The presence of red deer may be a contributing factor within the 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 become a reservoir for vector-borne disease. Three spatial modelling techniques were used to generate an ensemble model describing the proportion of suitable red 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:** red deer; *Cervus elaphus*; 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 VMERGE002 (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**

Red deer, *Cervus elaphus*.

---

**(2) Methods**

**Steps**

**Binary presence and absence**

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

- The Global Biodiversity Information Facility (GBIF) [3]
- IUCN Red List Dataset [4]
- The National Biodiversity Network UK 10k Data [5]
- Spanish Ministry of Agriculture National Inventory of Biodiversity [6]

**Habitat definition**

For much of the indicated range the distributions 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 being more than 25% Woodland or Moorland according to Tapper(1999) [8], and are thus somewhat UK centric.

The 300m GLOBCOVER dataset was reclassified as woodland or moorland = 1 and other = 0 as per Table 1. The data was then aggregated to 1km and those cells with greater than 25% woodland or moorland were then classed as suitable. 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

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

| Table 1 | Reclassed values defining the GLOBCOVER suitability. |
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 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 for each species was then produced as an ensemble consensus product for each species.

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.

Folder structure

- quicklooks - JPEG maps for viewing only
- tiff - GeoTIFF data 0.008333 degree (~1km) 32bit floating point
- geojPG2k - GeoJP2 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. 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 red 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
euredmodel.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

License

This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Embargo

N/A

Repository location
http://dx.doi.org/10.6084/m9.figshare.1008334

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 red 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 red 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.

Thank you also to the following experts in the fields of Deer ecology that helped us identify the key habitats and gave us feedback on the outputs: Dr. Kate Searle, Ecological Modeller, Centre for Ecology & Hydrology, Penicuik, UK.

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.
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