open health data

Alexander, N S et al 2016 The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster – Where are All the Boars? An Attempt to Gain a Continental Perspective. *Open Health Data,* 4: e1, DOI: http://dx.doi.org/10.5334/ohd.24

DATA PAPER

The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster – Where are All the Boars? An Attempt to Gain a Continental Perspective

Neil Stewart Alexander[1], Giovanna Massei[2] and William Wint[3]

[1] Analyst, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK
neil.alexander@zoo.ox.ac.uk

[2] Senior Ecologist, National Wildlife Management Centre, Animal and Plant Health Agency (APHA), Sand Hutton, York, YO41 1LZ, UK
Giovanna.Massei@apha.gsi.gov.uk

[3] Senior Analyst, Environmental Research Group Oxford (ERGO), Department of Zoology, Oxford, OX1 3PS, UK
william.wint@zoo.ox.ac.uk

Corresponding author: Neil Stewart Alexander

Wild boar is a host of a number of arthropod-vectored diseases and its numbers are on the rise in mainland Europe. The species potentially impacts ecosystems, humans and farming practices and so its distribution is of interest to policy makers in a number of fields beyond that of the primarily epidemiological goal of this study.

Three statistical model outputs describing the distribution and abundance of the species *Sus scrofa* (Wild boar) are included in this data package. The extent of this dataset covers continental Europe. These data were presented as a poster [1] at the conference Genes, Ecosystems and Risk of Infection (GERI 2015).

The first of the three models provide a European map presenting the probability of presence of *Sus scrofa*, which can be used to describe the likely geographical distribution of the species. The second and third models provide indices to help describe the likely abundance across the continent. The two indices include "the proportion of suitable habitat where presence is estimated" and a simple classification of boar abundance across Europe using quantiles of existing abundance data and proxies.

Keywords: *Sus scrofa*; Distribution; Abundance, Random Forest; Statistical modelling; Europe
Funding statement: This study was funded by EU grant FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext416 (http://www.edenext.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

Introduction/Study Description

Wild boar *Sus scrofa* are an important component of the ecological and epidemiological systems within which vector-borne diseases persist. Wild boar are hosts to a number of vector species, and they can therefore impact on disease cycles as reservoirs of pathogens. Information on wild boar distribution and abundance could therefore make an important contribution to models of vector-borne disease risk.

With a single exception [2], the many studies that have focussed on the distribution, abundance and habitat-use of wild boar were generally carried out in relatively small areas such as national parks or at country level. Given the broader, continental scale required for effectively advising European policy on disease management, an attempt has been made to produce a continental scale distribution and abundance map.

This study combines a review of the existing literature along with abundance-related data from a range of sources, including national hunting organisations, international and national distribution databases, to provide a continental dataset and perspective of boar distribution and abundance.

To create the final European 1km resolution boar map, the combined quantitative data described above were constrained using a habitat suitability mask derived from the GlobCover land cover database informed by published descriptions of habitat preference as well as expert opinion. A number of spatial distribution modelling tools

available from the VECMAP [3] Modelling suite were used to produce three final modelled distribution outputs for Europe using the Random Forest approach. These comprise a 1km probability of presence/absence layer, a 1km abundance index based on presence and habitat availability, and a 1km ranked abundance map based on regional abundance studies and national hunting figures.

2. Context

Spatial coverage

Description: Continental Europe, including European Russia.

Northern boundary: 72.
Southern boundary: 10.
Eastern boundary: –24.5.
Western boundary: 60.

Temporal coverage

2015.

Species

Sus scrofa, wild boar, pig (feral).

3. 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 [4]: Mapping Europe's mammals using data from the Atlas of European Mammals.
- The Global Biodiversity Information Facility (GBIF) [5].
- IUCN Red List Dataset [6].
- The National Biodiversity Network [7] UK 10k Data.
- Spanish Ministry of Agriculture National Inventory of Biodiversity [8].

Habitat definition

For much of the indicated range, the distributions detailed above were, by their nature indications of current 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 [9] at 1 km resolution Downloaded from the EDENext Data Portal [10]. These suitability definitions are recorded in **Table 1**.

The presence absence data described in the previous section were combined with the suitability layer and aggregated to a 10km grid as a proportion of suitable habitat. The values of which were sampled and offered up to the Random Forest modelling framework within VECMAP [3] outlined later in this paper.

Boar Abundance Inputs

A comprehensive literature review of *Sus scrofa* abundance studies was undertaken [11–33] which unearthed a piecemeal collection of abundance data focused mainly on small areas such as national parks or in some cases up to country level. These were recorded by different methods and across different time periods and has a spatial coverage across Europe which was far from regular. A notable exception was a recent review of wild boar population trends in 18 countries in Europe, based on hunting statistics [2].

To complement these abundance data, hunting figures were also identified for a number of countries at both national level and sub-national level [34–38]. After discussion with boar specialists it was agreed that, at least within a single country, hunting data could be considered as a valid proxy for abundance. In order to get the most complete coverage across the continent, it was decided to convert the available data to relative abundance indices that could be compared across countries by normalising the available number according to known national abundance figures.

The data were thus categorised into quantiles, with a fifth category of 0 or negligible boar numbers where known or inferred in areas defined as unsuitable habitat. The resulting database provided categorical boar abundance ranging from 0–4 (0 = none/negligible boar abundance to 4 = high abundance).

Model predictor suite

A suite of spatial covariate layers of environmental data were used by the VECMAP [3] model tools to define statistical relationships with the variable to be modelled. This predictor suite 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 [39].
- Digital Elevation from the Shuttle Radar Topography Mission, together with derived aspect and ruggedness [40].
- Temporal Fourier Analysis (TFA) of Precipitation, and allied Bioclimatic Indicator (Bioclim) precipitation variables from the WORLDCLIM datasets [41].
- Length of Growing Period from United Nations Food and Agriculture Organisation [42].
- Travel Time to major towns from the Joint Research Centre at ISPRA [43].
- Human population density derived from the Global Rural Urban Mapping project at CEISIN [44].
- A distance weighted human population index layer [45] representing the likelihood of human visits based on the population within 30km.

Random Forest Spatial Modelling

Three measures of distribution/abundance were offered to the Random Forest module [46] using R-project [47] modules embedded within the VECMAP [3software. This flexible modelling framework can utilise either categorical or continuous input. In this case a presence absence (Boolean data) layer was chosen which resulted in: a probability surface output; a percentage of suitable habitat where presence is recorded, which resulted in a direct RF

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

Table 1: Reclass values defining the GLOBCOVER suitability layer for *Sus scrofa.*

regression continuous output; a classified boar abundance index, which resulted in a RF categorical model output.

Sampling strategy

Sample points were extracted for input into the three different Random Forest models from a 10km matrix defining each of the three input variables within known distributions. Overall there were ~12000 random points used across Europe. The following VECMAP [3] default sample parameters were used to define the Random Forest prediction for each of the models:

- Prediction forest forest size: 100.
- Prediction forest sample size: 90.
- Prediction forest node size: 7.

Quality Control

These models are a first attempt at quantifying the boar distribution at this scale and there has been no ground truth validation of these maps so far. All the model outputs l, however, satisfy standard accuracy metrics (R squared or Cohen's kappa coefficient where relevant) assuring statistical reliability. Model outputs have also been informally reviewed by project boar experts.

Constraints

There were no constraints involved in data production.

Privacy

N/A.

Ethics

N/A.

Research involving human participants should be approved by your institutional review board or equivalent committee(s) and that board must be named here. In addition, the research must have been conducted in accordance with the Declaration of Helsinki.

Non-human research on vertebrates must comply with institutional, national, or international guidelines, and where available should have been approved by an appropriate ethics committee.

4. Dataset description

Object name
euroboar.zip.

Data type
primary data, processed data, interpretation of data.

Ontologies
NONE.

Format names and versions
TIF, JPEG, JPEG2000, XML.

Creation dates
18/06/2015.

Dataset creators
As per author list.

Language
English.

Programming language
None.

Licence
CC0.

Accessibility criteria
All three layers have been provided as a quick look map in JPEG format to view from any image viewer.

The data itself are 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) [48] or R-project [47] raster package). If the user has no suitable software already installed the authors suggest downloading the open source QGIS software free of charge from http://www.qgis.org to view these data.

Repository location
http://dx.doi.org/10.6084/m9.figshare.1502662
Retrieved 12:12, Aug 05, 2015 (GMT).

Publication date
(05/08/2015).

5. Reuse potential

Wild boar is a large mammal and a species for which numbers and distribution are increasing in mainland Europe. The species' potential impact to environment, human activities and farming practices ensure the model outputs will be of interest to ecologists, human and animal health authorities and policy makers in a number of fields beyond that of the epidemiological goal of this study.

Competing Interests
The authors declare that they have no competing interests.

Acknowledgements
For identifying data & providing translations where required, Cornelia Silaghi, Gábor Földvári, Maria Kazimirova, Heidi Hauffe & Jonas Kindberg.

References

1. **Alexander, N S** and **Wint, W** 2015 Where are all the boars? An attempt to gain a continental perspective. Available at: http://geri2015.edenext.eu/content/download/4185/31369/version/1/file/P+2.17.pdf [Last accessed 18 June 2015].
2. **Massei, G, Kindberg, J, Licoppe, A, Gačić, D, Šprem, N, Kamler, J, Baubet, E, Hohmann, U, Monaco, A, Ozoliņš, J, Cellina, S, Podgórski, T, Fonseca, C, Markov, N, Pokorny, B, Rosell, C** and **Náhlik, A** 2015 Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Management Science*, 71: 492–500. DOI: http://dx.doi.org/10.1002/ps.3965. PMid: 25512181.
3. **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.
4. **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.
5. **The Global Biodiversity Information Facility (GBIF)** Available at: http://www.gbif.org/ [Last accessed 08 October 2014].
6. **IUCN** 2012 The IUCN Red List of Threatened Species. Version 2012.2. Available at: http://www.iucnredlist.org [Last accessed 13 June 2013].
7. **NBN Gateway.** Available at: http://data.nbn.org.uk [Last accessed 23 October 2014]. 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 nterpretation of this material, data and/or information.
8. **Spanish Ministry of Agriculture.** National Inventory of Biodiversity.
9. **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
10. **EDENext Data Management Team EDENext Data Portal.** Available at: http://www.edenextdata.com [Last accessed 18 June 2015].
11. **Acevedo, P, Quirós-Fernández, F, Casal, J** and **Vicente, J** 2014 Spatial distribution of wild boar population abundance: Basic information for spatial

epidemiology and wildlife management. *Ecological Indicators*, 36(2014): 594–600. DOI: http://dx.doi.org/10.1016/j.ecolind.2013.09.019

12. **Boitani, L, Trapanese, P** and **Mattei, L** 1995 Methods of population estimates of a hunted wild boar (Sus scrofa L.) population in Tuscany (Italy). *Ibex Journal of Mountain Ecology*, 3: 204–208.
13. **Bosch, J, Peris, S, Fonseca, C, Martinez, M, De La Torre, A, Iglesias, I** and **Munoz, M J** 2012 : Distribution, abundance and density of the wild boar on the Iberian Peninsula, based on the CORINE program and hunting statistic. *Folia Zool*, 61(2): 138–151.
14. **Dardaillon, M** 1986 Seasonal variations in habitat selection and spatial distribution of wild boar (Sus scrofa) in the Camargue, Southern France. *Behavioural Processes*, 13: 251–268. DOI: http://dx.doi.org/10.1016/0376-6357(86)90088-4
15. **EDEN FP7 Tick Borne Group Host Database** 2008 Available at www.edendatasite.com (last accessed 2008 no longer online).
16. **Fadeev, E V** 1973 Population dynamics of wild boar (Sus scrofa) in European Russia. *Zoologicheskii Zhurnal*, 52: 1214–1219 (in Russian with English summary).
17. **Fernandez-Llario, P, Matoes-Quesada, P M, Silverio, A** and **Santos, P** 2003 Habitat effects and shooting techniques on two wild boar (Sus scrofa) populations in Spain and Portugal. *Zeitschrift fur Jagdwissenschaft*, 49: 120–129. DOI: http://dx.doi.org/10.1007/bf02190452
18. **Fernandez-Llario, P, Parra, A, Cerrato, R** and **de Mendoza, J H** 2004 Spleen size variations and reproduction in a Mediterranean population of wild boar (Sus scrofa). *European Journal of Wildlife Research*, 50: 13–17. DOI: http://dx.doi.org/10.1007/s10344-003-0028-6
19. **Hebeisen, C, Fattebert, J, Baubet, E** and **Fischer, C** 2007 Estimating wild boar (Sus scrofa) abundance and density using capture–resites in Canton of Geneva, Switzerland. *Eur J Wildl Res*. DOI: http://dx.doi.org/10.1007/s10344-007-0156-5
20. **Herrero, J, Garcia-Serrano, A** and **Garcia-Gonzalez, R** 1995 Wild boar (Sus scrofa L.) hunting in South-Western Pyrenees (Spain): preliminary data. *Journal of Mountain Ecology*, 3: 228–229.
21. **Kanzaki, N, Perzanowski, K** and **Nowosad, M** 1998 Factors affecting wild boar (Sus scrofa) population dynamics in Bieszczady, Poland. *Gibier Faune Savage*, 15: 1171–1178. DOI: http://dx.doi.org/10.1046/j.1439-0450.1999.00214.x. PMid: 10085775.
22. **Kern, B, Depner, K R, Letz, W, Rott, M, Thaleheim, S, Nitschke, B, Plagemann, R** and **Liess, B** 1999 Incidence of classical swine fever (CSF) in wild boar in a densely populated area indicating CSF virus persistence as a mechanism for virus perpetuation. *Journal of Veterinary Medicine B*, 46: 63–67.
23. **Kozlo, P G** 1970 Factors determining the population dynamics of wild boar in Belovezhskyi forest. *Zoologicheskii Zhurnal*, 49: 422–430 (in Russian with English summary).
24. **Lavov, M A** 1981 Dinamika i regulirovanie chislennosti kabana w Berezinskom zapovednike [Dynamics and regulation of wild boar abundance in the Berezinskii Reserve]. Zapovedniki Belarusii, pp. 93–98. Izdatelstvo Uradzhai, Minsk (in Russian).
25. **Litvinov, V P** 1981 The wolf (Canis lupus) and wild boar (Sus scrofa) in the Kyzyl-Agach State Reserve. Zoologicheskii. PMCid: PMC2396115.
26. **Melis, C, Szafranska, P A, Jedrzejewska, B** and **Barton, K** 2006 Biogeographical variation in the population density of wild boar (Sus scrofa) in western Eurasia. *Journal of Biogeography*. DOI: http://dx.doi.org/10.1111/j.1365-2699.2006.01434.x
27. **Massei, G, Genov, P V, Staines, B W** and **Gorman, M L** 1997 Mortality of wild boar, Sus scrofa, in a Mediterranean area in relation to sex and age. *Journal of Zoology, London*, 242: 394–400. DOI: http://dx.doi.org/10.1111/j.1469-7998.1997.tb05813.x
28. **Mattioli, L, Apollonio, M, Mazzarone, V** and **Centofanti, E** 1995 Wolf food habits and wild ungulate availability in the Foreste Casentinesi National Park, Italy. *Acta Theriologica*, 40: 387–402. DOI: http://dx.doi.org/10.4098/AT.arch.95-36
29. **Monaco, A, Pedrotti, L** and **Franzetti, B** 1999 Population estimate using wild boar (Sus scrofa) harvest data: testing three different methods. Abstracts of the 24th Congress of International Union of Game Biologists, Thessaloniki, Greece, p. 87.
30. **Pucek, Z, Bobek, B, Labudzki, L, Miłkowski, L, Morow, K** and **Tomek, A** 1975 Estimates of density and number of ungulates. *Polish Ecological Studies*, 1: 121–135.
31. **Spitz, F** and **Janeau, G** 1990 Spatial strategies: an attempt to classify daily movements of wild boar. *Acta Theriologica*, 35: 129–149. DOI: http://dx.doi.org/10.4098/AT.arch.90-14
32. **Telleria, J L** and **Saez-Royuela, C** 1989 Ecologia de una poblacion iberica de lobos (Canis lupus). *Acta Vertebrata Donana*, 16: 105–122.
33. **Tupicina, L F** 1988 Dinamika chislennosti i razmeshchenie kabana (Sus scrofa) v Darvinskom Zapovednike [Population dynamics and distribution of wild boar (Sus scrofa) in the Darvinskii Reserve]. Populyacionnye issledovanya zhivotnykh v zapovednikakh [Populational research in reserves], pp. 128–139. Izdatelstwo Nauka, Moscow (in Russian).
34. **Deutscher Jagdverband, Handbuch (Germany)** 2014 Available at: http://www.jagdverband.de (Last accessed 18/06/2015).
35. **Csányi, S, Tóth, K, Kovács, I** és **Schally, G** (szerk.) 2014 Vadgazdálkodási Adattár – 2013/2014. Vadászati év. Országos Vadgazdálkodási Adattár, Gödöllö, 48pp: Hungarian Game Management Database 2013/2014. Available at: http://www.vmi.szie.hu/adattar/pdf/VA-2013-2014.pdf (Last accessed 18/06/2015).
36. **Poľovnícka štatistická ročenka Slovenskej republiky** 2013 National Forest Centre (Slovak Republic). Available at: http://www.forestportal.sk/sitepages/lesne_hospodarstvo/doc/Po%C4%BEovn%C3%ADcka%20

%C5%A1tatistick%C3%A1%20ro%C4%8Denka%20SR%202013.pdf (Last accessed 18/06/2015).

37. **Réseau Ongulés Sauvages ONCFS/FNC/FDC (France)** 2014. Available at: http://www.oncfs.gouv.fr/IMG/file/mammiferes/ongules/tableau/FS304_tableaux_chasse_ongules_2013_2014.pdf (Last accessed 18/06/2015).
38. **The Swedish Association for Hunting and Wildlife Management**, Wildlife Monitoring: Wild boar harvest per county for 2012/2013 courtesy Jonas Kindberg.
39. **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
40. **SRTM** 2012 The Shuttle Radar Topography Mission (SRTM) homepage. Available at: http://www2.jpl.nasa.gov/srtm/.
41. **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
42. **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.
43. **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.
44. **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].
45. **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
46. **R-Project randomForest package** 2012 randomForest: Breiman and Cutler's random forests for classification and regression. Available at: http://cran.rproject.org/web/packages/randomForest/index.hml.
47. **R Core Team** 2012 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.Rproject.org/.
48. **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, Massei, G and Wint, W 2016 The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster – Where are All the Boars? An Attempt to Gain a Continental Perspective. *Open Health Data* 4: e1, DOI: http://dx.doi.org/10.5334/ohd.24

Published: 28 January 2016