PALE-Blu work package 4, deliverable report (August 2017)

D4.1 List of required environmental and landscape factors needed (with WP3) due M3

Introduction and WP4 overview

One of the PALE-Blu activities is to identify pathways and mechanisms for BTV spread into and within regions covered by the project, including Europe, Maghreb and Western Africa. This will be achieved by the definition and the mapping of spatial epi-zones for BTV circulation, which should correspond to phylogeographic structure in the virus. These epi-zones could be connected, i.e. virus could be introduced from one to another epi-zone, by national and international animal movement data or by the movements of infected *Culicoides*, which could be active or driven by winds.

Within the project, Work Package 4 is dedicated to the study of *Culicoides* populations because their structure may contribute to the definition of epi-zones and their movements are involved in the connectivity of epi-zones. The WP4 aims to establish the structure of *Culicoides* populations at two levels: at the species community and at the population, and to assess the ability of *Culicoides* dispersal using mainly indirect methods.

These analyses require an assessment of the factors driving these processes, using multivariate analysis for species community and microbiome data and population genetic tools for data on population structure. These analyses will use spatial covariate (or predictor) datasets to identify the environmental factors driving these structures. Activities to assemble the requisite spatial covariate datasets are part of the Work Package 3. The applicable deliverables are D3.1 *(A list of spatial data requirements from WP1)* and D4.1 (*List of required environmental and landscape factors needed) (with WP3)* both due in project month 3. The scope of D3.1 has been deliberately expanded to include anticipated spatial data needs from the whole project. The objective of D4.1 is to validate the list proposed by D3.1. Thus, the D3.1 and D4.1 were constructed conjointly with the same timing.

These activities are a core element of task 3.1 (Assess the need, acquire, process and standardise spatial data) as follows, which are reported in the following pages:
   a) **Identify what spatial data are available**: Inform partners of what data are already available for project use
   b) **Establish future spatial data needs**: Ask partners to identify data that will be needed for their analyses
   c) **Define future spatial data acquisition strategy**: Outline future data acquisition and update strategies in the light of a) and b) above
   d) **Ensuring spatial data access**: Providing partners with the information needed to access the spatial data

Within that, we inform WP4 partners of what data are already available and identify with WP4 partners data that will be needed for their analyses.

Deliverable Activities

The structure of *Culicoides* species community, *Culicoides* population and *Culicoides* microbiome may be driven by:

- **Climate**

It is probably one of the main drivers structuring communities, populations and microbiomes. We need to have timely aggregated variables, as structure will be analysed against space and not against time. The most classical approach is to use average over 30 years of climatic parameters such as annual mean temperature, mean diurnal temperature range, maximum temperature of warmest month, minimum temperature of coldest months, annual precipitation and precipitation of the wettest and driest months, and other composite climatic indices. Otherwise, Fourier analysis could also be used to extract climatic synthetic indices from time series (for instance from MODIS images). The current climatic normalized period is 1981-2011. Current climate data is essential to model *Culicoides* distribution and would be quite convenient to have the new version of WorldClim as suggested in the ERGOs list, especially if database is at daily resolution, and cover the extent of the project from Senegal to northern Europe. In other case, other data source must be found. As an option, API developments could be used to download climate data from Internet repositories as: accuweather.com, wunderground.com, etc. Currently, partner 12 (UCM) has an API to download climate records from apixu.com (temp, humidity, wind, precipitations, etc). Some works suggest also that wind speed (nocturnal and daytime) could be closely related to *Culicoides* activity (Mellor *et al.*, 2000).

> Mellor PS *et al.* (2000) *Culicoides* biting midges: their role as arbovirus vectors. Annu Rev Entomol, **45**:307-40

- **Soil**

This is another important factor, and the "soil water index data" would be useful if covering the project extent. Some authors have used the Harmonized World Soil Database (HWSD), which provide the percentages of sand, silt, clay and organic carbon in topsoil (0–30 cm) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009). This could be useful, I do not know if soils datasets from Oak Ridge National Laboratories are equivalent.

- **Topography**

Elevation and slope are important factors determining the species distribution. I guess data already available in the "Elevation" category are sufficient.

- **Solar radiation**

The intensity of light is cited as a factor determining the species distribution. However, it may be not possible to use such as factor at a continental scale. May be worth to test some parameters such number of sunny days, or any other assessment of the solar radiation. Otherwise, information could be extracted from the Surface Radiation Data Set - Heliosat (SARAH) - Edition 2 (https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V002) which covers Europe, Africa, parts of South America as well as the surrounding ocean areas from 1983 to 2015.

- **Vegetation**

It should be the result of climate, soil, topography and solar radiation. However, it could be useful to test the effect of vegetation typology on the structures as the link between vegetation and species distribution could be more direct and easier to understand. France has produced recently a harmonised map of the

potential vegetation, in a vector format, at 1/1,000,000 such as database could not be available for the whole area. However apart from the usual vegetation indexes from MODIS used to model vectors distribution (NDVI and EVI), it could be possible to include the *Leaf Area Index* (also available from MODIS data set; Tatem *et al.*, 2003) to assess the suitability of the parameter for modelling *Culicoides* distribution. Moreover, it could be interesting to consider the inclusion of an *aridity index* like in previous *Culicoides* models (Conte *et al.*, 2004 and 2007).

Tatem AJ *et al.* (2003) Prediction of bluetongue vector distribution in Europe and North Africa using satellite imagery. Vet Microbiol, **97**:13–29

Conte A *et al.* (2004) Towards the identification of potential infectious sites for bluetongue in Italy: a spatial analysis approach based on the distribution of *Culicoides imicola*. Vet Ital, **40**(3):311-5

Conte A *et al.* (2007) Influence of biotic and abiotic factors on the distribution and abundance of *Culicoides imicola* and the Obsoletus Complex in Italy. Vet Parasitol, **150**(4):333-44

- **Natural barriers**

Mountains, deserts, or seas (and locally some landscape) could act as barriers for insect dispersal and then be responsible for structure of communities, populations and microbiomes. All information needed may be derived from the current available data in 'Hydrography', 'Climate' and 'Elevation' categories.

- **Host density**

Host densities have been linked to species distribution. Thus, data on cattle (dairy and beef), goat, sheep and horse would be useful. Red deer and roe deer are also available. Data could be sufficient if covering the project extent. There are also some endemic species in Mediterranean islands and in northern Africa, but developing distribution models for these species may be tricky. Apart from host density data, some authors (Ducheyne *et al.*, 2013) have used human density to model presence and abundance of *Culicoides*.

Ducheyne E *et al.* (2013) Abundance modelling of invasive and indigenous *Culicoides* species in Spain. Geospat Health, **8**(1):241-54

- **Landscape**

Definitively an important factor for species distribution, but also for population and microbiome structures. The equivalent of recent Corine Land Cover data for the project extent would be perfect. The CCI Land cover (ESA) could be the best option since it covers the project extent; the period is from 1992 to 2015. Updated land cover data is essential to model *Culicoides* distribution-abundance.

Establish future spatial data needs:

Partner responses were largely consistent in that

a) They considered the list of data currently available to be comprehensive and sufficient for their anticipated short-term needs, with the proviso that they reserved the option to request additional data during the project if necessary.

b) They pointed out that most of the data archive was continental or global in extent, and that there may be a future as yet unspecified requirement for more local, high resolution datasets covering the same parameters and covariates

c) A limited number of specific parameters were identified as possible additions to complement the existing suites – notably GMTED digital elevation, and consensus land use/landcover, wind speed and predominant direction, current climate and vegetation. The European Copernicus satellite

imagery products were also suggested for attention if they become readily accessible to the research community.

There are number of parameters that are specified in the project proposal or that partner 6 has identified as potentially valuable additions to the existing archive. These include continental coverage of livestock densities; trade in livestock and livestock products; and animal movements, as well as a wide range of processed Earth Observation satellite imagery from the MODIS platform.

In addition to these, partners also provided suggestions of potentially useful external datasets, which will be added to the Palebludata.com links page to make the suggestions available to all project partners (see D3.1).

Appendix 1: Data request sent to PALE Blu Partners in Project month 1
PALE-Blu spatial data requirements for WP4 activities

The main objective of the WP4 is to compare the structure of *Culicoides* species communities, the structure of some species populations using genetic data, and the structure of their microbiomes using metagenomic approaches, and to determine the environmental drivers of these structures. The first deliverable is due by late August (before the kick-off meeting). This deliverable needs to list the requested spatial data to be tested against vector and microbiome structure. This document presents a list of datasets currently available to the project partners, so that partners can then advise us of any other data that may be required.

Could you please have a look of the list of the available data, and complete/comment/correct my suggestions and comments below, preferentially before your summer break and ideally before mid-July?

Regards and we look forward to hearing from you

Thomas

Thomas Balenghien, Cirad

SUMMARY DATA LIST

Data category	Key datasets, Resolution and date	Other data available	Data available shortly
Land cover	• Corine Land Cover 100m/250m/1km 1990, 2000, 2006 – EU only • Globcover 2.3 2009 300m – Global • Consensus Land cover/land use. 1km, 2010	• Multiple forest cover datasets and change in forest cover. Suorces: European Forestry Institute, POSTEL, HarvestChoice, GLCF. 300m upwards • Irrigated areas IWMI • Snow cover MODIS/ISCCP 1km	• Corine Land Cover 2012 100m & 1km • CCI Land cover - 300 m annual global land cover time series from 1992 to 2015(ESA)
Transport networks	• JRC Accessibility Map: Estimated travel time to the nearest city of population 50,000 • Eurostat: Selected transport statistics - passenger and freight volumes	• Roads: Vector Map Level 0 (DIGITAL CHART OF THE WORLD)	
Elevation	• MODIS 1km Digital Elevation Model and Land/Water Mask V5 1km • The Global Land 1km Base Elevation (GLOBE) 1km Digital Elevation Model (DEM) • Slope from GLOBE DEM – 1km • SRTM 100m Digital Elevation VMERGE extent	• Terrain Ruggedness Index (TRI)	
Orthoimagery	• MODIS: Fourier Processed Imagery 2001-15 1km: land surface temperature; Vegetation indices • MODIS v5: Synoptic Months 2001-2012 (smooth-fill) 1km	• MODIS: Fourier Processed Imagery time slice 2001-02, 2014-15 1km • 1km Land Water Mask derived from MODIS • Night time lights datasets (NOAA/NGDC) 1km • Night time lights averaged by admin level 2 areas - indicator of GDP 1km • PROBA-V Level3 Top Of Canopy NDVI Ten-Daily Synthesis at 333M resolution - VMERGE Extent 333m • VMERGE: MODIS Phenology 2006-2012 1km	• MODIS: Fourier Processed Imagery projected to 2020, 2050, 2080 rcp6.5 1km in development • Updated Synoptic Series 2000-2016, version 6 imagery
Human health and safety		• Eurostat Healthcare Indicators (2000-09) at NUTS2 • Projected healthcare spending for the EC (1970-2050) at NUTS0 • UN Maternal mortality ratio by country and Infant mortality rates by country	
Species Distribution	• FAO Livestock Densities: Gridded Livestock of the World (GLW v2) 1km • EDENext Preliminary Livestock Distribution Models 1km • Beta Horse Density Models 1km	• EDENext Vole Species and Biodiversity Input Models 1km • Distribution of Major Crops (SAGE) 10km • HarvestChoice Spatial Production Allocation Model (SPAM) 10km	• FAO Livestock Densities: Gridded Livestock of the World (GLW v3, 2010) 1km and 10km updates

Data category	Key datasets, Resolution and date	Other data available	Data available shortly
	• EDENextRed Deer Models 1km • EDENext Roe Deer Models 1km • Current JRC Tree Species Data 1km • JRC Tree Species Suitability & Climate Change 1km • VBORNET/VECTORNET Culicoides vector distirubution Models 1km for Eu and Mediterranean basin	• IUCN Species distribution: EDENext selected species 1:1,000,000 • The EMMA Database: Gridded Mammal Distributions for Europe 50km	•
Current Climate Data	• Worldclim 1km baseline norms 1950-2000 • BIOCLIM: WorldClim Bioclimatic variables 2000 1km • Cmorph Daily Rainfall data Global and West Africa 2010 - Dec 2013 0.25 degree	• CRUCL2 (10') Baseline norms 1961-1990	Worldclim data Likely to be updated 2017/2018
Future Climate Projections	• Ensenmbled and downscaled IPCC5 Europe projections using Worldclim base 1km, 4 RCPS. Temp, precipitation, RH • Futureclim 1km climate projections (2035,55,85) 1km • BIOCLIM: WorldClim Bioclimatic variables (2000, 20 ,50 ,80) 1km	• CLIMOND: Bioclimatic Variables (2030 and 2070) 5km – provides wider no of variables than BIOCLIM at a coarser resolution	
Other Climate Related Datasets	• Relative Humidity 5km & 10km present data monthly and annual min and max values, Global • Relative Humidity 5km & 10km present data summer season min, mean and max summaries, Global • MODIS Land Cover Phenology: Version 5 - 500m • FAO Length of Growing Period 5km • TAMSAT African Rainfall Climatology And Time-series TARCAT V2 •	• MODIS Snow cover: Mean percent of snow cover 2000-2008 by month 0.05 deg • MODIS Snow Cover Mean Monthly Percentage Cover by year 0.05 deg • Days above 105 degrees Fahrenheit (E-OBS data) 1km • MODIS Net Primary Productivity V5 - 1km • VEGETATION Small Water Bodies VITO/JRC - 1km	Summarised Daily wind speed (NOT direction) 25km resolution, from 1974, EU and north Africa, Agri4cast
Utility and governmental services		• Waterbase - UWWTD summarised at NUTS2 • Eurostat: Education, selected statistics • Downscaled Improved Water statistics for Europe at NUTS2 2005, 2030, 2050 • UN UIS: Public expenditure on education, percent of GDP by country for Europe • UN Millenium Development Goals: Proportion of the population using improved sanitation facilities	
Hydrography	• JRC Flood return rate projections 2 year return rates for 2000/25/35/85 – 100m • JRC Flood return rate projections 100 year return rates for 2000/25/35/85 – 100m • VEGETATION Small Water Bodies VITO/JRC - 1km	• Global Lakes and Wetlands Database - Level 3 (WWF) - 30 seconds • Global Lakes and Wetlands Database - Level 3 (WWF) - Seasonal Irregular coverage - 30 seconds • AVHRR Water Layer -1km	

Data category	Key datasets, Resolution and date	Other data available	Data available shortly
	• West Africa Land Surface Water Index MODIS V5 Oct 2010 - Dec 2013 1km • Quantitative ground water maps for Africa – BGS 5km	• Rivers: Vector Map Level 0 (DIGITAL CHART OF THE WORLD) 1:1,000,000	
Soil	• The soil water index data (SWI) Mean synoptic monthly values - 25km • Soils datasets from Oak Ridge National Laboratories - 10km	• POSTEL Soil Moisture 25-04-2004: Soil Moisture (ERS Scatterometer) - 50km	
Bio-geographical regions	• WWF Terrestrial Ecoregions and Biomes 1km • Anthropogenic Biomes 10km • West African Livelihood Zones (FEWS) – 1km	• Olson ecological zones 10k • Köppen-Geiger climate classification: Observed and projected climate shifts 1976-2100 1km	
Population distribution and Demographics	• GPW Populated Places -1km • Human Population Dataset GPW -1km • Eurostat: Population projections by NUTS2 admin areas for 2008, 2010, 2030 • GRUMP Rural Urban Extent and Masks -1km	• Populated places: Vector Map Level 0 (DIGITAL CHART OF THE WORLD) 1:1,000,000 • Night time lights datasets (NOAA/NGDC) 1km • Global Human Footprint Dataset • Night time lights averaged by admin level 2 areas - indicator of GDP • Geographically based Economic data (G-Econ): Purchasing parity GDP 5km • Distance weighted population proximity index v0.3 • UN Maternal mortality ratio by country and Infant mortality rates by country	• WorldPop Africa population and demographics at 100m. Likely updates 2018
Trade and economics	• Global GDP. 2000, 2005. Gridded	•	• TRACES, UNFAO, and ETIS summaries