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

Neil S. Alexander,¹ David Morley,¹ Jolyon Medlock,² Kate Searle³ and William Wint⁴

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 0QF, 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.
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]

Table 1: Reclassed values defining the GLOBCOVER suitability layers.
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
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.