

## DATA PAPER

## Four Rodent and Vole Biodiversity Models for Europe

William Wint<sup>1</sup>, David Morley<sup>2</sup>, Neil S. Alexander<sup>2</sup><sup>1</sup> Senior Research Associate, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, United Kingdom<sup>2</sup> Research Assistant, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, United Kingdom

The mechanisms underlying associations between rodent diversity and the spread of rodent borne diseases are still unclear. The number of potential host species may influence a disease by either spreading the pathogen (or vectors) more rapidly or conversely by reducing contact with other hosts through the effects of dilution. In either case the number of potential hosts may impact on the distribution of a disease or its vectors. Four spatially modelled indices of rodent species richness have been generated to support distribution modelling of rodent borne diseases specifically initially focusing on Hantaviruses and tick borne diseases.

**Keywords:** rodent, vole, biodiversity, tick-borne, hantavirus, linear regression, random forest, generalised linear modelling

**Funding statement**

EU grant GOCE-2003-010284 EDENext and is catalogued by the EDENext Steering Committee as EDENextD002 (<http://www.edenext.eu>). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

**(1) Overview****Context****Spatial coverage**

Description: This dataset is clipped to the EDENext<sup>1</sup> extent that 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**

01/01/2012 (Current)

**Species**

Striped Field Mouse, *Apodemus agrarius*; Yellow-necked Mouse, *Apodemus flavicollis*; Broad-toothed Field Mouse, *Apodemus mystacinus*; Wood Mouse, *Apodemus sylvatica*; Bank vole, *Clethrionomys glareolus*; Common vole, *Microtus arvalis*; European pine vole, *Microtus subterraneus*; Norway rat, *Rattus norvegicus*; Black Rat, *Rattus rattus*; Common Shrew, *Sorex araneus*; Eurasian Pygmy Shrew, *Sorex minutus*

**(2) Methods****Steps****Input Datasets**

- The rodent distribution layers that were modelled models were compiled from several sources as follows:
- The EMMA Database<sup>2</sup>: Mapping Europe's mammals using data from the Atlas of European Mammals.
- IUCN Red List Dataset<sup>3</sup>
- The National Biodiversity Network<sup>4</sup> UK 10k Data- Making all biological records freely and easily available to everyone
- Globcover 2009 Land cover map<sup>5</sup>

Four biodiversity indices - the number of species recorded at a single location - were defined according to different species lists suggested by various experts identified in the Acknowledgements section. The species for each index are identified in Table 1.

Species\Index	sp04no	sp05n0	Hanta10	ters4
<i>Apodemus agrarius</i>	X	X	X	
<i>Apodemus flavicollis</i>	X	X	X	X
<i>Apodemus mystacinus</i>			X	
<i>Apodemus sylvatica</i>		X	X	X
<i>Clethrionomys glareolus</i>	X	X	X	

Species\Index	sp04no	sp05n0	Hanta10	ters4
<i>Microtus arvalis</i>	X	X	X	X
<i>Microtus subterraneus</i>			X	
<i>Rattus norvegicus</i>			X	
<i>Rattus rattus</i>			X	
<i>Sorex araneus</i>			X	X
<i>Sorex minutus</i>			X	

**Table 1:** Species included in each index are marked with an "X". For more details please see the supporting document (edenextvolelayers.docx) within the data download.

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

1. 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<sup>6</sup>
2. Digital Elevation from the Shuttle Radar Topography Mission, together with derived aspect and ruggedness<sup>7</sup>
3. Fourier Processed Precipitation, and allied Bioclimatic Indicator (Bioclim) precipitation variables from the WORLDCLIM datasets<sup>8</sup>
4. Length of Growing Period from United Nations Food and Agriculture Organisation<sup>9</sup>
5. Travel Time to major towns from the Joint Research Centre at Ispra<sup>10</sup>
6. Human population density derived from the Global Rural Urban Mapping project at CEISIN<sup>11</sup>
7. A weighted human population proximity layer<sup>12</sup>

The four different rodent biodiversity index layers were modelled using a range of techniques:

1. zoned random forest<sup>14</sup> method calculated via the VECMAP<sup>15</sup> modelling system using a point input dataset
2. zoned bootstrapped General Linear Model (GLM)<sup>16</sup> method calculated via VECMAP<sup>15</sup> modelling system using a point input dataset
3. zoned regression calculated using the FARMS<sup>17</sup> modelling suite using an polygon input dataset

An ensemble mean was taken of the three model outputs for each Biodiversity variable.

#### Output Datasets

Each of the four Biodiversity indices has 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. Please note that as the GeoJP2 format is an integer-only format, the values found in this format have been multiplied by 100 to minimise the loss of data.

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) or R-project raster package)

#### 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. \*\*\*This data has been multiplied by 100 to save as an integer value with minimum loss of data.\*\*\*

#### Sampling Strategy

Sample points were extracted for input into the three different models from a 20km matrix defining the number of key rodent species for each index (as defined in table 1). Depending on the model 1000-2000 sample points were used in each of 25 bootstraps.

#### Quality Control

These models are a first attempt at quantifying the rodent biodiversity 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, R squared and Kappa statistic) assuring statistical reliability. They have also been informally reviewed by project Rodent-Borne disease experts.

#### Constraints

There were no constraints involved in data production.

#### Privacy

N/A

### (3) Dataset Description

#### Object Name

volebiodiv2.zip

#### Data Type

Primary data, Processed data, Interpretation of data.

#### Format Names and Versions

JPG, TIF, TFW, XML

#### Creation Dates

15/08/2012 - 15/08/2012

#### Dataset Creators

William Wint, David Morley, Neil S. Alexander.

#### Repository Location

<http://dx.doi.org/10.5061/dryad.771gr>

#### Publication Date

08/07/2013

#### Language

English

#### License

CC0

#### (4) Reuse potential

The subject of the effect of rodent biodiversity upon the spread of vector-borne disease is a fairly new topic. These layers are a first attempt to provide a description of rodent biodiversity at a continental scale. They have been developed in the hope they will aid epidemiologists test hypotheses relating to the role of rodent biodiversity' in the spread of infectious 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; the application of this approach to different groups of species thought to be related to the spread of different diseases.

#### Acknowledgements

Thank you to the following experts in the fields of Rodent-Borne Disease that helped us identify the key species to include in this analysis: Katrien Tersago, Evolutionary Ecology Group, University of Antwerp, Belgium; Liina Voutilainen, METLA and University of Helsinki, Finland; Gert Olsson, Swedish Institute for Infectious Disease Control, Stockholm, Sweden.

#### References

1. **EDENext Data Management Team** EDENext Data Portal. Accessed 27/04/2012. Available from: <http://www.edenextdata.com>
2. **University of Paris 7 Denis Diderot on behalf of UMS2414 RIATE, France** ESPON2013 database. Accessed 27 April 2012. Available from: [http://www.espon.eu/main/Menu\\_Projects/Menu\\_ScientificPlatform/espondatabase2013.html](http://www.espon.eu/main/Menu_Projects/Menu_ScientificPlatform/espondatabase2013.html)
3. **UN Dept of Economic and social affairs, Population Division** 2010 World Population Prospects, the 2010 Revision. Accessed 27 April 2011. Available from: <http://esa.un.org/unpd/wpp/Excel-Data/population.htm>
4. **National Biodiversity Network** NBN Gateway. Accessed 7 July 2008. Available from: <http://data.nbn.org.uk>. The data providers and NBN Trust bear no responsibility for the further analysis or interpretation of this material, data and/or information.
5. **Arino O**, et al. 2009 Global Land Cover Map for 2009. Pangea., DOI: <http://dx.doi.org/10.1594/PANGAEA.787668>
6. **Scharlemann, J P W** 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>
7. **The Shuttle Radar Topography Mission** The Shuttle Radar Topography Mission (SRTM) Home Page. <http://www2.jpl.nasa.gov/srtm/>
8. **Hijmans R J**, et al 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>
9. **UN Dept of Economic and social affairs, Population Division** Length of growing period (LGP) zones of the world (FGGD). Available from: <http://www.fao.org/geonetwork/srv/en/metadata.show?id=14057>
10. **UN Dept of Economic and social affairs, Population Division** 2010 World Population Prospects, the 2010 Revision. Available from: <http://www.fao.org/geonetwork/srv/en/metadata.show?id=14057>
11. **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)** Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Accessed 27 April 2011. Available from: <http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-count>
12. **Alexander N S and Wint W** 2013 Projected Population Proximity Indices (30km) for 2005, 2030 and 2050. *Journal of Open Public Health Data* 1(1) DOI: <http://dx.doi.org/http://dx.doi.org/10.5334/jophd.ab>
13. **R Core Team** 2012 R: A language and environment for statistical computing. 2nd ed. Vienna: R Foundation for Statistical Computing.
14. **randomForest: Breiman and Cutler's random forests for classification and regression** R-Project randomForest package. Available from: <http://cran.r-project.org/web/packages/randomForest/index.html>
15. **Avia-GIS Zoersel, Belgium, ERGO Ltd Oxford, UK and MEDES, Toulouse, France, for the European Space Agency** Modelling Module in development for the VECMAP system. Available from: [http://www.medes.fr/home\\_en/telemedicine/tele\\_epidemiology/vecmap.html](http://www.medes.fr/home_en/telemedicine/tele_epidemiology/vecmap.html)
16. **R-Project glm** Fitting Generalized Linear Models. Available from: <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html>
17. **Robinson T** et al FARMS: Mapping the Global Distribution of Livestock. In preparation.

**How to cite this article:** Wint, W, Morley, D and Alexander, N S 2013 Four Rodent and Vole Biodiversity Models for Europe. *Journal of Open Public Health Data* 1(1):e3, DOI: <http://dx.doi.org/10.5334/jophd.ac>

**Published:** 11 July 2013

**Copyright:** © 2013 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.