DATA PAPER

VBORNET Gap Analysis: Mosquito Vector Distribution Models Utilised to Identify Areas of Potential Species Distribution in Areas Lacking Records

Francis Schaffner[1], Veerle Versteirt[2], Wim Van Bortel[3], Herve Zeller[3], William Wint[4] and Neil Stewart Alexander[4]

[1] Francis Schaffner Consultancy, Riehen, Switzerland; & University of Zurich, Zurich, Switzerland

[2] Avia-GIS, Zoersel, Belgium

[3] European Centre for Disease Prevention and Control, Stockholm, Sweden

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

Corresponding author: Neil Stewart Alexander (neil.alexander@zoo.ox.ac.uk)

This is the second of a number of planned data papers presenting modelled vector distributions produced originally during the ECDC funded VBORNET project. This work continues under the VectorNet project now jointly funded by ECDC and EFSA. Further data papers will be published after sampling seasons when more field data will become available allowing further species to be modelled or validation and updates to existing models.

The data package described here includes those mosquito species first modelled in 2013 & 2014 as part of the VBORNET gap analysis work which aimed to identify areas of potential species distribution in areas lacking records. It comprises three species models together with suitability masks based on land class and environmental limits. The species included as part of this phase are the mosquitoes *Aedes vexans, Anopheles plumbeus* and *Culex modestus*.

The known distributions of these species within the area covered by the project (Europe, the Mediterranean Basin, North Africa, and Eurasia) are currently incomplete to a greater or lesser degree. The models are designed to fill the gaps with predicted distributions, to provide a) assistance in targeting surveys to collect distribution data for those areas with no field validated information, and b) a first indication of the species distributions within the project areas.

1. Overview

Introduction/Study Description

VBORNET [1] was an initiative of the European Centre for Disease Prevention and Control (ECDC), which ran from 2009 to 2014. The project established a European network of entomological and public health specialists in order to assist ECDC in its preparedness activities on vector-borne diseases. As part of this work a database collating validated records of key vector species distributions was commissioned. In this data paper we describe work done on *Aedes vexans* and *Culex modestus*, both vectors of West Nile fever virus, and *Anopheles plumbeus*, a potential vector of malaria parasite.

VectorNet [2] is continuing this work and builds upon VBORNET by supporting the collection of data on vectors and pathogens in vectors, related to both animal and human health. VectorNet is a joint initiative of the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC), which started in May 2014.

Whilst VBORNET and VectorNet have made substantial progress collating European data on key vector species, the coverage is still incomplete. The 'Gap Analysis' work within these projects aims to identify those areas of likely species distribution within the project extent where there are no current data. These estimates produced by spatial modelling techniques are intended to meet two objectives: firstly to help direct extensive VectorNet sampling efforts in the field; and secondly to provide first indications of the current likely extent and distribution of key vector species within continental Europe and its surrounding regions. It is hoped that publishing these models will aid

the VectorNet network of experts to engage the wider research and professional community in the drive to expand and validate the VectorNet database. Readers are encouraged to visit the VectorNet website [2] or to directly contact the authors to report complementary data.

For each species, probability of presence maps at the resolution of 1 km were generated using a variety of well-established spatial modelling techniques available through the VECMAP system [3]. Both the input data and the resulting models were iteratively assessed by project experts and the best performing are included in this data package.

2. Context

Spatial coverage
Description: Continental Europe
Northern boundary: 71.8
Southern boundary: 33.5
Eastern boundary: 62.3
Western boundary: −19.2

Temporal coverage
Known presence up to 31/01/2013.

Species
The inland floodwater mosquito, *Aedes vexans*, vector of West Nile fever and Rift Valley fever viruses.

Anopheles plumbeus, a potential vector of *Plasmodium falciparum*.

Culex modestus, a vector of West Nile fever virus.

3. Methods
For each of the species the following method was followed.

Steps
Identifying presence and absence training data
The reported distributions of each of the three mosquito species by VBORNET were used as the basis for species present training data. Data reported from the VBORNET map published January 2013 were utilised for *Aedes vexans* and *Culex modestus*. Data reported from January 2014 were used for *Anopheles plumbeus*. Maps of the recorded known distributions at that time are presented in Appendix 1 available within this data package. These reported distributions were recorded in VBORNET at a coarse NUTS 3 polygon scale. The data originates from a combination of aggregated data contributed by the authors and listed contributors, as well as a literature review completed by the VBORNET vector group leaders. The full data set and sources are available to contributors of VBORNET and VectorNet.

Habitat suitability and environmental limits
The recorded distribution at a NUTS 3 scale was too coarse to be utilised by the model framework. In addition, the selected modelling methods required information on both presence and absence to calibrate the modelling process. It was therefore necessary to identify areas of absence within NUTS 3 regions assigned as present. To do this a suitability mask at 1 km resolution was compiled by requesting experts within the network (see the Data Creators section) to identify primary, secondary and unsuitable land cover classes. Where available environmental limiting factors such as altitude or precipitation limits which are derived from remotely sensed imagery. Land cover masks were defined using the 100 m Corine land cover dataset [4] and the 300 m GLOBCOVER [5] product where no Corine data was available. Definitions of land class suitability for each species as defined by experts can be found in **Tables 1** and **2**.

Modelling procedure
A range of modelling techniques available in the VECMAP™ [3] system including Non Linear Discriminant Analysis [6], Logistic Regression [7] and Random Forests [8], using 10–25 repeated bootstraps per run, were used to provide a range of outputs for expert assessment.

The covariates offered to the modelling procedures were drawn from a standardised set of ecological parameters, and in particular a suite of Fourier processed MODIS satellite imagery [9] which provides a range of biologically interpretable variables related to levels and seasonality of temperature and vegetation related factors during the period 2000–2012. These are summarised in **Table 3**, and are all available to registered members of the VMerge/EDENext Data Website (www.vmergedata.com) [10].

Output layers
The suitability masked modelled outputs are produced in the form of probability maps at the pixel level with a resolution of 1 kilometre for each species. Quick view for each vector species is available in Appendix 2 available within this data package.

Sampling strategy
Training sample point data for the model was extracted as follows:

- Random present points were created from any area within a NUTS 3 polygon recorded as present and where the suitability masked did not indicate unsuitability.
- Random absence points were selected areas from identified in the mask as unsuitable.

Quality Control
The model outputs were initially evaluated using the standard, and extensive, accuracy metrics (e.g. R-squared, AIC, Kappa, Confusion matrices) provided by the VECMAP™ [3] software. Providing the accuracy metrics indicated sufficient statistical reliability.

The range of models were then sent to the relevant experts who were asked to choose from the selection provided. These included paper authors themselves and individuals listed in the Data Creator section of this paper. This feedback is critical as experts can comment

CORINE LABEL	AEVE	ANPL	CUMO
Continuous urban fabric	0	0	0
Discontinuous urban fabric	0	0	0
Industrial or commercial units	0	0	0
Road and rail networks and associated land	0	0	0
Port areas	0	0	0
Airports	0	0	0
Mineral extraction sites	0	0	0
Dump sites	0	0	0
Construction sites	0	0	0
Green urban areas	1	1	0
Sport and leisure facilities	0	1	0
Non-irrigated arable land	0	0	0
Permanently irrigated land	0	0	1
Rice fields	1	0	1
Vineyards	0	0	0
Fruit trees and berry plantations	0	0	0
Olive groves	0	1	0
Pastures	1	0	0
Annual crops associated with permanent crops	0	0	0
Complex cultivation patterns	0	0	0
Land principally occupied by agriculture, with significant areas of natural vegetation	1	1	1
Agro-forestry areas	1	1	1
Broad-leaved forest	1	1	0
Coniferous forest	0	0	0
Mixed forest	1	1	0
Natural grasslands	1	0	0
Moors and heathland	1	0	0
Sclerophyllous vegetation	0	0	0
Transitional woodland-shrub	0	1	0
Beaches, dunes, sands	0	0	0
Bare rocks	0	0	0
Sparsely vegetated areas	0	1	0
Burnt areas	0	0	0
Glaciers and perpetual snow	0	0	0
Inland marshes	1	0	1
Peat bogs	1	0	0
Salt marshes	0	0	0
Salines	0	0	0
Intertidal flats	0	0	0
Water courses	0	0	0
Water bodies	0	0	0
Coastal lagoons	0	0	0
Estuaries	1	0	1
Sea and ocean	0	0	0

Table 1: Reclassed values defining the Corine suitability layers 1 = suitable and 0 = unsuitable. AEVE = *Aedes vexans*, ANPL = *Anopheles plumbeus* and CUMO = *Culex modestus*.

GLOBCOVER LABEL	AEVE	ANPL	CUMO
Post-flooding or irrigated croplands (or aquatic)	1	0	1
Rainfed croplands	0	0	0
Mosaic cropland (50–70%)/vegetation (grassland/shrubland/forest) (20–50%)	1	0	1
Mosaic vegetation (grassland/shrubland/forest) (50–70%)/cropland (20–50%)	1	1	1
Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	1	1	0
Closed (>40%) broadleaved deciduous forest (>5m)	1	1	0
Open (15–40%) broadleaved deciduous forest/woodland (>5m)	0	1	0
Closed (>40%) needleleaved evergreen forest (>5m)	0	0	0
Open (15–40%) needleleaved deciduous or evergreen forest (>5m)	1	1	0
Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	1	1	0
Mosaic forest or shrubland (50–70%)/grassland (20–50%)	1	0	1
Mosaic grassland (50–70%)/forest or shrubland (20–50%)	1	0	1
Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	0	0	0
Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	1	0	0
Sparse (<15%) vegetation	0	0	0
Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) – Fresh or brackish water	1	0	1
Closed (>40%) broadleaved forest or shrubland permanently flooded – Saline or brackish water	0	1	0
Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil – Fresh, brackish or saline water	1	0	1
Artificial surfaces and associated areas (Urban areas >50%)	0	1	0
Bare areas	0	0	0
Water bodies	0	0	1
Permanent snow and ice	0	0	0
No data (burnt areas, clouds, . . .)	0	0	0

Table 2: Reclassed values defining the Globcover suitability layers 1 = suitable and 0 = unsuitable. AEVE = *Aedes vexans*, ANPL = *Anopheles plumbeus* and CUMO = *Culex modestus*.

further on how the maps compare to species prevalence on the ground. This can look very different from the presence/absence picture reported at NUTS 3 polygons by VBORNET. This is most obvious in areas such as central Spain where the hot arid environment means large areas may be unsuitable for certain vector species. But presence can be recorded from suitable microenvironments which registers a strong visual signature in that area on the VBORNET present/absence maps. On these occasions we use the expert opinion to validate where we set the environmental limits we refer to earlier in the paper.

In the first phase of modelling (*Aedes vexans* and *Culex modestus*) the best model selected by the experts was used as the final model for that species. During phase 2 of the modelling (*Anopheles plumbeus*), Ensembles of the different model techniques were preferred to attempt to iron out any inherent bias within individual modelling methods. Naturally if a model was not approved by the network experts it was not included in the ensemble.

Ground truthing has yet to be completed on these models although fieldwork has been subsequently sponsored by the VectorNet project which will visit areas which have been modelled, but currently have no data available. So retrospective quality assessments should be completed in the future.

Constraints
There were no constraints in the data production.

Privacy
N/A

Ethics
N/A

1 ED1803A0: Middle infra-red mean
2 ED1803A1: Middle infra-red amplitude 1
3 ED1803A2: Middle infra-red amplitude 2
4 ED1803A3: Middle infra-red amplitude 3
5 ED1803MN: Middle infra-red minimum
6 ED1803MX: Middle infra-red maximum
7 ED1803P1: Middle infra-red phase 1
8 ED1803P2: Middle infra-red phase 2
9 ED1803P3: Middle infra-red phase 3
10 ED1803VR: Middle infra-red variance
11 ED1807A0: Daytime LST mean
12 ED1807A1: Daytime LST amplitude 1
13 ED1807A2: Daytime LST amplitude 2
14 ED1807A3: Daytime LST amplitude 3
15 ED1807MN: Daytime LST minimum
16 ED1807MX: Daytime LST maximum
17 ED1807P1: Daytime LST phase 1
18 ED1807P2: Daytime LST phase 2
19 ED1807P3: Daytime LST phase 3
20 ED1807VR: Daytime LST variance
21 ED1808A0: Nighttime LST mean
22 ED1808A1: Nighttime LST amplitude 1
23 ED1808A2: Nighttime LST amplitude 2
24 ED1808A3: Nighttime LST amplitude 3
25 ED1808MN: Nighttime LST minimum
26 ED1808MX: Nighttime LST maximum
27 ED1808P1: Nighttime LST phase 1
28 ED1808P2: Nighttime LST phase 2
29 ED1808P3: Nighttime LST phase 3
30 ED1808VR: Nighttime LST variance
31 ED1814A0: NDVI mean
32 ED1814A1: NDVI amplitude 1
33 ED1814A2: NDVI amplitude 2
34 ED1814A3: NDVI amplitude 3
35 ED1814MN: NDVI minimum
36 ED1814MX: NDVI maximum
37 ED1814P1: NDVI phase 1

38 ED1814P2: NDVI phase 2
39 ED1814P3: NDVI phase 3
40 ED1814VR: NDVI variance 41 ED1815A0: EVI mean
42 ED1815A1: EVI amplitude 1
43 ED1815A2: EVI amplitude 2
44 ED1815A3: EVI amplitude 3
45 ED1815MN: EVI minimum
46 ED1815MX: EVI maximum
47 ED1815P1: EVI phase 1
48 ED1815P2: EVI phase 2
49 ED1815P3: EVI phase 3
50 ED1815VR: EVI variance
51 EDBC2K12: BioClim Annual Precipitation
52 EDBC2K13: BioClim Precipitation of Wettest Month
53 EDBC2K14: BioClim Precipitation of Driest Month
54 EDBC2K15: BioClim Precipitation Seasonality (Coefficient of Variation)
55 EDBC2K16: BioClim Precipitation of Wettest Quarter
56 EDBC2K17: BioClim Precipitation of Driest Quarter
57 EDBC2K18: BioClim Precipitation of Warmest Quarter
58 EDBC2K19: BioClim Precipitation of Coldest Quarter
59 EDV590AS: DEM (Aspect)
60 EDV590EL: DEM (Elevation)
61 EDV590RG: DEM (Ruggedness)
62 EDWC57A0: WORLDCLIM precipitation mean
63 EDWC57A1: WORLDCLIM precipitation amplitude 1
64 EDWC57A2: WORLDCLIM precipitation amplitude 2
65 EDWC57A3: WORLDCLIM precipitation amplitude 3
66 EDWC57MN: WORLDCLIM precipitation minimum
67 EDWC57MX: WORLDCLIM precipitation maximum
68 EDWC57P1: WORLDCLIM precipitation phase 1
69 EDWC57P2: WORLDCLIM precipitation phase 2
70 EDWC57P3: WORLDCLIM precipitation phase 3
71 EDWC57VR: WORLDCLIM precipitation variance
72 EDXXGRPD: GRUMP Population density
73 EDXXGRPW: GRUMP Population weighted
74 EDXXJRCA: JRC Access
75 EDXXLPG1: Length of Growing Period LGP

LST = Land Surface Temperature. NDVI Normalised Difference vegetation Index; EVI Enhanced Vegetation Index. DEM Digital Elevation. All files starting with ED18 are Fourier processed MODIS Satellite Imagery produced by the TALA research Group Oxford [9].
Files with Bioclim and Worldclim in filename derived from WORLDCLIM datasets [11].
GRUMP derived from population layers produced by [12].
JRC Accessibility downloaded from [13].
Length of growing Period derived from data provided by FAO, Rome. Available from www.vmerge.com [10].
All layers extracted and standardised by ERGO for EDENext (www.edenextdata.com) [14].

Table 3: Covariates offered to modelling procedures.

4. Dataset description
Object name
https://dx.doi.org/10.6084/m9.figshare.3426068

Data type
Processed data; Interpretation of data.

Ontologies
N/A

Format names and versions
JPG. JP2, TIF, TFW, XML.

Creation dates
(13/04/2013).

Dataset creators
The following table lists VBORNET contributors who directly contributed to the VBORNET Mosquito database that was used as training data in the models presented in this data paper. While Francis Schaffner's extensive experience of research in the field of mosquitoes in and around Europe were extremely useful in the land cover suitability exercise and when assessing the maps and the success of the model outputs.

Contributor	Affiliation
Albieri, Alessandro	Centro Agricoltura Ambiente "Giorgio Nicoli", Bologna, Italy
Alten, Bulent	Hacettepe University, Ankara, Turkey
Alves, Maria Joao	Minesterio da Saude, Lisbon, Portugal
Antunes, Ana	Faculdade de Medicina Veterinária – Universidade de Lisboa, Lisbon, Portugal
Aranda, Carles	Consell Comarcal del Baix Llobregat, Servei de Control de Mosquits, Barcelona, Spain
Beeuwkes, Jacob	Laboratory of Entomology, Wageningen, The Netherlands
Bødker, Rene	National Veterinary Institute (DTU), Frediksberg, Denmark
Bucher, Edith	Biological Laboratory, Laives, Italy
Bueno Mari, Ruben	Laboratorios Lokímica, Valencia, Spain
Collantes, Francisco	Universidad de Murcia, Murcia, Spain
Dikolli, Enkelejda	Institute of Public Health, Tirana, Albania
Eritja, Roger	Consell Comarcal del Baix Llobregat – Servei de Control de Mosquits, Barcelona, Spain
Falcuta, Elena	Cantacuzino Institute, Bucharest, Romania
Fontenille, Didier	IRD/Directeur de l'Institut Pasteur du Cambodge, Cambodge
Gewehr, Sandra	Ecodevelopment, Thessaloniki, Greece
Gunay, Filiz	Hacettepe University, Ankara, Turkey
Hristovski, Slavco	Faculty of Natural Sciences and Mathematics, Skopje, Macedonia
Hufnagl, Peter	Austrian Agency for Health and Food Safety (AGES), Vienna, Austria
Ibañez-Justicia, Adolfo	Centre for Monitoring of Vectors, Wageningen, the Netherlands
Kalan, Katja	University of Primorska, Koper, Slovenia
Kampen, Helge	Friedrich-Loeffler-Institut, Greifswald – Insel Riems, Germany
Kavur, Hakan	Cukurova University, Dept of Medical Parasitology, Adana, Turkey
Klobucar, Ana	Institute of public health "Dr. Andrija Stampar", Zagreb, Croatia
Krüger, Andreas	Berhard Nocht Institut für Tropenmedizin, Hamburg, Germany
Medlock, Jolyon	Public Health England, Porton Down, UK
Miranda Chueca, Miguel Angel	University of the Balearic Islands, Department of Biology, Palma de Mallorca, Mallorca
Montalvo, Tomas	Agència de Salut Pública de Barcelona, Barcelona, Spain
Mosca, Andrea	IPLA, Turin Area, Italy
Ognyan, Mikov	National Centre of Infectious and Parasitic Diseases, Parasitology and Tropical Medicine, Sofia, Bulgaria
Pajovic, Igor	University of Montenegro, Biotechnical Faculty, Montenegro
Perrin, Yvon	Centre National d'Expertise sur les Vecteurs, Montpellier, France
Petrić, Dusan	Faculty of Agriculture, University of Novi Sad, Serbia
Piazzi, Mauro	IPLA, Turin Area, Italy
Plenge-Bönig, Anita	Div. Hygiene and Infectious Diseases, Institute for Hygiene and Environment of the City of Hamburg, Hamburg, Germany
Prioteasa, Liviu	Cantacuzino Institute, Bucharest, Romania
Regan, Eugenie	National Biodiversity Data Centre, Ireland
Sousa, Carla A.	Instituto de Higiene e Medicina Tropical, Lisbon, Portugal
Sulesco, Tatiana	Academy of Sciences of Moldova, Chisinau, Moldova
Walder, Gernot	Medizinische Universität Innsbruck, Division of Hygiene and Medical Microbiology, Innsbruck, Austria
Zamburlini, Renato	University of Udine, Dept. of Agricultural and Environmental Science, Udine, Italy
Zygutiene, Milda	Centre for Communicable diseases and AIDS, Vilnius, Lithuania

Language
English.

Programming language
N/A

Licence
The open licence under which the data has been deposited CC-BY.

Accessibility criteria
The data are distributed as GIS raster GeoTIFF formats. Which is a standard proprietary GIS raster format. To access and analyse the raster data directly GeoTIFFs 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). 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.

A simple schematic of the data layers and directories found within this data package is shown below with descriptions where filenames are not self-explanatory:

- **Appendices** – Directory containing the appendices for this document.
  ◦ **ohd_VBNMBD_SchaffnerEtAl_Appendix1. pdf**
  ◦ **ohd_VBNMBD_SchaffnerEtAl_Appendix2. pdf**
- **Quickview** – Directory containing small JPEG files allowing the reader to view the data visually without specialist software.
  ◦ **appendix1mapsAEVE.jpg** – VBORNET Status *Aedes vexans*
  ◦ **appendix1mapsANPL.jpg** – VBORNET Status *Anopheles plumbeus*
  ◦ **appendix1mapsCUMO.jpg** – VBORNET *Culex modestus*
  ◦ **appendix2mapsAEVE.jpg** – VBORNET Status *Aedes vexans*
  ◦ **appendix2mapsANPL.jpg** – Model output *Anopheles plumbeus*
  ◦ **appendix2mapsCUMO.jpg** – Model output *Culex modestus*
- **Tiff** – Directory containing model output data for display and interrogation within GIS and geostatistical software.*
  ◦ **aevemodelMsk.tif** – Model output *Aedes vexans*
  ◦ **anplMskensNFL.tif** – Model output *Anopheles plumbeus*
  ◦ **cumomodelMsk.tif** – Model output *Culex modestus*

*Only the tif files within this directory are listed. Other file formats of the same name within the directory are ancillary files that provide additional data to the GIS software and as a rule should be copied along with the TIFF file of the same name if you are moving the data between directories.

Repository location
https://dx.doi.org/10.6084/m9.figshare.3426068

Publication date
(23/08/2016).

5. Reuse potential

These layers have been created in an attempt to identify probable areas of species distribution where there are currently no sample data. These maps therefore could be useful in identifying suitable areas for further sampling in an attempt to identify the true distribution of the species. The VectorNet project [2] plans to utilise these datasets in such a way.

The covariates of the models are also mainly climate orientated. A possible avenue of further work therefore could be to use the models to assess the potential change in distribution after a shift in climate parameters.

Additional Files
The additional files for this article can be found as follows:

- **Additional File 1: Appendix 1.** VBORNET Reported Distributions 2013 & 2014. http://dx.doi. org/10.5334/ohd.27.s1
- **Additional File 2: Appendix 2.** VBORNET model output quick view. http://dx.doi.org/10.5334/ ohd.27.s2

Competing Interests
The authors have no competing interests to declare.

References
1. **VBORNET Website** Available from: http://www. vbornet.eu/.
2. **VectorNet Website** Available from: http://ecdc. europa.eu/en/healthtopics/vectors/VectorNet/ Pages/VectorNet.aspx.
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. **Corine Land Cover** 2006 Raster data. Available from: http://www.eea.europa.eu/data-and-maps/ data/corine-land-cover-2006-raster-1.
5. **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
6. **Rogers, D J** 2006 Models for vectors and vector-borne diseases. *Advances in Parasitology*, 62: 1–35. DOI: https://doi.org/10.1016/S0065-308X(05)62001-5

7. **Lindsey, J K** 2007 Applying Generalized Linear Models. Springer, New York.

8. **Breiman, L** 2001 Random Forests. *Machine Learning*, 45: 5–32. DOI: https://doi.org/10.1023/A:1010933404324

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. **VMERGE Data Portal Website** Available from: http://www.vmergedata.com

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: 1965–1978. DOI: https://doi.org/10.1002/joc.1276

12. **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 [accessed 02/03/2016].

13. **Nelson, A** (2008) Estimated travel time to the nearest city of 50,000 or more people in year 2000. Global Environment Monitoring Unit – Joint Research Centre of the European Commission, Ispra Italy. Available at: http://forobs.jrc.ec.europa.eu/products/gam/ (accessed 02/03/2016)

14. **EDENext Data Portal** Available from http://www.edenextdata.com