africa: Vector species' niche models and relative risk maps', *PLoS ONE*, 2(9), pp. 1–14. doi: 10.1371/journal.pone.0000824.
- Moiroux, N., Bio-Bangana, A. S., Djènantin, A., Chandre, F., Corbel, V. and Guis, H. (2013) 'Modelling the risk of being bitten by malaria vectors in a vector control area in southern Benin, west Africa', *Parasites & Vectors*, 6(71), pp. 1–13. doi: 10.1186/1756-3305-6-71.
- Mowbray, F., Amlôt, R. and Rubin, G. J. (2014) 'Ticks and Tick-borne Diseases Predictors of protective behaviour against ticks in the UK : A mixed methods study', *Ticks and Tick-borne Diseases*, 5(4), pp. 392–400. doi: 10.1016/j.ttbdis.2014.01.006.
- Mulder, S., Vliet, A. J. H. Van, Bron, W. A., Gassner, F. and Takken, W. (2013) 'High Risk of Tick Bites in Dutch Gardens', *Vector-Borne and Zoonotic Diseases*, 13(12), pp. 865–871. doi: 10.1089/vbz.2012.1194.
- Müller, R., Reuss, F., Kendrovski, V. and Montag, D. (2019) 'Vector-Borne Diseases', in Marselle, M. R., Stadler, J., Korn, H., Irvine, K. N., and Bonn, A. (eds) *Biodiversity and Health in the Face of Climate Change*. Springer Open, pp. 67–90. doi: 10.1007/978-3-030-02318-8.
- Murindahabi, M. M., Asingizwe, D., Poortvliet, P. M., van Vliet, A. J. H., Hakizimana, E., Mutesa, L., Takken, W. and Koenraadt, C. J. M. (2018) 'A citizen science approach for malaria mosquito surveillance and control in Rwanda', *NJAS - Wageningen Journal of Life Sciences*, 86(87), pp. 101–110. doi: 10.1016/j.njas.2018.07.005.
- Murray, K., Godbole, S., Natarajan, L., Full, K., Hipp, J. A., Glanz, K., Mitchell, J., Laden, F., James, P., Quante, M. and Kerr, J. (2017) 'The relations between sleep, time of physical activity, and time outdoors among adult women', *PLoS ONE*, 12(9), pp. 1–13. doi: 10.1371/journal.pone.0182013.
- Musenge, E., Chirwa, T. F., Kahn, K. and Vounatsou, P. (2013) 'Bayesian analysis of zero inflated spatiotemporal HIV/TB child mortality data through the INLA and SPDE approaches: Applied to data observed between 1992 and 2010 in rural North East South Africa', *International Journal of Applied Earth Observation and Geoinformation*, 22(100), pp. 86–98. doi: 10.1016/j.jag.2012.04.001.
- Musso, D., Cao-Lormeau, V. M. and Gubler, D. J. (2015) 'Zika virus: Following the path of dengue and chikungunya?', *The Lancet*, 386(9990), pp. 243–244. doi: 10.1016/S0140-6736(15)61273-9.
- Myer, M. H., Campbell, S. R. and Johnston, J. M. (2017) 'Spatiotemporal modeling of ecological and sociological predictors of West Nile virus in Suffolk County, NY, mosquitoes', *Ecosphere*, 8(6), pp. 1–17. doi: 10.1002/ecs2.1854/full.
- Mysterud, A., Hatlegjerde, I. L. and Sørensen, O. J. (2014) 'Attachment site

- selection of life stages of *Ixodes ricinus* ticks on a main large host in Europe, the red deer (*Cervus elaphus*)', *Parasites & Vectors*, 7(510), pp. 1–5. doi: 10.1186/s13071-014-0510-x.
- Mysterud, A., Stigum, V. M., Jaarsma, R. I. and Sprong, H. (2019) 'Genospecies of *Borrelia burgdorferi* sensu lato detected in 16 mammal species and questing ticks from northern Europe', *Nature Scientific Reports*, 9(5088), pp. 1–8. doi: 10.1038/s41598-019-41686-0.
- Napp, S., Petrić, D. and Busquets, N. (2018) 'West Nile virus and other mosquito-borne viruses present in Eastern Europe', *Pathogens and Global Health*, 112(5), pp. 233–248. doi: 10.1080/20477724.2018.1483567.
- NASA (2016) *MODIS Evapotranspiration*. Available at: <https://modis.gsfc.nasa.gov/data/dataproduct/mod16.php> (Accessed: 20 September 2020).
- Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., Huang, A., Rosenberg, A., Greenberg, A., Sherman, M., Wong, S. and Layton, M. (2001) 'The outbreak of West Nile Virus Infection in the New York City Area in 1999', *The New England Journal of Medicine*, 344(24), pp. 1807–1814. doi: 10.1097/00019048-200106000-00036.
- Nelli, L., Ferguson, H. M. and Matthiopoulos, J. (2020) 'Achieving explanatory depth and spatial breadth in infectious disease modelling: Integrating active and passive case surveillance', *Statistical Methods in Medical Research*, 29(5), pp. 1273–1287. doi: 10.1177/0962280219856380.
- Neuwirth, E. (2014) 'RColorBrewer: ColorBrewer Palettes.' R package version 1.1-2.
- Nieto, N. C., Porter, W. T., Wachara, J. C., Lowrey, T. J., Martin, L., Motyka, P. J. and Salkeld, D. J. (2018) 'Using citizen science to describe the prevalence and distribution of tick bite and exposure to tick-borne diseases in the United States', *PLoS ONE*, 13(7), pp. 1–14. doi: 10.1371/journal.pone.0199644 Editor:
- Nilsson, K., Lindquist, O. and Pålsson, C. (1999) 'Association of *Rickettsia helvetica* with chronic perimyocarditis in sudden cardiac death', *The Lancet*, 354(9185), pp. 1169–1173. doi: 10.1016/S0140-6736(99)04093-3.
- Norman, R. A., Worton, A. J. and Gilbert, L. (2016) 'Past and future perspectives on mathematical models of tick-borne pathogens', *Parasitology*, 143(7), pp. 850–859. doi: 10.1017/S0031182015001523.
- Núñez, O., Fernández-Navarro, P., Martín-Méndez, I., Bel-Lan, A., Locutura, J. F. and López-Abente, G. (2016) 'Arsenic and chromium topsoil levels and cancer mortality in Spain', *Environmental Science and Pollution Research*, 23(17), pp. 17664–17675. doi: 10.1007/s11356-016-6806-y.
- O'Connell, S., Granström, M., Gray, J. S. and Stanek, G. (1998) 'Epidemiology of European Lyme borreliosis', *Zentralblatt für Bakteriologie*, 287(3), pp. 229–240. doi: 10.1016/S0934-8840(98)80124-2.

- Oechslin, C. P., Heutschi, D., Lenz, N., Tischhauser, W., Péter, O., Rais, O., Beuret, C. M., Leib, S. L., Bankoul, S. and Ackermann-Gäumann, R. (2017) 'Prevalence of tick-borne pathogens in questing *Ixodes ricinus* ticks in urban and suburban areas of Switzerland', *Parasites & Vectors*, 10(558), pp. 1–18. doi: 10.1186/s13071-017-2500-2.
- Ogden, N. H., Bouchard, C., Kurtenbach, K., Margos, G., Robbin Lindsay, L., Trudel, L., Nguon, S. and Milord, F. (2010) 'Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of lyme disease risk emergence in Canada', *Environmental Health Perspectives*, 118(7), pp. 909–914. doi: 10.1289/ehp.0901766.
- Ogden, N. H., Cripps, P., Davison, C. C., Owen, G., Parry, J. M., Timms, B. J. and Forbes, A. B. (2000) 'The ixodid tick species attaching to domestic dogs and cats in Great Britain and Ireland', *Medical and Veterinary Entomology*, 14, pp. 332–338. doi: 10.1046/j.1365-2915.2000.00244.x.
- Ogden, N. H., Nuttall, P. A. and Randolph, S. E. (1997) 'Natural Lyme disease cycles maintained via sheep by cofeeding ticks', *Parasitology*, 115(6), pp. 591–599. doi: 10.1017/S0031182097001868.
- Ogden, N. H., Trudel, L., Artsob, H., Barker, I. K., Beauchamp, G., Charron, D. F., Drebot, M. A., Galloway, T. D., O'Handley, R., Thompson, R. A. and Lindsay, L. R. (2006) 'Ixodes scapularis ticks collected by passive surveillance in Canada: Analysis of geographic distribution and infection with lyme borreliosis agent *Borrelia burgdorferi*', *Journal of Medical Entomology*, 43(3), pp. 600–609. doi: 10.1603/0022-2585(2006)43[600:ISTCBP]2.0.CO;2.
- OIE (2020) *OIE-Listed diseases, infections and infestations in force in 2020*. Available at: <https://www.oie.int/animal-health-in-the-world/oie-listed-diseases-2020/> (Accessed: 13 October 2020).
- Olden, J. D., Lawler, J. J. and Poff, N. L. (2008) 'Machine Learning Methods Without Tears: A Primer for Ecologists', *The Quarterly Review of Biology*, 83(2), pp. 171–193. doi: 10.1086/587826.
- Olivier, A., Ramos Perez, J., Kalogirou, V., Bontemps, S., Defourny, P. and Van Bogaert, E. (2012) *Global Land Cover Map for 2009 (GlobCover 2009)*. doi: 10.1594/PANGAEA.787668.
- Ostfeld, R. S. and Keesing, F. (2000) 'Biodiversity and disease risk: The case of Lyme disease', *Conservation Biology*, 14(3), pp. 722–728. doi: 10.1046/j.1523-1739.2000.99014.x.
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E. and Peterson, A. T. (2013) 'Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas', *Ecological Modelling*, 263(2013), pp. 10–18. doi: 10.1016/j.ecolmodel.2013.04.011.
- Palmer, J. R. B., Oltra, A., Collantes, F., Delgado, J. A., Lucientes, J., Delacour, S., Bengoa, M., Eritja, R. and Bartumeus, F. (2017) 'Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes', *Nature Communications*, 8(916), pp. 1–12. doi:

10.1038/s41467-017-00914-9.

- Paul, R. E. L., Cote, M., Le Naour, E. and Bonnet, S. I. (2016) 'Environmental factors influencing tick densities over seven years in a French suburban forest', *Parasites & Vectors*, 9(309), pp. 1–10. doi: 10.1186/s13071-016-1591-5.
- Pearce, J. L. and Boyce, M. S. (2006) 'Modelling distribution and abundance with presence-only data', *Journal of Applied Ecology*, 43(3), pp. 405–412. doi: 10.1111/j.1365-2664.2005.01112.x.
- Pebesma, E. and Bivand, R. (2005) *Classes and Methods for Spatial Data: the sp Package*, *R News*. Available at: <http://cran.r-project.org/doc/Rnews/> (Accessed: 10 January 2020).
- Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., PerBaston, T. L. and Baston, D. (2020) 'sf: Simple Features for R.' R package version 0.9-5.
- Pepin, K. M., Eisen, R. J., Mead, P. S., Piesman, J., Fish, D., Hoen, A. G., Barbour, A. G., Hamer, S. and Diuk-Wasser, M. A. (2012) 'Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking Ixodes scapularis nymphs in the Eastern United States', *American Journal of Tropical Medicine and Hygiene*, 86(6), pp. 1062–1071. doi: 10.4269/ajtmh.2012.11-0630.
- Perret, J.-L., Guigoz, E., Rais, O. and Gern, L. (2000) 'Influence of saturation deficit and temperature on Ixodes ricinus tick', *Parasitology Research*, 86, pp. 554–557. doi: 10.1007/s004360000209.
- Perret, J.-L., Rais, O. and Gern, L. (2004) 'Influence of Climate on the Proportion of Ixodes ricinus Nymphs and Adults Questing in a Tick Population', *Journal of Medical Entomology*, 41(3), pp. 361–365. doi: 10.1603/0022-2585-41.3.361.
- Pfäffle, M., Littwin, N., Muders, S. V and Petney, T. N. (2013) 'The ecology of tick-borne diseases', *International Journal for Parasitology*, 43(12–13), pp. 1059–1077. doi: 10.1016/j.ijpara.2013.06.009.
- PHE (2019) *Human Animal Infections and Risk Surveillance ( HAIRS ) group Qualitative assessment of the risk that tick-borne encephalitis presents to the UK population*. London. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/832172/Tick\\_borne\\_encephalitis\\_risk\\_assessment.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/832172/Tick_borne_encephalitis_risk_assessment.pdf) (Accessed: 20 June 2020).
- Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006) 'Maximum entropy modeling of species geographic distributions', *Ecological Modelling*, 190(2006), pp. 231–259. doi: 10.1016/j.ecolmodel.2005.03.026.
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009) 'Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data', *Ecological Applications*, 19(1), pp. 181–197. doi: 10.1890/07-2153.1.
- Piesman, J. and Dolan, M. C. (2002) 'Protection against lyme disease spirochete transmission provided by prompt removal of nymphal Ixodes scapularis (Acari: Ixodidae).', *Journal of Medical Entomology*,



- 39(3), pp. 509–512. doi: 10.1603/0022-2585-39.3.509.
- Piesman, J. and Gern, L. (2004) 'Lyme borreliosis in Europe and North America', *Parasitology*, 129(7), pp. S191–S220. doi: 10.1017/S0031182003004694.
- Pietzsch, M. E., Medlock, J. M., Jones, L. and Avenell, D. (2005) 'Distribution of *Ixodes ricinus* in the British Isles : investigation of historical records', *Medical and Veterinary Entomology*, 19, pp. 306–314. doi: 10.1111/j.1365-2915.2005.00575.x.
- Poggio, L., Gimona, A., Spezia, L. and Brewer, M. J. (2016) 'Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA', *Geoderma*, 277, pp. 69–82. doi: 10.1016/j.geoderma.2016.04.026.
- Porretta, D., Mastrantonio, V., Amendolia, S., Gaiarsa, S., Epis, S., Genchi, C., Bandi, C., Otranto, D. and Urbanelli, S. (2013) 'Effects of global changes on the climatic niche of the tick *Ixodes ricinus* inferred by species distribution modelling', *Parasites & Vectors*, 6(271), pp. 1–8. doi: 10.1186/1756-3305-6-271.
- Porter, W. T., Motyka, P. J., Wachara, J., Barrand, Z. A., Hmood, Z., McLaughlin, M., Pemberton, K. and Nieto, N. C. (2019) 'Citizen science informs human-tick exposure in the Northeastern United States', *International Journal of Health Geographics*, 18(9), pp. 1–14. doi: 10.1186/s12942-019-0173-0.
- Porterfield, J. S. (1975) 'The basis of arbovirus classification', *Medical Biology*, 53(5), pp. 400–405.
- Portillo, A., Santibáñez, P., Palomar, A. M., Santibáñez, S. and Oteo, J. A. (2018) "'Candidatus Neoehrlichia mikurensis" in Europe', *New Microbes and New Infections*, 22, pp. 30–36. doi: 10.1016/j.nmni.2017.12.011.
- Prasad, A. M., Iverson, L. R. and Liaw, A. (2006) 'Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction', *Ecosystems*, 9(2), pp. 181–199. doi: 10.1007/s10021-005-0054-1.
- Purse, B. V., Mellor, P. S., Rogers, D. J., Samuel, A. R., Mertens, P. P. C. and Baylis, M. (2005) 'Climate change and the recent emergence of bluetongue in Europe', *Nature Reviews Microbiology*, 3(2), pp. 171–181. doi: 10.1038/nrmicro1090.
- Qviller, L., Viljugrein, H., Loe, L. E., Meisingset, E. L. and Mysterud, A. (2016) 'The influence of red deer space use on the distribution of *Ixodes ricinus* ticks in the landscape', *Parasites & Vectors*, 9(545), pp. 1–9. doi: 10.1186/s13071-016-1825-6.
- R Core Team (2013) 'R: A language and environment for statistical computing, v3.4.4.' Vienna, Austria: R Foundation for Statistical Computing.
- R Core Team (2019a) 'R: A Language and Environment for Statistical Computing, v3.5.3.' Vienna, Austria: R Foundation for Statistical Computing.
- R Core Team (2019b) 'R: A Language and Environment for Statistical Computing, v3.6.2.' Vienna, Austria: R Foundation for Statistical Computing.

Computing.

- Ramilo, D. W., Nunes, T., Madeira, S., Boinas, F. and Da Fonseca, I. P. (2017) 'Geographical distribution of Culicoides (DIPTERA: CERATOPOGONIDAE) in mainland Portugal: Presence/absence modelling of vector and potential vector species', *PLoS ONE*, 12(7), pp. 1–18. doi: 10.1371/journal.pone.0180606.
- Rand, P. W., Lacombe, E. H., Dearborn, R., Cahill, B., Elias, S., Lubelczyk, C. B., Beckett, G. A. and Smith, R. P. (2007) 'Passive surveillance in Maine, an area emergent for tick-borne diseases', *Journal of Medical Entomology*, 44(6), pp. 1118–1129. doi: 10.1603/0022-2585(2007)44[1118:PSIMAA]2.0.CO;2.
- Randolph, S. E. (1998) 'Ticks are not insects: Consequences of contrasting vector biology for transmission potential', *Parasitology Today*, 14(5), pp. 186–192. doi: 10.1016/S0169-4758(98)01224-1.
- Randolph, S. E. (2004) 'Tick ecology: Processes and patterns behind the epidemiological risk posed by ixodid ticks as vectors', *Parasitology*, 129, pp. S37–S65. doi: 10.1017/S0031182004004925.
- Randolph, S. E. (2009) 'Tick-borne disease systems emerge from the shadows: The beauty lies in molecular detail, the message in epidemiology', *Parasitology*, 136, pp. 1403–1413. doi: 10.1017/S0031182009005782.
- Randolph, S. E., Green, R., Hoodless, A. and Peacey, M. (2002) 'An empirical quantitative framework for the seasonal dynamics of *Ixodes ricinus*', *International Journal for Parasitology*, 32(2002), pp. 979–989. doi: 10.1016/s0020-7519(02)00030-9.
- Randolph, S. E. and Steele, G. M. (1985) 'An experimental evaluation of conventional control measures against the sheep tick, *Ixodes ricinus* (L.) (Acari: Ixodidae). II. The dynamics of the tick-host interaction', *Bulletin of Entomological Research*, 75(3), pp. 501–518. doi: 10.1017/S0007485300014607.
- Rath, P. M., Ibershoff, B., Mohnhaupt, A., Albig, J., Eljaschewitsch, B., Jürgens, D., Horbach, I. and Fehrenbach, F. J. (1996) 'Seroprevalence of Lyme borreliosis in forestry workers from Brandenburg, Germany', *European Journal of Clinical Microbiology and Infectious Diseases*, 15(5), pp. 372–377. doi: 10.1007/BF01690092.
- Ribeiro, R., Eze, J. I., Gilbert, L., Wint, G. R. W., Gunn, G., MacRae, A., Medlock, J. M. and Auty, H. (2019) 'Using imperfect data in predictive mapping of vectors: A regional example of *Ixodes ricinus* distribution', *Parasites & Vectors*, 12(536), pp. 1–13. doi: 10.1186/s13071-019-3784-1.
- Ribeiro, R., Wilson, A. J., Nunes, T., Ramilo, D. W., Amador, R., Madeira, S., Baptista, F. M., Harrup, L. E., Lucientes, J. and Boinas, F. (2015) 'Spatial and temporal distribution of culicoides species in mainland Portugal (2005-2010). Results of the Portuguese entomological surveillance programme', *PLoS ONE*, 10(4), pp. 1–19. doi: 10.1371/journal.pone.0124019.
- Richman, R., Diallo, D., Diallo, M., Sall, A. A., Faye, O., Diagne, C. T., Dia, I.,

- Weaver, S. C., Hanley, K. A. and Buenemann, M. (2018) 'Ecological niche modeling of Aedes mosquito vectors of chikungunya virus in southeastern Senegal', *Parasites & Vectors*, 11(255), pp. 1–17. doi: 10.1186/s13071-018-2832-6.
- Rijpkema, S. G. T., Tazelaar, D. J., Molkenboer, M. J. C. H., Noordhoek, G. T., Plantinga, G., Schouls, L. M. and Schellekens, J. F. P. (1997) 'Detection of *Borrelia afzelii*, *Borrelia burgdorferi sensu stricto*, *Borrelia garinii* and group VS116 by PCR in skin biopsies of patients with erythema migrans and acrodermatitis chronica atrophicans.', *Clinical Microbiology and Infection*, 3(1), pp. 109–116. doi: 10.1111/j.1469-0691.1997.tb00259.x.
- Ripoche, M., Gasmi, S., Adam-Poupart, A., Koffi, J. K., Lindsay, L. R., Ludwig, A., Milord, F., Ogden, N. H., Thivierge, K. and Leighton, P. A. (2018) 'Passive tick surveillance provides an accurate early signal of emerging lyme disease risk and human cases in Southern Canada', *Journal of Medical Entomology*, 55(4), pp. 1016–1026. doi: 10.1093/jme/tjy030.
- Robertson, J. N., Gray, J. S. and Stewart, P. (2000) 'Tick bite and Lyme borreliosis risk at a recreational site in England', *European Journal of Epidemiology*, 16, pp. 647–652. doi: 10.1023/a:1007615109273.
- De Roeck, E., Van Coillie, F., De Wulf, R., Soenen, K., Charlier, J., Vercruyse, J., Hantson, W., Ducheyne, E. and Hendrickx, G. (2014) 'Fine-scale mapping of vector habitats using very high resolution satellite imagery: A liver fluke case-study', *Geospatial Health*, 8(3), pp. S671–S683. doi: 10.4081/gh.2014.296.
- Rosà, R., Pugliese, A., Norman, R. and Hudson, P. J. (2003) 'Thresholds for disease persistence in models for tick-borne infections including non-viraemic transmission, extended feeding and tick aggregation', *Journal of Theoretical Biology*, 224(3), pp. 359–376. doi: 10.1016/S0022-5193(03)00173-5.
- Rowland, C., Morton, R., Carrasco, L., McShane, G., O'Neil, A. and Wood, C. (2017a) *Land Cover Map 2015 (1 km dominant target class, GB)*. NERC Environmental Information Data Centre. doi: 10.5285/c4035f3d-d93e-4d63-a8f3-b00096f597f5.
- Rowland, C., Morton, R., Carrasco, L., McShane, G., O'Neil, A. and Wood, C. (2017b) *Land Cover Map 2015 (1 km percentage target class, GB)*. NERC Environmental Information Data Centre. doi: 10.5285/505d1e0c-ab60-4a60-b448-68c5bbae403e.
- Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. 1st edn, *Monographs on statistics and applied probability 104*. 1st edn. New York: Chapman & Hall/CRC. doi: 10.1201/9780203492024.
- Rue, H., Martino, S. and Chopin, N. (2009) 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations', *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(2), pp. 319–392. doi: 10.1111/j.1467-9868.2008.00700.x.
- Rue, H., Martino, S., Lindgren, F., Simpson, D. and Riebler, A. (2013) 'R-

- INLA: Approximate Bayesian Inference Using Integrated Nested Laplace Approximations.' Trondheim, Norway. Available at: <http://www.r-inla.org/>.
- Ruiz-Fons, F., Fernández-de-Mera, I., Acevedo, P., Gortázar, C. and de la Fuente, J. (2012) 'Factors driving the abundance of *Ixodes ricinus* ticks and the prevalence of zoonotic *I. ricinus*-borne pathogens in natural foci', *Applied and Environmental Microbiology*, 78(8), pp. 2669–2676. doi: 10.1128/AEM.06564-11.
- Ruiz-Fons, F. and Gilbert, L. (2010) 'The role of deer as vehicles to move ticks, *Ixodes ricinus*, between contrasting habitats', *International Journal for Parasitology*, 40(9), pp. 1013–1020. doi: 10.1016/j.ijpara.2010.02.006.
- Sakamoto, J. M. (2018) 'Progress, challenges, and the role of public engagement to improve tick-borne disease literacy', *Current Opinion in Insect Science*, 28, pp. 81–89. doi: 10.1016/j.cois.2018.05.011.
- Salkeld, D. J., Porter, W. T., Loh, S. M. and Nieto, N. C. (2019) 'Time of year and outdoor recreation affect human exposure to ticks in California, United States', *Ticks and Tick-borne Diseases*, 10(5), pp. 1113–1117. doi: 10.1016/j.ttbdis.2019.06.004.
- Sallam, M. F., Xue, R. De, Pereira, R. M. and Koehler, P. G. (2016) 'Ecological niche modeling of mosquito vectors of West Nile virus in St. John's County, Florida, USA', *Parasites & Vectors*, 9(371), pp. 1–14. doi: 10.1186/s13071-016-1646-7.
- Santino, I., Cammarata, E., Franco, S., Galdier, F., Oliva, B., Sessa, R., Cipriani, P., Tempera, G. and Piano, M. D. E. L. (2004) 'Multicentric study of seroprevalence of *Borrelia burgdorferi* and *Anaplasma phagocytophila* in high-risk groups in regions of central and southern Italy', *International Journal Immunopathology and Pharmacology*, 17, pp. 219–223. doi: 10.1177/039463200401700214.
- Santos, J. and Meneses, B. M. (2017) 'An integrated approach for the assessment of the *Aedes aegypti* and *Aedes albopictus* global spatial distribution, and determination of the zones susceptible to the development of Zika virus', *Acta Tropica*, 168(2017), pp. 80–90. doi: 10.1016/j.actatropica.2017.01.015.
- Sarkar, S., Strutz, S. E., Frank, D. M., Rivaldi, C. L., Sissel, B. and Sánchez-Cordero, V. (2010) 'Chagas disease risk in Texas', *PLoS Neglected Tropical Diseases*, 4(10), pp. 1–14. doi: 10.1371/journal.pntd.0000836.
- Scharlemann, J., Johnson, P., Smith, A., Macdonald, D. and Randolph, S. (2008) 'Trends in ixodid tick abundance and distribution in Great Britain', *Medical and Veterinary Entomology*, 22, pp. 238–247. doi: 10.1111/j.1365-2915.2008.00734.x.
- Schulz, M., Mahling, M. and Pfister, K. (2014) 'Abundance and seasonal activity of questing *Ixodes ricinus* ticks in their natural habitats in southern Germany in 2011', *Journal of Vector Ecology*, 39(1), pp. 56–65. doi: 10.1111/j.1948-7134.2014.12070.x.
- Schwarz, A., Hönig, V., Vavrušková, Z., Grubhoffer, L., Balczun, C., Albring, A. and Schaub, G. A. (2012) 'Abundance of *Ixodes ricinus* and

- prevalence of *Borrelia burgdorferi* s.l. in the nature reserve Siebengebirge, Germany, in comparison to three former studies from 1978 onwards.', *Parasites & vectors*, 5(268), pp. 1–15. doi: 10.1186/1756-3305-5-268.
- Schwarz, A., Maier, W. A., Kistemann, T. and Kampen, H. (2009) 'Analysis of the distribution of the tick *Ixodes ricinus* L. (Acari: Ixodidae) in a nature reserve of western Germany using Geographic Information Systems', *International Journal of Hygiene and Environmental Health*, 212(2009), pp. 87–96. doi: 10.1016/j.ijheh.2007.12.001.
- Scottish Government (2020) *Scotland is now - Weather*. Available at: <https://www.scotland.org/about-scotland/weather> (Accessed: 10 July 2020).
- Seifert, V. A., Wilson, S., Toivonen, S., Clarke, B. and Prunuske, A. (2016) 'Community Partnership Designed to Promote Lyme Disease Prevention and Engagement in Citizen Science', *Journal of Microbiology & Biology Education*, 17(1), pp. 63–69. doi: 10.1128/jmbe.v17i1.1014.
- Signorini, M., Stensgaard, A. S., Drigo, M., Simonato, G., Marcer, F., Montarsi, F., Martini, M. and Cassini, R. (2019) 'Towards improved, cost-effective surveillance of *Ixodes ricinus* ticks and associated pathogens using species distribution modelling', *Geospatial Health*, 14(745), pp. 46–52. doi: 10.4081/gh.2019.745.
- Silva, L. D., De Azevedo, E. B., Elias, R. B. and Silva, L. (2017) 'Species distribution modeling: Comparison of fixed and mixed effects models using INLA', *ISPRS International Journal of Geo-Information*, 6(12), pp. 1–35. doi: 10.3390/ijgi6120391.
- Silvertown, J. (2009) 'A new dawn for citizen science', *Trends in Ecology and Evolution*, 24(9), pp. 467–471. doi: 10.1016/j.tree.2009.03.017.
- Simons, R. R. L., Croft, S., Rees, E., Tearne, O., Arnold, M. E. and Johnson, N. (2019) 'Using species distribution models to predict potential hot-spots for Rift Valley Fever establishment in the United Kingdom', *PLoS ONE*, 14(12), pp. 1–18. doi: 10.1371/journal.pone.0225250.
- Sinka, M. E., Bangs, M. J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C. M., Hemingway, J., Patil, A. P., Temperley, W. H., Gething, P. W., Kabaria, C. W., Burkot, T. R., Harbach, R. E. and Hay, S. I. (2012) 'A global map of dominant malaria vectors', *Parasites & Vectors*, 5(69), pp. 1–11. doi: 10.1186/1756-3305-5-69.
- Šmit, R. and Postma, M. J. (2015) 'The burden of tick-borne encephalitis in disability-adjusted life years (DALYs) for Slovenia', *PLoS ONE*, 10(12), pp. 1–17. doi: 10.1371/journal.pone.0144988.
- SOA (2020) *Scottish Orienteering Association*. Available at: <https://www.scottish-orienteering.org/> (Accessed: 20 March 2020).
- Sormunen, J. J., Klemola, T., Vesterinen, E. J., Vuorinen, I., Hytönen, J., Hänninen, J., Ruohomäki, K., Sääksjärvi, I. E. and Tonteri, E. (2016) 'Ticks and Tick-borne Diseases Assessing the abundance, seasonal questing activity, and *Borrelia* and tick-borne encephalitis virus (TBEV) prevalence of *Ixodes ricinus* ticks in a Lyme borreliosis

- endemic area in Southwest Finland', *Ticks and Tick-borne Diseases*, 7(2016), pp. 208–215. doi: 10.1016/j.ttbdis.2015.10.011.
- Soucy, J. P. R., Slatculescu, A. M., Nyiraneza, C., Ogden, N. H., Leighton, P. A., Kerr, J. T. and Kulkarni, M. A. (2018) 'High-Resolution Ecological Niche Modeling of Ixodes scapularis Ticks Based on Passive Surveillance Data at the Northern Frontier of Lyme Disease Emergence in North America', *Vector-Borne and Zoonotic Diseases*, 18(5), pp. 235–242. doi: 10.1089/vbz.2017.2234.
- Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1999) 'WinBUGS version 1.2. User Manual.', *MRC Biostatistics Unit*.
- Sprong, H., Azagi, T., Hoornstra, D., Nijhof, A. M., Knorr, S., Baarsma, M. E. and Hovius, J. W. (2018) 'Control of Lyme borreliosis and other Ixodes ricinus-borne diseases', *Parasites & Vectors*, 11(145), pp. 1–16. doi: 10.1186/s13071-018-2744-5.
- Sprong, H., Hofhuis, A., Gassner, F., Takken, W., Jacobs, F. and Vliet, A. J. H. Van (2012) 'Circumstantial evidence for an increase in the total number and activity of borrelia-infected Ixodes ricinus in the Netherlands', *Parasites & Vectors*, 5(294), pp. 1–11. doi: 10.1186/1756-3305-5-294.
- Stanek, G. and Reiter, M. (2011) 'The expanding Lyme Borrelia complex--clinical significance of genomic species?', *Clinical Microbiology and Infection*, 17(4), pp. 487–493. doi: 10.1111/j.1469-0691.2011.03492.x.
- Stanton, M. C., Esterhuizen, J., Tirados, I., Betts, H. and Torr, S. J. (2018) 'The development of high resolution maps of tsetse abundance to guide interventions against human African trypanosomiasis in northern Uganda', *Parasites and Vectors*, 11(1), pp. 1–12. doi: 10.1186/s13071-018-2922-5.
- Steere, A. C., Malawista, S. E., Snyderman, D. R., Shope, R. E., Andiman, W. a, Ross, M. R. and Steele, F. M. (1977) 'Lyme Arthritis: An epidemic of oligoarticular arthritis in children and adults in three Connecticut communities', *Arthritis and Rheumatism*, 20(1), pp. 7–17. doi: 10.1002/art.1780200102.
- Steere, A. C., Strle, F., Wormser, G. P., Hu, L. T., Branda, J. A., Hovius, J. W. R., Li, X. and Mead, P. S. (2016) 'Lyme borreliosis.', *Nature reviews. Disease primers*, 2(16090), pp. 1–44. doi: 10.1038/nrdp.2016.90.
- Stensgaard, A. S., Vounatsou, P., Onapa, A. W., Simonsen, P. E., Pedersen, E. M., Rahbek, C. and Kristensen, T. K. (2011) 'Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: Predictors of risk and geographical patterns of co-endemicity', *Malaria Journal*, 10(298), pp. 1–14. doi: 10.1186/1475-2875-10-298.
- Stjernberg, L. and Berglund, J. (2002) 'Risk of Acquiring Tick Bites in South-eastern Sweden', *Scandinavian Journal of Infectious Diseases*, 34, pp. 840–844. doi: 10.1080/0036554021000026955.
- Stone, B., Tourand, Y. and Brissette, C. (2017) 'Brave New Worlds: The Expanding Universe of Lyme Disease', *Vector-Borne and Zoonotic Diseases*, 17(9), pp. 619–629. doi: 10.1089/vbz.2017.2127.

- Swart, A., Ibañez-Justicia, A., Buijs, J., van Wieren, S., Hofmeester, T., Sprong, H. and Takumi, K. (2014) 'Predicting Tick Presence by Environmental Risk Mapping', *Frontiers in Public Health*, 2(238), pp. 1–8. doi: 10.3389/fpubh.2014.00238.
- Switters, J. and Osimo, D. (2019) *Citizen Science in the Surveillance and Monitoring of Mosquito-Borne Diseases. Open Science Monitor Case Study*. Brussels. doi: 10.2777/431775.
- Sykes, R. A. and Makiello, P. (2017) 'An estimate of Lyme borreliosis incidence in Western Europe', *Journal of Public Health*, 39(1), pp. 74–81. doi: 10.1093/pubmed/fdw017.
- Tack, W., Madder, M., Baeten, L., Vanhellemont, M., Gruwez, R. and Verheyen, K. (2012) 'Forest Ecology and Management Local habitat and landscape affect Ixodes ricinus tick abundances in forests on poor , sandy soils', *Forest Ecology and Management*, 265, pp. 30–36. doi: 10.1016/j.foreco.2011.10.028.
- Takken, W., Van Vliet, A. J. H., Verhulst, N. O., Jacobs, F. H. H., Gassner, F., Hartemink, N., Mulder, S. and Sprong, H. (2017) 'Acarological risk of borrelia burgdorferi sensu lato infections across space and time in the Netherlands', *Vector-Borne and Zoonotic Diseases*, 17(2), pp. 99–107. doi: 10.1089/vbz.2015.1933.
- Tälleklint, L. and Jaenson, T. G. T. (1998) 'Increasing Geographical Distribution and Density of Ixodes ricinus (Acari: Ixodidae) in Central and Northern Sweden', *Journal of Medical Entomology*, 35(4), pp. 521–526. doi: 10.1093/jmedent/35.4.521.
- Thuiller, W. (2003) 'BIOMOD - Optimizing predictions of species distributions and projecting potential future shifts under global change', *Global Change Biology*, 9(10), pp. 1353–1362. doi: 10.1046/j.1365-2486.2003.00666.x.
- Tiago, P., Pereira, H. M. and Capinha, C. (2017) 'Using citizen science data to estimate climatic niches and species distributions', *Basic and Applied Ecology*, 20(2017), pp. 75–85. doi: 10.1016/j.baae.2017.04.001.
- Tiffin, H. S., Peper, S. T., Wilson-Fallon, A. N., Haydett, K. M., Cao, G. and Presley, S. M. (2019) 'The influence of new surveillance data on predictive species distribution modeling of aedes aegypti and aedes albopictus in the United States', *Insects*, 10(400), pp. 1–15. doi: 10.3390/insects10110400.
- Tobler, W. (1970) 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*, 46, pp. 234–240. doi: 10.2307/143141.
- Tomkins, J., Aungier, J., Hazel, W. and Gilbert, L. (2014) 'Towards an evolutionary understanding of questing behaviour in the tick Ixodes ricinus', *PLoS ONE*, 9(10), pp. 1–11. doi: 10.1371/journal.pone.0110028.
- Tregidgo, D. J., West, S. E. and Ashmore, M. R. (2013) 'Can citizen science produce good science? Testing the OPAL Air Survey methodology, using lichens as indicators of nitrogenous pollution', *Environmental Pollution*, 182(2013), pp. 448–451. doi: 10.1016/j.envpol.2013.03.034.

- Tulloch, J. S. P., Semper, A. E., Brooks, T. J. G., Russell, K., Halsby, K. D., Christley, R. M., Radford, A. D., Vivancos, R. and Warner, J. C. (2019) 'The demographics and geographic distribution of laboratory-confirmed Lyme disease cases in England and Wales (2013-2016): An ecological study', *BMJ Open*, 9(8), pp. 1–8. doi: 10.1136/bmjopen-2018-028064.
- Tyson, E., Bowser, A., Palmer, J., Kapan, D., Bartumeus, F. and Brocklehurst, M. (2018) *Global Mosquito Alert : Building Citizen Science Capacity for Surveillance and Control of Disease-Vector Mosquitoes*. Washington. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1471492218300928> (Accessed: 12 July 2020).
- Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Nilsen, K., Takahashi, S., Metcalf, C. J. E., Lessler, J. and Tatem, A. J. (2019) 'A spatial regression model for the disaggregation of areal unit based data to high-resolution grids with application to vaccination coverage mapping', *Statistical Methods in Medical Research*, 28(10–11), pp. 3226–3241. doi: 10.1177/0962280218797362.
- Venturi, G., Di Luca, M., Fortuna, C., Remoli, M. E., Riccardo, F., Severini, F., Toma, L., Del Manso, M., Benedetti, E., Caporali, M. G., Amendola, A., Fiorentini, C., De Liberato, C., Giammattei, R., Romi, R., Pezzotti, P., Rezza, G. and Rizzo, C. (2017) 'Detection of a chikungunya outbreak in central Italy, August to September 2017', *Eurosurveillance*, 22(39), pp. 11–14. doi: 10.2807/1560-7917.ES.2017.22.39.17-00646.
- Vianna, G. M. S., Meekan, M. G., Bornovski, T. H. and Meeuwig, J. J. (2014) 'Acoustic telemetry validates a citizen science approach for monitoring sharks on coral reefs', *PLoS ONE*, 9(4), pp. 1–12. doi: 10.1371/journal.pone.0095565.
- vonHedemann, N., Robbins, P., Butterworth, M. K., Landau, K. and Morin, C. W. (2016) 'Managing mosquito spaces: Citizen self-governance of disease vectors in a desert landscape.', *Health and Place*, 43, pp. 41–48. doi: 10.1016/j.healthplace.2016.11.004.
- Vourc'h, G., Abrial, D., Bord, S., Jacquot, M., Masségli, S., Poux, V., Pisanu, B., Bailly, X. and Chapuis, J. L. (2016) 'Mapping human risk of infection with *Borrelia burgdorferi sensu lato*, the agent of Lyme borreliosis, in a periurban forest in France', *Ticks and Tick-borne Diseases*, 7(5), pp. 644–652. doi: 10.1016/j.ttbdis.2016.02.008.
- Wagemakers, A., Staarink, P. J., Sprong, H. and Hovius, J. W. R. (2015) 'Borrelia miyamotoi: A widespread tick-borne relapsing fever spirochete', *Trends in Parasitology*, 31(6), pp. 260–269. doi: 10.1016/j.pt.2015.03.008.
- Walker, A. R., Alberdi, M. P., Urquhart, K. A. and Rose, H. (2001) 'Risk factors in habitats of the tick *Ixodes ricinus* influencing human exposure to *Ehrlichia phagocytophila* bacteria', *Medical and Veterinary Entomology*, 15, pp. 40–49. doi: 10.1046/j.1365-2915.2001.00271.x.
- Westblade, L. F., Simon, M. S., Mathison, B. A. and Kirkman, L. A. (2017) 'Babesia microti: from Mice to Ticks to an Increasing Number of Highly



- Susceptible Humans', *Journal of Clinical Microbiology*, 55(10), pp. 2903–2912. doi: 10.1128/JCM.00504-17.
- WHO (2014) *A global brief on vector-borne diseases*. Geneva. Available at: [http://apps.who.int/iris/bitstream/10665/111008/1/WHO\\_DCO\\_WHD\\_2014.1\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/111008/1/WHO_DCO_WHD_2014.1_eng.pdf) (Accessed: 14 September 2020).
- WHO (2016) *World Malaria Report 2016*. Geneva. Available at: <https://apps.who.int/iris/bitstream/handle/10665/252038/9789241511711-eng.pdf?sequence=1> (Accessed: 14 September 2020).
- WHO (2017) *Global vector control response 2017–2030*. Geneva. Available at: <https://apps.who.int/iris/bitstream/handle/10665/259205/9789241512978-eng.pdf?sequence=1> (Accessed: 14 September 2020).
- WHO (2020) *Vector-borne diseases*. Available at: <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases> (Accessed: 12 October 2020).
- Wickham, H. (2019) 'tidyverse: Easily Install and Load the "Tidyverse".' R package version 1.3.0.
- Wickham, H. (2020) 'tidyr: Tidy Messy Data.' R package version 1.1.2.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H. and Dunnington, D. (2020) 'ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics'. R package version 3.3.0.
- Wickham, H., François, R., Henry, L. and Müller, K. (2020) 'dplyr: A Grammar of Data Manipulation.' R package version 1.0.2.
- Wickham, H. and Grolemund, G. (2017) *R for data science: chapter 10 Tibbles*. Available at: <https://r4ds.had.co.nz/tibbles.html> (Accessed: 2 June 2020).
- Wiggins, A., Newman, G., Stevenson, R. D. and Crowston, K. (2011) 'Mechanisms for data quality and validation in citizen science', in *Proceedings - 7th IEEE International Conference on e-Science Workshops, eScienceW 2011*, pp. 14–19. doi: 10.1109/eScienceW.2011.27.
- van den Wijngaard, C. C., Hofhuis, A., Harms, M. G., Haagsma, J. A., Wong, A., de Wit, G. A., Havelaar, A. H., Lugnér, A. K., Suijkerbuijk, A. W. M. and van Pelt, W. (2015) 'The burden of Lyme borreliosis expressed in disability-adjusted life years.', *European Journal of Public Health*, 25(6), pp. 1071–1078. doi: 10.1093/eurpub/ckv091.
- van den Wijngaard, C. C., Hofhuis, A., Simões, M., Rood, E., van Pelt, W., Zeller, H. and Van Bortel, W. (2017) 'Surveillance perspective on Lyme borreliosis across the European Union and European Economic Area', *Eurosurveillance*, 22(27), pp. 1–9. doi: 10.2807/1560-7917.ES.2017.22.27.30569.
- Wilhelmsson, P., Lindblom, P., Fryland, L., Nyman, D., Jaenson, T. G. T. and Forsberg, P. (2013) 'Ixodes ricinus ticks removed from humans in Northern Europe : seasonal pattern of infestation , attachment sites and duration of feeding', *Parasites & vectors*, 6(362), pp. 1–11. doi: 10.1186/1756-3305-6-362.
- Wint, W., Morley, D., Medlock, J. and Alexander, N. S. (2014) 'A First

- Attempt at Modelling Red Deer (*Cervus elaphus*) Distributions Over Europe', *Open Health Data*, 2(1), pp. 1–5. doi: 10.5334/ohd.ah.
- Wise, J. (2020) 'Covid-19: Study reveals six clusters of symptoms that could be used as a clinical prediction tool', *BMJ (Clinical research ed.)*, 370(m2911), pp. 1–2. doi: 10.1136/bmj.m2911.
- Worton, A. (2016) *Using Mathematical Models to Understand the Impact of Climate Change on Tick-Borne Infections across Scotland*. PhD thesis. University of Stirling. Available at: <https://dspace.stir.ac.uk/bitstream/1893/24918/1/thesis.pdf> (Accessed: 10 January 2019).
- Wozniak, E. J., Lawrence, G., Gorchakov, R., Alamgir, H., Dotson, E., Sissel, B., Sarkar, S. and Murray, K. O. (2015) 'The Biology of the Triatomine Bugs Native to South Central Texas and Assessment of the Risk They Pose for Autochthonous Chagas Disease Exposure', *Journal of Parasitology*, 101(5), pp. 520–528. doi: 10.1645/15-748.
- Xiao, Y. and Zou, X. (2014) 'Transmission dynamics for vector-borne diseases in a patchy environment', *Journal of Mathematical Biology*, 69, pp. 113–146. doi: 10.1007/s00285-013-0695-1.
- Xu, G., Mather, T. N., Hollingsworth, C. S. and Rich, S. M. (2016) 'Passive Surveillance of *Ixodes scapularis* (Say), Their Biting Activity, and Associated Pathogens in Massachusetts', *Vector-Borne and Zoonotic Diseases*, 16(8), pp. 520–527. doi: 10.1089/vbz.2015.1912.
- Xu, G., Pearson, P., Dykstra, E., Andrews, E. S. and Rich, S. M. (2019) 'Human-Biting *Ixodes* Ticks and Pathogen Prevalence from California, Oregon, and Washington', *Vector-Borne and Zoonotic Diseases*, 19(2), pp. 106–114. doi: 10.1089/vbz.2018.2323.
- Yee, T. W. and Mitchell, N. D. (1991) 'Generalized additive models in plant ecology', *Journal of Vegetation Science*, 2(5), pp. 587–602. doi: 10.2307/3236170.
- Yoshioka, K. (2013) 'Impact of a community-based bug-hunting campaign on Chagas disease control: A case study in the department of Jalapa, Guatemala', *Memórias do Instituto Oswaldo Cruz*, 108(2), pp. 205–211. doi: 10.1590/0074-0276108022013013.
- Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., Rue, H. and Gerrodette, T. (2016) 'Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales', *Annals of Applied Statistics*, 11(4), pp. 2270–2297. doi: 10.1214/17-AOAS1078.
- Zanotto, P. M. de A. and Leite, L. C. de C. (2018) 'The Challenges Imposed by Dengue, Zika, and Chikungunya to Brazil', *Frontiers in Immunology*, 9(1964), pp. 1–6. doi: 10.3389/fimmu.2018.01964.
- Zhang, G. and Zhu, A. X. (2018) 'The representativeness and spatial bias of volunteered geographic information: a review', *Annals of GIS*, 24(3), pp. 151–162. doi: 10.1080/19475683.2018.1501607.
- Zhioua, E., Gern, L., Aeschlimann, A., Sauvain, M. J., Van der Linden, S. and Fahrner, H. (1998) 'Longitudinal study of Lyme borreliosis in a high risk population in Switzerland', *Parasite*, 5(4), pp. 383–386. doi: 10.1051/parasite/1998054383.

Zintl, A., Mulcahy, G., Skerrett, H. E., Taylor, S. M. and Gray, J. S. (2003) 'Babesia divergens, a Bovine Blood Parasite of Veterinary and Zoonotic Importance', *Clinical Microbiology Reviews*, 16(4), pp. 622–636. doi: 10.1128/CMR.16.4.622-636.2003.

ZOE Global Limited & King's College London (2020) *COVID Symptom Study*. Available at: <https://covid.joinzoe.com/> (Accessed: 27 November 2020).

## Appendix A: Supplementary material for Chapter 1

**Table A-S1.1:** Non-exhaustive list of CS projects implemented in the US and Canada, with the objective of generating data and information on ticks and TBDs.

The projects are based mainly on *I. scapularis*, their dominant vector species. This table includes information about the project, the data collected, and a link for the study and/or published references.

Country or region or state	Project name and developers	Description of the project and data collected	Link for the study and/or references
<b>Northeast US</b>	Bay Area Lyme Foundation and Northern Arizona University	From January 2016 until August 2017, Northern Arizona University offered a free tick identification and testing service made available to the general public. Citizens emailed information about the tick encounter (bite or crawling): location, date, habitat type, host (human, pet or other animals), activity of the person who encountered the tick. Ticks were sent to Northern Arizona University to test to pathogenic agents.	<a href="http://www.bayarealyme.org/">www.bayarealyme.org/</a> (Nieto <i>et al.</i> , 2018; Porter <i>et al.</i> , 2019; Salkeld <i>et al.</i> , 2019)
<b>Massachusetts and then extended to other areas of the US</b>	Project Acari, University of Massachusetts Medical School	Designed in 2017 to collect microbiome and pathogen data from large numbers of individual ticks collected from both broad geographic areas and over extended time periods, from humans, pets and other animals. Citizen scientists mail unattached ticks they find during their normal activities, each labelled by geographic location, using 'Tick Kit'.	<a href="https://projectacari.org/">https://projectacari.org/</a> (Chauhan <i>et al.</i> , 2020)
<b>Vermont</b>	Vermont Tick tracker, Vermont Department of Health	Website which started in 2013 and allows citizens to contribute with tick data (report a tick location and tick species). Results are uploaded in a map with a layer of the tick reports and a layer of disease incidence.	<a href="https://apps.health.vermont.gov/vttracking/ticktracker/2019/d/index.html">https://apps.health.vermont.gov/vttracking/ticktracker/2019/d/index.html</a>
<b>All US</b>	The Tick App project, Columbia University and the University of Wisconsin – Madison, and members of the Centers for Disease Control and Prevention	This project has an app and a website, with information to educate citizens about ticks, and which also allows citizens to report tick encounters (in humans and pets). The final objective is to understand how people's activities and practices influence their exposure to ticks. People engaged in the study need to complete daily logs (activity reports). Tick reports are followed by the questions about the host, location, type of activity, tick species, and picture of the tick. During the submission, the app records	<a href="https://thetickapp.org/">https://thetickapp.org/</a> (Fernandez <i>et al.</i> , 2019)

		date, time and weather. The app will also track user's location by GPS as frequently as every 15 minutes and as precise as 10 ft within people's actual location. This information will be used to determine how much time people spend in high and low risk environments.	
<b>All US</b>	'Ticktracker', LivLyme Foundation	'Ticktracker' is a mobile app developed to help educate people to prevent tick borne diseases and a CS initiative where people report tick sightings and bites. Results are displayed in a map shared with participants, which displays a tick severity index for different areas across US.	<a href="https://ticktracker.com/">https://ticktracker.com/</a>
<b>Rhode Island</b>	'TickEncounter' Resource Center, University of Rhode Island	The 'TickEncounter' Resource Center promotes tick-bite protection and TBD prevention. They developed 'TickSpotters' which tracks tick activity across North America. Citizen scientists submit tick reports from ticks removed from themselves or pets and submit tick photos through the website.	<a href="http://www.tickencounter.org">http://www.tickencounter.org</a> <a href="http://www.tickencounter.org/tickspotters">http://www.tickencounter.org/tickspotters</a>
<b>Connecticut</b>	'TickNET', Centers for Disease Control and Prevention through the Emerging Infections Program Cooperative Agreement	Created in 2007, aims to identify risk factors and preventive measures against TBDs. In one of the studies, Connecticut's citizens completed a log diary each night for one week during June 2013, recording where they were and what they did, and if they had a tick (bite or crawling), which enabled researchers to identify risk factors for tick exposure.	(Mead <i>et al.</i> , 2015, 2018)
<b>Canada</b>	Passive surveillance	Passive surveillance for the occurrence of <i>I. scapularis</i> ticks and their infection with <i>B. burgdorferi</i> (s.l.) has taken place in Canada since 1990. Citizens submit ticks found on themselves and pets. Some of these passive surveillance programmes are incentivising citizens to report photographs of the ticks they found (date, location, host) through etick.ca platform (website and app).	<a href="https://www.etick.ca/en">https://www.etick.ca/en</a> (Ogden <i>et al.</i> , 2006, 2010; Koffi <i>et al.</i> , 2012; Gasmi <i>et al.</i> , 2016; Ripoche <i>et al.</i> , 2018)
<b>Massachusetts, California, Oregon, Washington</b>	Passive surveillance, University of Massachusetts Amherst	Passive surveillance of ticks and tick-borne pathogens was implemented in 2006. Tick specimens are sent by mail with correspondent information on the location, data and host (human or animals), age, gender and attachment site. Currently to help the process of surveillance, this is implemented via the 'TickReport' program with website and app.	<a href="https://ag.umass.edu/news-events/highlights/2019-tick-box-mail-tick">https://ag.umass.edu/news-events/highlights/2019-tick-box-mail-tick</a> <a href="https://www.tickreport.com/">https://www.tickreport.com/</a> (Xu <i>et al.</i> , 2016, 2019)
<b>Indiana</b>	'Tick Insiders',	'Tick Insiders' is a CS project aiming to help prevention, diagnosis and treatment of TBDs in Indiana, by determining the range and incidence of tick vector and the pathogens in the ticks. Experts	<a href="https://tickinsiders.org/">https://tickinsiders.org/</a>

	Purdue University, Entomology College of Agriculture	conduct tick surveillance with standardised methods of tick collection. Since 2019 that the project is open to citizens who can send ticks crawling on themselves or pets, with the additional information of date and location of the tick (s) found. People are also incentivised to collect ticks in their property using standard methods.	
<b>Arkansas</b>	The Arkansas Tick borne Disease Project, Arkansas Biosciences Institute	This project started in 2017 with the objective of identifying risk areas and what disease-causing agents are present in Arkansas, to then inform decision making and educate citizens. Experts are actively collecting ticks around the state, identifying the ticks, and then screening the ticks for disease causing agents. Citizens can send alive and intact ticks (crawling on themselves and/or on animals, or collected from the vegetation), with date and location of collection. Project distribute kit to citizens and are looking for people who wants to volunteer to collect ticks actively.	<a href="https://adowling.hosted.uar.k.edu/ARTicks/index.html">https://adowling.hosted.uar.k.edu/ARTicks/index.html</a> <a href="https://www.uaex.edu/farm-ranch/special-programs/arkansas-tick-project.aspx">https://www.uaex.edu/farm-ranch/special-programs/arkansas-tick-project.aspx</a>
<b>Maine</b>	Tick Submission Program, Maine Department of Health and Human Services	This passive surveillance project started in 1989, with citizens submitting ticks found on themselves and animals, with additional information: date and location of collection, host, age of the person the tick was removed.	(Rand <i>et al.</i> , 2007)
	The Maine Forest Survey, School of Biology and Ecology and University of Maine	Recent CS project (2020) aiming to understand the growing risks of TBDs in Maine. The objective is to determine how forest land management practices impact tick populations and disease risk. Information will be gathered with the help of CS landowners, where volunteers are engaged to collect tick on their wooded properties.	<a href="https://umaine.edu/forestticksurvey/">https://umaine.edu/forestticksurvey/</a>

Abbreviations: App, application; CS, citizen science; GPS, global positioning system; s.l., sensu lato; TBD, tick-borne disease; US, United States

# Appendix B: Supplementary materials for Chapter 2

## S1: Using imperfect data in predictive mapping of vectors: a regional example of *Ixodes ricinus* distribution.

Ribeiro et al. *Parasites Vectors* (2019) 12:536  
<https://doi.org/10.1186/s13071-019-3784-1>

Parasites & Vectors

RESEARCH

Open Access

### Using imperfect data in predictive mapping of vectors: a regional example of *Ixodes ricinus* distribution



Rita Ribeiro<sup>1,2\*</sup>, Jude I. Eze<sup>1,3</sup>, Lucy Gilbert<sup>4</sup>, G. R. William Wint<sup>5</sup>, George Gunn<sup>1</sup>, Alastair Macrae<sup>2</sup>, Jolyon M. Medlock<sup>6</sup> and Harriet Auty<sup>1</sup>

#### Abstract

**Background:** Knowledge of *Ixodes ricinus* tick distribution is critical for surveillance and risk management of transmissible tick-borne diseases such as Lyme borreliosis. However, as the ecology of *I. ricinus* is complex, and robust long-term geographically extensive distribution tick data are limited, mapping often relies on datasets collected for other purposes. We compared the modelled distributions derived from three datasets with information on *I. ricinus* distribution (quantitative *I. ricinus* count data from scientific surveys; *I. ricinus* presence-only data from public submissions; and a combined *I. ricinus* dataset from multiple sources) to assess which could be reliably used to inform Public Health strategy. The outputs also illustrate the strengths and limitations of these three types of data, which are commonly used in mapping tick distributions.

**Methods:** Using the Integrated Nested Laplace algorithm we predicted *I. ricinus* abundance and presence-absence in Scotland and tested the robustness of the predictions, accounting for errors and uncertainty.

**Results:** All models fitted the data well and the covariate predictors for *I. ricinus* distribution, i.e. deer presence, temperature, habitat, index of vegetation, were as expected. Differences in the spatial trend of *I. ricinus* distribution were evident between the three predictive maps. Uncertainties in the spatial models resulted from inherent characteristics of the datasets, particularly the number of data points, and coverage over the covariate range used in making the predictions.

**Conclusions:** Quantitative *I. ricinus* data from scientific surveys are usually considered to be gold standard data and we recommend their use where high data coverage can be achieved. However in this study their value was limited by poor data coverage. Combined datasets with *I. ricinus* distribution data from multiple sources are valuable in addressing issues of low coverage and this dataset produced the most appropriate map for national scale decision-making in Scotland. When mapping vector distributions for public-health decision making, model uncertainties and limitations of extrapolation need to be considered; these are often not included in published vector distribution maps. Further development of tools to better assess uncertainties in the models and predictions are necessary to allow more informed interpretation of distribution maps.

**Keywords:** Data quality, Decision making, *Ixodes ricinus*, Predictive maps, Public health, Uncertainty, Vector-borne diseases

\*Correspondence: [rita.ribeiro@ruc.ac.uk](mailto:rita.ribeiro@ruc.ac.uk)

<sup>1</sup> Department of Veterinary and Animal Science, Northern Faculty  
Scotland's Rural College, An Lochran, 10 Inverness Campus, Inverness IV2  
SNA, UK

Full list of author information is available at the end of the article



© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

## Background

Tick-borne diseases affect the health of both humans and animals and impact on the economy [1]. Lyme borreliosis, caused by the *Borrelia burgdorferi* (*sensu lato*) complex of bacteria, is the most prevalent arthropod-borne disease of humans in the Northern Hemisphere [2]. Since the early 1990s, the number of reported cases of Lyme borreliosis is rising, and the geographical distribution of cases is expanding [3]. One of the reasons for these shifts is the expansion of the distribution of the main vector of *B. burgdorferi* in Europe, the tick *Ixodes ricinus* [4].

*Ixodes ricinus* is the most abundant and widespread tick species in western Europe. As well as *Borrelia burgdorferi* (*s.l.*) [5], it transmits other pathogens responsible for causing diseases of humans and animals. This species is now found at higher northern latitudes and higher altitudes than previously reported [6, 7] and is more abundant in several areas [8, 9]. The expansion of *I. ricinus* could be partly due to changes in host animal availability such as an increase in deer abundance and distribution [10], since deer densities frequently correlate with tick abundance [11] and perhaps also partly due to abiotic environmental changes, such as climate warming, since higher temperatures can increase interstadial development rate, oviposition rate and egg development rates [12], and the proportion of active ticks [13, 14].

Understanding the drivers of the distribution and abundance of *I. ricinus* is one of the critical steps in assessing the risk of tick-borne diseases and informing policy on awareness and control strategies [15]. Reliable maps of *I. ricinus* distribution are essential to understand and identify changes in the pattern of *I. ricinus* and diseases it transmits [16], and to identify hot-spots of vector occurrence that will inform policy makers in allocating resources to high risk areas, including targeting education and preventive measures [3] or management of important tick population hosts as deer [11].

Several *I. ricinus* distribution models and maps have been published, aiming to predict current and future distribution of *I. ricinus* on different geographic scales, ranging from European to country or local levels [17–20]. The purpose of the study will determine the geographical scale of the map and the resolution will determine the degree of precision, realism and applicability of the models and maps [21]. Therefore, if the objective is to make decisions at country or regional levels, finer resolution maps can detect high variability in tick distribution patterns.

However, predicting *I. ricinus* distribution and abundance is challenging due to the complex ecology of *I. ricinus* (with multiple tick stages and multiple hosts), the limited availability of detailed, long-term and

geographically extensive tick distribution data, and a wide range of environmental variables that may influence tick distribution. Reliable data on *I. ricinus* presence and absence or abundance can be collected during field surveys which use standardized sampling methods, such as the blanket-dragging technique [22]. However, the resources required for field sampling (trained personnel, cost and time required) mean that data are often not available at meaningful spatial and seasonal scales [17]. Other sources of data that were not collected with the purpose of predictive mapping are therefore often used instead. Data submitted by the public can be used to improve the knowledge of *I. ricinus* distribution [23] but usually comprise presence-only data so are subject to biases. An alternative approach, often undertaken by large-scale projects such as VectorNet [22], is to combine available data sources into one composite dataset.

Although Lyme borreliosis is an important public health concern in Scotland [24], published predictive maps of *I. ricinus* distribution in Scotland are limited, particularly at an appropriate scale for national and local decision-making. Although some (as yet unpublished) predictive maps have been made [25, 26], the only peer reviewed publication is a mechanistic model predicting the distribution of infected *I. ricinus* nymphs now and under climate warming [17]. Large-scale presence-absence maps at the European level [27] do not have sufficient resolution for targeting public health resources within Scotland, where *I. ricinus* is endemic.

The main aim of this study was to compare the performance of three datasets to predict *I. ricinus* distribution in Scotland, in order to produce predictive maps for use by decision-makers. We generated model, map and uncertainty outputs of predicted tick abundance and distribution over Scotland from three datasets: (i) quantitative (abundance) *I. ricinus* data from scientific surveys; (ii) *I. ricinus* presence-only data resulted from public submissions plus absence points; and (iii) a composite dataset that combines presence data from public submissions, presence and absence from scientific tick surveys, literature reviews and expert opinion and, absence from a habitat suitability mask for *I. ricinus*. These datasets, which comprise the only data available on tick distributions at a national scale for Scotland, also represent three data types commonly used in mapping tick distributions (i.e. surveyed abundance; surveyed presence and absence; and public submission). We assessed the outputs derived from these different inputs to highlight the strengths and limitations of each data type, and compared the performance of these different types of data in predicting tick distribution, in order to make recommendations for future tick mapping for use in a Public Health context.



## Methods

### Tick data

We used three datasets with information on *I. ricinus* occurrence or abundance in Scotland. As is often the case with predictive mapping exercises, none of these datasets were collected with the main objective of predicting tick distribution at the national level, but they represent the most extensive datasets currently available for mainland Scotland.

#### Dataset 1: “quantitative *I. ricinus* survey data”

Dataset 1 (Fig. 1a) is quantitative tick data, and consists of counts of questing *I. ricinus* ticks (nymphs and adults) in sampled environments in mainland Scotland between 2006 and 2017. Questing ticks were sampled using the standard technique of dragging a white blanket of 1 m<sup>2</sup> across the ground vegetation area of 10 × 10 m, with an average of approximately 15 drags per site [28–30]. During this 11-year period, 687 sites were visited, with varying frequency (1–4 visits), and a total of 10,611 drags were performed.

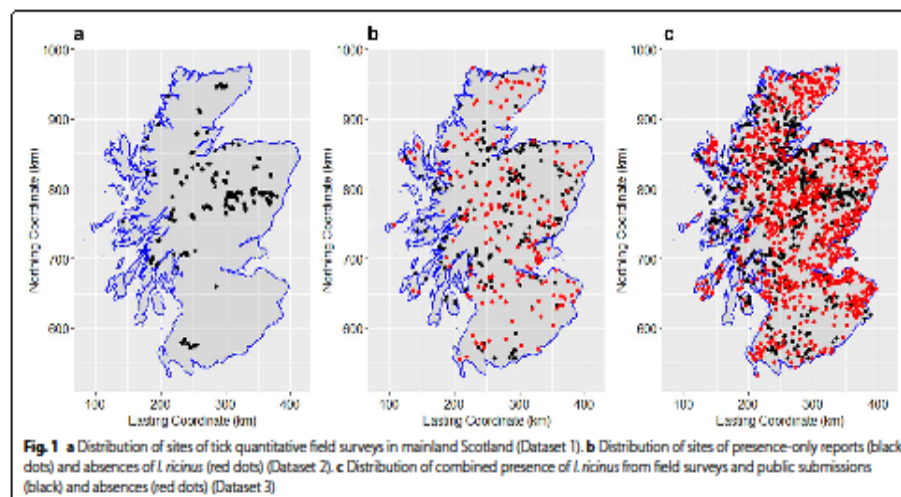
#### Dataset 2: “*I. ricinus* public tick submissions”

Dataset 2 (Fig. 1b) comprised tick submissions by the public to Public Health England (PHE) made through the Tick Surveillance Scheme between 1998 and 2016 in mainland Scotland. The scheme focuses on tick surveillance in England and Wales; ticks are also submitted from Scotland, but in relatively small numbers. The number of ticks (adults and nymphs) submitted per geographical location

were transformed to presence-only data originating 198 data points. Due to inherent limitations of modelling presence-only data, we added a similar number of data points on absence. We therefore randomly selected 200 absence and pseudo-absence points from Dataset 3 to include in Dataset 2. We recognize that other methods could be applied in the selection of the pseudo-absence points as described by [31]. However, we used the sample of absence points for Scotland that was validated by a group of entomologists and public health experts within VectorNet project [32] and offered with Dataset 3 (details below).

#### Dataset 3: “*I. ricinus* combined dataset”

Dataset 3 (Fig. 1c) consists of presence and absence records of *I. ricinus* (adult and nymphs) in Scotland and is part of a large dataset with *I. ricinus* records for all Europe, produced for VectorNet project [22] by a team of tick experts (a network of entomologists and public health professionals supported by the European Centre for Disease Control and Prevention, ECDC). The full methodology is described in an ECDC internal document (manuscript in preparation) [32]. In VectorNet project tick records were assembled from different sources, from public submissions (including presence-only data from Dataset 2 for Scotland) and from scientific tick surveys (including Dataset 1 for Scotland) and then validated. Due to the small amount of absence data in comparison with presence data, absence points were assigned using a mask of suitable habitats for *I. ricinus*.



The habitat suitability mask was defined by the same tick experts from VectorNet project as primary, secondary and unsuitable habitat types (land classes where a species is unlikely to be found except in exceptional circumstances such as continuous and discontinuous urban fabric, industrial or commercial units, port areas, post-flooding or irrigated croplands (or aquatic) or closed to open (> 15%) broadleaved forest regularly flooded) from two land cover maps: CORINE 2006 and GLOBCOVER 2009 [33, 34] and by adding additional information about *I. ricinus* environmental limits (e.g. the fact that *I. ricinus* is only present in areas with less than 150 days of snow cover per year and where the vegetation period is greater than 145 days). Inferred absences were then extracted from unsuitable areas defined by the habitat [27, 32]. The dataset used in this study, after data management, included 1102 presence points and 1058 absence points.

For consistency between datasets, only mainland Scotland was considered. The extraction of points in mainland Scotland and the random selection of 200 absence points were conducted using ArcGIS version 10.2.2 [35].

#### Georeferenced environmental data and variable selection

Ecologically relevant climatic, topographic, land cover and host-related variables for *I. ricinus* occurrence and abundance were selected. These variables were collated as GIS-based raster maps. The variables used were: (i) Moderate Resolution Imaging Spectroradiometer (MODIS) variables (monthly averages, 2001–2013): Normalized Difference Vegetation Index (NDVI, 1 km resolution), land surface temperature (LST, 1 km resolution), cumulative land surface temperature (end of May 2010–2012, MODIS derived data, 0.01 degree resolution (~1.1 km)) and length of vegetation growth period (2008–2014, 0.01 degree resolution); (ii) topographic variables: elevation above sea level (90 m resolution); (iii) long-term average climate data from UK Met Office (from 1981–2010, 5 km resolution): monthly maximum, mean and minimum temperature, number of consecutive dry days (annual average), extreme temperature range (annual average), rainfall (monthly total precipitation), days of air and ground frost (monthly average), mean relative humidity (monthly average); (iv) host-related variables: an index of presence of roe [36] and red deer [37] [both at resolution of 0.008333 degree (~1 km)], and red deer density for 2016 (head per square km, based on a 10 km radius smoothing of the Deer Management Unit density figures [38]); and (v) Land Cover 2006 (0.008333 degree resolution). Monthly derived variables were extracted from each month.

For data extraction compatibility and modelling purposes, all variables were converted to a standardised extent (mainland Scotland), format (tif), resolution (1

km) and projection (British National Grid). Environmental data were extracted for each of the sites of tick collection and reporting [687 sites with counts of *I. ricinus* (Dataset 1); 398 presence-absence points (Dataset 2); and 2160 presence-absence points (Dataset 3)] using the tool extract multiple values to points from ArcGIS version 10.2.2 [35].

Before model implementation, a correlation analysis and a univariate regression analysis were performed with each response variables. If a variable was strongly correlated with another variable (correlation coefficient higher than 0.6), one of them was dropped (variables with correlation coefficient between 0.5 and 0.6 were kept for analysis but under observation for possible interactions). Following with univariate analysis, biologically relevant variables with a *P*-value less than 0.10 were considered as model candidates.

In general, due to issues of autocorrelation and collinearity, satellite-derived covariates were preferred when compared with similar interpolated climatic variables [39].

#### Statistical model, model validation and predictive map

Models were fitted using the Integrated Nested Laplace Approximation (INLA) R package. This Bayesian approach was selected due to its ability to account for irregular sampling intensity, spatial dependency and to quantify uncertainty in data and variables, attributing to each variable a distribution of values [40]. We recognize that other methods could be used but our objective was not to compare different modelling techniques for species distribution models, but instead compare dataset types using the same modelling technique.

The response variables were the count of *I. ricinus* ticks (nymphs and adults) per drag, site of collection and visit and *I. ricinus* (nymphs and adults) presence and absence. A model for predicting tick relative abundance was first created considering just the counts of nymphs per drag, site and visit because nymphs of *I. ricinus* pose the greatest risk of tick bites of humans [17]. However, for consistency with Datasets 2 and 3 which include higher reports of adult ticks, it was decided to model *I. ricinus* relative abundance considering the total count of adult and nymphs per drag, site and visit. This model did not differ significantly from the model using nymphs only.

The fixed effects were the previously selected set of most suitable environmental variables, including the spatial location of the data (as an interaction term between latitude and longitude). A zero-inflated Poisson distribution was chosen to model *I. ricinus* abundance (Model 1) due to a high number of survey drags with 0 counts. Presence and absence of *I. ricinus* (Model 2 and Model 3) was modelled as a binomial distribution.

The selected model for predicting *I. ricinus* relative abundance (Model 1) had two random effects: the effect of the site to capture the unstructured heterogeneity in the distribution of tick abundance among sites, and the effect of each data point (each drag) in order to account for overdispersion not captured by the zero-inflated Poisson and also to account for possible serial correlation in the data arising due to repeated sampling or drags in each site. Tick presence and absence (Model 2 and Model 3) was modelled without random effects because the inclusion of random effects did not improve model fit and predictive power.

The models were evaluated using the Deviance Information Criteria (DIC) as a measure for goodness-of-fit and a parameter from the cross-validation leave-one-out, namely the negative of the sum of the log-conditional predictive ordinates (log-CPO score) as a measure for the predictive quality of the model [41]. A backward stepwise procedure was used to select the most parsimonious model. For all three datasets, the most suitable models were selected based on the lowest values of DIC and log-CPO, amongst competing models with various covariate combinations.

The model posterior means were used to produce the predictive maps of *I. ricinus* abundance (Model 1/Dataset 1) and presence-absence (Model 2/Dataset 2 and Model 3/Dataset 3). The difference between the 97.5% and 2.5% quartiles of the predicted values were used to create uncertainty maps. The resolution of all maps was 1 km<sup>2</sup> each pixel.

A matrix of boxplots, comparing the interquartile range of the models' covariates over mainland Scotland with the interquartile range of the same covariates covered by the data points in each model, was developed.

Descriptive analyses, plots, models and maps were made using R software version 3.4.4 [42].

**Results**

Figure 1 presents the spatial distribution of the three datasets of *I. ricinus* counts per drag (Dataset 1, Fig. 1a) and presence and absence (Datasets 2 and 3, respectively, Fig. 1b, c). Figure 1a shows an uneven distribution of tick collection sites over mainland Scotland, with aggregation of collection sites in the east, particularly Aberdeenshire and in opposite, lack of sampling points in the west coast. The distribution of data points in Dataset 2 (Fig. 1b) is sparse compared to Dataset 3 (Fig. 1c).

**Model 1 (tick relative abundance, using Dataset 1: quantitative survey data)**

A spatial model of the count of ticks (adult and nymphs) per drag, visit and site was run initially (DIC of the most suitable spatial model is 29786.66, log-CPO is 20427.23). Subsequently, month was added in the model as a categorical variable, improving model predictive power (DIC 29774.49; log-CPO 19686.78). The model fitted the data well (goodness-of-fit plot in Additional file 1: Figure S1). The results of Model 1 are presented in Table 1. A map for the month with highest predicted tick abundance (April) was created (Fig. 2a).

Land surface temperature in July, presence of roe deer and deciduous and coniferous forest were associated with an increase in *I. ricinus* questing tick abundance, whilst a higher number of frost days in September lead to a decrease in tick abundance. April had the highest questing tick abundance (Table 1). The interaction term between latitude and longitude was included in the model because, although not significant, it decreased

**Table 1** Model 1: posterior mean, standard deviation, 2.5% and 97.5% quartiles and estimates of fixed and random effects for the seasonal model of tick abundance, Dataset 1

Fixed effects	Mean	SD	2.5% quartile	97.5% quartile
Intercept	-150.7016	32.7201	-214.9447	-86.5187
April	2.3606	0.3390	1.7198	3.0520
May	1.9424	0.3174	1.3467	2.5944
June	1.8192	0.3178	1.2227	2.4718
July	1.2388	0.3149	0.6485	1.8863
August	1.3438	0.3151	0.7530	1.9916
September	1.4308	0.3186	0.8325	2.0850
Land surface temperature in July	0.0103	0.0022	0.0059	0.0147
No. days of frost in September	-0.4035	0.0954	-0.5910	-0.2167
Roe deer	0.0096	0.0034	0.0030	0.0163
% cover of deciduous woodland	2.5341	0.7380	1.0837	3.9806
% cover of coniferous woodland	0.9053	0.2138	0.4848	1.3240
Interaction between latitude and longitude	0.0010	0.0018	-0.0026	0.0045

Abbreviation: SD, standard deviation

model residual variance and can help take into account spatial effects in questing tick abundance, such as spatial autocorrelation (Table 1).

The predictive map of tick abundance (Fig. 2a) shows that tick abundance increases from the north and west to the south and east of Scotland with highest predicted tick abundance in Aberdeenshire and the central belt. Areas of average to high uncertainty are present over all of the east, centre and south of Scotland whereas all the west shows a lower level of uncertainty in predicted values (Fig. 2b).

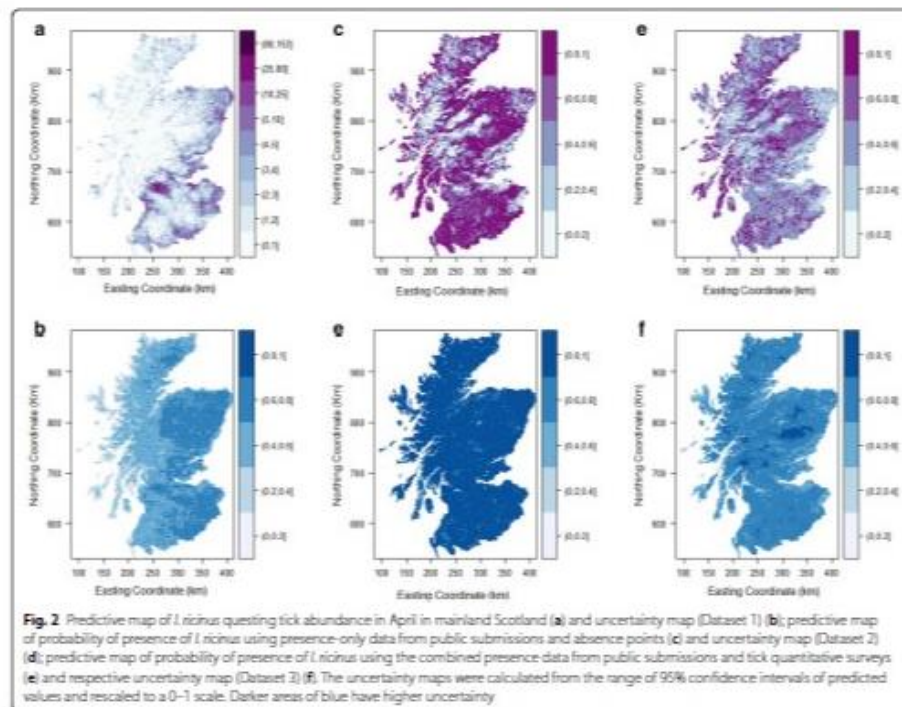
**Model 2 (tick presence-absence, using Dataset 2: tick public submissions)**

Model 2 (DIC of 388.61, log-CPO of 195.81) fitted the data well (plot of model goodness-of-fit in Additional file 2: Figure S2). Presence of *I. ricinus* was correlated with an increase in NDVI and some measure of habitat composition. An increased number of days with air frost in November and increased precipitation in April were

associated with tick absence. The site location of tick submission was important (Table 2). The predicted map for Model 2 (Fig. 2c) does not capture well areas of lower probability of tick presence and shows very high levels of uncertainty for most of Scotland (Fig. 2d).

**Model 3 (tick presence-absence, using Dataset 3: combined dataset)**

The adopted model (Model 3) gave the lowest values of DIC of 2614.61 and a log-CPO of 1307.74 (plot of model goodness-of-fit in Additional file 3: Figure S3). Model 3 presented very similar covariates as Model 2 but deciduous forest and deer density became significant predictors, likely due to the increased number of points used to model tick presence-absence (Table 2). Figure 2e shows a similar pattern of *I. ricinus* probability of presence as Fig. 2c, but the predictive map using Model 3 has more detailed definition. The uncertainty is lower for the east of Scotland and in the north and centre of the Highlands (Fig. 2f).



**Table 2** Posterior mean, standard deviation, 2.5% and 97.5% quartiles for the binomial models of tick presence-absence with the data from public submissions (Dataset 2) and the combined dataset (Dataset 3)

Model	Fixed effects	Mean	SD	2.5% quartile	97.5% quartile
Model 2: Presence-absence model with presence points from public submissions plus absence points	Intercept	-6.2657	1.0232	-8.3326	-4.3135
	NDVI August <sup>a</sup>	0.1373	0.0176	0.1040	0.1732
	No. days of air frost November	-0.1729	0.0521	0.2784	-0.0738
	Rain April	-0.0148	0.0053	-0.0255	-0.0045
	% cover of coniferous woodland	5.1989	1.2015	3.0921	7.8095
	% cover of moorland	2.2180	0.5656	1.1499	3.3725
	Interaction between latitude and longitude	0.0053	0.0036	-0.0017	0.0123
Model 3: Presence-absence model with composite dataset	Intercept	-3.4700	0.4771	-4.4160	-2.5424
	NDVI August	0.0005	0.0001	0.0004	0.0006
	Deer density	0.0336	0.0100	0.0139	0.0533
	No. days of air frost November	-0.0527	0.0207	-0.0936	-0.0122
	Rain April	-0.0123	0.0020	-0.0163	-0.0085
	% cover of moorland	1.3920	0.1640	1.0726	1.7161
	% cover of deciduous woodland	3.1762	0.6757	1.9203	4.5770
	% cover of coniferous woodland	2.1861	0.2128	1.7753	2.6100
	Interaction between latitude and longitude	-0.0029	0.0013	-0.0054	-0.0004

<sup>a</sup> The posterior mean of NDVI was divided by 100

Abbreviation: SD, standard deviation

We assessed how well the three datasets cover the range of the covariates used in the models to explore the validity of the predictions. The interquartile range of each covariate in mainland Scotland was compared to the interquartile range of each covariate in the models for the data points included (Fig. 3). Although the models fitted the data well, the predictions of the three models were associated with uncertainty that was not captured in the uncertainty measures in Fig. 2, because the tick data did not cover all the range of the covariates used. Dataset 3 covered the covariate range used in the predictions better than Dataset 1 or 2. Dataset 1 was mainly collected in predominantly forest areas. The covariate index of presence of roe deer was found important for *I. ricinus* presence in Dataset 2 and it was included in the first selected Model 2 (see Additional file 4: Table S1, first model). However as can be seen in Fig. 3, the range of the covariate index of roe deer presence is not well covered by Dataset 2, contributing to higher uncertainty in the predictions (see Additional file 5: Figure S4a, b). The covariate roe deer was therefore removed from the final model (Table 2). Using all the covariates of Model 3 for fitting a model with Dataset 2 helped to corroborate how the covariates (type and range) are important in predictive mapping and can be a source of error for model predictions (also shown in Additional file 4: Table S1 and Additional file 5: Figure S4c, d).

## Discussion

Predictive maps of tick distribution are essential for understanding human disease risk and allocating resources for prevention and control. However, they require extensive data on tick distribution, and robust long-term and geographically extensive datasets are often difficult to obtain. Therefore datasets are often used that were not collected for this purpose, or do not cover the entire extent of the required predicted distribution. The main aim of this study was to compare the performance of three different types of data to predict *I. ricinus* distribution in Scotland. We critically assessed modelled distributions to conclude how useful each dataset may be to inform policy, future data collection and risk mapping, both for Scotland and more widely.

### Predictors for *I. ricinus* relative abundance and presence-absence

The predictors identified in the models are consistent with the current knowledge on *I. ricinus* ecology. *Ixodes ricinus* activity is seasonal, and in Scotland peaks in April and then generally declines over the season as ticks die or find a host (Model 1). It is not surprising that roe deer presence and red deer density are correlated with both tick abundance and presence as deer are important hosts maintaining *I. ricinus* populations [11]. Many other studies have found deer abundance as a predictor of tick presence and/or abundance [11, 29, 43]. Temperature affects tick behaviour, interstadial development rate,



fertility, survival and the proportion of active ticks [12–14]. Warmer climates are frequently associated with tick presence and abundance (e.g. [28]). Land surface temperature was also used in other studies to estimate *I. ricinus* presence–absence and abundance [39, 44]. In accordance, in this study we found that areas with warmer climates and lower number of frost days (minimum temperature below 0 °C) were associated with higher tick abundance, consistent with previous studies [25, 45].

Forested areas, particularly mixed and deciduous forest, as well as other habitats that provide a sheltered canopy over the ground, such as bracken and deep heather in moorland that are characterized by thick ground vegetation or shrub and deep leaf litter layers, provide moist and mild microclimates which aid tick survival and activity. Forests are also a source of food and shelter for many species of tick hosts [46]. NDVI, which quantifies the level of photosynthetic activity of the vegetation, has been previously identified as being an important physical parameter that correlates with *I. ricinus* presence and abundance [39]. More questing *I. ricinus* are predicted in areas with warmer climate and higher rainfall or higher humidity [47]. However, the negative effect of annual precipitation that we found in this study, was also found by James et al. [29] in Scotland and Schulz et al. [48] in Germany. These findings suggest that the wettest conditions in Scotland are probably too wet for *I. ricinus* to quest, while drier areas of Scotland are still wet enough for good survival rates.

For all three datasets plausible predictors were identified and predictive maps were created. Although the models fitted the data well, the three predictive maps do not present a consistent pattern of *I. ricinus* distribution and the predictions are associated with a large amount of uncertainty, particularly for Model 2 and to a lesser degree for Model 1. The uncertainty presented in Fig. 2 (b, d, f; difference between 2.5% and 97.5% values), provides a measure of the predicted uncertainty at each pixel. However, the predicted uncertainty is correlated with the model posterior mean for each covariate. In addition, this measure does not include uncertainty associated with extrapolation outside the covariate range. We therefore explored factors that could influence the validity of the predictions for each dataset, and identified when each dataset might be most appropriately used.

#### Use of quantitative *I. ricinus* survey data (Dataset 1)

The relative abundance of questing *I. ricinus* ticks is generally measured by dragging a blanket over the vegetation. This technique does not measure the absolute density of the whole tick population in an area because it does not count ticks that are moulting, resting, feeding or in diapause [44]. It is also worth noting that the blanket

drag method's efficiency is affected by ground vegetation height and density [43]. Ground vegetation height and density should be included in statistical models and repeated sampling of a site is recommended due to the impact of weather conditions on tick activity on the day of sampling. Another limitation from this technique is that tick absence data cannot be considered completely free of error because some of the zeros could in reality indicate very low tick densities rather than true absolute absence, due to the finite number of blanket drag transects per site and conditions on the day of surveying [44]. However, this is a standard scientific technique which provides a useful comparable index of abundance of questing nymphs between sites [22].

Conducting blanket drags is time-consuming which makes it resource-intensive to perform large-scale, long-term field studies using this technique. However, estimating questing *I. ricinus* relative abundance gives more information about this species distribution when compared with presence-only and presence–absence data. Abundance data are necessary to calculate the density of infected ticks, which is important in estimating disease risk [44], as well as providing more information on *I. ricinus* dynamics. Abundance data also improve model accuracy, predictive performance and ability to discriminate trends at finer scales, compared to presence–absence data. This improvement is particularly important for species of high abundance compared to "rare" species [49]. When the objective is to create a predictive map for a country such as Scotland where *I. ricinus* is endemic, abundance models will provide more meaningful distribution maps.

The predictive map of questing *I. ricinus* relative abundance enables us to clearly identify areas with high and low tick abundance, and shows an increasing trend of tick abundance from the west to the east coast of Scotland. However, although the predictions of relative abundance of questing ticks had lower uncertainty compared to both of our predictive maps of *I. ricinus* probability of presence, it is clear that the sample sites are clustered and do not cover all of mainland Scotland, and that the covariate range covered by mainland Scotland is not fully represented in the data. Since there is an ecological gradient from the west (higher temperatures, higher rainfall) to the east, it is concerning that there may be insufficient data for accurate predictions in the west. Dataset 1 was collected mainly (although not exclusively) in forest areas, reflecting the data collection, which was aimed at specific ecological studies [29, 30]. Hence the dataset does not have good coverage for some areas of mainland Scotland and therefore the reliability of the predictions is likely to be lower outside the core survey areas. These issues of low coverage are common to these type of data,

due to the resources needed to collect quantitative survey data over a wide scale. Quantitative survey data are often considered the "gold standard", but this is only true for models that make predictions in the same geographical area and covariate range from which the surveys were conducted.

#### Use of *I. ricinus* presence data obtained through public submissions plus absence points (Dataset 2)

Presence-only data, generated from submissions by the public and often obtained from citizen science studies, are frequently used to map species distribution. These data usually require fewer resources to collect than scientific surveys [50] but include random error associated with uncertainty in the location of ticks were collected, variability in sampling (e.g. variability between different people reporting) and in effort (e.g. some people contribute more data than others and effort can also change over time) [50]. Bias is also associated with the fact that people report from places that are visited frequently or are more accessible [51]. This type of data lacks information on where the species is absent, which limits the predictive power of the inference and also restricts the type of questions that can be asked [52]. In this study, information of where the tick was absent (true absences from the scientific surveys from Dataset 1 and pseudo-absence points from habitat unsuitability mask from Dataset 3) was added to the presence-only records from submissions to improve the predictive power of the model. This process is not free of error since *I. ricinus* is not confirmed to be absent at all the points used as absence [52]. A general disadvantage of this type of distribution data is that all presences are treated as equal, regardless of the abundance of *I. ricinus* ticks that the habitat supports, which may not provide enough information to enable the model to differentiate a scarce habitat classified as having the species present from a habitat where the species is in fact established [49].

The predictive map resulting from Model 2 presents lower spatial detail compared to the other two maps, indicates high probability of *I. ricinus* presence over much of mainland Scotland and does not reflect the known vector habitat preferences, as presence is predicted in some unsuitable areas. This does not provide particularly useful information for targeting public health interventions and illustrates the challenges of using sparse presence-absence data in areas where ticks are endemic. The predictive map presents high uncertainty demonstrating low confidence in the predictions, likely due to the small sample size. In addition, there is uncertainty relating to the predictions as Dataset 2 does not cover all the covariate range (Fig. 3).

Although the potential biases of submission data are common to similar studies, they can often be minimised

if sufficient sample sizes are obtained. The dataset used in this survey was not collected for the purpose of mapping tick distribution, so the sample size was low (~200). In England and Wales where the submission scheme has been promoted, over 4000 data points were collected for the same period, giving more capacity for predictive mapping.

These results should not rule out the use of data from public submissions that can be used to infer range limits of *I. ricinus* after careful analysis to account for adventitious ticks dispersed by hosts [32].

#### Use of *I. ricinus* combined datasets (Dataset 3)

The predictive model based on a dataset that combined data from scientific studies, public submissions data and absences of *I. ricinus* increased the spatial coverage of the data in mainland Scotland (Fig. 1c) and produced a more detailed predictive map. In addition this dataset had the best coverage for the covariate range used in predictions (Fig. 3). It is not surprising that the spatial trend of the predictions from both presence-absence models were comparable. However, the model developed using the combined data (Model 3) provides a better description of the presence and absence of *I. ricinus* not only because of the higher number of points but also because it includes presence and absence data from quantitative tick surveys (Dataset 1). This method of adding information from different datasets can be more easily applied at country and continental levels to obtain distribution maps. However, because composite datasets combine different types of data, it is more challenging to understand how the different errors, bias and limitations of each dataset might affect the model outputs and the predicted uncertainty.

#### Predictive *I. ricinus* maps for Scotland

The three datasets used in this study are the only *I. ricinus* datasets that are available at a national scale in Scotland (as far as the authors are aware). As discussed above, although quantitative survey data are usually regarded as a gold standard, the data used in this study did not have good coverage, both geographically and over the covariate range, for the whole of Scotland. This dataset is appropriate for making decisions that require detailed distribution data only in areas where the coverage is good. Outputs from Dataset 2, comprising public submission data, were limited by the small sample sizes in this dataset, which gave high model uncertainty. Therefore, Dataset 3, which uses data from multiple sources, provides the most convincing predictive map and is recommended for decision-making at national scale.

It is conceivable that any of these maps could be used alone for decision-making, without further consideration of the limitations of the data inputs. The differences



between the three maps highlight the importance of exploring sources of uncertainty in models and in predictions and presenting this alongside predictive maps. Although there are a high number of published papers on *I. ricinus* predicted distribution, uncertainty is rarely presented (a rare example is [18]). For other vectors, when uncertainty is considered, the uncertainty metric used in this study is commonly reported (such as [53]), but its value is limited because the uncertainty values correlate with the posterior mean. In addition, this measure does not include uncertainty associated with extrapolation outside the covariate range. Further development of methodological approaches to quantify this uncertainty, such as statistical tools for the diagnosis of model prediction reliability or to limit predictions to the range and covariates encountered during surveys would be beneficial [54].

Although we conducted this exercise with the aim of improving tick distribution and Lyme borreliosis management, the findings are relevant to other vector-borne disease systems for animal and human health.

### Conclusions

The choice of the most suitable model and map of *I. ricinus* distribution in Scotland depends on the objective. For local-level decision-making, Model 1 and map 1 (using quantitative *I. ricinus* survey data, Dataset 1) are more appropriate, with a good coverage for the east coast of Scotland. For decision-making at national level, Model 3 and map 3 (using combined Dataset 3) provide a better coverage of the country and the range of the covariates. Although tick surveys provide detailed data on questing tick relative abundance, the resources required often limit the number of areas that can be sampled, which makes it challenging to make predictions for extensive areas. If available at larger spatial and temporal resolution, relative abundance data will result in finer scale maps that are more effective for risk management and communication at national and regional levels. The analysis in this study highlights the need for additional surveying in areas with poor previous coverage. Future maps of *I. ricinus* abundance could be improved by adding *a priori* information of habitat preferences into the model structure [54]. For large-scale mapping at lower resolution, or if there are few tick data from quantitative surveys, data on *I. ricinus* presence-only should be combined with data from field surveys and absence data for modelling presence-absence. To overcome the problems inherent in the use of presence-only data from public submissions, it is necessary to decrease associated errors and bias by accounting for observer effort and expertise [55] or to find approaches by which absence data are

also reported [44]. When predictive maps are needed for public health decision making, such as allocation of resources for awareness campaigns, information on uncertainty should be included with vector distribution maps. However, because map uncertainty reflects a single source of uncertainty (the spatial model), improved statistical techniques are required to quantify uncertainties relating to predictions.

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13071-019-3784-1>.

**Additional file 1: Figure S1.** Plot of fitted (blue line) versus observed values (dots) for the seasonal model of nymph and adult abundance, Dataset 1. The observed number of nymphs plus adults per drag has a minimum of 0, a mean of 2.34 and a maximum of 109. The predicted number of nymphs plus adults per drag has a minimum of 0, a mean of 2.5 and a maximum of 106.

**Additional file 2: Figure S2.** Goodness of model fit, model 2, Dataset 2. The plot is presented as a histogram and curve for binomial regression.

**Additional file 3: Figure S3.** Goodness of model fit, model 3, Dataset 3. The plot is presented as a histogram and curve for binomial regression.

**Additional file 4: Table S1.** Posterior mean, standard deviation, 2.5% and 97.5% quantiles for the binomial models of tick presence-absence with the data from public submissions (Dataset 2), first model selected based on the Bayesian criteria; secondly using covariates from model 3 to predicted Dataset 2 distribution. None of these two models were selected for the predictions of Dataset 2.

**Additional file 5: Figure S4.** Predictive maps of binomial models of tick presence-absence with the data from public submissions (Dataset 2): predictive map from first model selected based on the Bayesian criteria (a) and uncertainty map (b); predictive map using covariates from Model 3 to predicted Dataset 2 distribution (c) and respective uncertainty map (d).

### Abbreviations

PHE: Public Health England; ECDC: European Centre for Disease Control and Prevention; GIS: geographical information system; MODIS: moderate resolution imaging spectroradiometer; NDVI: normalized difference vegetation index; LST: land surface temperature; INLA: integrated nested Laplace approximation; DIC: deviance information criteria; CPO: conditional predictive ordinate.

### Acknowledgements

We would like to thank Jim McLeod, working at The James Hutton Institute, for developing the red deer density layer.

### Authors' contributions

RR, JE and HA conceived the ideas and designed methodology. Data were provided by LG, WW and JM. RR and JE analysed the data. All authors have written the manuscript. All authors read and approved the final manuscript.

### Funding

RR was supported by Scotland's Rural College (SRUC) Internal studentship. LG was supported by the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS) and as part of the Scottish Government's Centre of Expertise on Animal Disease Outbreaks (EPIC).

### Availability of data and materials

Data supporting the conclusions of this article are included within the article and its additional files. The code is available via Zenodo: <https://doi.org/10.5281/zenodo.3476288>. Dataset 1 is available on request; please contact Professor Lucy Gilbert (Lucy.Gilbert@glasgow.ac.uk). Dataset 2 is freely available. Data from Public Health England can be downloaded at <https://rtrnrat>

las.org/ VectorNET data (Dataset 3) is available on request from ECDC. The request needs to include the purpose of the data request. More information available at <https://ecdc.europa.eu/en/publications-data/european-surveillance-system-issy>. Any question, please contact: [data.access@ecdc.europa.eu](mailto:data.access@ecdc.europa.eu).

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Veterinary and Animal Science, Northern Faculty Scotland's Rural College, An Lochran, 10 Inverness Campus, Inverness IV2 5NA, UK. <sup>2</sup>The Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. <sup>3</sup>Biostatistics and Statistics Scotland, ICMB, The King's Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. <sup>4</sup>Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK. <sup>5</sup>Environmental Research Group Oxford, c/o Department of Zoology, South Parks Road, Oxford OX1 3PS, UK. <sup>6</sup>Medical Entomology & Zoonoses Ecology, Emergency Response Department-Science & Technology, Public Health England (PHE), Porton Down, Salisbury SP4 0JG, UK.

Received: 17 July 2019 Accepted: 4 November 2019

Published online: 11 November 2019

#### References

- Randolph S. Tick-borne disease systems emerge from the shadows: the beauty lies in molecular detail, the message in epidemiology. *Parasitology*. 2009;136:1403–13.
- Stone B, Tourand Y, Brisette C. Brave new worlds: the expanding universe of Lyme disease. *Vector Borne Zoonotic Dis*. 2017;17:619–29.
- Lindgren E, Jaenson TGT. Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures. Geneva: World Health Organization; 2006. [http://www.euro.who.int/\\_data/assets/pdf\\_file/0006/96819/889522.pdf](http://www.euro.who.int/_data/assets/pdf_file/0006/96819/889522.pdf). Accessed 13 Apr 2018.
- Gray JS, Dautel H, Estrada-Peña A, Kahl O, Lindgren E. Effects of climate change on ticks and tick-borne diseases in Europe. *Interdiscip Perspect Infect Dis*. 2009;2009:593232.
- Burgdorfer W, Barbour A, Hayes S, Peter C, Aeschlimann A. Erythema chronicum migrans - a tickborne spirochetosis. *Acta Trop*. 1983;40:79–83.
- Jone S, Viljgrain H, Hofshagen M, Brun-Hansen H, Kristoffersen AB, Nygård K, et al. Multi-source analysis reveals latitudinal and altitudinal shifts in range of *Ixodes ricinus* at its northern distribution limit. *Parasit Vectors*. 2011;4:84.
- Tälkklint L, Jaenson TGT. Increasing geographical distribution and density of *Ixodes ricinus* (Acari: Ixodidae) in central and northern Sweden. *J Med Entomol*. 1998;35:521–6.
- Scharlemann J, Johnson P, Smith A, Macdonald D, Randolph S. Trends in Ixodid tick abundance and distribution in Great Britain. *Med Vet Entomol*. 2008;22:238–47.
- Schwarz A, Hönig V, Vavrusková Z, Grubhoffer L, Balczun C, Albring A, et al. Abundance of *Ixodes ricinus* and prevalence of *Borrelia burgdorferi* s.l. in the nature reserve Siebengebiet, Germany, in comparison to three former studies from 1978 onwards. *Parasit Vectors*. 1978;2012(5):268.
- Clutton-Brock T, Coulson T, Milner J. Red deer stocks in the Highlands of Scotland. *Nature*. 2004;429:11–2.
- Gilbert I, Maffey GL, Ramsay SL, Hester AJ. The effect of deer management on the abundance of *Ixodes ricinus* in Scotland. *Ecol Appl*. 2012;22:658–67.
- Randolph S, Green R, Hoodless A, Piacoy M. An empirical quantitative framework for the seasonal dynamics of *Ixodes ricinus*. *Int J Parasitol*. 2002;32:979–89.
- Gilbert I, Aungler J, Tomkins J. Climate of origin affects tick (*Ixodes ricinus*) host-seeking behavior in response to temperature: implications for resilience to climate change? *Ecol Evol*. 2014;4:1186–98.
- Tomkins J, Aungler J, Hazel W, Gilbert I. Towards an evolutionary understanding of questing behaviour in the tick *Ixodes ricinus*. *PLoS ONE*. 2014;9:10.
- Estrada-Peña A. Distribution, abundance, and habitat preferences of *Ixodes ricinus* (Acari: Ixodidae) in northern Spain. *J Med Entomol*. 2001;38:361–70.
- Madlock JM, Hansford KM, Borriano A, Dardakova M, Estrada-peña A, George J, et al. Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. *Parasit Vectors*. 2013;6:1.
- Li S, Gilbert I, Harrison PA, Rousewell MDA. Modelling the seasonality of Lyme disease risk and the potential impacts of a warming climate within the heterogeneous landscapes of Scotland. *J R Soc Interface*. 2016;13:20160140.
- Alkshie AA, Peterson AT, Sarmy AM. Climate change influences on the potential geographic distribution of the disease vector tick *Ixodes ricinus*. *PLoS ONE*. 2017;12:12.
- Baughniet F, Chalvet-Monfray K, Loukos H. FleaTickRisk: a meteorological model developed to monitor and predict the activity and density of three tick species and the cat flea in Europe. *Geospat Health*. 2009;4:97–113.
- Estrada-Peña A, Venzal J. Changes in habitat suitability for the tick *Ixodes ricinus* (Acari: Ixodidae) in Europe (1900–1999). *Ecohealth*. 2006;3:154–62.
- Kitron U. Risk Maps: transmission and burden of vector-borne diseases. *Parasitol Today*. 2000;16:2.
- European Centre for Disease Prevention and Control (ECDC). The importance of vector abundance and seasonality. Results from an expert consultation; 2018. <https://doi.org/10.2900/37171>.
- Jameson LJ, Madlock JM. Tick surveillance in Great Britain. *Vector Borne Zoonotic Dis*. 2011;11:403–12.
- Mavin S, Watson EJ, Evans R. Distribution and presentation of Lyme borreliosis in Scotland—analysis of data from a national testing laboratory. *J R Coll Physicians Edinb*. 2015;45:196–200.
- Braga JE. Predicting current and future tick abundance across Scotland. M.Sc. thesis, University of Aberdeen; 2012.
- Worton A. Using mathematical models to understand the impact of climate change on tick-borne infections across Scotland. Ph.D. thesis, University of Stirling; 2016. <https://dspace.stir.ac.uk/bitstream/1893/24918/1/thesis.pdf>. Accessed 10 Jan 2019.
- Braks M, Mulder A, Swart A, Wirt W. Grasping risk mapping. In: Braks M, van Wieren S, Takken W, Sprong H, editors. Ecology and prevention of Lyme borreliosis. Wageningen: Wageningen Academic Publishers; 2016. p. 351–71.
- Gilbert I. Altitudinal patterns of tick and host abundance: a potential role for climate change in regulating tick-borne diseases? *Oecologia*. 2010;162:217–25.
- James M, Bowman A, Forbes K, Lewis F, McLeod J, Gilbert I. Environmental determinants of *Ixodes ricinus* ticks and the incidence of *Borrelia burgdorferi sensu lato*, the agent of Lyme borreliosis, in Scotland. *Parasitology*. 2012;140:237–46.
- Millins C, Gilbert I, Johnson P, James M, Kilbride E, Birtles R, et al. Heterogeneity in the abundance and distribution of *Ixodes ricinus* and *Borrelia burgdorferi sensu lato* in Scotland: implications for risk prediction. *Parasit Vectors*. 2016;9:595.
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol*. 2012;3:327–38.
- European Centre for Disease Prevention and Control (ECDC). A spatial modelling approach to guide surveillance for vector distribution. Technical Report. Stockholm.
- CORINE Land Use Dataset. 2006. <https://land.copernicus.eu/pan-europ/ear/corine-land-cover>.
- Olivier A, Ramos Perez, Julio J, Kalogrou V, Bontemps S, Defourny P, Van Bogaert E. Global Land Cover Map for 2009 (GlobCover 2009); 2012. <https://doi.org/10.1594/PANGAEA.787668>.
- Environmental Systems Research Institute (ESRI). ArcGIS Release 10.2.2. Redlands; 2012.

36. Alexander NS, Morley D, Medlock J, Searle K, Wint W. A first attempt at modelling roe deer (*Capreolus capreolus*) distributions over Europe. *Open Health Data*. 2014;2:e2.
37. Wint W, Morley D, Medlock J, Alexander NS. A first attempt at modelling roe deer (*Capreolus capreolus*) distributions over Europe. *Open Health Data*. 2014;2:e2.
38. Albon S, McLeod J, Potts J, Brewer M, Irvine J, Towers M, et al. Estimating national trends and regional differences in roe deer density on open-hill ground in Scotland: identifying the causes of change and consequences for upland habitats. *SNH Commun Rep*. 2017;981:1–29.
39. Estrada-Peña A, Estrada-Sánchez A, Estrada-Sánchez D. Methodological caveats in the environmental modelling and projections of climate niche for ticks, with examples for *Ixodes ricinus* (ixodidae). *Vet Parasitol*. 2015;208:14–25.
40. Rue H, Martino S, Chopin N. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J R Stat Soc Ser B Stat Methodol*. 2009;71:319–92.
41. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. 2014;24:997–1016.
42. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
43. Ruiz-Fons F, Gilbert L. The role of deer as vehicles to move ticks, *Ixodes ricinus*, between contrasting habitats. *Int J Parasitol*. 2010;40:1013–20.
44. Swart A, Ibañez-Justicia A, Bujs J, van Wieren S, Hofmeester T, Sprong H, et al. Predicting tick presence by environmental risk mapping. *Front Public Health*. 2014;2:238.
45. Paul REL, Cote M, Le Naour E, Bonnet SI. Environmental factors influencing tick densities over seven years in a French suburban forest. *Parasit Vectors*. 2016;9:309.
46. Gray JS, Kahl O, Robertson JN, Daniel M, Estrada-Peña A, Gettinby G, et al. Lyme borreliosis habitat assessment. *Zentralblatt für Bakteriol*. 1998;287:211–28.
47. Ruiz-Fons F, Fernández-de-Mera L, Acavado P, Gortázar C, de la Fuente J. Factors driving the abundance of *Ixodes ricinus* ticks and the prevalence of zoonotic *I. ricinus*-borne pathogens in natural foci. *Appl Environ Microbiol*. 2012;78:2669–76.
48. Schulz M, Mahling M, Pfister K. Abundance and seasonal activity of questing *Ixodes ricinus* ticks in their natural habitats in southern Germany in 2011. *J Vector Ecol*. 2014;39:56–65.
49. Howard C, Stephens PA, Pearce-Higgins JW, Gregory RD, Willis SG. Improving species distribution models: the value of data on abundance. *Methods Ecol Evol*. 2014;5:506–13.
50. Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, et al. Statistical solutions for error and bias in global citizen science datasets. *Biol Conserv*. 2014;173:144–54.
51. Boakes EH, McGowan PIK, Fuller RA, Chang-Qing D, Clark NE, O'Connor K, et al. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol*. 2010;8:e6.
52. Pearce JL, Boyce MS. Modelling distribution and abundance with presence-only data. *J Appl Ecol*. 2006;43:405–12.
53. Lord JS, Torr SJ, Auty HK, Brock PM, Byamungu M, Hargrove JW, et al. Geostatistical models using remotely-sensed data predict savanna tree decline across the interface between protected and unprotected areas in Serengeti, Tanzania. *J Appl Ecol*. 2018;55:1997–2007.
54. Corns PJ, Johnson DS, Bowers PL. On extrapolating past the range of observed data when making statistical predictions in ecology. *PLoS ONE*. 2015;10:10.
55. Palmer JR, Oltra A, Collantes F, Delgado JA, Lucientes J, Delacour S, et al. Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes. *Nat Commun*. 2017;8:916.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

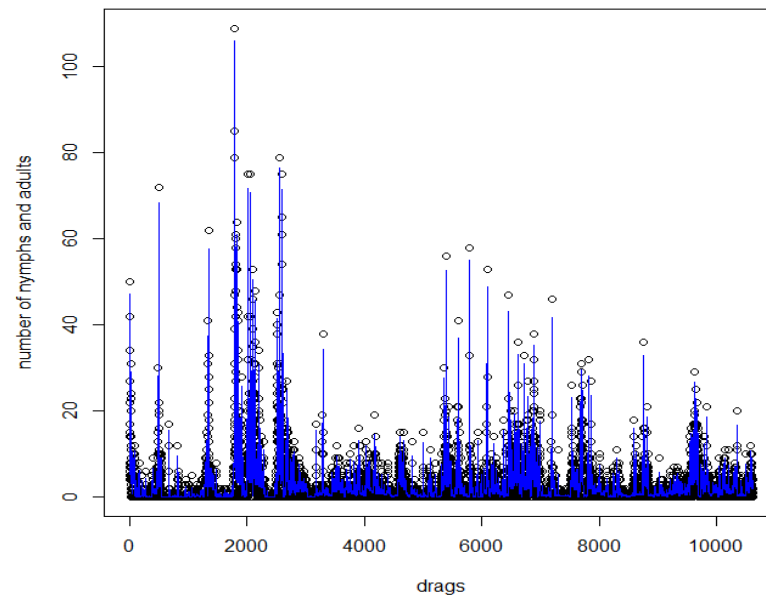
- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

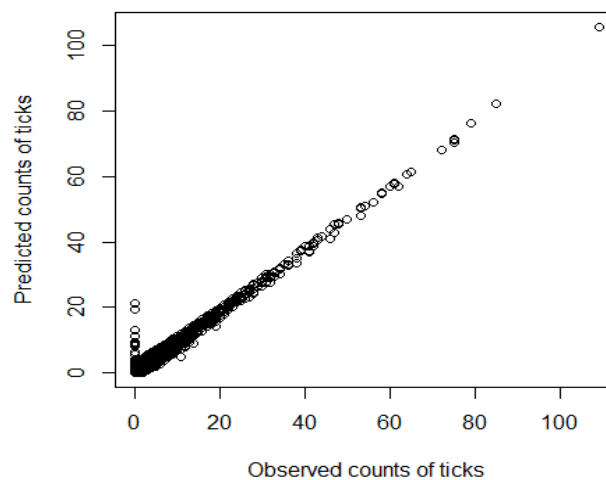


**S2: Goodness of model fit – Models 1 (Dataset 1), 2 (Dataset 2) and 3 (Dataset 3)**

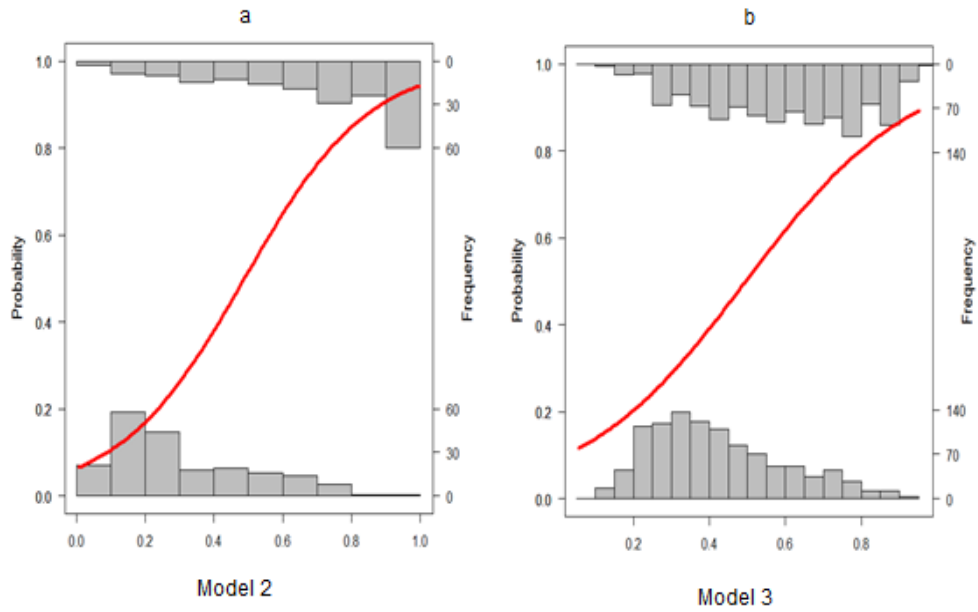


**Figure B-S2.1:** Plot of fitted (blue line) versus observed values (dots) for the seasonal model of nymph and adult abundance, Model 1, Dataset 1.

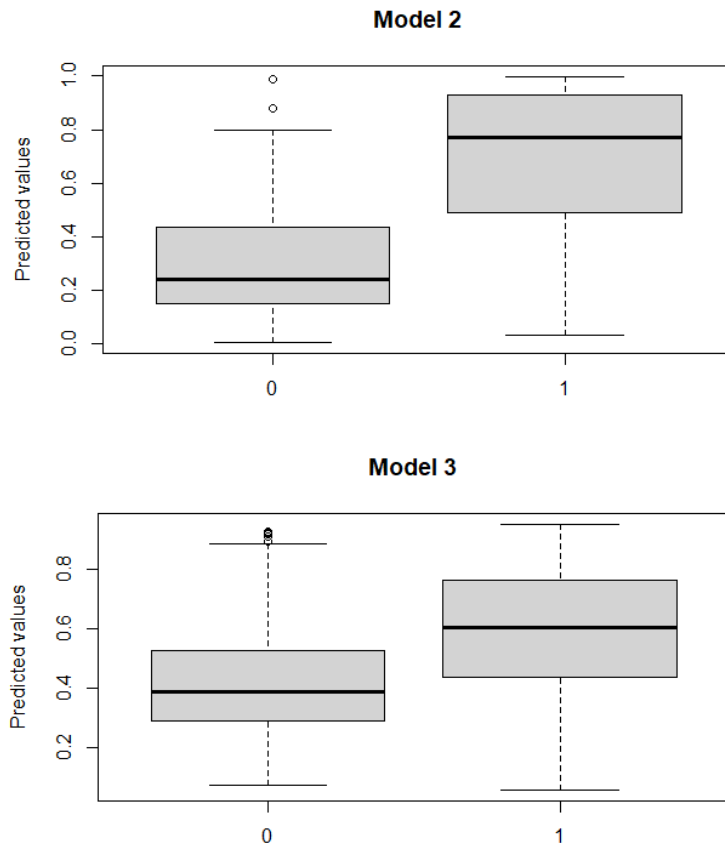
The observed number of nymphs plus adults per drag has a minimum of 0, a mean of 2.34 and a maximum of 109. The predicted number of nymphs plus adults per drag has a minimum of 0, a mean of 2.5 and a maximum of 106.



**Figure B-S2.2:** Plot of the observed versus the predicted counts of ticks, Model 1.



**Figure B-S2.3:** The two plots are presented as a histogram and curve for binomial regression. (a) Goodness of model fit, Model 2, Dataset 2. The distribution of observed values ranged from a minimum, first quartile and median of 0, mean of 0.498, third quartile and maximum of 1, and the predicted values ranged from a minimum and first quartile of 0, mean of 0.503, median, third quartile and maximum of 1; (b) Goodness of model fit, Model 3, Dataset 3. Observed values have a minimum and 1<sup>st</sup> quartile of 0, mean of 0.51, median, third quartile and maximum of 1, and predicted values have a minimum, first quartile and median of 0, a mean of 0.48 and a third quartile and maximum of 1.



**Figure B-S2.4:** Boxplot of the predicted values for observed absences (0) and presences (1), Model 2 and Model 3.

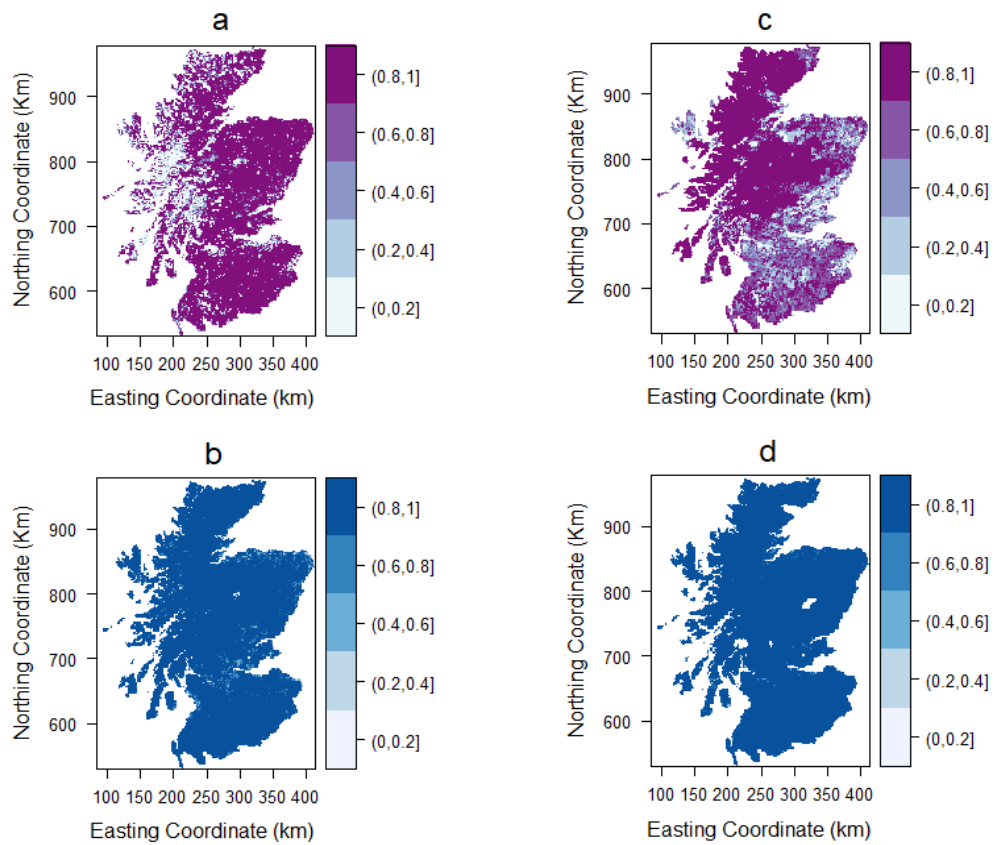
**S3: Results of two first models of *I. ricinus* presence-absence with respective predictive maps, with the data from public submissions (Dataset 2)**

**Table B-S3.1:** Posterior mean, SD, 2.5% and 97.5% quantiles for the binomial models of tick presence-absence with the data from public submissions (Dataset 2).

The first model was selected based on the Bayesian criteria (DIC and CPO); the second model was fitted using the significant covariates of Model 3. Neither of these models were selected as the final model for the predictions with Dataset 2.

Model	Fixed effects	Posterior mean	SD	2.5% quantile	97.5% quantile
Model 2: Presence-absence model selected based on the two Bayesian criteria.	Intercept	-4.9576	1.8414	-8.6884	-1.4563
	NDVI August <sup>a</sup>	0.1504	0.0296	0.0956	0.2120
	Roe deer presence	0.1234	0.0128	0.1001	0.1506
	Number of days of air frost November	-0.4190	0.1195	-0.6631	-0.1939
	Rain April	-0.0280	0.0089	-0.0462	-0.0111
	% cover of moorland	2.9543	1.0542	0.9449	5.0844
	Interaction between latitude and longitude	0.0166	0.0065	0.0042	0.0297
Model 2: Presence-absence model fitted with the significant covariates used in Model 3 (combined dataset).	Intercept	-3.8516	1.1225	-6.1032	-1.6939
	NDVI August	0.0012	0.0002	0.0008	0.0016
	Red deer density	8.3610	11.4365	-8.1275	35.2416
	Number of days of air frost November	-0.2962	0.0749	-0.4506	-0.1563
	Rain April	-0.0199	0.0074	-0.0350	-0.0057
	% cover of moorland	0.4616	0.8325	-1.2237	2.0459
	% cover of deciduous woodland	26.5532	15.9831	0.7997	62.9048
	% cover of coniferous woodland	5.5018	1.2241	3.3527	8.1592
	Interaction between latitude and longitude	0.0076	0.0041	-0.0003	0.0157

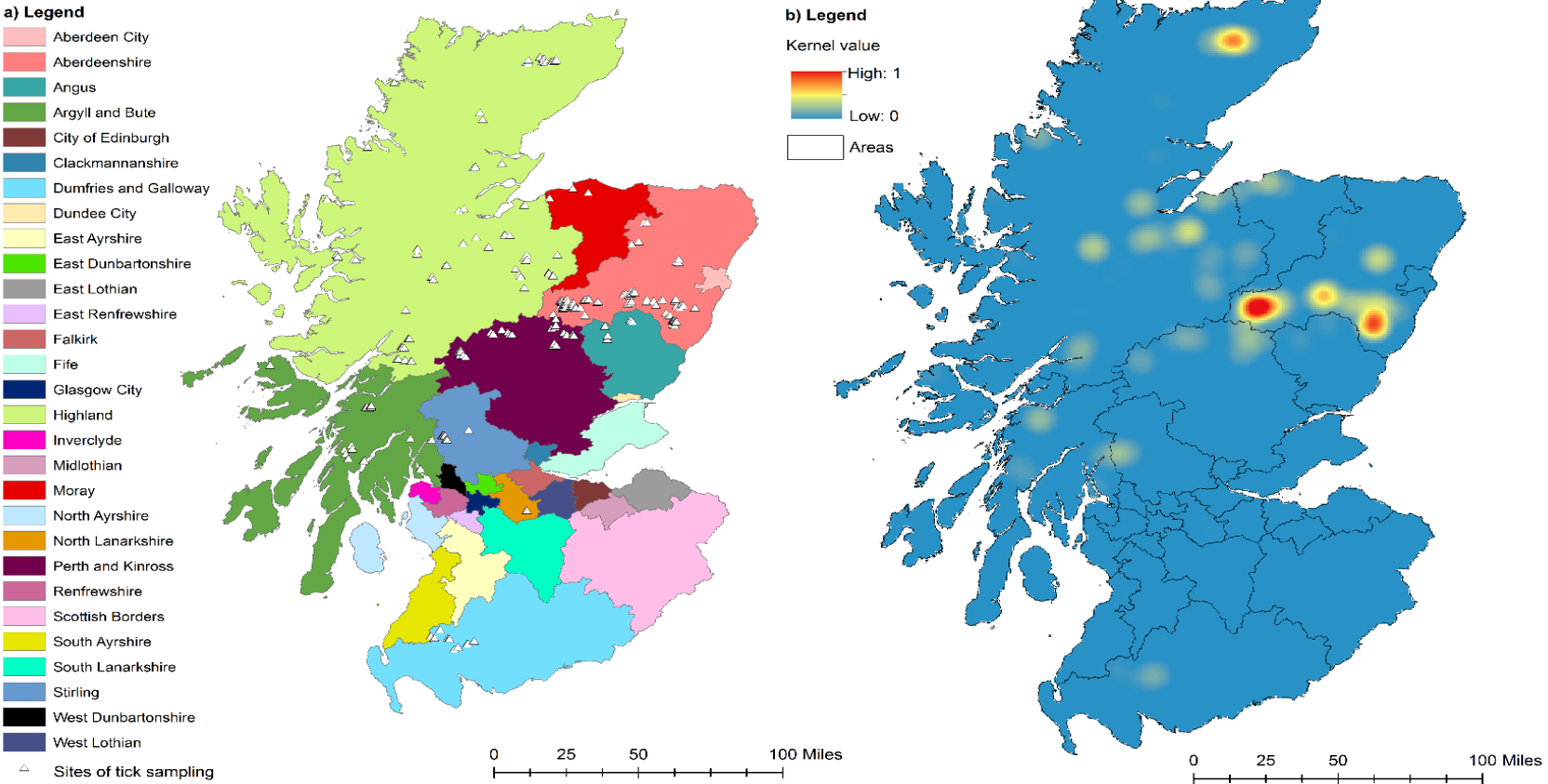
<sup>a</sup>In this model the posterior mean of NDVI was divided by 100. Abbreviations: SD, standard deviation; NDVI, normalised difference vegetation index



**Figure B-S3.1:** Predictive maps of binomial models of *I. ricinus* presence-absence with the data from public submissions (Dataset 2). (a) predictive map from first model selected based on the Bayesian criteria and (b) uncertainty map; (c) predictive map using covariates from Model 3 to predict Dataset 2 distribution and (d) respective uncertainty map.

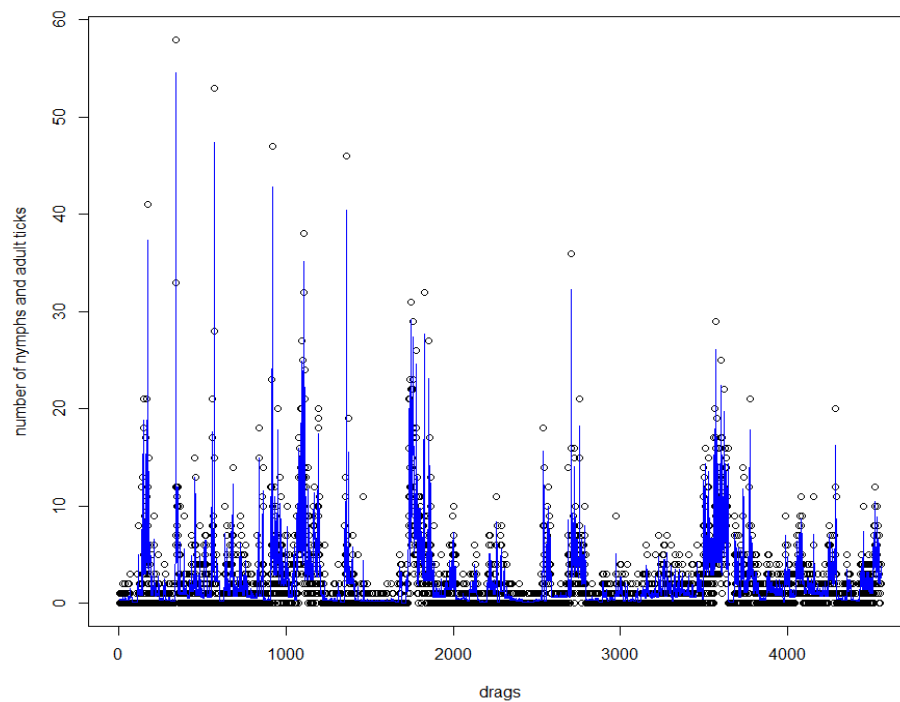


**S4: Kernel density analysis**



**Figure B-S4.1:** (a) Map of the council areas in mainland Scotland with the sites for questing tick surveys; (b) map output of the kernel density analysis for the sites of questing tick surveys.

## S5: Goodness of model fit – Model for subset of Dataset 1



**Figure B-S5.1:** Plot of fitted (blue line) versus observed values (dots) for the seasonal model of nymph and adult abundance, for Aberdeenshire (Dataset 1).

The observed number of nymphs plus adults per drag has minimum and 1<sup>st</sup> quartile of 0, median of 1, mean of 1.98, 3<sup>rd</sup> quartile of 2 and maximum of 58. In the predicted number of nymphs plus adults per drag, the mean changed for 2.1 and the maximum changed for 55.

# Appendix C: Supplementary materials for Chapter 4

## S1: Software for the website wireframes and website screenshots

The software used to develop the wireframes was mockflow (<https://mockflow.com/>).

SRUC: Epidemiology Resources Home About the Epi Unit Models Applications

---

Welcome to TickApp and thank you for taking the time to log your outdoor activities and tick sightings.

Please report if you have seen a tick or ticks during your activity – either crawling on your body, on your clothes or if you have a tick bite (attached tick). If you have not seen any ticks, please still report your activity – it is as useful for us to know where ticks aren't as where they are.

**Guidance for completing the form:**

- Please record your outdoor activity whether you find ticks or not.
- Please do not report ticks found on or attached to pets.
- Please do not report ticks seen on vegetation.

[Show/Hide Terms and Conditions](#)

**1a. What date did the activity take place?**

**1b. What time did the activity start?**

**2. How long did you spend doing the activity (hours)?**  
 10

**3. What was the predominant type of activity that you were doing?**


**4. During your activity, how much have you been exposed to vegetation?**

**5. Are you answering as an individual or on behalf of a group?**  
 Group  Individual


**6. How many were in your group?**

---

During your activity, if you saw any ticks please record them on the sliders opposite. If you did not see any ticks, please leave the sliders at zero and continue to the map below.



**7. How many ticks have you seen crawling on your body or clothes?**  
 10




**8. How many ticks were attached to your body (tick bites)?**  
 10

Please use the map to report the location/route of your activity. You can zoom in and out to find the area of your activity. Then click the map at different points along the route you took to lay down a marker. These markers will record the latitude and longitude of your activity. Please provide at least one marker per km of your route, if possible.

If you make a mistake you can clear the last point marked or all the points using the buttons below.

If you would prefer to upload a GPX file, you can do so below.

[Clear Last Point](#) [Clear All Markers](#)



## S2: Website consent form

### Terms and conditions - Ticks and Lyme Disease Project

In Europe the tick *Ixodes ricinus* is an abundant and widespread tick species and is also the primary vector of Lyme disease. Scotland's Rural College (SRUC) and the University of Edinburgh are carrying out a citizen science project which aims to improve predictive mapping of the risk of tick bites and Lyme disease. 'Citizen science' is the term that is used to describe a wide range of activities in which citizens participate actively in a scientific project. Citizen science can provide a valuable way of improving data collection on rate of human-tick encounters.

The project involves recruiting volunteers to report the number of ticks or tick bites they have found on themselves whilst performing outdoor activities, such as camping, walking, running, cycling, climbing, orienteering, or forestry work. We have developed this website – *TickApp* – so that volunteers can easily report this information. The volunteer reporting the information should be 18 years of age or older and may answer either as an individual or on behalf of a group (if answering on behalf of a group, children can be included in the group). At regular intervals (at least annually) all volunteers will be informed about the results of the project and what the results mean.

This project meets the requirements of the Data Protection Act of 1998 and, from 25<sup>th</sup> May 2018, the General Data Protection Regulation (GDPR). GDPR provides increased rights to individuals on (1) what data pertaining to living people organisations hold; (2) how these organisations process it; (3) how transparent these organisations must be regarding such data and its use. Our obligation according to GDPR is to ensure that any personal data is going to be used is looked after appropriately. This website does not collect any personal data. This website does not collect any personal data. Information on your location and activities are stored anonymously on secure servers at SRUC.

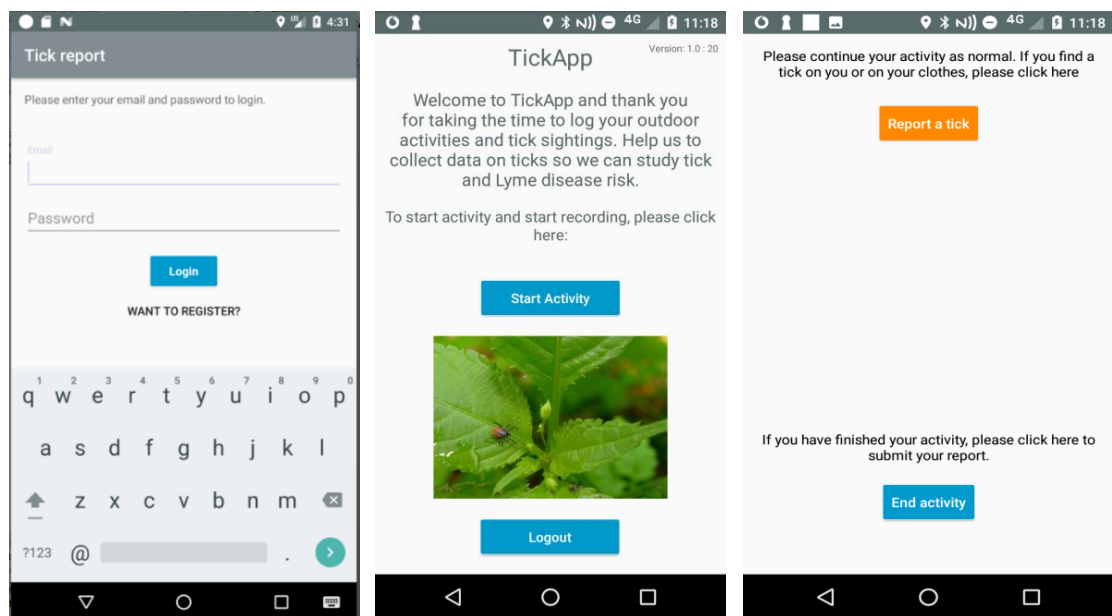
You are free to choose whether to be part of this study or not and may withdraw at any time. If you would like to participate in this study, and consent to the sharing of data as outlined above, please use the *TickApp* website to submit data. Your use of the website in this way is assumed to demonstrate your consent.

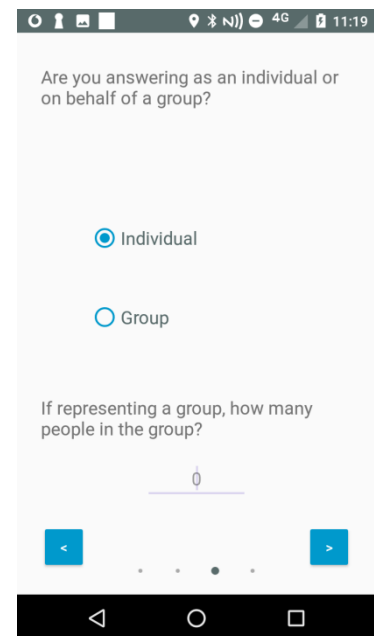
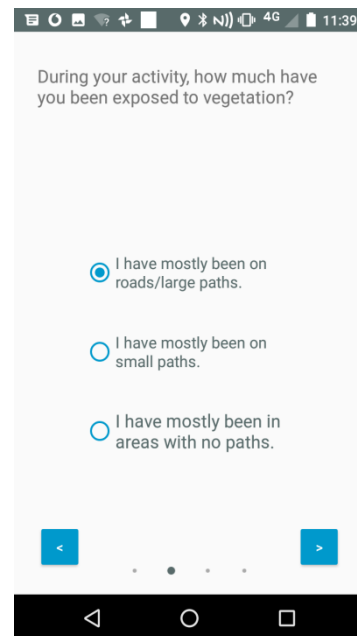
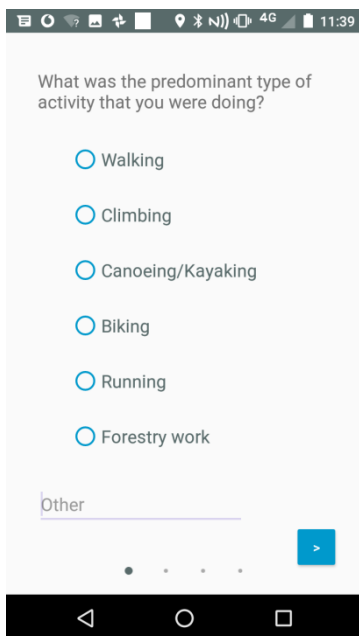
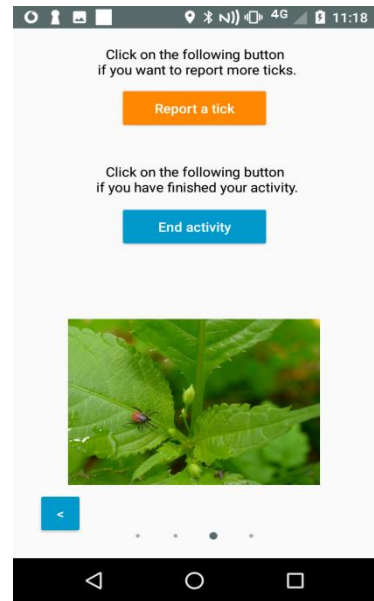
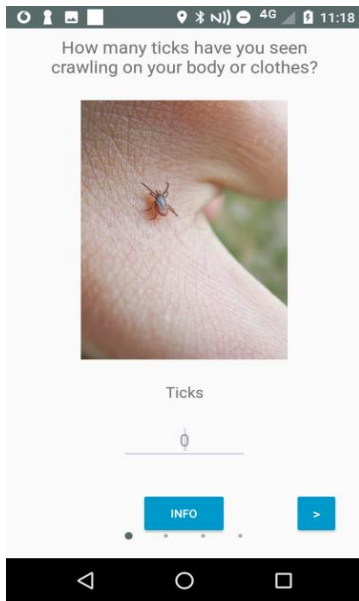
If you have any questions, please contact project staff at [tick.project@sruc.ac.uk](mailto:tick.project@sruc.ac.uk). This study has received ethical clearance from the University of Edinburgh.

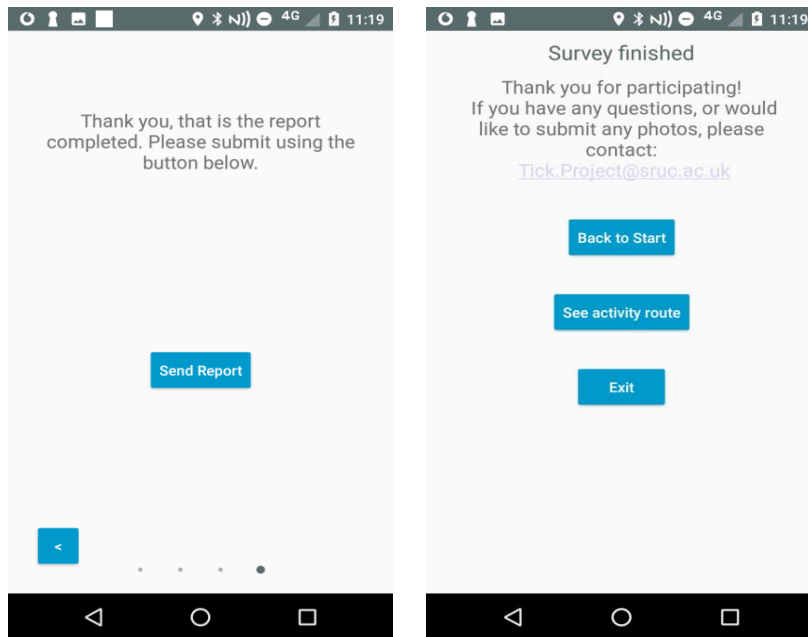
## Appendix D: Supplementary materials for Chapter 5

### S1: App software and screenshots

The software used for the app development included: Android Studio 3.1.2 (<https://developer.android.com/studio/>); Kotlin programming language (<https://kotlinlang.org/>); PHP programming language (<http://www.php.net/>); SQL language (<https://www.w3schools.com/sql/>); PostgreSQL (<https://www.postgresql.org/>); PgAdmin (<https://www.pgadmin.org/>). The software used to develop the wireframes was mockflow (<https://mockflow.com/>).







## S2: App consent form

### Terms and conditions - Ticks and Lyme Disease Project

In Europe the tick *Ixodes ricinus* is an abundant and widespread tick species and is also the primary vector of Lyme disease. Scotland's Rural College (SRUC) and the University of Edinburgh are carrying out a citizen science project which aims to improve predictive mapping of the risk of tick bites and Lyme disease. 'Citizen science' is the term that is used to describe a wide range of activities in which citizens participate actively in a scientific project. Citizen science can provide a valuable way of improving data collection on rate of human-tick encounters.

The project involves recruiting volunteers to report the number of ticks or tick bites they have found on themselves whilst performing outdoor activities, such as camping, walking, running, cycling, climbing, orienteering, or forestry work. We have developed this app – *TickApp* – so that volunteers can easily report this information. The volunteer reporting the information should be 18 years of age or older and may answer either as an individual or on behalf of a group (if answering on behalf of a group, children can be included in the group). At regular intervals (at least annually) all volunteers will be informed about the results of the project and what the results mean.

This project meets the requirements of the Data Protection Act of 1998 and, from 25<sup>th</sup> May 2018, the General Data Protection Regulation (GDPR). GDPR provides increased rights to individuals on (1) what data pertaining to living people organisations hold; (2) how these organisations process it; (3) how

transparent these organisations must be regarding such data and its use. Our obligation according to GDPR is to ensure that any personal data is going to be used is looked after appropriately. The only personal data required for this project is the email address used to download the app. This information will be securely held by SRUC and anonymised for data analysis. At the end of the project all personal data will be deleted. The app will also collect data from your phone on your location. These data will be anonymised and will not be linked to any of your personal information.

You are free to choose whether to be part of this study or not and may withdraw at any time. If you would like to participate in this study, and consent to the sharing of data as outlined above, please press the button below to continue to the *TickApp* app.

If you have any questions, please contact project staff at tick.project@sruc.ac.uk. This study has received ethical clearance from the University of Edinburgh.

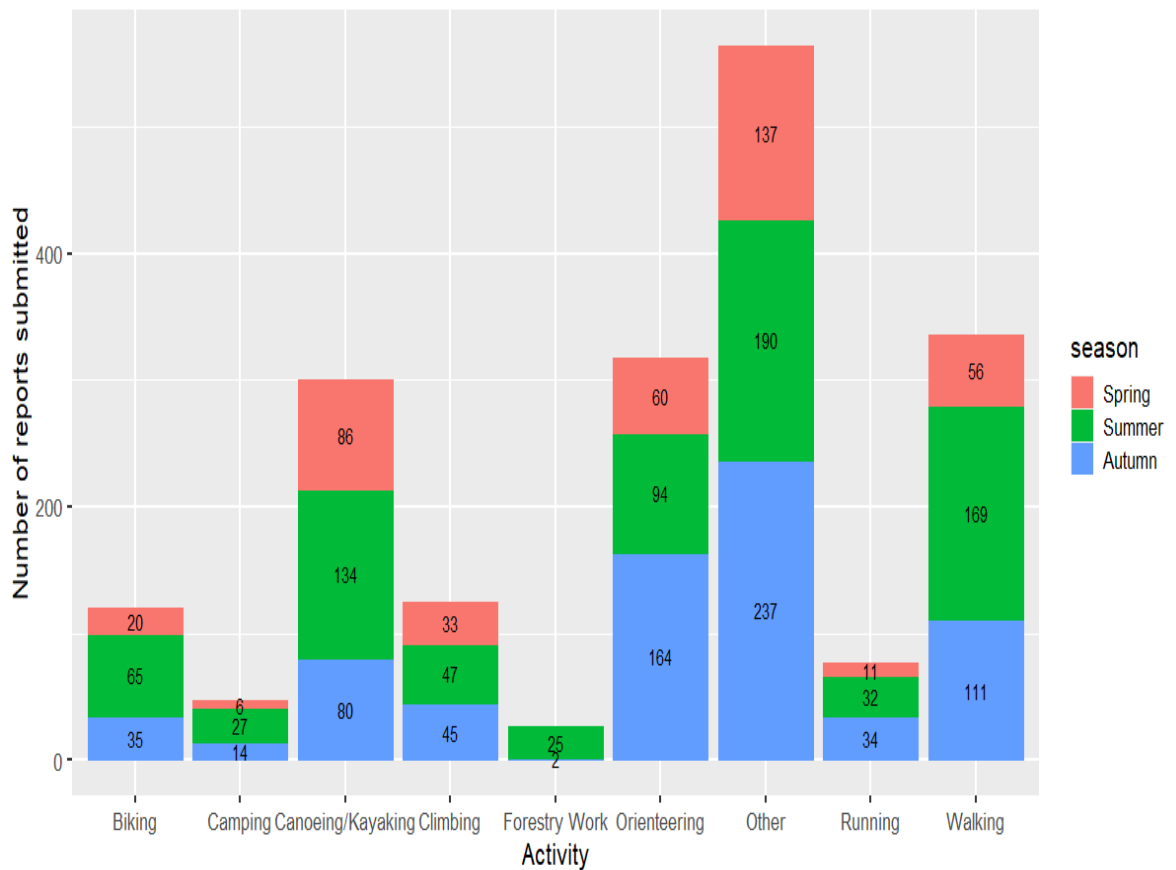
### **S3: Descriptive analysis of the citizen science data in Scotland**

**Table D-S3.1:** Summary statistics of the number of reports submitted to the CS project in Scotland (includes the Cairngorms and Lochaber). This includes information regarding the number of tick bites, tick encounters, and tick bite and tick encounter rates.

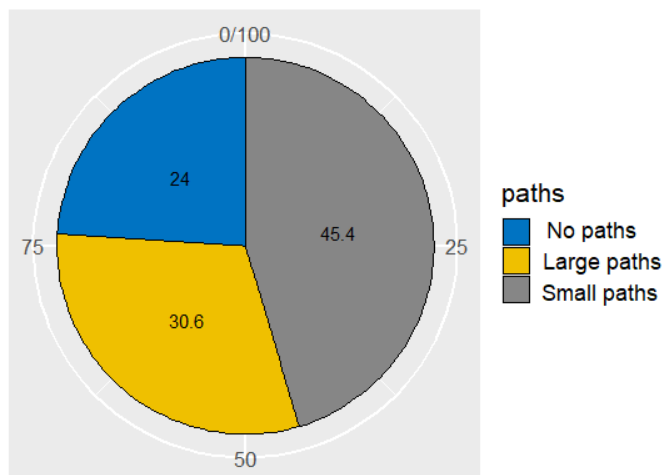
	<b>Scotland</b>
Number (and percentage) of reports with at least one tick bite	200 (10.5%)
Number (and percentage) of reports with at least one tick encounter	245 (12.8%)
Number of volunteers	65
Total number of people included*	15,646
Number (and percentage) of volunteers who were outdoor recreational organisations	2 (3.1%)
Number (and percentage) of reports from outdoor recreational organisations	1,266 (66.1%)
Number and percentage of volunteers who reported one time only	24 (36.9%)
Total number of tick bites	564
Number of tick encounters (i.e. tick bites + ticks crawling)	1,183
Tick bite rate (per person per hour): average, median and range	0.098; 0; 0 – 20.5
Tick encounter rate (per person per hour): average, median and range	0.203; 0; 0 - 33

\*Since volunteers could report as individual or as a group of N people; this value has duplicate information regarding each specific volunteer who reported more than one time.





**Figure D-S3.1:** Frequency and seasonal distribution of the reports submitted by activities in Scotland. Reports submitted between May and November 2018 and March and November 2019.



**Figure D-S3.2:** Percentage of reports submitted by type of path used in Scotland. Reports submitted between May and November 2018 and March and November 2019.

**Table D-S3.2:** Mean and SD of tick bite and tick encounter rates, per activity, season and type of path in Scotland. Rates were estimated with the data submitted to the 'TickApp' project (May and November 2018 and March and November 2019).

Variable	Category	Mean (SD) tick bite rate	Mean (SD) tick encounter rate
Activity	Biking	0.05 (0.17)	0.167 (1.23)
	Camping	0.003 (0.015)	0.065 (0.39)
	Canoeing/Kayaking	0.0009 (0.01)	0.009 (0.16)
	Climbing	0.018 (0.08)	0.04 (0.19)
	Forestry Work	0	0.015 (0.08)
	Orienteering	0.376 (1.35)	0.67 (2.21)
	Other	0.034 (0.25)	0.05 (0.39)
	Running	0.213 (0.88)	0.82 (3.90)
	Walking	0.07 (0.41)	0.154 (0.86)
Season	Spring	0.04 (0.23)	0.09 (0.50)
	Summer	0.136 (0.91)	0.273 (1.82)
	Autumn	0.09 (0.36)	0.191 (0.96)
Paths	No paths	0.338 (1.21)	0.647 (2.45)
	Small paths	0.028 (0.25)	0.08 (0.66)
	Large paths	0.013 (0.11)	0.04 (0.45)

Abbreviation: SD, standard deviation

#### S4: Descriptive analysis of the questing tick data

**Table D-S4.1:** Mean and respective SD of the total number of questing ticks (adults and nymphs) collected in each land cover type, in the two study areas.

Area	Land cover type	Number of sites	Mean questing ticks (SD)
Cairngorms	Coniferous forest	4	2.24 (5.4)
	Deciduous forest	3	7.8 (8.4)
	Heather/Moorland	5	0.83 (2.0)
	Heather grassland	3	0.4 (1.25)
	Improved grassland	3	0.013 (0.115)
	Total	18	2.48 (5.59)
Lochaber	Coniferous forest	4	0.72 (1.44)
	Deciduous forest	3	2.30 (5.57)
	Heather/Moorland	3	3.01 (5.84)
	Heather grassland	3	0.32 (0.92)
	Acid grassland	5	1.36 (2.95)
	Total	18	1.47 (3.82)

Abbreviation: SD, standard deviation

**S5: Correlation between predicted questing tick abundance and tick bite and tick encounter rates in the Cairngorms and in Lochaber**

**Table D-S5.1:** Results of the two univariable models for tick bite and tick encounter rates and the predicted questing tick abundance, in the Cairngorms and Lochaber.

Estimates for the posterior mean, SD, 2.5% and 97.5% quantiles are presented.

Area	Response variable	Fixed effects	Mean	SD	2.5% quartile	97.5% quartile
Cairngorms	Tick bite rate	Intercept	-0.194	0.082	-0.354	-0.034
		Tick abundance	0.103	0.017	0.069	0.136
	Tick encounter rate	Intercept	-0.862	0.387	-1.623	-0.102
		Tick abundance	0.458	0.081	0.298	0.617
Lochaber	Tick bite rate	Intercept	0.025	0.006	0.015	0.036
		Tick abundance	0.005	0.003	-0.001	0.010
	Tick encounter rate	Intercept	0.040	0.010	0.020	0.060
		Tick abundance	0.006	0.005	-0.004	0.016

Abbreviation: SD, standard deviation