



Project ID: 727393

Work Package 3 Deliverable Report: June 2018

Deliverable 3.2. *Updated website with environment/anthropogenic data:* Due Month 12 (May 18)

Summary: The spatial data site is populated with several hundred spatial datasets. Considerable progress has been made in sourcing, processing and standardising the many of the new and updated datasets specified in the spatial data acquisition strategy set out in Deliverable 3.1, August 2017. Most particularly three major datasets covering greater Europe, the Near East and North Africa have been produced for Project use: an updated archive of Fourier processed MODIS satellite imagery; monthly wind speed and direction; and accessibility to roads – express as road density. These are all world firsts and can be downloaded from the Pale-BLU spatial data site <http://www.palebludata.com>

Palebludata.com provides access to hundreds of spatial covariate datasets falling into a number of thematic categories as shown in the screengrab below. A section dedicated to PALE-BLU datasets can be seen at the top of the list.

The screenshot displays the Pale-Blu website interface. On the left, there is a 'User login' section with fields for 'Username' and 'Password', a 'Request new password' link, and a 'Log in' button. Below this is a 'Data Uploads' section explaining the process of uploading data via email or FTP. At the bottom left is an 'In collaboration with...' section. The central 'Data Archive' section features a list of downloadable datasets, with 'PALE-Blu Data' highlighted as a top-level category. This category includes sub-items like 'Publication Data', 'Population Distribution and Demographics', 'PALE-Blu Meetings', 'Land Cover', 'PALE-Blu Masks', 'PB Soils', 'PB Species Distribution', 'PB Elevation', 'PB Orthoimagery', and 'PB Transport Networks'. Other categories include 'Land Cover', 'Transport networks', 'Elevation', 'Orthoimagery', 'Human health and safety', 'Species Distribution', 'Atmospheric Conditions and Meteorological Geographical Features', 'Training Program Presentations and Data', 'Utility and governmental services', 'Hydrography', 'Soil', 'Bio-geographical regions', 'Population distribution and Demographics', 'Data only available on request', 'Sister Projects', 'AGM Downloads', 'EDENext Group Data', 'EDENext Masks', and 'VMERGE Data'. On the right side, there are three promotional boxes: 'Keep Informed...' with social media icons, 'The PaleBlu Project' with the project logo and website URL, and 'News from the projects' with a list of recent events like the 'International Conference on El Niño' and 'Rift Valley fever newsletter'.

All visitors can display the content lists, but access required registration. Access to these data falls into several categories – public domain, available to all registered users; available on request; available with permission of the data owners; available only to project members; and available only to specified registered users.

Users can also provide data for the site administrators to add to the site – and can define precise access conditions. An example of such data will be the *Culicoides* vector distributions produced as part of Deliverable 4.2, details of which will be posted in the coming months.

Deliverable 3.1 set out the anticipated spatial data requirements solicited from project Work Packages 1-4 as follows: *“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”*.

The spatial data acquisition strategy was further set out as follows: *“WP3 will therefore focus its initial spatial data acquisition efforts on the datasets identified above. The topography, consensus land use and recent (2015) land cover data have been acquired and standardised, and will be added to the project archive in the near future. Coarse resolution wind data have been identified and the feasibility (and cost) of acquisition and processing higher resolution data are being investigated. Leaf Area Index will be added to imagery archive; the availability of an aridity index (or a proxy such as Embergers Index) will be investigated; and the feasibility of extracting appropriate summary information from the SARA solar radiation archive will be assessed.*

As well as providing new data as requested, the WP3 spatial data management specialists will devote substantial efforts to upgrading the existing data to the latest versions. A major effort in this context is to convert all the Fourier processed MODIS V5 imagery to the new Version 6 which has better quality control, fewer data gaps, and improved pre-processing algorithms. This involves reprocessing 16 years of data (2000 to 2016) for six parameters with image frequencies of 8 – 16 days – approximately 9 TB. This is a substantial task, anticipated to take several person months, which is currently in progress. Other upgrades will include human and livestock populations and projected climate. As a matter of routine, watching briefs will be maintained for newly produced spatial data that are likely to be relevant to project objectives”

Whilst many of these data may be fairly readily available, they are often provided in highly specialised formats, projections or coverages that are not suitable as covariates for spatial modelling procedures. Typically such modelling requires a series of covariate images as a standard resolution, with a fixed extent, and in a common GIS format such as geotiff. The data management teams aim therefore not only to source and acquire the datasets required but also to process them into standardised formats, as well as calculating the required parameters from the raw data.

Progress to date on new data

Progress on these objectives has been substantial as follows:

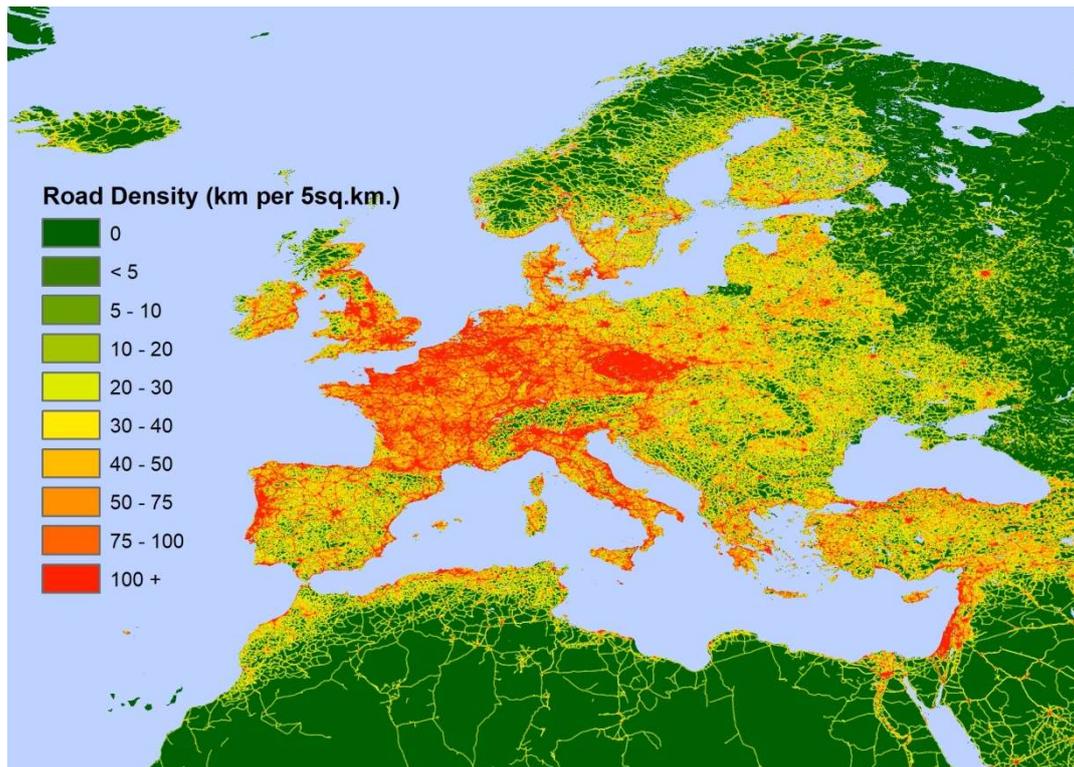
- A) The datasets for GMTED digital elevation, consensus land cover (earthenv.com), current climate (worldclim.com) and vegetation (ESA CCI 2015) have been acquired, standardised for format and extent, and are now incorporated into the Project archive.
- B) The entire 2000-2016 MODIS V6 imagery archive, amounting to just over 80,000 images, has been acquired and successfully been subjected to Temporal Fourier Analysis. Data sets for Channel 3, Day-time and night-time Land Surface Temperature, and for both Normalised Difference and Enhanced Vegetation Indices. These data were produced at 1km resolution with global coverage, but have also been provided for the project area (Europe and its Eastern neighbouring areas, North Africa and the Sahel, the near East. This is a major upgrade of the spatial data that has underpinned much of the arthropod vector and disease host distribution modelling for many research groups for a number of years, and will be central to future PaleBLU spatial modelling efforts of the Culicoides vectors and of BTV. An example of the timing of the peak Enhanced Vegetation Index is illustrated below:



- C) In the context of satellite imagery, the new European Copernicus will inevitably become the main source of earth observation data in the coming years. Whilst these data have now been reliably produced by the Copernicus system, and are confirmed to be as good or better than the MODIS data they will effectively replace, the processing chain needed to produce usable climate and vegetation covariates with continental or global extents for general use has not yet been established. As a result these data are yet to be available at the continental scales needed for PaleBLU spatial analysis. Contacts have however been made with two European image processing groups (VITO and Mundialis) who are both likely to be able to provide the required data streams within the project lifetimes. In the mean time it is fortunate that the MODIS datasets potentially relevant to spatial analysis of the BTV outbreaks in Europe (i.e. imagery for 2002 – 2012) are available in the PALE-BLU archives
- D) Monthly wind speed and direction from 2008. Daily windspeed is available for the EU 28 from agri4cast.com, but this does not include direction. In addition, as stated in the Deliverable 3.1 documentation, the Danish Global Wind Atlas is able to provide medium (1 to 5km) resolution gridded data for windspeed and direction derived from a daily archive stretching back ten or more years and for the whole project area. The cost of producing the monthly averages for these parameters for a number of years was, however, is well beyond the resources available (in excess of €30K). As a result, the data management team acquired the only feasible alternative multi year (2002-2012) time series from the ECWMF ERA Interim datasets (http://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc/?month_years=2009¶m=165.128,166.128,207.128), as component vertical and horizontal wind velocity parameters from which the direction and wind speed were produced. Whilst this is a coarser resolution product than that available from the Global Wind Atlas, Work Package colleagues agreed that it would be sufficient for their needs and that the length of the time series it represented compensated for the reduced level of detail. An example for the average direction from which the wind (0-360) was blowing in October 2008 is shown below.



- E) At the request of WP1 partners, the project Data Management Team were tasked with providing continent wide measures of accessibility to the road network for the entire project area. Whilst on the face of it, a relatively simple request, initial searches revealed that whilst there are road network datasets available for a number of European countries, they are either not available or incomplete for much of the continent. The only consistent data available were found to be the open source Open Street Data which underpin the web mapping services like Google and Bing maps. These datasets incorporate not only roads, but also the 'points of interest' (hospitals, shops, schools, businesses) as well as paths, railways and the like. As a result they are large (several tens of GB per country), and the roads need to be extracted from the larger archives. This turns out to be a non trivial exercise as the dataset sizes are at the limits of what commercially available GIS software (ESRI ArcMap or QGIS) can reliably process and combine into continental extent databases. In any event, road network databases for a number of road types: motorways, trunk roads, secondary roads, tertiary roads, and unclassified roads. From these, a series of road density maps were produced (measures as kilometres of road per square km). As far as can be ascertained by extensive searches, these data are the first consistent continental scale road density maps ever produced. An example for road density of all classified roads in 2017 is shown in Figure xxxx.



A number of potential covariates specified in the acquisition strategy remain to be acquired though some, such as the SARAH solar radiation have been sourced from EUMETSAT CMSAF (<https://wui.cmsaf.eu/safira/action/viewProduktList?dId=2>) and so can be downloaded when requested by project members. The same applies to Leaf Area Index, available as a MODIS version 6 8-day product at 500m resolution (MOD15A2H, <http://earthdata.nasa.gov>). Investigations of an aridity index continue. Livestock numbers are available from the FAO Gridded Livestock of the World, and it is envisaged that some or all the ruminant species distributions will be updated by collaborators in the coming years.