DATA PAPER

A First Attempt at Modelling Roe Deer (*Capreolus capreolus*) Distributions Over Europe

Neil S. Alexander,[1] David Morley,[1] Jolyon Medlock,[2] Kate Searle[3] and William Wint[4]

[1] Research Assistant, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK

[2] Head of Medical Entomology at Public Health England, Porton Down, Salisbury, Wiltshire, SP4 0JG, UK

[3] Ecological Modeller, Centre for Ecology & Hydrology, Bush Estate, Penicuik, EH26 0QB, UK

[4] Senior Research Associate, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK

The presence of roe deer can be an important component within ecological and epidemiological systems contributing to the risk and spread of a range of vector-borne diseases. Deer are important hosts for many vectors, and may therefore serve as a focal point or attractant for vectors or may themselves act as a reservoir for vector-borne disease. Three spatial modelling techniques were used to generate an ensemble model describing the proportion of suitable roe deer habitat within recorded distributions for Europe as identified from diverse sources. The resulting model is therefore an index of presence, which may be useful in supporting the modelling of vector-borne disease across Europe.

Keywords: roe deer; *Capreolus capreolus*; tick-borne; culicoides-borne; mosquito-borne; distribution; disease; habitat; linear regression; Random Forest; generalised linear modelling

Funding Statement: This study was partially funded by EU grant FP7-613996 VMERGE and is catalogued by the VMERGE Steering Committee as VMERGE001 (http://www.vmerge.eu). The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission.

(1) Overview

Context

Spatial coverage

Description: This dataset is clipped to the EDENext [1] extent which covers the continent of Europe and parts of North Africa down to 34 degrees latitude. The projection is WGS84 (ESPG:4326).

- Northern boundary: 72.3
- Southern boundary: 34.0
- Eastern boundary: -12.0
- Western boundary: 47.6

Temporal coverage

01 April 2014 (current).

Species

Roe deer, *Capreolus capreolus.*

(2) Methods

Steps

Binary presence and absence

Five independent sets of distribution data were combined to produce a single presence absence mask. The data sets used were as follows:

- The EMMA Database [2]: Mapping Europe's mammals using data from the Atlas of European Mammals
- The Global Biodiversity Information Facility (GBIF) [3]
- IUCN Red List Dataset [4]
- The National Biodiversity Network [5] UK 10k Data
- Spanish Ministry of Agriculture National Inventory of Biodiversity [6]

Habitat definition

For much of the indicated range the distributions detailed above were, by their nature, simple presence limits. Within these designated boundaries there was no indication of absence. In order to introduce absences within these limits, suitability masks were defined using species-specific habitat preferences derived from land cover classes, using GLOBCOVER [7] at 1 km resolution. The habitats were defined as more than 10% Woodland, and neither urban nor peri-urban, according to Tapper(1999) [8], and is thus somewhat UK centric To allow for behaviours where deer utilise pasture/heathland/grassland close to woodland shelter we also defined as suitable habitat areas where grassland/heathland occurred within 1km of a cell with sufficient woodland (Searle, personal communication).

The 300m GLOBCOVER dataset was reclassified three times for woodland = 1 and other = 0; for urban areas = 0

Value	Label	Grass Pasture	No Urban & Urban Fringe	Roe
11	Post-flooding or irrigated croplands (or aquatic)	0	1	0
14	Rainfed croplands	0	1	0
20	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	1	1	1
30	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	1	1	1
40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	0	1	1
50	Closed (>40%) broadleaved deciduous forest (>5m)	0	1	1
60	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	0	1	1
70	Closed (>40%) needleleaved evergreen forest (>5m)	0	1	1
90	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	0	1	1
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	0	1	1
110	Mosaic forest or shrubland (50-70%) / grassland (20-50%)	1	1	1
120	Mosaic grassland (50-70%) / forest or shrubland (20-50%)	1	1	1
130	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	0	1	0
140	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	1	1	0
150	Sparse (<15%) vegetation	0	1	0
160	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily)	0	1	1
170	Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water	0	1	0
180	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil	1	1	0
190	Artificial surfaces and associated areas (Urban areas >50%)	0	0	0
200	Bare areas	0	1	0
210	Water bodies	0	1	0
220	Permanent snow and ice	0	1	0
230	No data (burnt areas, clouds,...)	0	1	0

Table 1: Reclassed values defining the GLOBCOVER suitability layers.

other = 1 and for grassland & pasture = 1 other = 0 as per **Table 1**. The three layers were each aggregated to 1km, and then Suitable habitat was defined as a) those cells containing more than 10% woodland but no urban area; or b) grassland cells next to otherwise suitable habitat. All data processing was undertaken in ESRI ArcGIS 10.0.

The 1km resolution habitat suitability masked data was then combined with the presence data and converted to a percentage of suitable habitat at a 20km resolution.

Model predictor suite

The spatial modelling requires a comprehensive predictor variable suite that included a wide range of remotely sensed variables as follows:

- Remotely sensed climatic indicators derived by Temporal Fourier Analysis (TFA) of MODIS satellite imagery of several temperature parameters, and vegetation indices for the period 2001-2008 [9]
- Digital Elevation from the Shuttle Radar Topography Mission, together with derived aspect and ruggedness [10]
- Temporal Fourier Analysis (TFA) of Precipitation, and allied Bioclimatic Indicator (Bioclim) precipitation variables from the WORLDCLIM datasets [11]
- Length of Growing Period from United Nations Food and Agriculture Organisation [12]
- Travel Time to major towns from the Joint Research Centre at Ispra [13]

- Human population density derived from the Global Rural Urban Mapping project at CEISIN [14]
- A distance weighted human population index layer [15] representing the likelihood of human visits based on the population within 30km.

Habitat suitability modelling

The percentage of suitable habitat layer was then offered to three modelling techniques: GLM [16] multivariate regression and Random Forest [17], both using R-project [18] modules embedded within the VECMAP [19] software suite, and the FAO FARMS [20] regression tool developed for livestock density modelling. All three methods were bootstrapped at least 25 times, and models were further refined by using a zoned approach whereby separate models were produced for a series of 50 eco-climatic zones based on climate, vegetation and seasonality. Such zonation tends to produce more accurate sub-models, which can then be combined into a single output.

The average of the three models was produced as an ensemble consensus product.

Output datasets

A copy of both the presence/absence layer and the ensembled modelled habitat suitability have been provided as a quick look map in JPEG format to view from any image viewer. The data itself is distributed as GIS Raster data in two formats. GeoTIFFs which is a standard proprietary GIS raster format. GeoJP2 (JPEG 2000 format) which is a non-proprietary format.

To access and analyse the Raster data directly GeoTIFFs and GeoJPGs can be read by most GIS software and some other software packages These formats are compatible with proprietary (ESRI ArcGIS) and open source Quantum GIS (QGIS) [21] or R-project [18] raster package).

If the reader has no suitable software already installed the authors suggest downloading the opensource QGIS software free of charge from http://www.qgis.org to view these data.

Folder structure

- quicklooks - JPEG maps for viewing only
- tiff - GeoTIFF data 0.008333 degree (~1km) 32bit floating point
- geoJPG2k - GeoJPG 2000, 0.008333 degree (~1km) 16bit unsigned Integer data

Sampling strategy

Sample points were extracted for input into the three different models from a 20km matrix defining the percentage of habitat suitability within known distributions. Depending on the model 1000-3000 sample points were used in each of 25 bootstraps.

Quality control

These models are a first attempt at quantifying the roe deer distribution at this scale and there has been no ground truth validation of these maps so far. The model outputs all, however, satisfy standard accuracy metrics (AIC and R squared) assuring statistical reliability. They have also been informally reviewed by project deer experts.

Constraints

There were no constraints involved in data production.

Privacy

N/A

(3) Dataset description

Object name

euroemodel.zip

Data type

Primary data, processed data, interpretation of data.

Format names and versions

JPG, JP2, TIF, TFW, XML.

Creation dates

28 April 2014

Dataset creators

- Dr. William Wint: Senior Research Associate, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK
- Dr. David Morley: Research Assistant, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK
- Neil S. Alexander: Research Assistant, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK. Corresponding email: neil.alexander@zoo.ox.ac.uk

Language

English

Embargo

N/A

Repository location

http://dx.doi.org/10.6084/m9.figshare.1008335

Publication date

If already known, the date the dataset was published in the repository (28 April 2014).

(4) Reuse potential

These layers are a first attempt to provide a description of roe deer habitat as a proxy for abundance at a continental scale. They have been developed in the hope they will aid epidemiologists test hypotheses relating to the role of roe deer in the spread of vector-borne disease.

Areas of future development on the dataset itself might be to: assess the accuracy of the maps through ground-truthing; a comparison of the three different models used in this analysis and an assessment of which model provides the most accurate outputs; An attempt at a more systems-based approach to modelling deer abundance at a country scale.

Acknowledgements
Particular thanks go to Stephen Tapper whose book *A question of balance: game animals and their role in the British countryside* was used as a basis for the habitat definitions for this work.

References

1. **EDENext Data Management Team** 2012 EDENext Data Portal. Available at: http://www.edenextdata.com [Last accessed 14 April 2014].
2. **Mitchel-Jones, A j, Amori, G, Bogdanowicz, W, Krystufek, B, Reijnders, P J H, Spitzenberger, F, Stubbe, M, Thissen, J B M, Vohralik, V** and **Zima, J** 1999 *The Atlas of European Mammals & EMMA Dataset*. London: Poyser. Available at: http://www.european-mammals.org
3. **The Global Biodiversity Information Facility (GBIF).** Available at: http://www.gbif.org/ [Last accessed 17 October 2012].
4. **IUCN** 2012 The IUCN Red List of Threatened Species. Version 2012.2. Available at: http://www.iucnredlist.org [Last accessed 17 October 2012].
5. **NBN Gateway** Available at: http://data.nbn.org.uk [Last accessed 07 July 2008]. The information used here was sourced through the NBN Gateway website and included multiple resources. The data providers and NBN Trust bear no responsibility for the further analysis or interpretation of this material, data and/or information.
6. Spanish Ministry of Agriculture National Inventory of Biodiversity.
7. **Arino, O, Ramos Perez, J, Julio, J, Kalogirou, V, Bontemps, S, Defourny, P** and **van Bogaert, E** 2012 Global Land Cover Map for 2009 (GlobCover 2009). © *European Space Agency (ESA) & Université catholique de Louvain (UCL)*. DOI: http://dx.doi.org/10.1594/PANGAEA.787668
8. **Tapper, S** (ed.) 1999 *A question of balance: game animals and their role in the British countryside.* United Kingdom: Game Conservancy Trust.
9. **Scharlemann, J P W, Benz, D, Hay, S I, Purse, B V, Tatem, A J, Wint, G R W** and **Rogers, D J** 2008 Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS ONE*, 3(1): e1408. DOI: http://dx.doi.org/10.1371/journal.pone.0001408
10. **SRTM** 2012 The Shuttle Radar Topography Mission (SRTM) homepage. Available at: http://www2.jpl.nasa.gov/srtm/
11. **Hijmans, R J, Cameron, S E, Parra, J L, Jones, P G** and **Jarvis, A** 2005 Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25(15): 1965-1978. DOI: http://dx.doi.org/10.1002/joc.1276
12. **United Nation Food and Agricultural Organisation** 2012 Length of growing period (LGP) zones of the world (FGGD). Available at: http://www.fao.org/geonetwork/srv/en/metadata.show?id=14057
13. **Joint Research Centre (JRC)** 2012 Travel time to major cities: A global map of Accessibility. Available at: http://bioval.jrc.ec.europa.eu/products/gam/sources.htm
14. **Center for International Earth Science Information Network (CIESIN)/Columbia University, International Food Policy Research Institute (IFPRI), The World Bank** and **Centro Internacional de Agricultura Tropical (CIAT)** 2011 *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at: http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-count [Last accessed 27 April 2011].
15. **Alexander, N** and **Wint, W** 2013 Data from: Projected population proximity indices (30km) for 2005, 2030 & 2050. *Dryad Digital Repository*. DOI: http://dx.doi.org/10.5061/dryad.12734
16. **R-Project glm** 2012 Fitting Generalized Linear Models. Available at: http://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html
17. **R-Project randomForest package** 2012 randomForest: Breiman and Cutler's random forests for classification and regression. Available at: http://cran.r-project.org/web/packages/randomForest/index.html
18. **R Core Team** 2012 *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/
19. Modelling Module in development for the VECMAP system. Produced by: Avia-GIS Zoersel Belgium; ERGO Ltd. Oxford, UK; MEDES, Toulouse, France for the European Space Agency.
20. **Robinson, T, Wint, W, Van Boeckel, T P, Conchedda, G, Ercoli, V, Palamara, E, Cinardi, G, D'Aietti, L** and **Gilbert, M** (in prep) FARMS: Mapping the Global Distribution of Livestock.
21. **QGIS Development Team** 2013 QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: http://qgis.osgeo.org

How to cite this article: 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): e2, DOI: http://dx.doi.org/10.5334/ohd.ah

Published: 14 July 2014

]u[*Open Health Data* is a peer-reviewed open access journal published by Ubiquity Press **OPEN ACCESS**