PALE-Blu work package 3, deliverable report (November 2018)

D3.3 *Refined spatial epizone maps*

Rationale

One of the objectives of WP3 is to disentangle the relative effects of environmental and anthropogenic factors in BTV spread and quantify connectivity between epi-zones over short periods of time. This will be achieved through phylogeographic analysis, by looking at how the spatial structure of the environment could have influenced the pattern of spread inferred from molecular data. Two different approaches can be considered to achieve this, continuous and discrete models. In the continuous model, space is divided into pixels and the spread or diffusion rate is quantified in each pixel and matched again environmental conditions. In the discrete model, between unit and spatial factors are converted into connectivity matrices. In the discrete model, space is divided between discrete spatial units where spread and evolutionary conditions are assumed to be similar and the spread and diffusion are quantified between these units, and tested against different connectivity matrices (typically trade or wind patterns in the case of BTV). The later model requires the definition of these epizones at the European scale.

Technical activities

We used and implemented a methodology that was developed by Ippoliti et al. (In revision) to map epizones of vector-borne diseases in Italy, but using spatial covariates at much coarser spatial resolution of 5 minutes of arc (0.083333 decimal degrees, corresponding to approximately 10 km at the equator) to be applicable at the scale of Europe.

Seven eco-climatic variables related to topography, temperature, rainfall and vegetation, and known to be relevant to BTV vectors, were selected. Topography influences the water runoff and the retention of water and nutrients, affecting the moisture of soil surface layer. Depending on the vector preferences for a mud, moist or aquatic breeding site, the topography contributes to promote or inhibit the proliferation of larval sites. Temperature influences the vector population dynamics and it drives the vector competence, by accelerating the virus replication within the insects and prolonging their breeding season.The effect of rainfall is more controversial in literature and is highly dependent on the habits of the vector species. Water pools are fundamental for the larval stage of mosquitoes and rainfall events favour their proliferation, but at the meantime dilute the content of nutrients decreasing their reproduction rate. Other vector species, as *Culicoides imicola*, needs moist soil but not flooded to breed, and high rainfall events could eliminate larval habitats and create unsuitable environmental conditions. Vegetation is a key parameter both to define vector habitat, and as a proxy of rainfall and humidity. The Normalised Difference Vegetation Index (NDVI) is a wide used index in remote sensing, representing the presence and density of green biomass. It is widely present in epidemiological models to explain disease occurrences or to describe the vector habitats.
In addition to the variables related to eco-climatic conditions, we also wanted to include data on the distribution of the main domestic ruminan hosts, cattle, sheep and goats.

In terms of spatial data, we used the GTOPO database at 1 km resolution to derive the standard deviation of altitude in each 10 km pixel (LDDAC 2005). We used Fourier-transformed MODIS estimates of land-surface temperature (LST) and NDVI from the database published by Scharlemann *et al*. (2008), with the mean LST, the mean amplitude of the annual cycle of LST, the peak timing of the annual cycle of the LST, the mean amplitude of the annual cycle of NDVI, and the average daily amount of rainfall in mm (Hijmans et al. 2005). For livestock, we used the latest version of the Gridded Livestock of the World database (Gilbert et al. 2018).

The analyses involved principal component analysis (PCA) of the spatial variables, followed by a multivariate geographic k-medoids cluster analysis, a classical partitioning method that aggregates the dataset of *n* objects into *k* clusters known *a priori*. The partition algorithm around medoids minimizes the average quadratic error between all points and the point centre of the cluster. We arbitrarily choose 15 clusters. The same analysis was repeated with two different sets of spatial predictors, the seven ecoclimatic variables in one hand (Fig. 1), and the seven ecoclimatic and three livestock species variables in the other hand (Fig. 2 and 3).

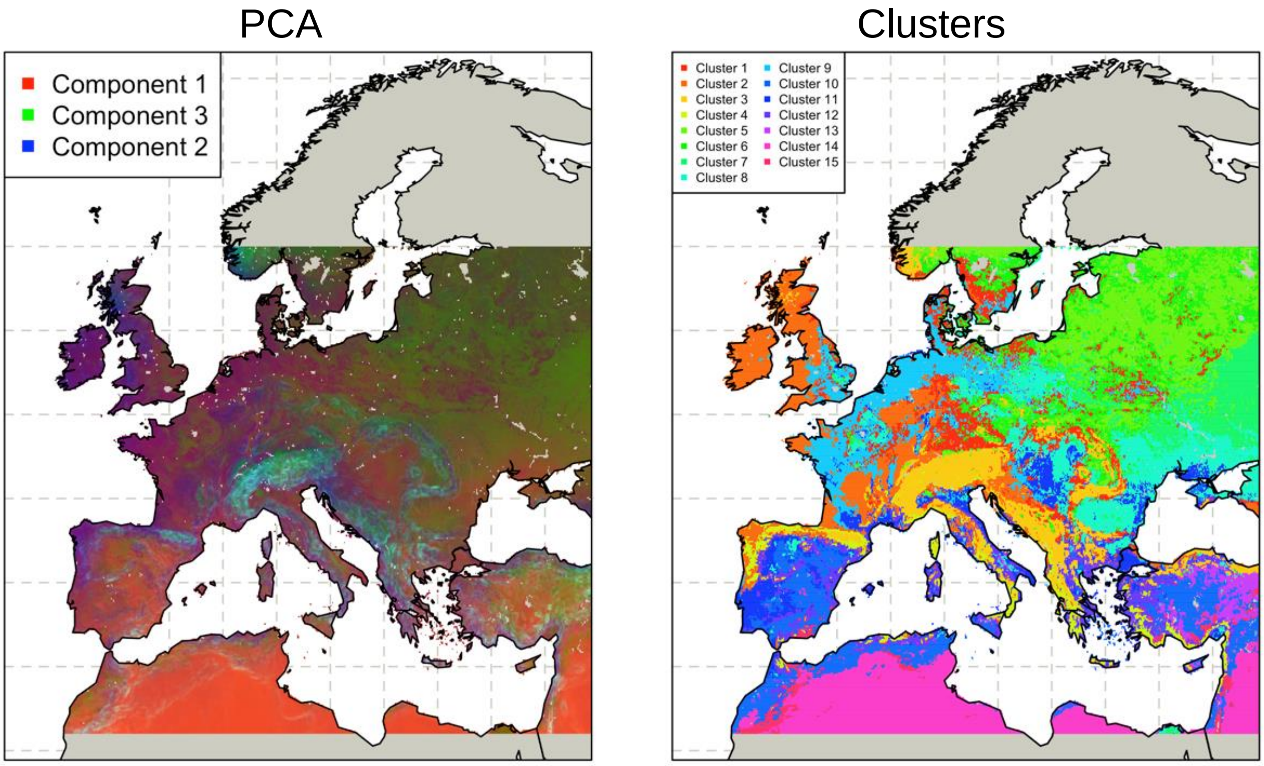


Fig. 1. PCA and cluster map using the seven eco-climatic variables.

Interestingly, the PCA of the eco-climatic and host variables showed that cattle density was aligned with areas with high mean annual NDVI and high precipitation (Fig. 2), whereas sheep and goat density were much more aligned with areas of high standard deviation of the elevation. The resulting cluster map shown in Fig. 3 depicts a number of clusters at the European scale. IN particular, one can highlight green and pale green clusters (4,5,6) corresponding to areas with low densities of all livestock species and located in eastern Europe, a typical cattle cluster (red, 1) that has the highest mean cattle density and spans across Denmark, German, Belgium, Switzerland and France. Cluster 9 (pale blue) depicts areas with high densities of both cattle and sheep, primarily located in Ireland, large parts of the UK, France and the Netherlands. Finally, the dark blue and pink cluster are linked to high densities of sheep and goats, as typically found around the Mediterranean basin.

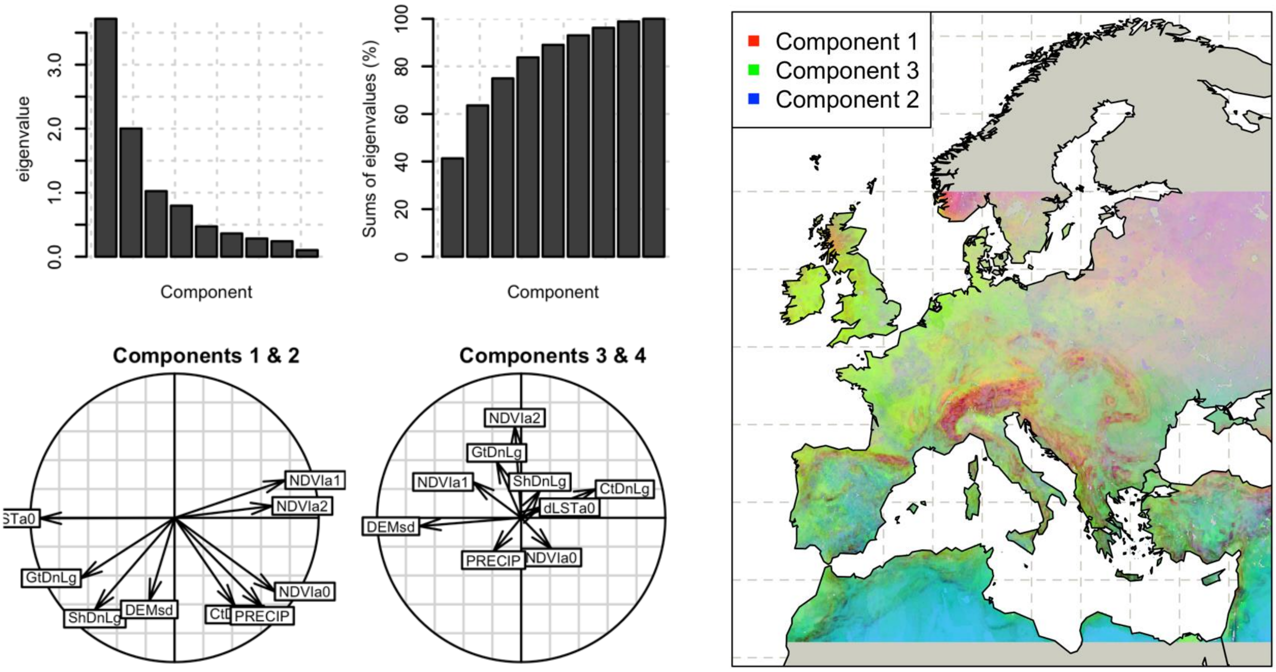
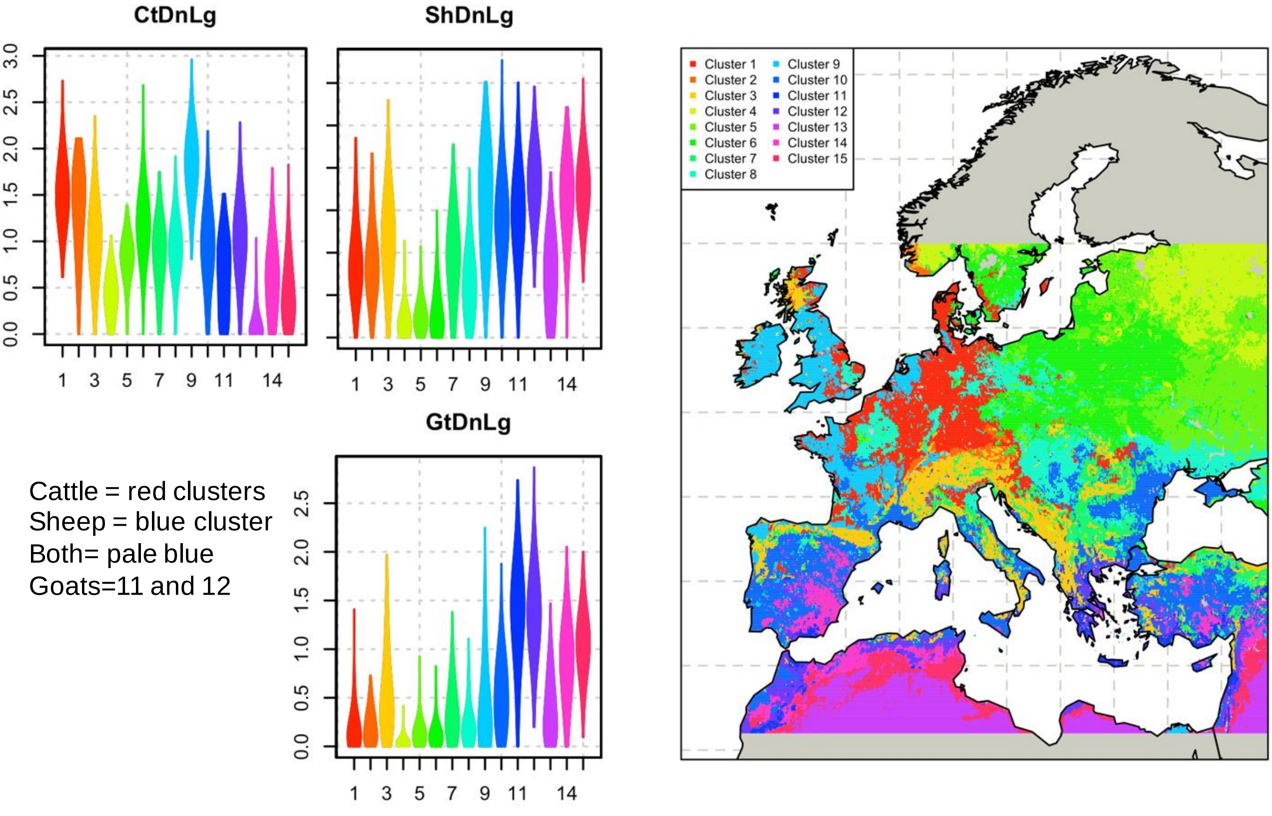


Fig. 2 PCA using the seven eco-climatic and livestock variables



Cattle = red clusters
Sheep = blue cluster
Both= pale blue
Goats=11 and 12

Fig. 4 Cattle (CtDnLg), sheep (ShDnLg) and goat (GtDnLg) density by cluster, as defined by the seven eco-climatic and three livestock variables.

This epizone map still has two areas of possible improvements that could be addressed through follow-up works. First, the clustering technique does not take spatial proximity into account, which results in salt and pepper effect of isolated pixels belonging to a particular cluster. Several techniques have been proposed in the literature to increase the spatial vicinity in the clustering algorithm and could be implemented. Second, and more importantly, the current epizone do not take BTV vector distribution into account. Future versions of the epizone maps should integrate the full set of environment, host and vectors variables to pinpoint areas where all these three components can be considered to be similar.

References

Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke SO, Wint GRW, et al. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Scientific Data. 2018 Oct 30;5:180227.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005 Dec 1;25(15):1965–78.

Ippolitti *et al*. (in revision)

LDAAC. Global 30 Arc-Second Elevation Data Set GTOPO30. Land Process Distributed Active Archive Center; 2004.

Scharlemann JP, Benz D, Hay SI, Purse BV, Tatem AJ, Wint GR, et al. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. PloS one. 2008;3(1):e1408.