PALE-Blu Understanding pathogen, livestock, environment interactions involving bluetongue

Horizon 2020

Project ID: *727393*

Work Package 4: December 2019

Task 4.5: Refine and update existing *Culicoides* host maps**. Deliverable 4.3.** *Culicoides host maps*

Summary: *Culicoides* hosts in Europe include domestic livestock and deer. Distribution data have been sourced and acquired to enable the revision of earlier versions of spatial distribution models for the densities of sheep and goats. Data have also been found for Sika deer, which have not been modelled before. All model outputs can be downloaded by registered members of the PALE-Blu spatial data site http://www.palebludata.com. Efforts will continue throughout the rest of the project to monitor the data available for other host species – including Red, Roe and Fallow deer, as well as Cattle, Horses and Buffaloes and Camels. Future priority focus is likely to be on dairy cattle rather than cattle as a whole.

The proposal for Work Package 4 sets out the following: *"Culicoides* host distributions are needed to help define the determinants of vector distributions. The starting point of these maps will be the outputs of the FP7 VMERGE project – namely for domestic ruminants, horses, and some deer species. These will be updated using new field distribution data as they become available in the public domain, to feed spatial distribution models produced by Random Forest and Boosted Regression Tree methods as appropriate. The availability of training data for other host species will also be investigated and spatial distribution models produced as permitted by the data available".

ERGO maintains a constant watching brief for *Culicoides* domestic and wildlife host distribution data. These include cattle, sheep, goats, camels, and horses as well as the main species of deer found in Europe – Red Deer (*Cervus elephus*), Roe deer (*Capreolus capreolus*), Fallow Deer (*Dama dama*) and Sika Deer (*Cervus nippon*). Muntjac (*Muntiacus reevesi*) has also been introduced into Europe, but is only widespread in mainland Great Britain, and is thus not significantly implicated in BTV occurrence.

Distributions for all these hosts except Sika were produced during the FP7 VMerge Project. In the two or three years since these models were produced, some new data have been made available: notable for cattle, sheep and goats thought the Food and Agriculture Organisation's Gridded Livestock of the World (GLW) project(with which ERGO is a founding collaborator). Distribution data has also become available for the Sika Deer – most importantly from the Global Biodiversity Information Facility (www.gbif.org) to which many national distribution databases contribute. No new significant data have been located for the other thee deer species, or for horses and camels.

Given these newly available data, it was decided to revise the existing distribution models for sheep and goats and to produce a new model for the Sika Deer, for which no models are yet. It was decided to delay any revision for Cattle distributions as the project requirements would be better served by focussing on Dairy Cattle rather than cattle numbers as a whole.

The deer models are only produced for Western Europe and the British Isles as they are absent from the rest of the Continent. The small ruminant models are produced at a resolution of 1 kilometre, for Europe, its neighbouring countries to the East, and the countries in the Middle East and North Africa that border the Mediterranean. All use the same well-established methodology which essentially establishes statistical relationships between the values of the target host parameter and a wide range of predictor covariates at a series of sample points covering the full geographic extent of the area to be modelled. These sample points are distributed in a stratified random fashion through the model extent at approximately four points for every 50 km square. The example for the livestock models is shown in Figure 1. These relationships are then used to calculate predicted values of the target parameters for every square kilometre within the modelled region. The sheep and goat values modelled were ln (density per square kilometre +1). The predicted logged values were converted to back to simple densities for display. Population numbers were not available for the Sika deer and so only its presence or absence was modelled, with the output presented as probability of presence.

Two methodologies are used to implement this process - Boosted Regression Tress and Random Forest techniques [1,2] implemented though the VECMAP software suite [3]. VECMAP provides a series of accuracy metrics for each method. Models are only 'adopted' if the AUC is at least 0.85. The models from the two methods are then ensembled to provide a 'consensus' output.

Figure 1. Sample points for the livestock models

The covariates used to drive the models are selected to represent likely drivers of host density namely human population, land use and land cover, elevation, temperature, rainfall and vegetation cover. These are shown in Table xx below. The climate and vegetation parameters are derived from a 15 year time series of MODIS satellite imagery using Temporal Fourier Analysis, which converts periodic (dekadal or monthly) imagery into a series of biologically relevant metrics relating to the mean levels, peaks and seasonality of the parameters concerned [1]. Maintenance and updating this entire predictor data suite has been another focus of ERGO's activities within PALE-Blu (Deliverables 3.6 and 12.4) which will continue throughout the project lifetime.

The outputs are shown in Figures 2, 3 and 4. The training data are inset to the top right of each graphic. The model outputs are available on the PaleBludata [4] website for all registered members of the site.

A new version of GLW is in the pipeline which may be able to provide sufficient dairy cattle training data for European subnational admin units. If so these will be used to provide a new dairy cattle density distribution. If not, the existing total cattle density layer will be revised, most probably during 2020.

Table 1: Covariates offered to modelling procedures.

1 ER011503A0: Middle infra-red mean 2 ER011503A1: Middle infra-red amplitude 1 3 ER011503A2: Middle infra-red amplitude 2 4 ER011503A3: Middle infra-red amplitude 3 5 ER011503MN: Middle infra-red minimum 6 ER011503MX: Middle infra-red maximum 7 ER011503P1: Middle infra-red phase 1 8 ER011503P2: Middle infra-red phase 2 9 ER011503P3: Middle infra-red phase 3 10 ER011503VR: Middle infra-red variance 11 ER011507A0: Daytime LST mean 12 ER011507A1: Daytime LST amplitude 1 13 ER011507A2: Daytime LST amplitude 2 14 ER011507A3: Daytime LST amplitude 3 15 ER011507MN: Daytime LST minimum 16 ER011507MX: Daytime LST maximum 17 ER011507P1: Daytime LST phase 1 18 ER011507P2: Daytime LST phase 2 19 ER011507P3: Daytime LST phase 3 20 ER011507VR: Daytime LST variance 21 ER011508A0: Nighttime LST mean 22 ER011508A1: Nighttime LST amplitude 1 23 ER011508A2: Nighttime LST amplitude 2 24 ER011508A3: Nighttime LST amplitude 3 25 ER011508MN: Nighttime LST minimum 26 ER011508MX: Nighttime LST maximum 27 ER011508P1: Nighttime LST phase 1 28 ER011508P2: Nighttime LST phase 2 29 ER011508P3: Nighttime LST phase 3 30 ER011508VR: Nighttime LST variance 31 ER011514A0: NDVI mean 32 ER011514A1: NDVI amplitude 1 33 ER011514A2: NDVI amplitude 2 34 ER011514A3: NDVI amplitude 3 35 ER011514MN: NDVI minimum 36 ER011514MX: NDVI maximum 37 ER011514P1: NDVI phase 1	38 ER011514P2: NDVI phase 2 39 ER011514P3: NDVI phase 3 40 ER011514VR: NDVI variance 41 ER011515A0: EVI mean 42 ER011515A1: EVI amplitude 1 43 ER011515A2: EVI amplitude 2 44 ER011515A3: EVI amplitude 3 45 ER011515MN: EVI minimum 46 ER011515MX: EVI maximum 47 ER011515P1: EVI phase 1 48 ER011515P2: EVI phase 2 49 ER011515P3: EVI phase 3 50 ER011515VR: EVI variance 51 EDV590EL: DEM (Elevation) 52 EDV590RG: DEM (Ruggedness) 53 ERPRECA0: WORLDCLIM precipitation mean 54 ERPRECA1: WORLDCLIM precipitation amplitude 1 55 ERPRECA2: WORLDCLIM precipitation amplitude 2 56 ERPRECA3: WORLDCLIM precipitation amplitude 3 57 ERPRECMN: WORLDCLIM precipitation minimum 58 ERPRECMX: WORLDCLIM precipitation maximum 59 ERPRECP1: WORLDCLIM precipitation phase 1 60 ERPRECP2: WORLDCLIM precipitation phase 2 61 ERPRECP3: WORLDCLIM precipitation phase 3 62 ERPRECVR: WORLDCLIM precipitation variance 63 ERXXGRPD: GRUMP Human Population density 64 ERV59EL500: SRTM Elevation 65 EREELCBARE: consensus % bare ground 66 EREELCDCBD: consensus % deciduous broadleaved forest 67 EREELCEVBD: consensus % evergreen broadleaved forest 68 EREELCEVBD: consensus % evergreen needleleaved forest 69 EREELCFLD: consensus % flooded 70 EREELCHERB: consensus % herbaceous cover 71 EREELCMANG: consensus % managed land 72 EREELCOTR: consensus % other land cover 73 EREELCSHR: consensus % shrub cover 74 EREELCURB: consensus % urban 75 EREELCSNOW: consensus % snow 76 EREELCWAT: consensus % water
LST = Land Surface Temperature. NDVI Normalised Difference vegetation Index; EVI Enhanced Vegetation Index. DEM Digital Elevation. All files starting with ER0115 are Fourier processed MODIS Satellite Imagery produced by the Environmental Research Group Oxford. [4] GRUMP derived from population layers produced by [6]. All Files with EREELC in file name were derived from the Earthenv consensus land cover data product [7] All layers extracted and standardised by ERGO for PALEBLU (www.palebludata.com) [5]	

Figure 2. Modelled Goat Density (number per square kilometre). Training data inset top right

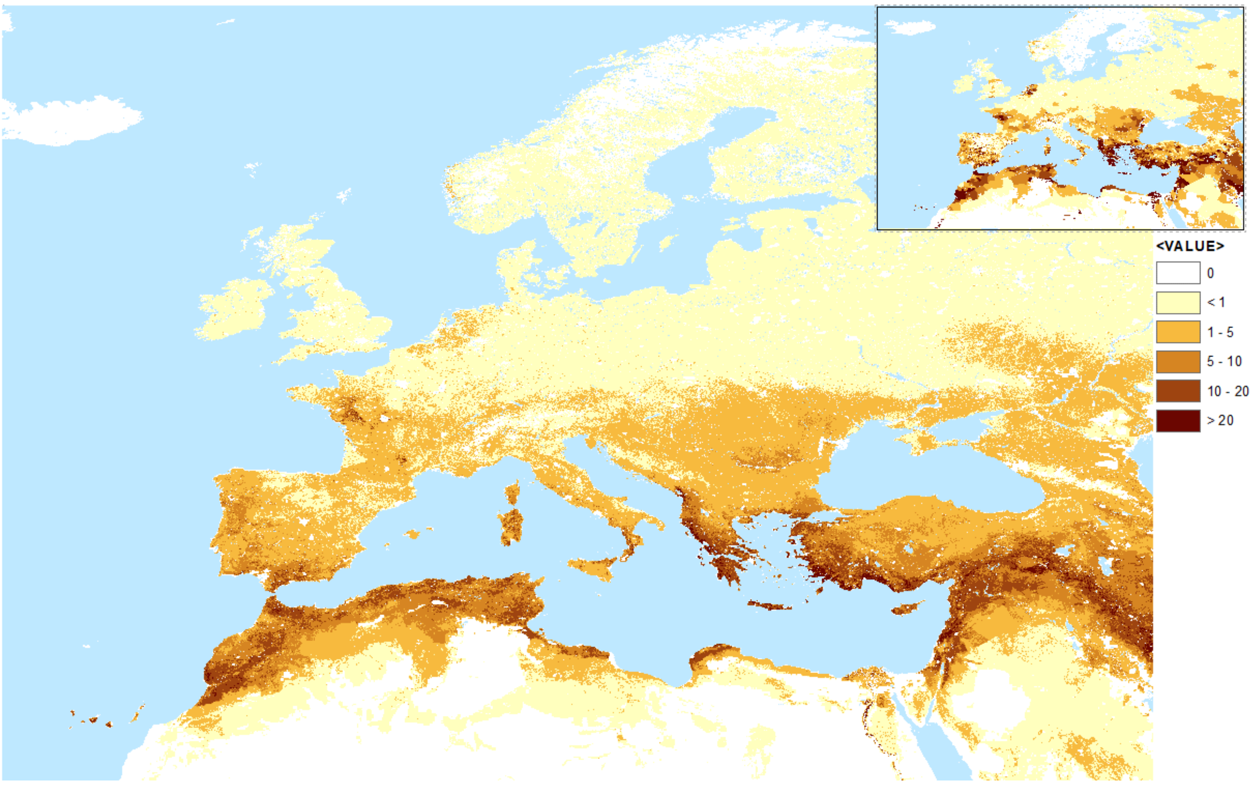


Figure 3. Modelled Sheep Density (number per square kilometre). Training data inset top right

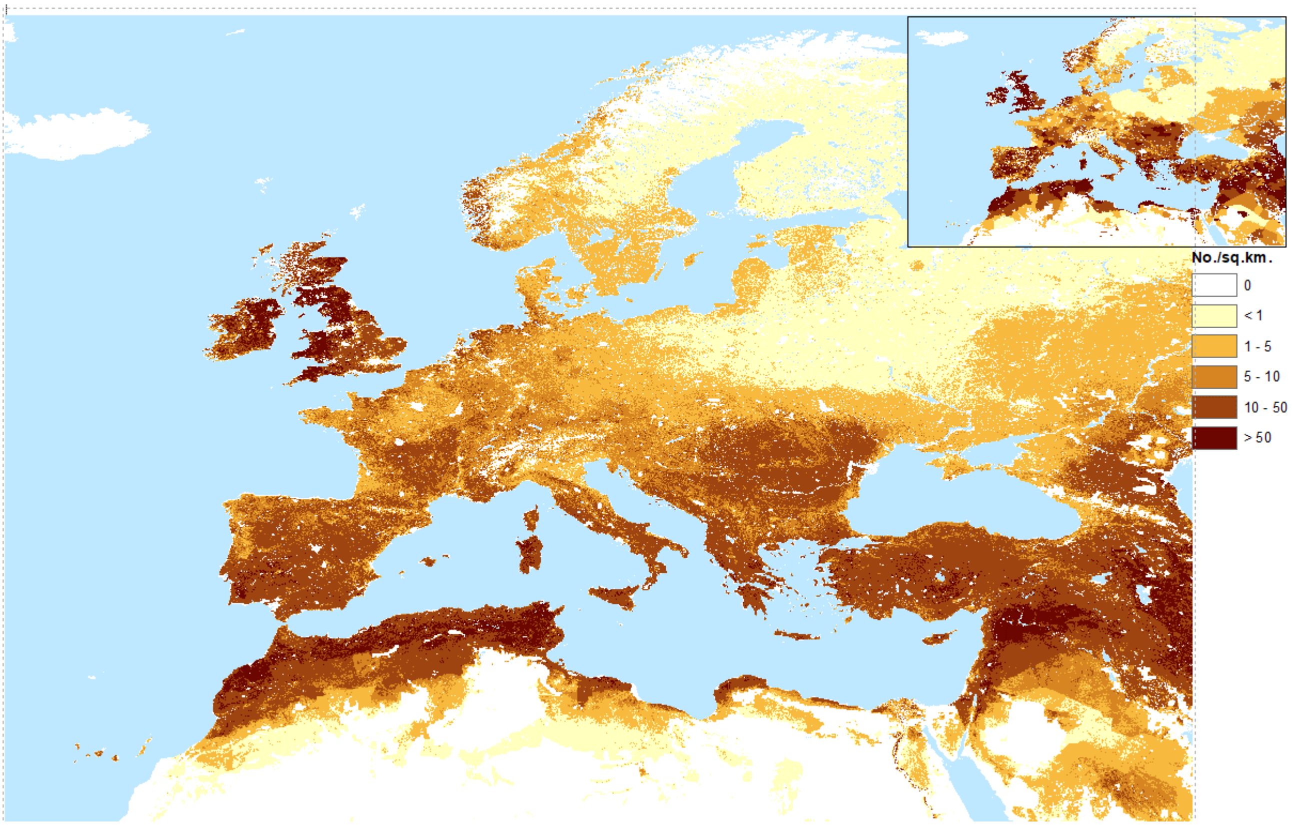
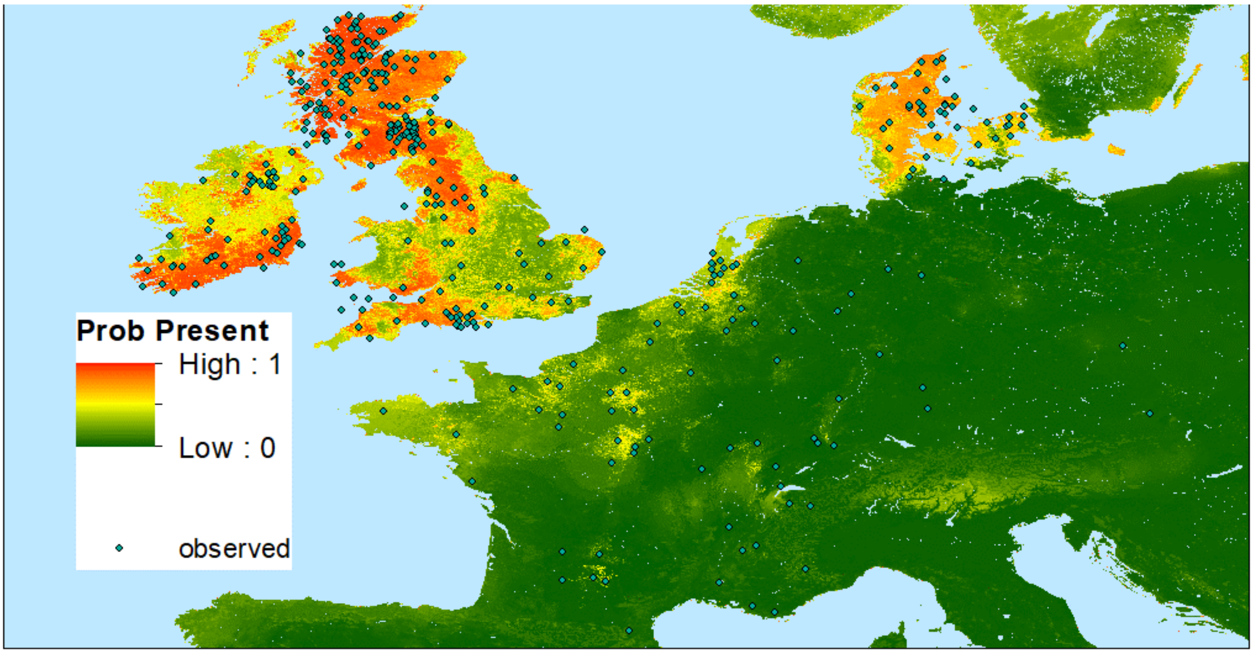


Figure 3. Modelled Probability of Presence for Sika Deer with Observed Present records



References

[1] Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A. Townsend Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann (2006). "Novel methods improve prediction of species' distributions from occurrence data". Ecography 29:129-151 https://doi.org/10.1111/j.2006.0906-7590.04596.x

[2] Breiman, L. 2001. Random Forests. Machine Learning. 45, 5-32.

[3] https://www.avia-gis.com/vecmap

[4] 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

[5] http://www.palebludata.com

[6] 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].

[7] http://www.earthenv.org