RESEARCH ARTICLE

Defining ecological regions in Italy based on a multivariate clustering approach: A first step towards a targeted vector borne disease surveillance

Carla Ippoliti[1]*, Luca Candeloro[1], Marius Gilbert[2], Maria Goffredo[1], Giuseppe Mancini[1], Gabriele Curci[3,4], Serena Falasca[3,4], Susanna Tora[1], Alessio Di Lorenzo[1], Michela Quaglia[1], Annamaria Conte[1]

1 Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise G. Caporale, Campo Boario, Teramo, Italy, 2 Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Brussels, Belgium, 3 Department of Physical and Chemical Sciences, University of L'Aquila, L'Aquila, Italy, 4 Center of Excellence in Telesensing of Environment and Model Prediction of Severe Events (CETEMPS), University of L'Aquila, L'Aquila, Italy

* c.ippoliti@izs.it

Abstract

Ecoregionalization is the process by which a territory is classified in similar areas according to specific environmental and climatic factors. The climate and the environment strongly influence the presence and distribution of vectors responsible for significant human and animal diseases worldwide. In this paper, we developed a map of the eco-climatic regions of Italy adopting a data-driven spatial clustering approach using recent and detailed spatial data on climatic and environmental factors. We selected seven variables, relevant for a broad set of human and animal vector-borne diseases (VBDs): standard deviation of altitude, mean daytime land surface temperature, mean amplitude and peak timing of the annual cycle of land surface temperature, mean and amplitude of the annual cycle of greenness value, and daily mean amount of rainfall. Principal Component Analysis followed by multivariate geographic clustering using the k-medoids technique were used to group the pixels with similar characteristics into different ecoregions, and at different spatial resolutions (250 m, 1 km and 2 km). We showed that the spatial structure of ecoregions is generally maintained at different spatial resolutions and we compared the resulting ecoregion maps with two datasets related to Bluetongue vectors and West Nile Disease (WND) outbreaks in Italy. The known characteristics of *Culicoides imicola* habitat were well captured by 2/22 specific ecoregions (at 250 m resolution). *Culicoides obsoletus/scoticus* occupy all sampled ecoregions, according to its known widespread distribution across the peninsula. WND outbreak locations strongly cluster in 4/22 ecoregions, dominated by human influenced landscape, with intense cultivations and complex irrigation network. This approach could be a supportive tool in case of VBDs, defining pixel-based areas that are conducive environment for VBD spread, indicating where surveillance and prevention measures could be prioritized in Italy. Also, ecoregions suitable to specific VBDs vectors could inform entomological surveillance strategies.

Introduction

Ecoregions have been defined as areas "*within which there are associations of interacting biotic and abiotic features*" [1]. Climate, orography, geological factors (abiotic) and vegetation (biotic), are the characteristics commonly used to define homogeneous land units, within which natural communities and species interact with the physical elements of the environment [2]. Ecoregions are typically applied in forest conservation [3], land management programs [4] and in environmental conservation strategies [2].

Ecoregionalization is then the process through which a territory is classified into similar areas according to specific environmental and climatic factors. There are several datasets already existing for Italy that reflect vegetation distribution (e.g. the CORINE Land Cover (CLC) map [5]), and climatic conditions or bioclimatic indices [6,7], but these datasets are still not fully integrated together. For Italy, the most recent classification available derives from a hierarchical, deductive and divisive expert-based method, vegetation-oriented, that, at its finer scale, divides the Italian land in 33 areas, homogeneous by climate, physiography, biogeography and vegetation [2]. This classification has been further evolved into an ecosystem mapping of Italy, where it was integrated with the CLC and the potential natural vegetation through an expert-based overlay [8].

A quantitative and objective classification (nor hierarchical or supervised) that includes land use, climatic and topographic features, which allows the identification of homogeneous areas at finer scale, using a raster-based approach is still missing. The recent and rapid advances in the performance and accessibility of computer hardware and statistical software facilitate the ecoregionalization process, taking also advantage of the large volume of spatial data availability, also remotely sensed [9].

Whether remotely sensed or ground collected, ancillary or in real time, the selection of data sets to be included in an ecoregionalization process responds to the specific purpose for which the similarity analysis is conducted.

In addition to the traditional fields of application explored so far, the use of the ecoregionalization approach can be of interest also in the veterinary field especially when dealing with vector-borne diseases, where the triangle vectors-host-pathogen is strictly linked to the environment [10–13]. Suitable conditions allow the proliferation of the vectors, providing rest and breeding sites, shapes the pathogen replication and facilitates the contact among the three actors promoting the infectious disease transmission. A recent assessment indicated that there is a common set of drivers to many human and domestic animal pathogens; in particular, the 82% of the most significant human and domestic animal pathogens occurring in Europe are influenced by temperature, rainfall, humidity or wind, and this is particularly true for VBDs [14].

Given the presence of susceptible reservoir hosts and competent vectors, areas with similar climatic and environmental conditions, are potentially exposed to similar disease risk, even if they are geographically very distant from each other. This simple assumption underlines much of the spatial modelling work that has been developed so far. Numerous studies have earlier examined the complex interactions among environment and vectors/diseases [15–18], also through the species distribution models (SDM) applied to vector occurrence [18–22]. They are then useful when the vectors are known and their distribution enough assessed with reliable field dataset so to feed the model and produce accurate predictions. When an incursion of a new VBD is observed in an area, species not previously known to be competent vector could result implicated in the transmission of the disease, as happened with bluetongue BTV8 in Northern Europe in 2006 [23,24] and this can limit the application of SDM so far developed. Emerging vector borne diseases could therefore benefit from a different approach starting

from the definition of generic ecoregions, mapping the climatic and environmental similarity so to inform the spatial context in which a VBD can potentially spread or persist. The idea of identifying similar areas in terms of climatic and environmental conditions does not aim to link a specific vector to specific conditions, but rather to highlight where the same characteristics (e.g. high/low temperatures, heavy/light rain, dense/sparse vegetation, etc.) are present so as to be able to identify *a priori* where a possible vector that prefers those specific conditions could spread.

In this paper, we identified similar eco-climatic regions of Italy following a data-driven spatial clustering approach and using the available, most detailed and up-to-date spatial data on climatic and environmental factors. The variables chosen, although not exhaustive, are among those that have been repeatedly found to be associated with a set of human and animal diseases [14], in particular with Bluetongue [25,26] and West Nile [27–29], two vector borne diseases affecting Italy in the last decades.

We also compare the resulting ecoregion maps with two dataset, i) the distributions of some BT vectors and ii) the distribution of WND outbreaks, to highlight pros and cons of such an approach.

Materials and methods

Environmental and climatic selected data

Seven variables related to topography, temperature, rainfall and vegetation, and known to be relevant to VBDs, available for the entire Italian territory, were selected for the ecoregionalization process.

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 [17,30,31].

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 [32,33].

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 [16,34]. Other vector species, as *Culicoides imicola*, needs moist soil but not flooded to breed [17], 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 [35], representing the presence and density of green biomass. It is widely present in epidemiological models to explain disease occurrences [36–38] or to describe the vector habitats [25,39–41].

In our study, temperature and vegetation were derived from remotely sensed archives and submitted to a Fourier decomposition: yearly time series in each pixel is decomposed in sine curves (harmonics), each characterised by specific amplitudes and phases whose sum recompose the original dataset [42]. Usually the first two harmonics are retained, as they explain the majority of variability and are easily biologically interpretable. The first component of the series, the harmonic 0, is an average of the values across each year and it gives the yearly mean value of the variable (A0). The harmonic 1 is characterised by amplitude (A1) and phase (ph1) and it describes the annual cycle of the variable. Once the yearly rasters of A0, A1 and ph1,

were calculated per each year in the period 2007–2016, an average was calculated across the 10 years.

The seven variables included in this study were:

- Standard deviation of altitude (AltSd)

- Mean daytime Land Surface Temperature in ˚C (LstdMn)

- Mean amplitude of annual cycle of LST in ˚C (LstdAmp1)

- Peak timing of the annual cycle of LST, i.e. the day when the first harmonic of the temperature series reaches its maximum value (LstdDPk1)

- Mean greenness value (NDVIMn)

- Mean amplitude of annual cycle of greenness values (NDVIAmp1)

- Daily mean amount of rainfall in millimetres (RainMn).

Data sources, collection and manipulation

Elevation was derived from the 20 m Digital Elevation Model of Italy (http://www.pcn. minambiente.it accessed 2nd March 2018). The standard deviation (**AltSd)** of the values falling in the pixel at the reference scale (see next paragraph for details) was calculated. The standard deviation of altitude points out the local variability of the altimetry and resulted more informative than the mean altitude that in Italy is highly correlated with mean temperature (Pearson correlation coefficient = 0.85, p-value <0.01).

From the product MOD11A2 (MODIS/Terra Land Surface Temperature and Emissivity 8-Day L3 Global 1km Version-5), the daytime LST images (LSTD) were downloaded for ten complete years, from 01.01.2007 to 12.31.2016 [43]. Each raster was converted into WGS84– UTM33 coordinate system and the values into degree Celsius (˚C). The four tiles covering Italy were mosaicked, and single raster files per date were archived, resulting in 46 images per year.

Linear temporal interpolation, spatial interpolation and climatology procedures [44] were run to fill missing pixel data. The dataset was then submitted to a Seasonal Trend Analysis to perform a harmonic regression on yearly data. The average over the decade of the yearly harmonics 0 and 1 were retained: amplitude 0 and amplitude 1 were named **LstdMn** and **LstdAmp1,** respectively. To facilitate interpretation, phase of the first harmonic was accounted as **LstdDPk1**: it is the day when the first harmonic of the temperature series reaches its maximum value and it was calculated applying the following formula

$$LstdDPk1 = MOD(450 - LSTD\_ph1, 360)$$

where $MOD(n, d) = n - d^*INT(n/d)$, MOD function returns the remainder of two numbers after division. $LSTD\_ph1$ is the phase of the first harmonic of temperature.

n = number, is the number to be divided; d = divisor, is the number to divide with. The parameter 450 in formula is set to always have the peak in the positive x-axis.

The NDVI dataset was downloaded for the reference period from the product MOD13Q1 (MODIS/Terra Vegetation Indices 16-Day L3 Global 250m Grid SIN V006) [45], geographically processed and resulting in 23 images per year. Each yearly NDVI dataset was submitted to a Fourier decomposition as described for LST; the average over the decade of the yearly harmonics of amplitudes 0 and 1, were retained and named **NdviMn** and **NdviAmp1,** respectively. The phase of the harmonic 1 (representing the day of the year when the greenness reach

its maximum value) is a variable highly dependent on the type of cultivation and averaging the values across the decade could mislead the ground truth. As the objective of this study is not to investigate the characteristics of field cultivation in Italy, we decide to not include it in the analyses.

Daily rainfall rasters for ten years, from 01.01.2007 to 12.31.2016, were downloaded from DEWETRA web application (http://dewetra.cimafoundation.org/dewetra/ accessed 2nd March 2018): these rasters derive from ground weather stations managed by the Italian Civil Protection, interpolated trough the GRISO method at a 0.02 degrees [46]. The daily mean amount of rainfall was considered as variable (**RainMn**), being representative of the total rainfall of the year [47].

Pixels corresponding to lakes and cities were set to NoData in all dataset to avoid spurious and artificial effects of reflectance wavelength on these surfaces.

The geographical manipulation was performed in ArcMap 10.3.1 ESRI with routines in Python language; the Fourier analyses in Idrisi Taiga software and with Idrisi Macro Language. Routines in R software (https://www.r-project.org/ accessed 2nd March 2018) and the Modis Reprojection Tool (https://lpdaac.usgs.gov/tools/modis_reprojection_tool accessed 2nd March 2018) were used to download and process the MODIS images. The processing archive size and number of files is reported in S1 Table.

The spatial resolutions for our analyses were: i) 2 km (driven by DEWETRA rainfall data spatial resolution); ii) 1 km (driven by LSTD data spatial resolution) and iii) 250 m (driven by NDVI spatial resolution). From the original resolution of each dataset, an aggregation was performed to obtain the datasets at the three resolutions (2 km, 1 km and 250 m). The S2 Table reports the aggregation methods.

Statistical analysis

For each of the three spatial resolutions (2 km, 1 km and 250 m), the following analyses were run.

All variables were standardized to a mean of zero and a standard deviation of one. A Principal Component Analysis (PCA) with a varimax rotation used to make the factors orthogonal and more interpretable, was first performed to take into account the correlation among the variables. The first three principal components were used as the axes for the environmental data space and the basis for the ecoregionalization and their scores were scaled in a 0–255 range for an RGB colour triplet representation [48,49]. All the values in the 10th percentile were assigned to zero; all the values higher than the 90th percentile were set to 255, linearly scaling all the others. In this way we mapped the first, second, and third principal component scores of each pixel to a red–green–blue (RGB) colour triplet. Pixels containing similar climatic and environmental combinations are coloured similarly.

The three components were then processed in a multivariate geographic k-medoids cluster analysis, a classical partitioning method that aggregates the dataset of *n* objects into *k* clusters known *a priori* [50]. The partition algorithm around medoids minimizes the average quadratic error between all points and the point centre of the cluster. The 'optimal' number of clusters was identified using the Silhouette technique on a priori groupings ranging between 10 and 50 [51]. This range was chosen to keep all landscape variability in Italy but avoiding a too fragmented clusterization.

Pixels belonging to the same cluster were displayed with the colour of the medoid of the corresponding cluster, thus colours reflect the distance in terms of information: similar colours reflect smaller distances between the cluster medoids, i.e. the pixels have similar environmental characteristics.

Once the ecoregionalization process was run at each spatial resolution, the three outputs were compared; for each pairwise combinations of resolutions i and j (250 m vs 1 km, 250 m vs 2 km, 1 km vs 2 km), we calculated:

- a confusion matrix with the percentage of the area that migrated from each cluster of resolution i to all clusters of resolution j;

- a confusion matrix with the Euclidean distance among the medoids in the 7-dimension space of the normalised variables to define the environmental similarity between each cluster of resolution i and clusters of resolution j.

This process provides a measure of the agreement between resolutions i and j.

All the statistical analyses were performed in R software version 3.3.2 (R Core team, https://www.r-project.org/) using *rasterVis* [52], *sp* [53] and *raster* packages for spatial analysis and visualization, *fpc* package [54] for cluster analysis, *rgl* package [55] and *vioplot* [56] for clusters representation.

Entomological and disease case studies

To discuss the possible use of the ecoregionalization process, two case studies were investigated verifying the overlapping between specific ecoregions (output result at 250 m resolution) and the spatial distribution of *C. imicola* and *C. obsoletus/scoticus* bluetongue vector and West Nile disease outbreaks. Although the WNV has a complex life cycle, involving several bird species as reservoir hosts, in Italy the West Nile virus circulation and transmission are indisputably closely linked to the presence and distribution of its main vector, *Culex pipiens s.l.* [57,58].

BT vectors dataset. An Entomological Surveillance National Plan for bluetongue is in place since 2000 in Italy [59], stating field protocols for the collection of *Culicoides* and laboratory protocols for insects identification [60]. Information on single night catches is collected in a centralised database, from which presence/absence and abundance of *Culicoides* species incriminated as BTV vectors were derived: *C. imicola* and *C. obsoletus/scoticus* (this taxon including the two cryptic species *C. obsoletus* and *C. scoticus*) [61,62]. Collections performed during winter period (December-March), with a number of insects lower than 5 per night and with no specimens belonging to one of the considered vector species were removed, and only correctly georeferenced sites were kept for the analysis. From the original 150,000 collections from August 2000 to August 2017, those analysed for *C. imicola* were 73,182, located in 3,630 sites across Italy; 6,898 collections were analysed in 1,511 sites for *C. obsoletus/scoticus*. For each cluster, the percentage of positive sites (a positive site is defined as a site with at least one specimen of the considered vector species collected at least in one night) and the average abundance in positive sites (through the whole period considered), were calculated. Since the number of sites was variable in the clusters, the uncertainty of the percentage was calculated using a Beta distribution and its 95% confidence intervals.

WN outbreaks dataset. The re-occurrence of WND in Italy in 2008 heavily affected the Po river delta in Northern Italy, leading to the establishment of an integrated national surveillance plan that combines human cases reported by local health authorities and animal surveillance data [63–65]. Besides, extensive entomological regional surveillance plans were put in place in the endemic areas of Northern Italy, nevertheless the entomological data are not available at national level. As reported by Calzolari *et al.* [57], in the endemic area of Pianura Padana, a robust entomological monitoring allowed the early detection of WNV in *Culex pipiens s.l.* pools before the occurrence of human cases. Thus we can assume that outbreaks locations represent a good proxy for *Culex pipiens s.l.* presence being this vector the species most involved in the WNV circulation in Italy [57,58].

WND outbreaks reported in Italy in the frame of the animal surveillance were then investigated to verify consistency with ecoregions. No human cases were included for privacy reason, but this does not affect the results as their distribution is within the geographical distribution of animal cases [64].

WN virus detection and seroconversions in horses and birds, equine clinical cases and positive entomological collections were derived from the Italian National Surveillance Information System (https://www.vetinfo.sanita.it/ accessed 2nd March 2018) collecting the notification of animal disease outbreaks in Italy [66].

Results

Ecoregionalization

Fig 1 shows the seven environmental and climatic input variables processed for PCA. The S1 Fig reports the correlation matrix among the seven variables.

The varimax rotation matrix highlights the relations between the input variables and the factors derived from PCA (Table 1 for 250 m resolution, S3 and S4 Tables for 1 km and 2 km respectively). The proportion of variance of the first three components was 79.35% at 250 m, 80.06% at 1 km, and 79.78% at 2 km spatial resolutions.

The first three factors of the PCA were associated to the colour triplet Blue-Green-Red: PC1 "Mountains", which was mainly driven by steep slopes, low mean temperature, high precipitations, was associated to the Blue colour in reverse mode (i.e. light blue is associated to higher values of PC1, darker blue goes to pixels with low PC1); PC2 "Vegetation greenness", characterised by low seasonal temperature variation and high vegetation presence, was associated to the Green colour in reverse mode; PC3 "Length of hot season", characterised by a long hot season was associated to the Red colour (S2 Fig). When applying this RGB colour scheme, the resulted map showed similar pixels in similar colours (Fig 2).

The multivariate k-medoids geographic clustering, applied to the optimal number of clusters identified by the Silhouette (S3 Fig), produced 11 clusters at 2 km spatial resolution, 10 clusters at 1 km and 22 clusters at 250 m (Fig 3). The total number of pixels in each ecoregion map is 79,619 (2 km), 312,673 (1 km) 4,869,825 (250 m) respectively (S5 Table).

The main characteristics of each ecoregion are shown through violin plots of the input variables (Fig 4 for 250 m resolution; S4 and S5 Figs for 1 and 2 km spatial resolutions); the shape and the size of the violins describe the distribution of the values around the centroid, i.e. it is a quantitative measure of variability of each factor within an ecoregion.

At 250 m resolution, clusters representing mountainous areas (1–7) have a prevailing purple-pink colour with low mean temperature, above average seasonal variance in temperature, low average values of vegetation greenness and high annual variations in vegetation, with consistent rains and high local variance in elevation. The green-based clusters (approx. 8–17) have a common high value of average vegetation, with the variance across the year depending on the land cover (low variation for forests, high for cultivated fields). The differences among the clusters of this group are mainly related to topography, rain and temperature variation. The group of clusters (approx. 18–22) are red-based, highlighting their characteristics of general high mean temperature, high variance across the year and presence of long hot season, low rainfall rates and flat areas. The clusters are differentiated by distinct crop types reflecting in different amount and seasonality of vegetation indices.

The robustness of the three output results was evaluated through the total area migrating from one cluster at resolution *i* to the more similar ones in resolution *j*. The three pairwise comparisons show that the 80% of the area of a classification migrated to the most similar and to the second most similar clusters of the other classification (Fig 5).
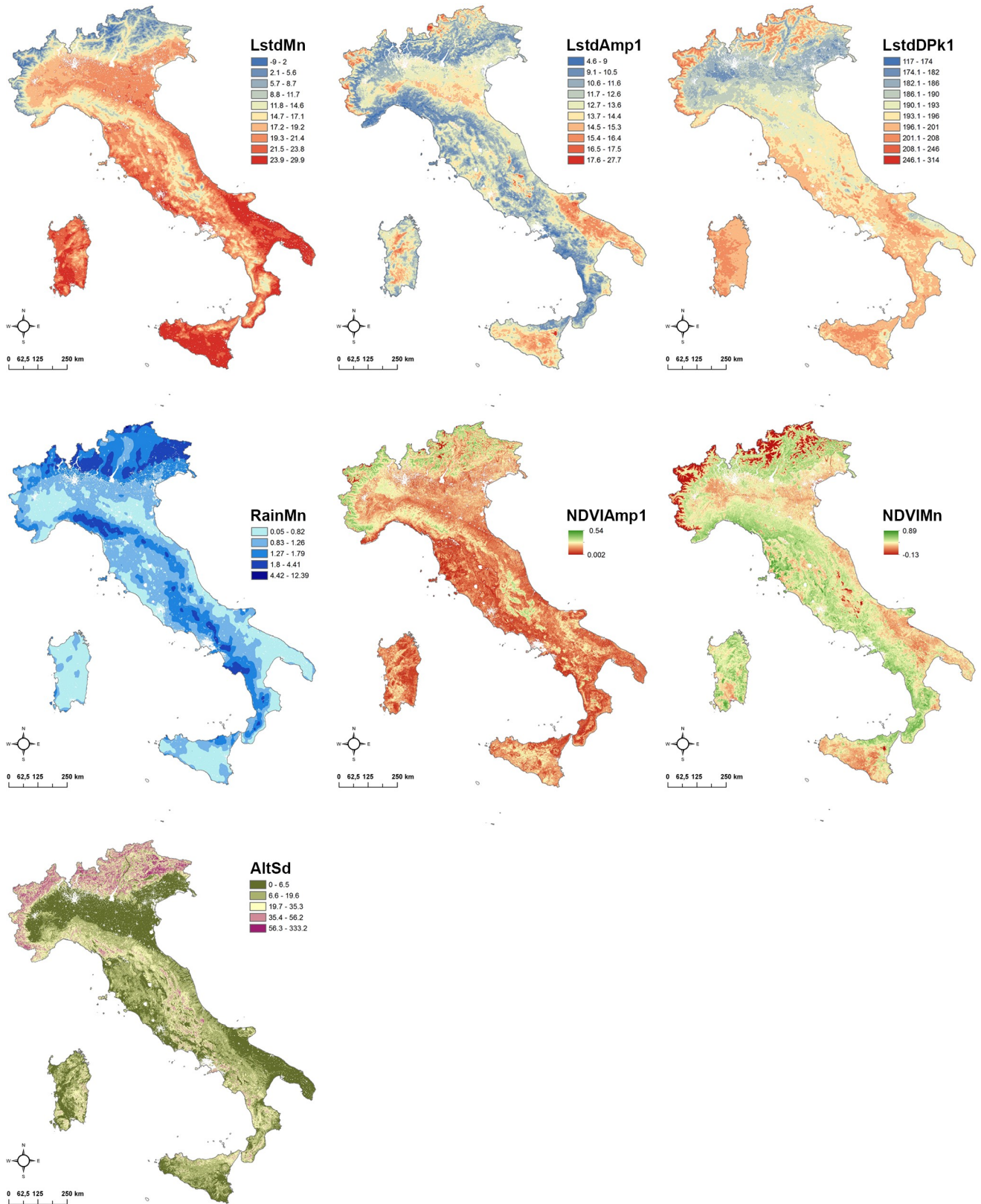
Fig 1. Raster input variables. The input variables are: mean daytime land surface temperature (LstdMn), mean amplitude of annual cycle of land surface temperature (LstdAmp1), peak timing of annual cycle of land surface temperature (LstdDPk1), daily mean amount of rainfall (RainMn), mean amplitude of annual cycle of greenness values (NDVIAmp1), mean greenness value (NDVIMn), standard deviation of altitude (AltSd).

https://doi.org/10.1371/journal.pone.0219072.g001

Entomological and disease case studies

As Fig 6 shows, *C. imicola* distribution is clustered in specific areas of Southern Italy. Conversely, *Culicoides obsoletus/scoticus* is widespread across the country. The number of trapping sites and collections performed per cluster is reported in Table 2. Ecoregions 1, 2 and 6 had no trapping site as they represent mountain areas with no (or few) livestock.

Fig 7 shows the vector taxa as quantified by the proportion of positive sites (sites in which the species was present) by cluster (left) and the maximum number of collected midges in one night collection (i.e. maximum number of specimens) per positive site (right). The figure shows how *C. imicola* is mostly distributed within the two reddish ecoregions 21 and 22, where the percentage of positive sites is 43% (95% confidence interval 39%-47%) and 61% (95% confidence interval 56%-66%) respectively. When considering only the positive *C. imicola* sites (n = 851), the majority of sites are either located in ecoregions 21 and 22 (487 presence sites) or within 2 km from these two ones (278 positive sites).

C. obsoletus/scoticus, are widespread among all the ecoregions identified, showing adaptability to a wide range of climatic and environmental conditions: for *C. obsoletus/scoticus* the percentage of positive sites ranged from 71% to 100% in all the clusters, although some of them (i.e. clusters 3, 4, 5, 11) have only few sites as proven by the wide confidence interval bar.

Fig 8 reports the spatial distribution of 1,259 WND outbreaks notified in the frame of the animal surveillance in Italy in the years 2008–2016, overlayed to the ecoregions 9, 14, 17 and 21. Despite the even distribution of equine farms across the country, the 87% of outbreaks are located in the four mentioned ecoregions.

Discussion

In this paper, a multivariate statistical clustering algorithm was applied to seven climatic and environmental variables to classify the Italian territory in similar regions at three spatial resolutions, 250 meters, 1 km and 2 km delineating 22, 10 and 11 ecoregions, respectively.

The coarser ecoregion division (2 km pixel resolution) accurately captures intuitively understood environmental differences in Italy. As in the most recent hierarchical classification available [2], although not properly comparable because of the differences in data and approach applied, the map distinguishes the high mountains of Alps (cluster 1), the pre-Alps and Apennines chain (clusters 2 and 3), the flat and cultivated areas of Po valley and of part of South peninsula (clusters 4 and 8), the hilly rainy areas on the Tyrrhenian coast with stable

Table 1. Relationship between the input variables and the factors of the PCA at 250 m spatial resolution.

	PC1 (blue)	PC2 (green)	PC3 (red)	PC4	PC5	PC6	PC7
LstdMn	0.53	-0.17	0.04	-0.22	0.23	-0.27	0.72
LstdAmp1	0.35	0.52	-0.14	0.07	-0.07	-0.68	-0.34
LstdDPk1	0.05	0.33	0.83	-0.15	0.38	0.12	-0.13
NDVIMn	-0.01	-0.67	0.16	-0.42	0.04	-0.38	-0.44
NDVIAmp1	-0.34	0.35	-0.35	-0.76	0.23	0.03	0.06
RainMn	-0.49	-0.08	-0.08	0.41	0.64	-0.39	0.12
AltSd	-0.48	0.08	0.36	-0.08	-0.58	-0.39	0.37
Cumulative proportion of variance	*0.39*	*0.64*	*0.79*	*0.87*	*0.93*	*0.97*	*1.00*

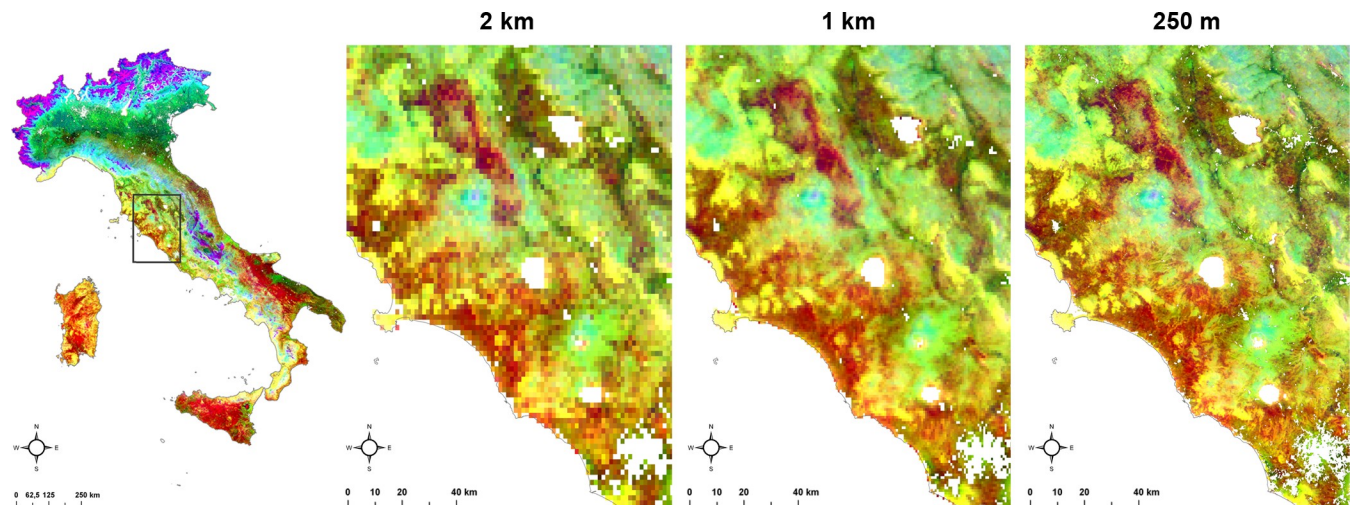https://doi.org/10.1371/journal.pone.0219072.t001

Fig 2. RGB classification: Pixels similar in terms of climate and environment are similar in colours. The three spatial resolutions 2 km, 1 km, 250 m are reported from left to right for a zoomed area (black box on the left map of Italy).

vegetation along the year (clusters 5 and 7), the pre-Apennines hills (most of the territory below 600 m a.s.l.) (cluster 6), the very hot areas of southern Italy and islands without rain with little vegetation fluctuating during the year (cluster 9) and very flat (cluster 10), the undulating very hot landscapes (cluster 11). This coarser classification is consistent with the one at 1 km resolution which defined 10 ecoregions. The number of classes increases to 22 at 250 m pixel resolution, giving further details nested in the broader patterns of the previous two classifications.

Rather than rely on expertise or follow a hierarchical deductive method [2], the process reads the environmental information in the 7-space data, and statistically groups them, assigning a colour to each group: similar colours (in the RGB palette) mean similar environment. The top of Alps in northern Italy and the top of Apennine chain are more purple/blue-related (Blue = "mountains"), and they are characterized by steep slopes, low mean temperature and high precipitations. Southern Italy, Sardinia and Sicily islands are more reddish (Red = "Length of hot season") being dominated by long hot season, with a mix of green (Green = "Vegetation greenness"), so getting to orange-brown. Green areas are related to high vegetation (either as cultivated areas or natural landscapes) and to moderate seasonal temperature variation, the latter mitigated by the presence of vegetation leaves.

Pixels with similar environments should be classified in the same ecoregion even if widely separated geographically [67]. For this reason, the geographical position of the pixels and the geographical neighbourhood were deliberately taken out from the clustering process so to identify clusters also spatially disjoint. Notwithstanding, ecoregions tend to be geographically cohesive because of the spatial autocorrelation usually present in the environmental data [48]. There is generally a gradient in colour among close pixels, highlighting similarity in the environmental conditions and information consistency, and when present, the salt and pepper effect is mainly due to the typical fragmented land use of Italian territory.

The ecoregions maps show a geographical coherence at the three spatial resolutions (250 m, 1 km, 2 km): the 80% of the pixels of a resolution migrated to the most similar and to the second most similar cluster of the other resolutions. This means that moving from a coarser to a finer spatial resolution (or vice-versa), the area covered by the corresponding pixels is classified similarly (in terms of colours, i.e. of eco-climatic characteristics). The choice of the resolution to be selected/to be produced depends mainly on the final use of the classification.
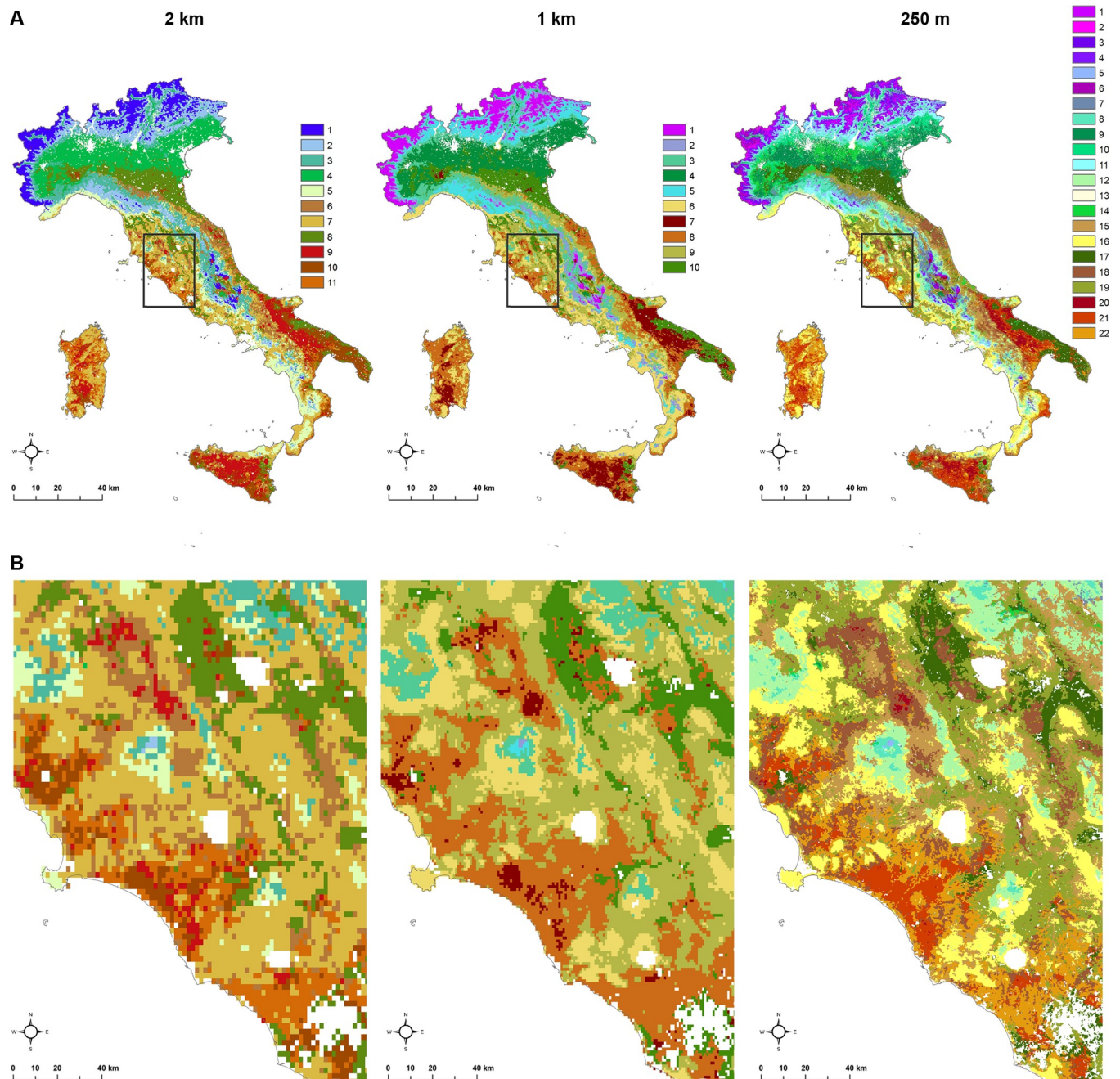
Fig 3. Cluster results at the spatial resolutions of 2 km, 1 km, 250 m from left to right, for whole Italy (A) and in a zoomed area (B).

https://doi.org/10.1371/journal.pone.0219072.g003

To our knowledge, this is the first attempt at producing a quantitative data-driven approach to delineate eco-climatic areas in Italy using variables associated to VBDs and VBDs vectors. The idea arises from the potential application of such a layer in support to surveillance and control management.

It is reasonable to assume a more probable spread of a VBD within each ecoregion, given the similar characteristics, rather than between different clusters. This is a first useful information from a public health perspective. Pixels of the same colour belong to the same cluster, and
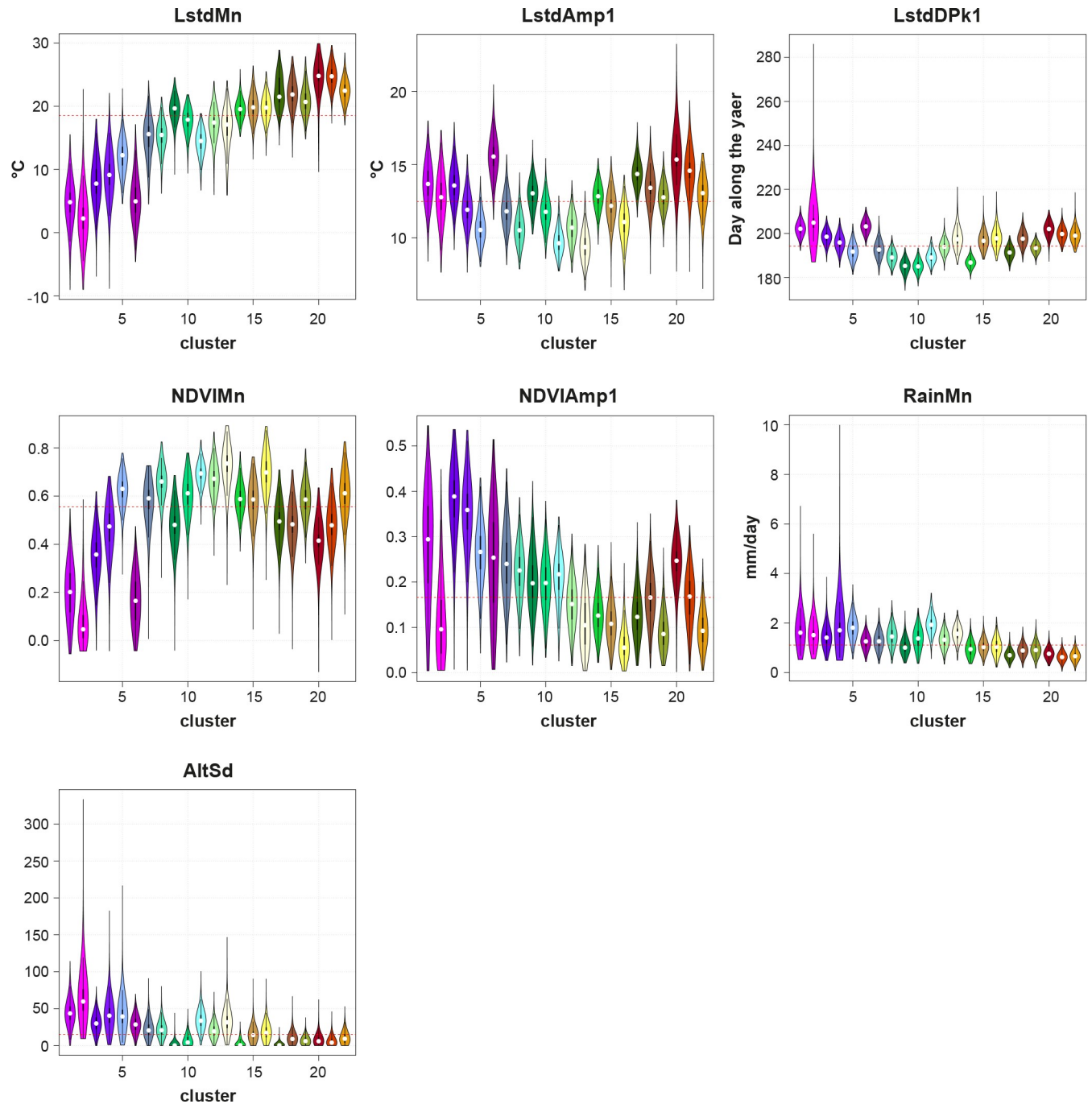
Fig 4. Violin plots depicting the characteristics of the seven input variables in the 22 ecoregions (250 m spatial resolution). The white marker is the median of the data, the box indicates the interquartile range, with a kernel density overlaid. The red line reports the average value across all clusters. Each violin has the same color of the corresponding ecoregion.

therefore have, on average, the same environmental characteristics. The WND case study highlights a strong correspondence between the location of WND cases and few greenish ecoregions (9, 14, 17 and 21 at 250 m spatial resolution). Since 2008, WND strongly hit Northern Italy and in particular the Po Valley [68]. This intensively cultivated area is characterized by the presence of rivers and wetlands where ornithophilic mosquitoes and migratory birds are in
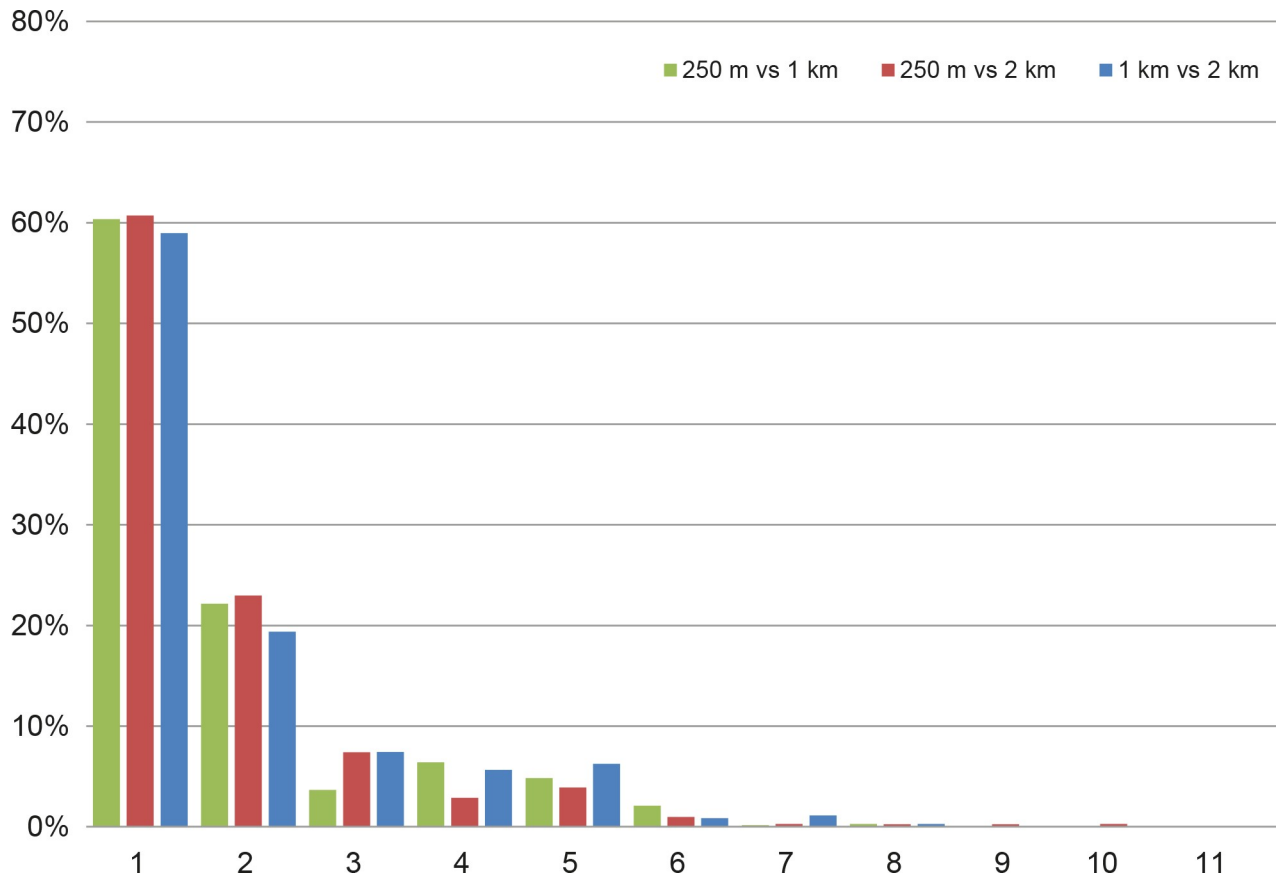
Fig 5. Migration area of the pixels from the resolution *i* to the more similar clusters in resolution *j*. The three pairwise comparisons show that the 80% of the area of a classification migrated to the most similar and to the second most similar clusters of the other classification.

Fig 6. Spatial distribution of *C. imicola*, *C. obsoletus/scoticus* in Italy. The maximum specimens catched in the most favourable period (April–November) in the period August 2000- August 2017.

Table 2. Number of trapping sites and collections per ecoregion for *Culicoides imicola, C. obsoletus/scoticus*.

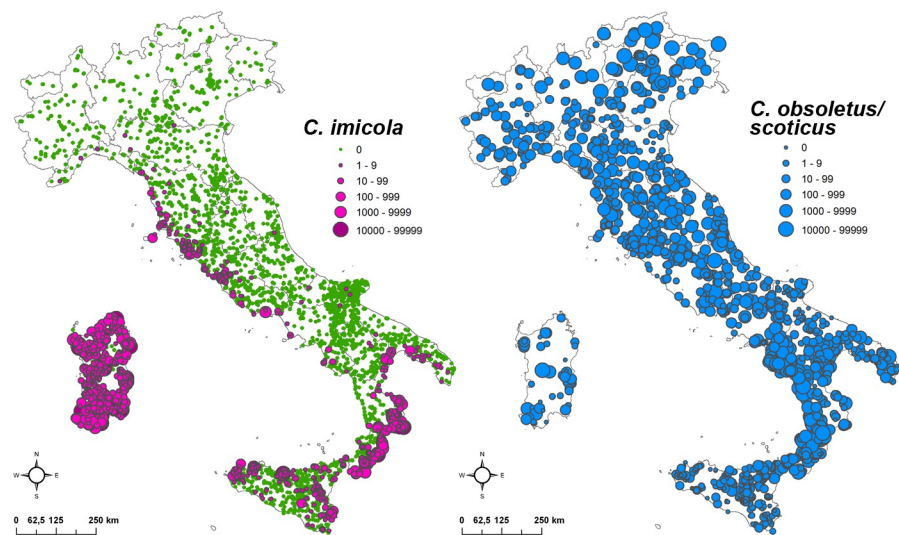| ecoregion	*C. imicola* dataset		*C. obsoletus/scoticus* dataset	
	Number of sites	Number of collections	Number of sites	Number of collections
1	0	0	0	0
2	0	0	0	0
3	10	643	3	16
4	3	9	2	2
5	12	373	5	16
6	0	0	0	0
7	47	1511	23	70
8	48	2717	32	137
9	100	6817	55	638
10	68	5244	48	382
11	5	9	2	2
12	151	2373	75	191
13	26	133	15	37
14	76	4756	46	395
15	422	6228	180	831
16	307	3549	162	369
17	320	10372	148	877
18	376	4963	139	955
19	484	8693	194	733
20	150	1561	52	156
21	684	7974	213	602
22	341	5257	117	489
TOTAL	3,630	73,182	1,511	6,898

close contact, allowing the transmission of WNV and the establishment of enzootic cycle of infection. The natural areas as well as the agricultural and urban areas, characterized by the constant presence of stagnant water provide potential breeding habitats for different mosquito species, including *Culex pipiens* s.l., considered the main vector of WNV, and of other *Flavivirus* such as USUTU virus, in the whole Italian Peninsula [57,58]. Although 80% of the WND outbreaks fall in the greener 3 ecoregions mentioned, the 7% are found in reddish cluster 21 representing risk areas in the southern part of Italy. The low percentage of cases in these areas and their clustered distribution inside the ecoregion, can be explained by the local presence of specific factors, additional to the ones included in the present ecoregion definition, such as humid areas and migratory bird settlements (e.g. ponds in Oristano province in Western Sardinia; Lentini Lake in Eastern Sicily, Stagnone Lagoon and Marsala salt ponds in Western Sicily).

Second, the spread of disease from one part of Italy to another geographically distant, but belonging to the same ecoregion, would be an important risk factor to take into account for disease control measures. The spatial structure of the ecoregions in relation to known infections could therefore be used to prioritize animal movement control such as to prevent infections from one geographic part of an ecoregion to another.

Third, this approach could also inform entomological surveillance strategies identifying similar places far apart in which to intensify or reduce the sampling activity. In the case of *C. imicola* in Italy, this BT vector revealed a strong association with two ecoregions, the 21 and 22. *Culicoides imicola* has already been proved to be associated with specific habitat characterised by high mean temperatures, dry environments, flatten and sunny surfaces, low vegetation
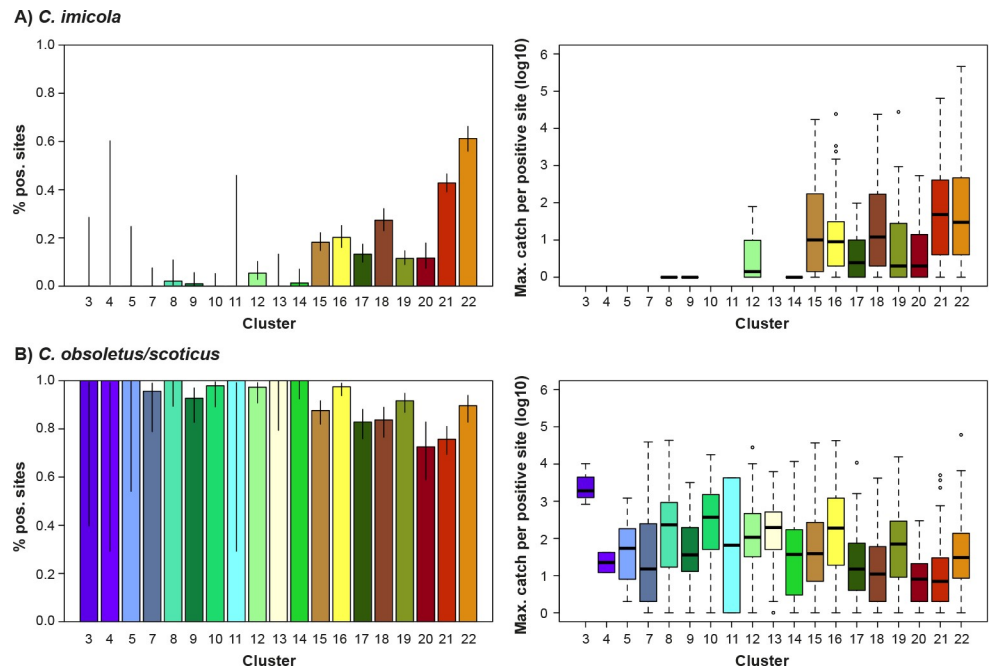
Fig 7. Distribution of *C. imicola* (A) and *C. obsoletus/scoticus* (B) in the 22 Italian ecoregions, as quantified by the proportion of positive sites by cluster (left) and the maximum catches per positive site (right).

[17,69,70] and in Italy the combination of these favourable factors clearly delineates the two specific ecoregions. These two ecoregions are mainly located in Sardinia, Sicily, Ionian part of toe of Italy, Tyrrhenian coast of Lazio and Tuscany, where the distribution of *C. imicola* is documented. The ecoregion map here produced highlights that these clusters (at risk) are also present in river valleys of Adriatic coast where few or none cases were reported, and in the area laying between Apulia and Basilicata regions; a deeper entomological investigation might verify if and which local conditions may locally decrease the vector population (type of crops, farming types, absence of water for larval sites, etc.).

If the approach captures the environmental preferences of *Culicoides imicola*, whose distribution in Italy is spatially constrained, it seems less accurate with the *Culicoides Obsoletus/scoticus*, a species with a broader distribution across the country. *Culicoides obsoletus/scoticus* is a taxon not showing evidence of prevalence in specific ecoregions, in agreement with previous studies [62,71]. In terms of abundances, the highest values were found in ecoregions characterized by high vegetation indices, but with different seasonal patterns. The two cryptic species *C. obsoletus* and *C. scoticus* are both related to vegetation, but with *C. scoticus* associated to more natural environments and *C. obsoletus* being a more widespread species, including urban areas [25,62]. *Culicoides obsoletus* in Italy is largely the most abundant of the two [62,72], thus it might have played a major role in the even distribution of positive sites. A more detailed set of variables [25] could derive a tailored ecoregionalization for these two species.

In general terms, the availability of specific environmental datasets is a key point in the development of such an approach. Some variables such as soil composition, permeability and capacity of retain water are key factors in defining vector habitat [71], but detailed information covering the entire Italian territory is not available and the ecoregionalization might be lacking in considering this aspect.

An additional consideration is the role of hosts in defining ecoregions. This aspect, deliberately taken out from our generic classification, it is important when tailoring the method to a
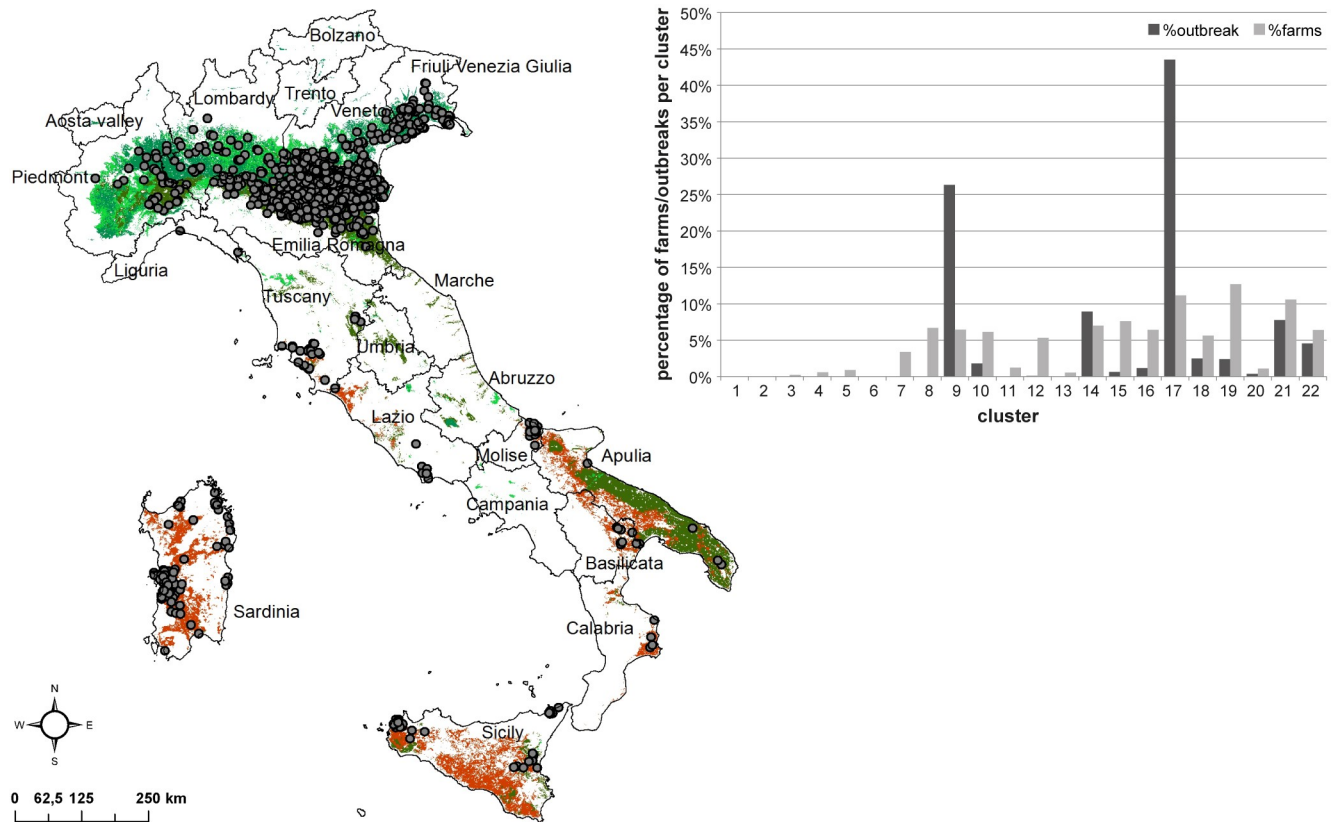
Fig 8. Spatial distribution of 2008–2016 WND outbreaks reported in Italy in the frame of animal surveillance overlaid to the clusters 9, 14, 17 and 21 where the 87% of outbreaks was located (map). In the graph, the percentage of outbreaks (in dark grey) and equine farms (in light grey) per ecoregion are reported.

specific pathogen/disease because the ecoregions will be based not only on climatic and environmental factors (suitability for the vector), but also on the availability of hosts (suitability for the disease).

Future perspectives

The multivariate clustering method can be extended to a geospatiotemporal clustering in which multi-dimensional datasets of temporal variables would be used for describing and tracking the similarity of ecosystem properties through time, either retrospective or forward. This could be used to detect changes in landscape features that may underline the areas of emergence of new diseases, the resurgence of old ones, or changes in transmission opportunities of established vector-borne pathogens [12].

Applied in Italy on weekly or monthly basis, a "dynamic ecoregion" definition would help to define which areas are similar in which period, and it might represent an added value in case of virus entrance in a specific time of the year so to adopt priority intervention and control measure in similar areas at higher risk. Moreover the method could be replicated including datasets on present and future climatic conditions, to investigate the evolution of ecoregions in Italy under climate change scenarios [21,26].

Although it still remains a rich and complex process requiring different expertise, ecoregionalization could contribute in the development of accurate early warning systems so to timely prevent and control VBDs spread through a harmonized and targeted surveillance systems, in space and time, for host and vectors.

Supporting information

S1 Table. Number of images processed and archive size of the climatic and environmental products.
(DOCX)

S2 Table. Aggregation methods of the original dataset to obtain the resolutions (2 km, 1 km and 250 m) of the factors used in the statistical analysis.
(DOCX)

S3 Table. Relationship between the input variables and the factors of the Principal Component Analysis at 1 km spatial resolution. The first factor of PCA is then associated to the Blue colour channel in reverse mode, the second component is associated to Green in reverse mode, the third component PC3 to the Red.
(DOCX)

S4 Table. Relationship between the input variables and the factors of the Principal Component Analysis at 2 km spatial resolution. The first factor of PCA is then associated to the Blue colour channel in reverse mode, the second component is associated to Green in reverse mode, the third component PC3 to the Red.
(DOCX)

S5 Table. Number of pixels in each ecoregion map, divided by cluster.
(DOCX)

S1 Fig. Correlation matrix among the seven variables at 2 km spatial resolution.
(TIF)

S2 Fig. Association between PCA components and RGB values at 250 m spatial resolution.
(TIF)

S3 Fig. The optimal number of clusters for each spatial resolution identified by the Silhouette technique: 11 clusters at 2 km, 10 clusters at 1 km and 22 clusters at 250 m spatial resolution.
(TIF)

S4 Fig. Violin plots depicting the characteristics of the seven input variables in the 10 ecoregions (1 km spatial resolution). The violin plots show the probability density of the data at different values; they include a marker (white point) for the median of the data and a box indicating the interquartile range, with a kernel density overlaid. The red line reports the average value across all clusters. Each violin has the same color of the corresponding ecoregion.
(TIF)

S5 Fig. Violin plots depicting the characteristics of the seven input variables in the 11 ecoregions (2 km spatial resolution). The violin plots show the probability density of the data at different values; they include a marker (white point) for the median of the data and a box indicating the interquartile range, with a kernel density overlaid. The red line reports the average value across all clusters. Each violin has the same color of the corresponding

ecoregion.
(TIF)

S1 File. Structure of entomological and farm data.
(XLSX)

S2 File. Raster data of ecoregion map with 22 classes at 250 meters spatial resolution in geotiff format.
(7Z)

Acknowledgments

Author Contributions

Conceptualization: Marius Gilbert, Annamaria Conte.

Data curation: Carla Ippoliti, Maria Goffredo, Giuseppe Mancini, Gabriele Curci, Serena Falasca, Susanna Tora, Alessio Di Lorenzo, Michela Quaglia.

Formal analysis: Carla Ippoliti, Luca Candeloro, Annamaria Conte.

Methodology: Carla Ippoliti, Luca Candeloro, Annamaria Conte.

Project administration: Annamaria Conte.

Software: Luca Candeloro, Marius Gilbert.

Supervision: Annamaria Conte.

Validation: Carla Ippoliti, Marius Gilbert, Maria Goffredo, Giuseppe Mancini, Annamaria Conte.

Visualization: Carla Ippoliti, Luca Candeloro, Marius Gilbert.

Writing – original draft: Carla Ippoliti, Marius Gilbert, Annamaria Conte.

Writing – review & editing: Carla Ippoliti, Luca Candeloro, Marius Gilbert, Maria Goffredo, Giuseppe Mancini, Gabriele Curci, Serena Falasca, Susanna Tora, Alessio Di Lorenzo, Michela Quaglia, Annamaria Conte.

References

1. Bailey RG. Identifying ecoregion boundaries. Environ Manage. 2004; 34 Suppl 1:S14–26.

2. Blasi C, Capotorti G, Copiz R, Guida D, Mollo B, Smiraglia D, et al. Classification and mapping of the ecoregions of Italy. Plant Biosystems—An International Journal Dealing with all Aspects of Plant Biology. 2014 Nov 2; 148(6):1255–345.

3. Bailey RG. Ecoregions: The Ecosystem Geography of the Oceans and Continents. 2nd ed. New York: Springer-Verlag; 2014 [cited 2019 Mar 13]. Available from: https://www.springer.com/fr/book/9781493905232

4. Bailey RG. Ecosystem Geography | Robert G. Bailey | Springer. 1996 [cited 2017 Nov 24]. Available from: //www.springer.com/us/book/9781461223580

5. European Environment Agency (EEA). CLC 2012—Copernicus Land Monitoring Service [Internet]. https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012. 2012 [cited 2019 Mar 26]. Available from: https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012

6. Rivas Martínez S, Rivas Sáenz S, Penas Merino A. Worldwide bioclimatic classification system. Global Geobotany. 2011 Nov; 1(1):1–638.

7. Pesaresi S, Biondi E, Casavecchia S. Bioclimates of Italy. Journal of Maps. 2017 Nov 30; 13(2):955–60.

8. Blasi C, Capotorti G, Alós Ortí MM, Anzellotti I, Attorre F, Azzella MM, et al. Ecosystem mapping for the implementation of the European Biodiversity Strategy at the national level: The case of Italy. Environmental Science & Policy. 2017 Dec 1; 78(Supplement C):173–84.

9. Jamison A, Tuttle E, Jensen R, Bierly G, Gonser R. Spatial ecology, landscapes, and the geography of vector-borne disease: A multi-disciplinary review. Applied Geography. 2015 Sep 1; 63(Supplement C):418–26.

10. Reisen WK. Landscape Epidemiology of Vector-Borne Diseases. Annual Review of Entomology. 2010 Jan; 55(1):461–83.

11. Reisen WK. Epidemiology. In: Marcondes CB, editor. Arthropod Borne Diseases [Internet]. Cham: Springer International Publishing; 2017 [cited 2019 Mar 8]. p. 7–34. Available from: https://doi.org/10.1007/978-3-319-13884-8_2

12. Kitron U. Landscape Ecology and Epidemiology of Vector-Borne Diseases: Tools for Spatial Analysis. J Med Entomol. 1998 Jul 1; 35(4):435–45. https://doi.org/10.1093/jmedent/35.4.435 PMID: 9701925

13. Lambin EF, Tran A, Vanwambeke SO, Linard C, Soti V. Pathogenic landscapes: Interactions between land, people, disease vectors, and their animal hosts. International Journal of Health Geographics. 2010 Oct 27; 9:54. https://doi.org/10.1186/1476-072X-9-54 PMID: 20979609

14. McIntyre KM, Setzkorn C, Hepworth PJ, Morand S, Morse AP, Baylis M. Systematic Assessment of the Climate Sensitivity of Important Human and Domestic Animals Pathogens in Europe. Scientific Reports. 2017 Aug 2; 7(1):7134.

15. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. eLife Sciences. 2015 Jun 30; 4:e08347.

16. Ozdenerol E, Taff G, Akkus C. Exploring the Spatio-Temporal Dynamics of Reservoir Hosts, Vectors, and Human Hosts of West Nile Virus: A Review of the Recent Literature. International Journal of Environmental Research and Public Health. 2013 Oct 25; 10(11):5399–432. https://doi.org/10.3390/ijerph10115399 PMID: 24284356

17. Conte A, Goffredo M, Ippoliti C, Meiswinkel R. Influence of biotic and abiotic factors on the distribution and abundance of *Culicoides imicola* and the Obsoletus Complex in Italy. Veterinary Parasitology. 2007 Dec 25; 150(4):333–44. https://doi.org/10.1016/j.vetpar.2007.09.021 PMID: 17997043

18. Tran A, Sudre B, Paz S, Rossi M, Desbrosse A, Chevalier V, et al. Environmental predictors of West Nile fever risk in Europe. International Journal of Health Geographics. 2014; 13:26. https://doi.org/10.1186/1476-072X-13-26 PMID: 24986363

19. Bisanzio D, Giacobini M, Bertolotti L, Mosca A, Balbo L, Kitron U, et al. Spatio-temporal patterns of distribution of West Nile virus vectors in eastern Piedmont Region, Italy. Parasit Vectors. 2011 Dec 9; 4:230.

20. Conte A, Candeloro L, Ippoliti C, Monaco F, Massis FD, Bruno R, et al. Spatio-Temporal Identification of Areas Suitable for West Nile Disease in the Mediterranean Basin and Central Europe. PLOS ONE. 2015 Dec 30; 10(12):e0146024.

21. Guichard S, Guis H, Tran A, Garros C, Balenghien T, Kriticos DJ. Worldwide Niche and Future Potential Distribution of *Culicoides imicola*, a Major Vector of Bluetongue and African Horse Sickness Viruses. Mans BJ, editor. PLoS ONE. 2014 Nov 12; 9(11):e112491.

22. Conte A, Ippoliti C, Calistri P, Pelini S, Savini L, Salini R, et al. Towards the identification of potential infectious sites for bluetongue in Italy: a spatial analysis approach based on the distribution of *Culicoides imicola*. Vet Ital. 2004 Sep; 40(3):311–5.

23. Meiswinkel R, van Rijn P, Leijs P, Goffredo M. Potential new *Culicoides* vector of bluetongue virus in northern Europe. Vet Rec. 2007 Oct 20; 161(16):564–5. PMID: 17951565

24. Dijkstra E, Ven IJK van der, Meiswinkel R, Hölzel DR, Rijn PA van, Meiswinkel R. *Culicoides chiopterus* as a potential vector of bluetongue virus in Europe. Veterinary Record. 2008 Mar 29; 162(13):422–422.

25. Ippoliti C, Goffredo M, Salini R, Bocci M, Quaglia M, Federici V, et al. Eco-climatic indicators for three *Culicoides* species of the Obsoletus complex in Italy. Vet Ital. 2016 Sep 30; 52(3–4):213–22.

26. Samy AM, Peterson AT. Climate Change Influences on the Global Potential Distribution of Bluetongue Virus. PLOS ONE. 2016 Mar 9; 11(3):e0150489.

27. Calistri P, Ippoliti C, Candeloro L, Benjelloun A, El Harrak M, Bouchra B, et al. Analysis of climatic and environmental variables associated with the occurrence of West Nile virus in Morocco. Prev Vet Med. 2013 Jul 1; 110(3–4):549–53. https://doi.org/10.1016/j.prevetmed.2013.02.011 PMID: 23453893

28. Marcantonio M, Rizzoli A, Metz M, Rosà R, Marini G, Chadwick E, et al. Identifying the Environmental Conditions Favouring West Nile Virus Outbreaks in Europe. PLOS ONE. 2015 Mar 24; 10(3):e0121158.

29. Stilianakis NI, Syrris V, Petroliagkis T, Pärt P, Gewehr S, Kalaitzopoulou S, et al. Identification of Climatic Factors Affecting the Epidemiology of Human West Nile Virus Infections in Northern Greece. PLoS ONE. 2016; 11(9):e0161510. https://doi.org/10.1371/journal.pone.0161510 PMID: 27631082

30. Nmor JC, Sunahara T, Goto K, Futami K, Sonye G, Akweywa P, et al. Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. Parasit Vectors. 2013 Jan 16; 6:14.

31. Purse BV, Carpenter S, Venter GJ, Bellis G, Mullens BA. Bionomics of Temperate and Tropical *Culicoides* Midges: Knowledge Gaps and Consequences for Transmission of *Culicoides*-Borne Viruses. Annual Review of Entomology. 2015; 60(1):373–92.

32. Blagrove MSC, Caminade C, Waldmann E, Sutton ER, Wardeh M, Baylis M. Co-occurrence of viruses and mosquitoes at the vectors' optimal climate range: An underestimated risk to temperate regions? PLoS Negl Trop Dis [Internet]. 2017 Jun 15 [cited 2017 Aug 9]; 11(6). Available from: http://www.ncbi. nlm.nih.gov/pmc/articles/PMC5487074/

33. Patz JA, Githeko AK, McCarty JP, Hussein S, Confalonieri U. Climate change and infectious diseases. In: Climate Change and Human Health—Risks and Responses. A. J. McMichael, D. H. Campbell-Lendrum, C. F. Corvalan, K. L. Ebi, A. K. Githeko, Scheraga J. D., and Woodward A. Geneva, Switzerland: WHO; 2003. p. 103–37.

34. Paz S, Semenza J. Environmental Drivers of West Nile Fever Epidemiology in Europe and Western Asia—A Review. International Journal of Environmental Research and Public Health. 2013 Aug 9; 10 (8):3543–62. https://doi.org/10.3390/ijerph10083543 PMID: 23939389

35. Wang J, Rich PM, Price KP. Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. International Journal of Remote Sensing. 2003 Jan; 24(11):2345–64.

36. Anyamba A, Linthicum KJ, Small JL, Collins KM, Tucker CJ, Pak EW, et al. Climate Teleconnections and Recent Patterns of Human and Animal Disease Outbreaks. PLOS Neglected Tropical Diseases. 2012 Jan 24; 6(1):e1465.

37. Caminade C, Ndione JA, Diallo M, MacLeod DA, Faye O, Ba Y, et al. Rift Valley Fever Outbreaks in Mauritania and Related Environmental Conditions. Int J Environ Res Public Health. 2014 Jan; 11 (1):903–18. https://doi.org/10.3390/ijerph110100903 PMID: 24413703

38. Ward MP. Equine West Nile virus disease occurrence and the Normalized Difference Vegetation Index. Preventive Veterinary Medicine. 2009 Mar 1; 88(3):205–12.

39. Estrada-Peña A, Estrada-Sánchez A, de la Fuente J. A global set of Fourier-transformed remotely sensed covariates for the description of abiotic niche in epidemiological studies of tick vector species. Parasit Vectors. 2014 Jul 2; 7:302.

40. Kalluri S, Gilruth P, Rogers D, Szczur M. Surveillance of Arthropod Vector-Borne Infectious Diseases Using Remote Sensing Techniques: A Review. PLOS Pathogens. 2007 Oct 26; 3(10):e116.

41. Roiz D, Ruiz S, Soriguer R, Figuerola J. Landscape Effects on the Presence, Abundance and Diversity of Mosquitoes in Mediterranean Wetlands. PLOS ONE. 2015 Jun 18; 10(6):e0128112.

42. Eastman JR, Sangermano F, Ghimire B, Zhu H, Chen H, Neeti N, et al. Seasonal trend analysis of image time series. International Journal of Remote Sensing. 2009 May 1; 30(10):2721–6.

43. Wan Z. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sensing of Environment. 2014 Jan; 140:36–45.

44. Eastman JR. IDRISI Taiga guide to GIS and image processing. Clark Labs Clark University, Worcester, MA. 2009;

45. Didan K. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Internet]. NASA EOSDIS Land Processes DAAC; 2015 [cited 2018 Dec 7]. Available from: https://lpdaac.usgs. gov/node/844

46. Rebora N, Pignone F, Silvestro F. GRISO: Spatial Interpolation Generator from Rainfall Observations. Geophysical Research Abstracts. 2015; 17(EGU2015-3002–1):1.

47. Piani C, Haerter JO, Coppola E. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor Appl Climatol. 2010 Jan 1; 99(1):187–92.

48. Hargrove WW, Hoffman FM. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environ Manage. 2004; 34 Suppl 1:S39–60.

49. Hargrove WW, Hoffman FM. Using multivariate clustering to characterize ecoregion borders. Computing in science & engineering. 1999; 1(4):18–25.

50. Kaufman L, Rousseeuw PJ. Cluster Analysis. Clustering by Means of Medoids. In: Statistical data analysis based on the L1-norm and related methods. Yadolah Dodge; 1987.

51. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987 Nov 1; 20(Supplement C):53–65.

52. Lamigueiro OP, Hijmans R. Visualization Methods for Raster Data [R package rasterVis version 0.40]. 2016 [cited 2018 May 15];http://CRAN.R-project.org/package=rasterVis. Available from: http://CRAN.R-project.org/package=rasterVis

53. Bivand RS, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R [Internet]. 2nd ed. Springer-Verlag New York; 2013 [cited 2018 May 15]. 405 p. (use R!; vol. 10). Available from: https://www.springer.com/it/book/9781461476177

54. Hennig C. fpc: Flexible Procedures for Clustering. Version: 2.1–11 [Internet]. 2015 [cited 2018 May 15]. Available from: https://CRAN.R-project.org/package=fpc

55. Adler D, Murdoch D, et al. rgl: 3D Visualization Using OpenGL. Version: 0.96.0 [Internet]. 2016 [cited 2018 May 15]. (https://CRAN.R-project.org/package=rgl). Available from: https://CRAN.R-project.org/package=rgl

56. Adler D. vioplot: Violin plot. Version: 0.2 [Internet]. 2005 [cited 2018 May 15]. (https://CRAN.R-project.org/package=vioplot). Available from: https://CRAN.R-project.org/package=vioplot

57. Calzolari M, Pautasso A, Montarsi F, Albieri A, Bellini R, Bonilauri P, et al. West Nile Virus Surveillance in 2013 via Mosquito Screening in Northern Italy and the Influence of Weather on Virus Circulation. PLoS ONE. 2015; 10(10):e0140915. https://doi.org/10.1371/journal.pone.0140915 PMID: 26488475

58. Mancini G, Montarsi F, Calzolari M, Capelli G, Dottori M, Ravagnan S, et al. Mosquito species involved in the circulation of West Nile and Usutu viruses in Italy. Vet Ital. 2017 Jun 30; 53(2):97–110. https://doi.org/10.12834/VetIt.114.933.4764.2 PMID: 28675249

59. Giovannini A, Paladini C, Calistri P, Conte A, Colangeli P, Santucci U, et al. Surveillance system of bluetongue in Italy. Veterinaria Italiana. 2004; 16.

60. Goffredo M, Meiswinkel R. Entomological surveillance of bluetongue in Italy: methods of capture, catch analysis and identification of *Culicoides* biting midges. Vet Ital. 2004; 40(3):260–265. PMID: 20419674

61. Goffredo M, Catalani M, Federici V, Portanti O, Marini V, Mancini G, et al. Vector species of *Culicoides* midges implicated in the 2012-2014 Bluetongue epidemics in Italy. Vet Ital. 2015 Jun; 51(2):131–8. https://doi.org/10.12834/VetIt.771.3854.1 PMID: 26129664

62. Goffredo M, Meiswinkel R, Federici V, Di Nicola F, Mancini G, Ippoliti C, et al. The '*Culicoides obsoletus* group' in Italy: relative abundance, geographic range, and role as vector for Bluetongue virus. Vet Ital. 2016 Sep 30; 52(3–4):235–41. https://doi.org/10.12834/VetIt.35.100.1 PMID: 27723032

63. Napoli C, Iannetti S, Rizzo C, Bella A, Di Sabatino D, Bruno R, et al. Vector Borne Infections in Italy: Results of the Integrated Surveillance System for West Nile Disease in 2013 [Internet]. BioMed Research International. 2015 [cited 2018 Oct 12]. Available from: https://www.hindawi.com/journals/bmri/2015/643439/

64. Rizzo C, Napoli C, Venturi G, Pupella S, Lombardini L, Calistri P, et al. West Nile virus transmission: results from the integrated surveillance system in Italy, 2008 to 2015. Euro Surveill [Internet]. 2016 Sep 15 [cited 2018 Oct 11]; 21(37). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5032855/

65. Rizzo C, Salcuni P, Nicoletti L, Ciufolini MG, Russo F, Masala R, et al. Epidemiological surveillance of West Nile neuroinvasive diseases in Italy, 2008 to 2011. Eurosurveillance. 2012; 17(20):20172. PMID: 22642945

66. Colangeli P, Iannetti S, Cerella A, Ippoliti C, Di Lorenzo A, Santucci U, et al. The national information system for the notification of animal diseases in Italy. Vet Ital. 2011 Sep; 47(3):303–12, 291–301.

67. Cheruvelil KS, Yuan S, Webster KE, Tan P-N, Lapierre J-F, Collins SM, et al. Creating multithemed ecological regions for macroscale ecology: Testing a flexible, repeatable, and accessible clustering method. Ecology and Evolution. 2017 May; 7(9):3046–58. https://doi.org/10.1002/ece3.2884 PMID: 28480004

68. Calistri P, Giovannini A, Hubalek Z, Ionescu A, Monaco F, Savini G, et al. Epidemiology of West Nile in Europe and in the Mediterranean basin. Open Virol J. 2010 Apr 22; 4:29–37.

69. Acevedo P, Ruiz-Fons F, Estrada R, Márquez AL, Miranda MA, Gortázar C, et al. A Broad Assessment of Factors Determining *Culicoides imicola* Abundance: Modelling the Present and Forecasting Its Future in Climate Change Scenarios. PLOS ONE. 2010 Dec 6; 5(12):e14236.

70. Ippoliti C, Gilbert M, Vanhuysse S, Goffredo M, Satta G, Wolff E, et al. Can landscape metrics help determine the *Culicoides imicola* distribution in Italy? Geospatial Health. 2013 Nov 1; 8(1):267–77.

71. Conte A, Ippoliti C, Savini L, Goffredo M, Meiswinkel R. Novel environmental factors influencing the distribution and abundance of *Culicoides imicola* and the Obsoletus Complex in Italy. Vet Ital. 2007 Sep; 43(3):571–80.

72. Goffredo M, Monaco F, Capelli G, Quaglia M, Federici V, Catalani M, et al. Schmallenberg virus in Italy: a retrospective survey in *Culicoides* stored during the bluetongue Italian surveillance program. Prev Vet Med. 2013 Sep 1; 111(3–4):230–6. https://doi.org/10.1016/j.prevetmed.2013.05.014 PMID: 23791122