The use of GIS for Spatial Analysis

Paolo Mulatti

Istituto Zooprofilattico Sperimentale delle Venezie
OIE Collaborating Centre for epidemiology, training and control of emerging avian diseases

World Organisation for Animal Health
OIE Headquarters

Overview

- Why to consider *space* in (Veterinary) Public Health?
- Where to retrieve spatial data?
- Which types of data in spatial epidemiology?
- How to visualize epidemiological data on a map?
- What is Spatial Analysis?
- Cases study

Spatiality in Public Health

- Many elements of interest in (Veterinary) Public Health present a spatial aspect:
  - Farms and other zootechnic premises
  - Demographic data
  - Potential sources of pollution
  - Outbreaks/cases of infectious diseases
  - …

- Infectious diseases spread ***geographically***

Spatiality in Public Health

- *Space* is important to:
  - Understand how a health issue spread
  - Identify potential risk sources
  - Define how risk factors are (geographically) distributed

Define risk areas

Keep trace of control measures

Risk-based control measures
(e.g. Risk Based Surveillance)

Sources for animal and epidemiological data

- Specific research investigations
  - High quality but usually limited spatial extension
- Surveillance data
  - Field data
  - Laboratory data
  - Etc.
- National\International reporting systems
  - Animal Disease Notification System
  - WAHIS

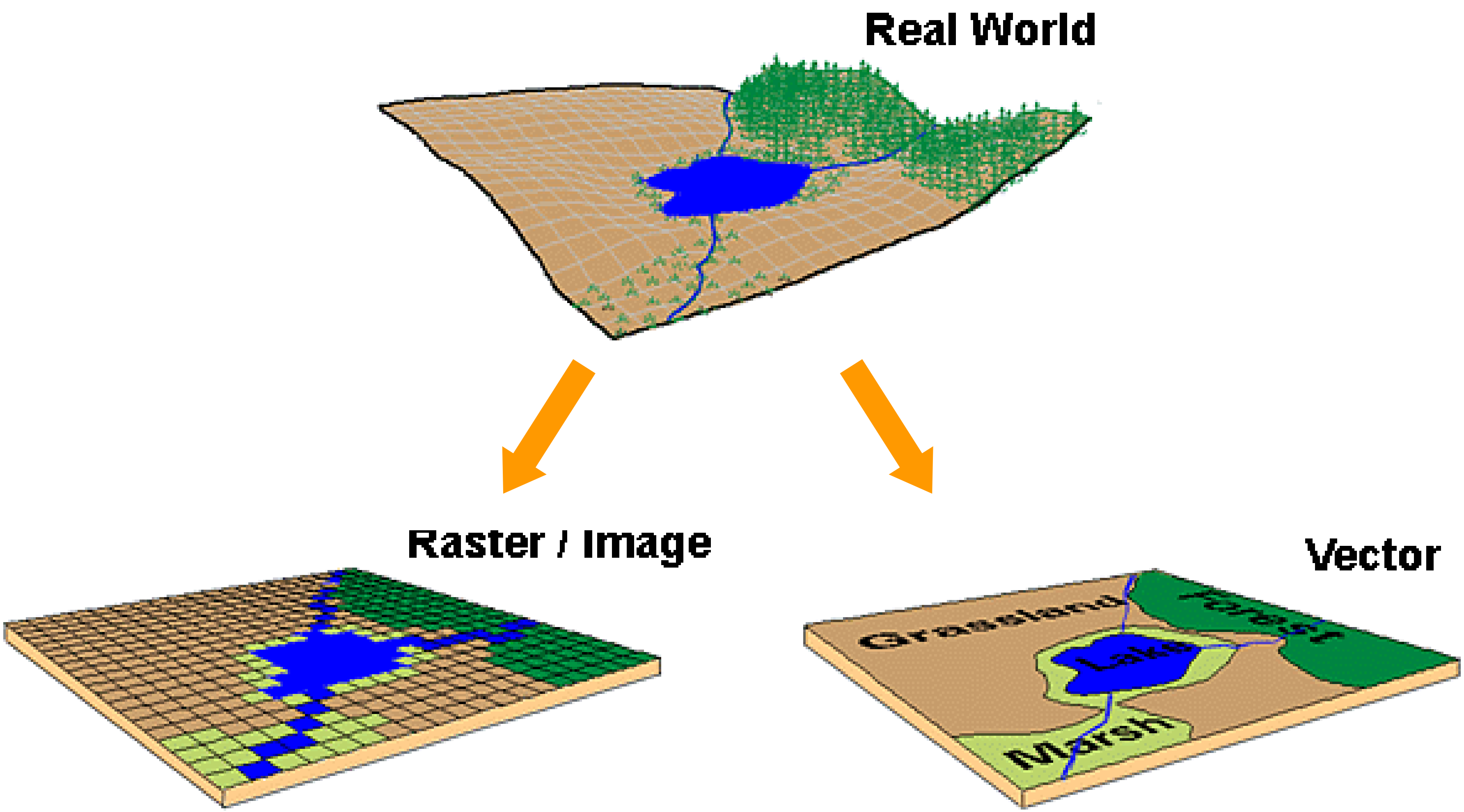
Sources for spatial data

- Field data
  - GPS
- Surveillance data
  - Animal Registry
  - GIS
- Environmental data
  - Remote sensing

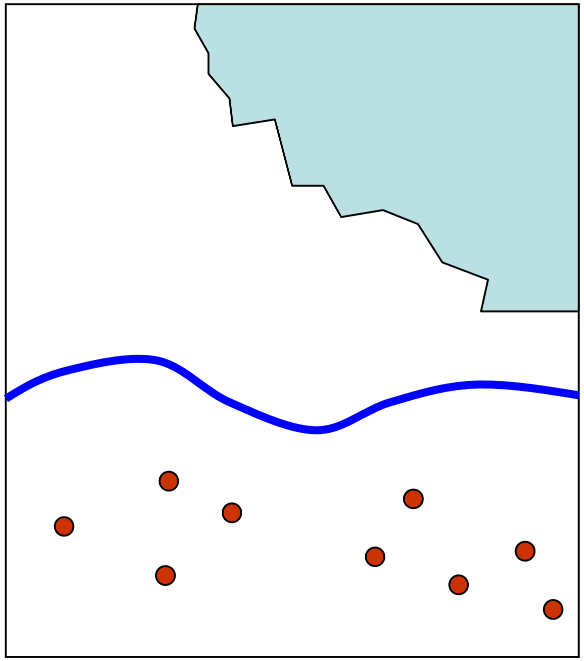
Spatial data formats

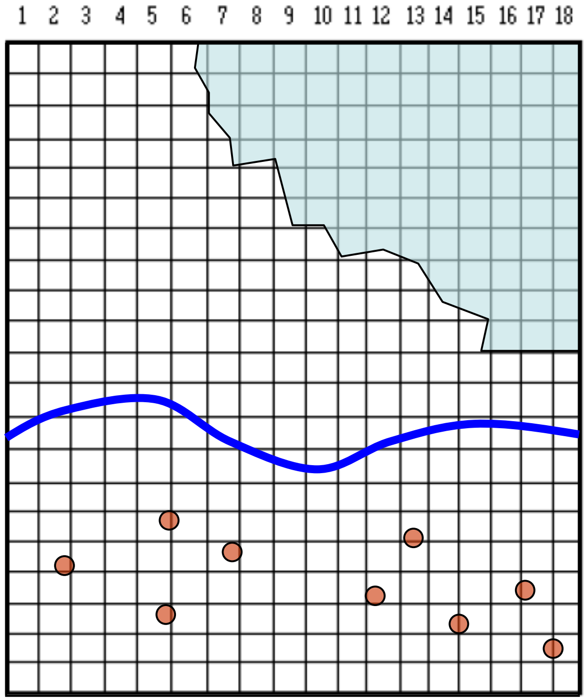
- Vector
  - o Different feature types:
    - Points (Precise geographic coordinates)
    - Lines (artificial or natural features as roads and rivers)
    - Polygons (administrative boundaries, surveillance zones)
  - o Aggregated vs Non-aggregated data
- Raster
  - o Environmental data
  - o Land Cover
  - o Geostatistical elaborations (e.g. spatial interpolations)

Remote Sensing

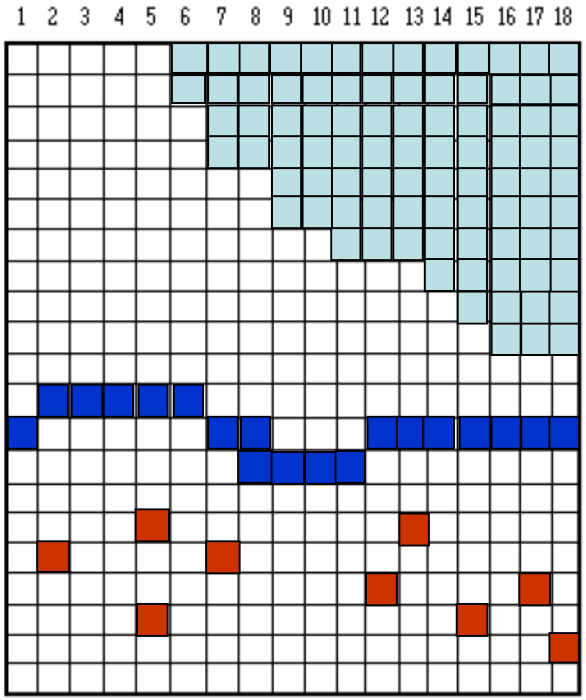

Real World
Raster / Image
Vector
Grassland
Forest
Lake
Marsh

Raster

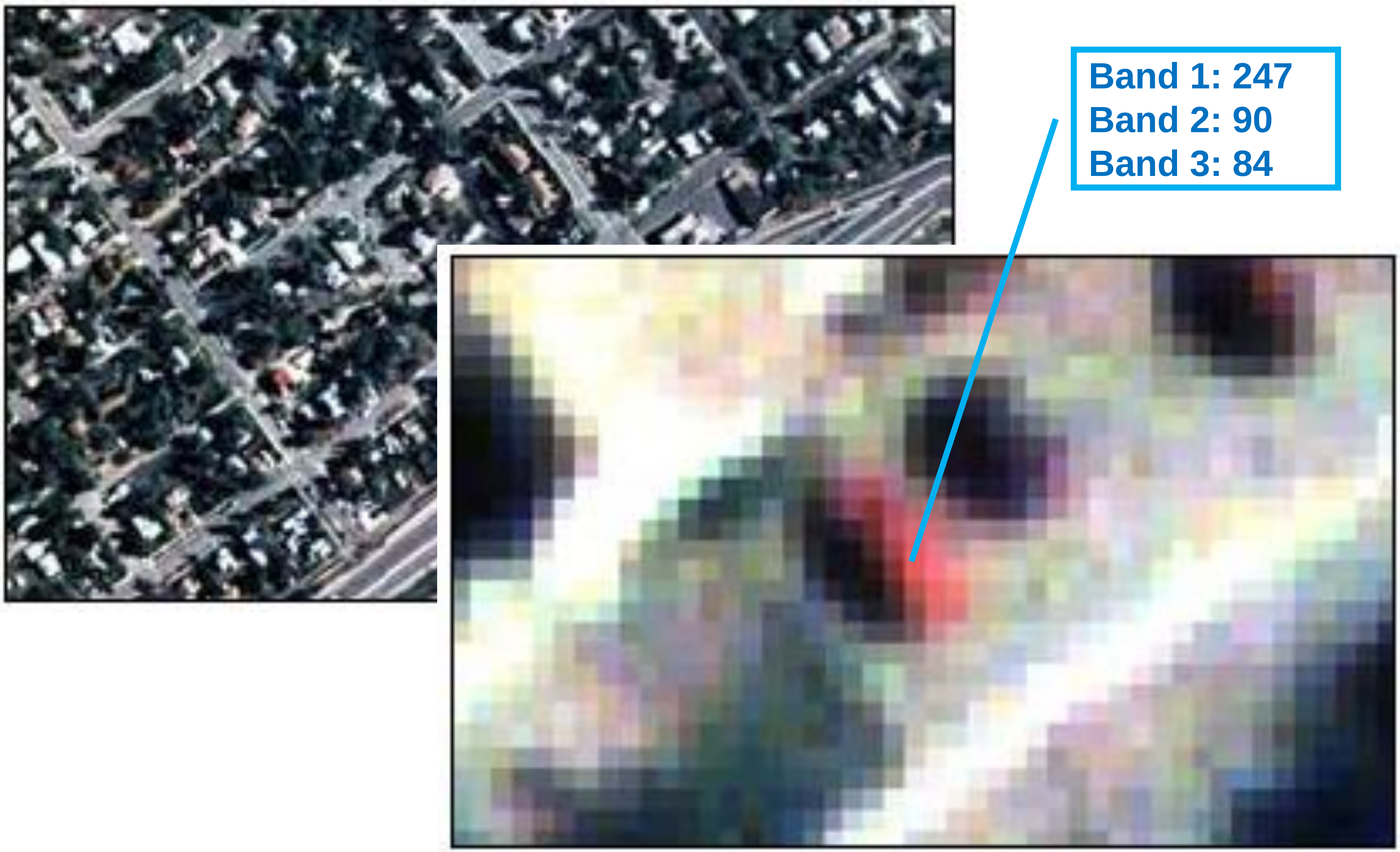




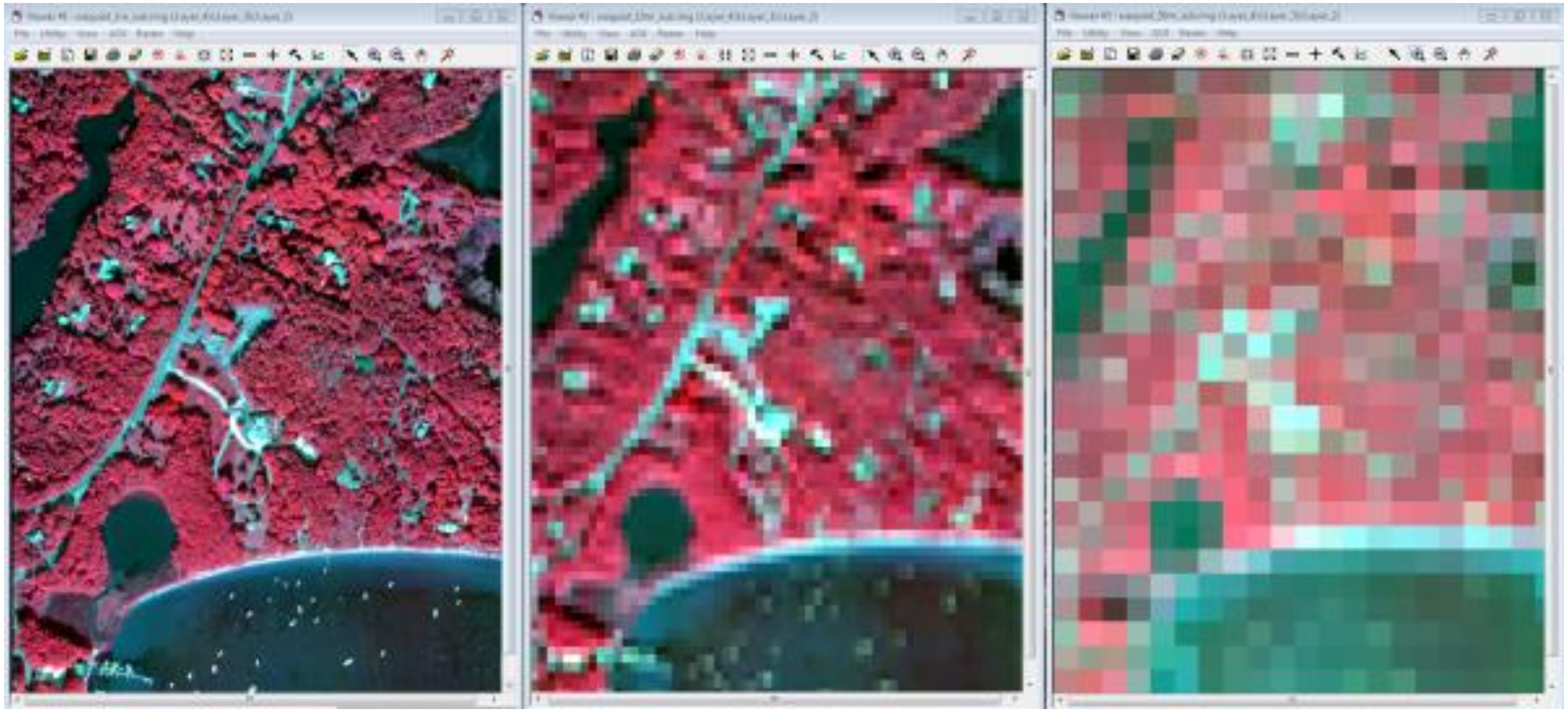
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18


Band 1: 247
Band 2: 90
Band 3: 84

Raster –Resolution

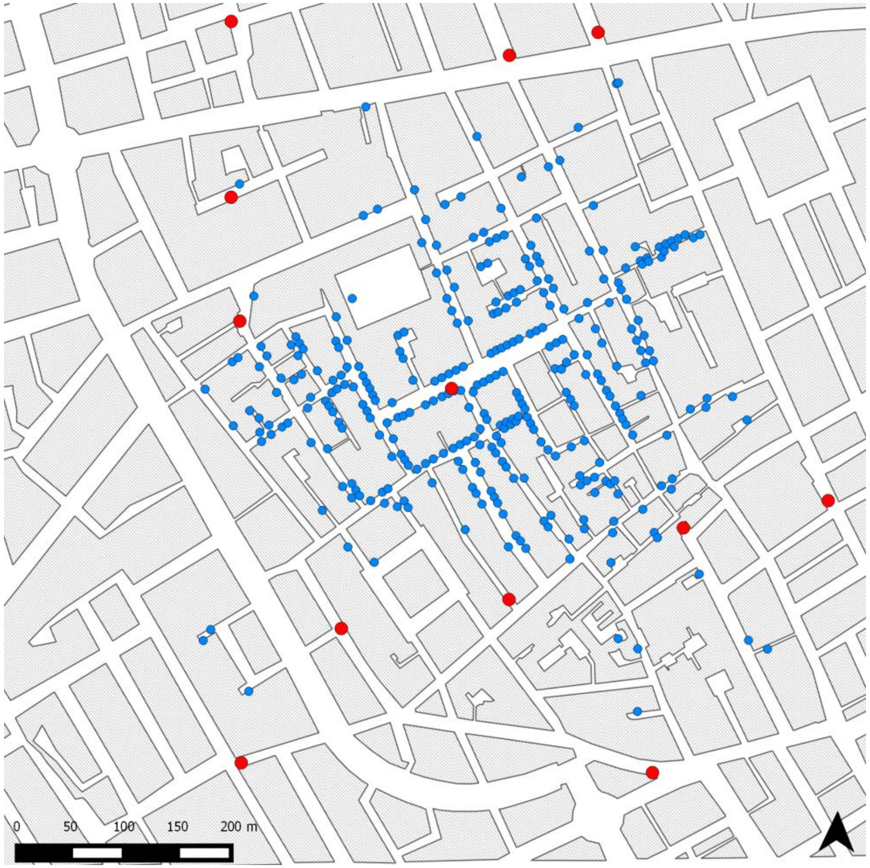


Spatial scale and temporal resolution

- Spatial scale
  - Ranging from a single farm/sampling up to continental level
    - Different representation (point vs polygon features)
- Temporal resolution
  - Frequence of update of the data
  - Caution: using data taken with different time-frame can lead to bias and misinterpretation
- Spectral resolution
  - Raster format

How to integrate animal and epidemiological data with space?

ID	Long	Lat	Cases
1	-0.140795	51.515913	1
2	-0.140379	51.512679	3
3	-0.140304	51.512809	1
4	-0.140501	51.516230	2
5	-0.140375	51.515406	1
6	-0.140454	51.516284	2
7	-0.140722	51.518504	1
8	-0.140403	51.517066	1
9	-0.140196	51.515577	1
10	-0.139714	51.511973	1
…	…	…	…



How to integrate animal and epidemiological data with space?

ID	Long	Lat
1	3.009669	45.64166
2	3.015011	45.34976
3	3.152603	45.05224
4	3.167533	45.95743
5	3.023017	45.44526
…	…	…

ID	Outbreak
1	1
2	0
3	1
4	1
5	0
…	…

ID	Population
1	150000
2	21000
3	90000
4	160000
5	800
…	…

ID	Species
1	Broiler
2	Turkey
3	Broiler
4	Laying hens
5	Turkey
…	…

How to integrate animal and epidemiological data with space?

ID	Long	Lat	Outbreak	Population	Species
1	3.009669	45.64166	1	150000	Broiler
2	3.015011	45.34976	0	21000	Turkey
3	3.152603	45.05224	1	90000	Broiler
4	3.167533	45.95743	1	160000	Laying hens
5	3.023017	45.44526	0	800	Turkey
…	…	…	…	…	…

How to integrate animal and epidemiological data with space?

Two distinct data formats $\begin{cases} \text{Geographical} \\ \text{Attributes} \end{cases}$

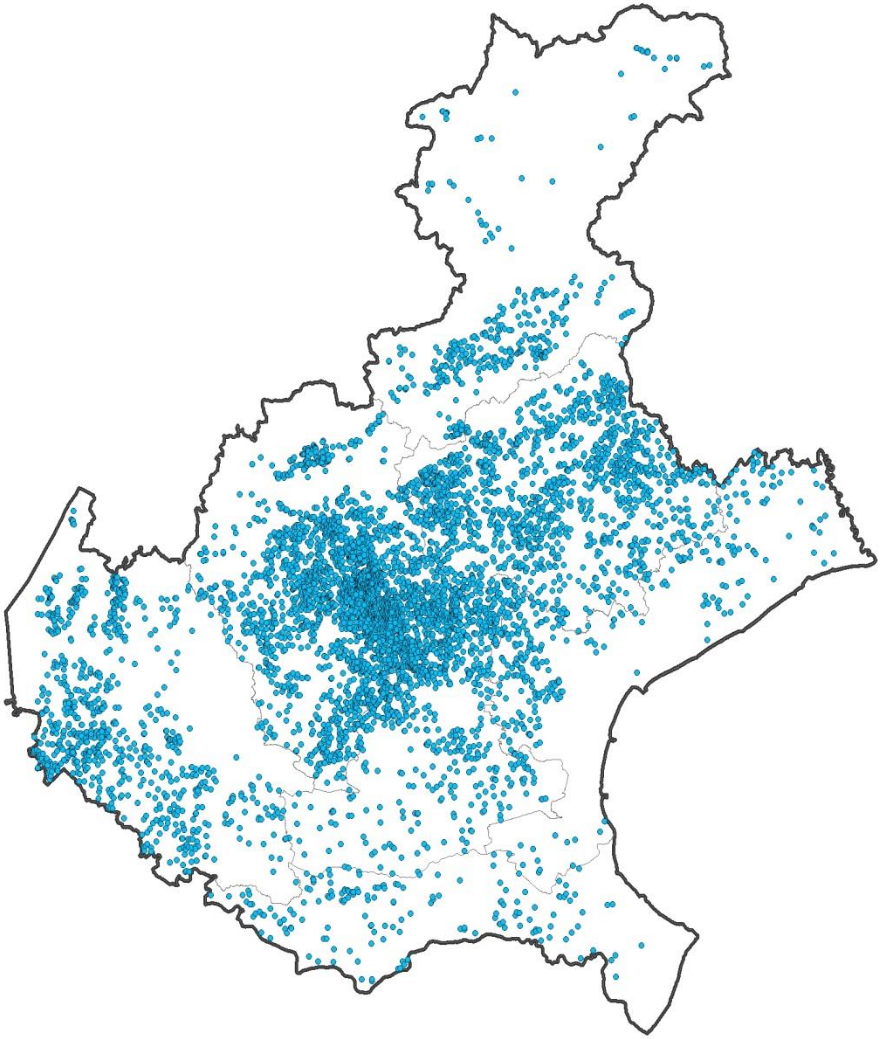
- Simplify management and maintenance procedures
- Exploit data interoperability

ID: **1**, *lat*: **DD.MMSS;** *lon*: **DD.MMSS**

ID: **1**
species: **bovine**
Productive type: **dairy**
…

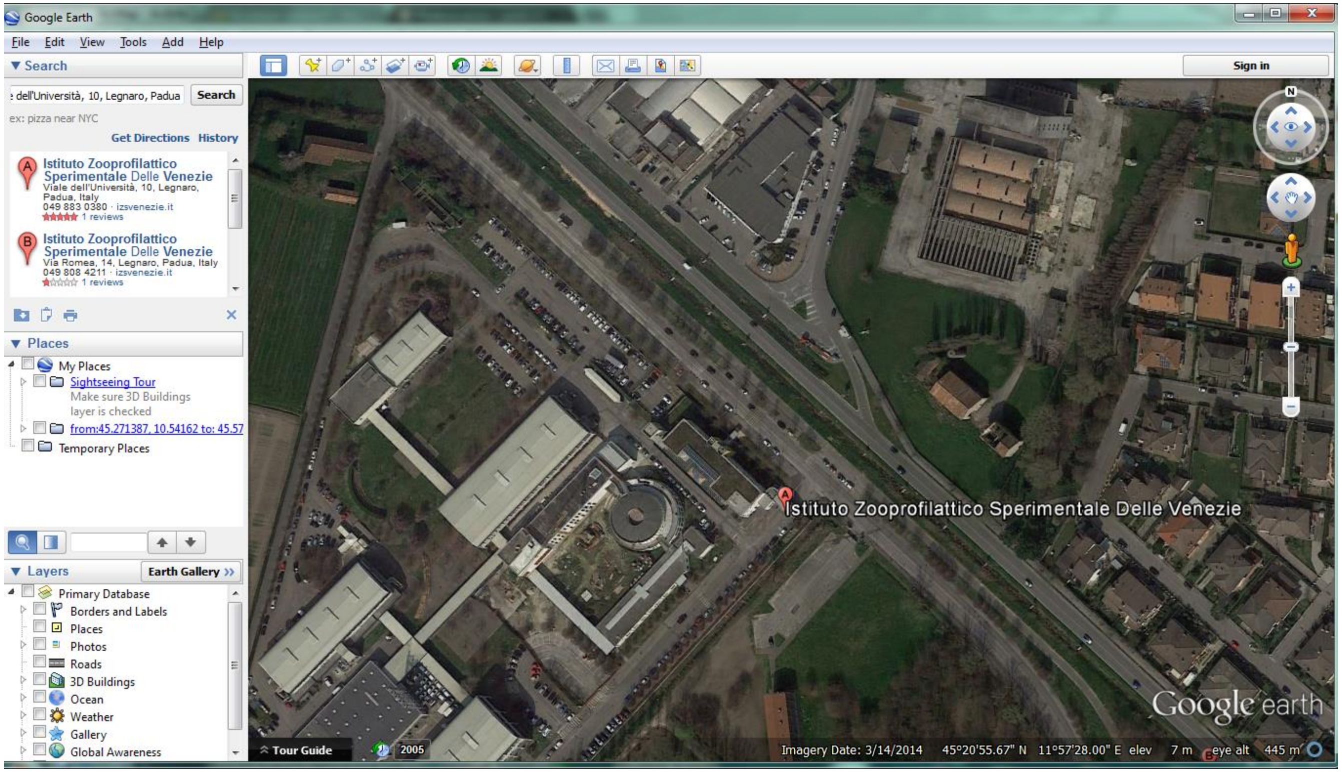
ID: **1**
Sampling date : **30/10/14**
Lab testing date: **03/11/14**
Test outcome: **negative**
…

ID: **1**
…



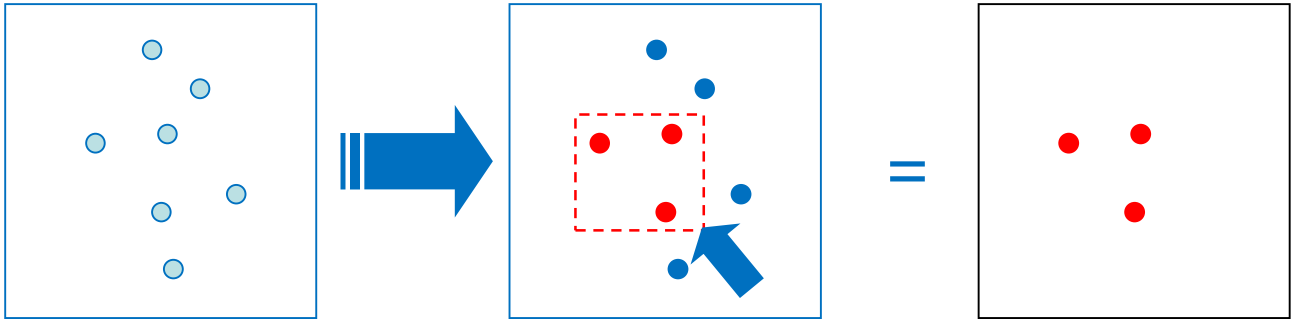
Geocoding

- No precise coordinates available
  - Use the Street Address to retrieve the geographical position



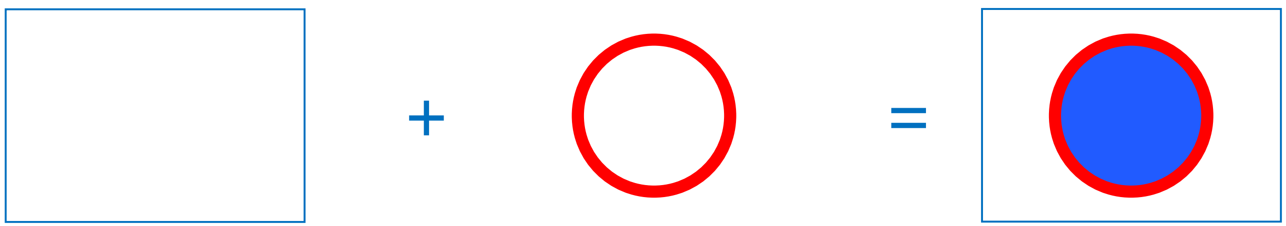
Useful GIS functions

- Selection



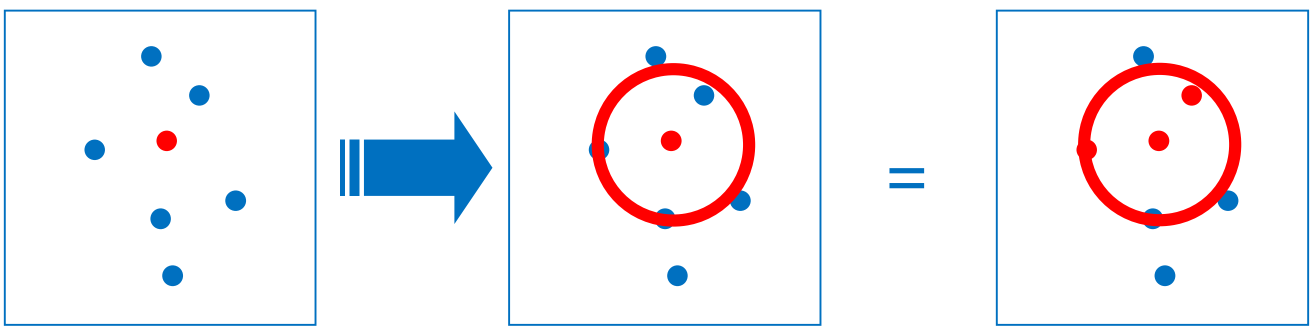
Useful GIS functions

- Selection
- Overlay



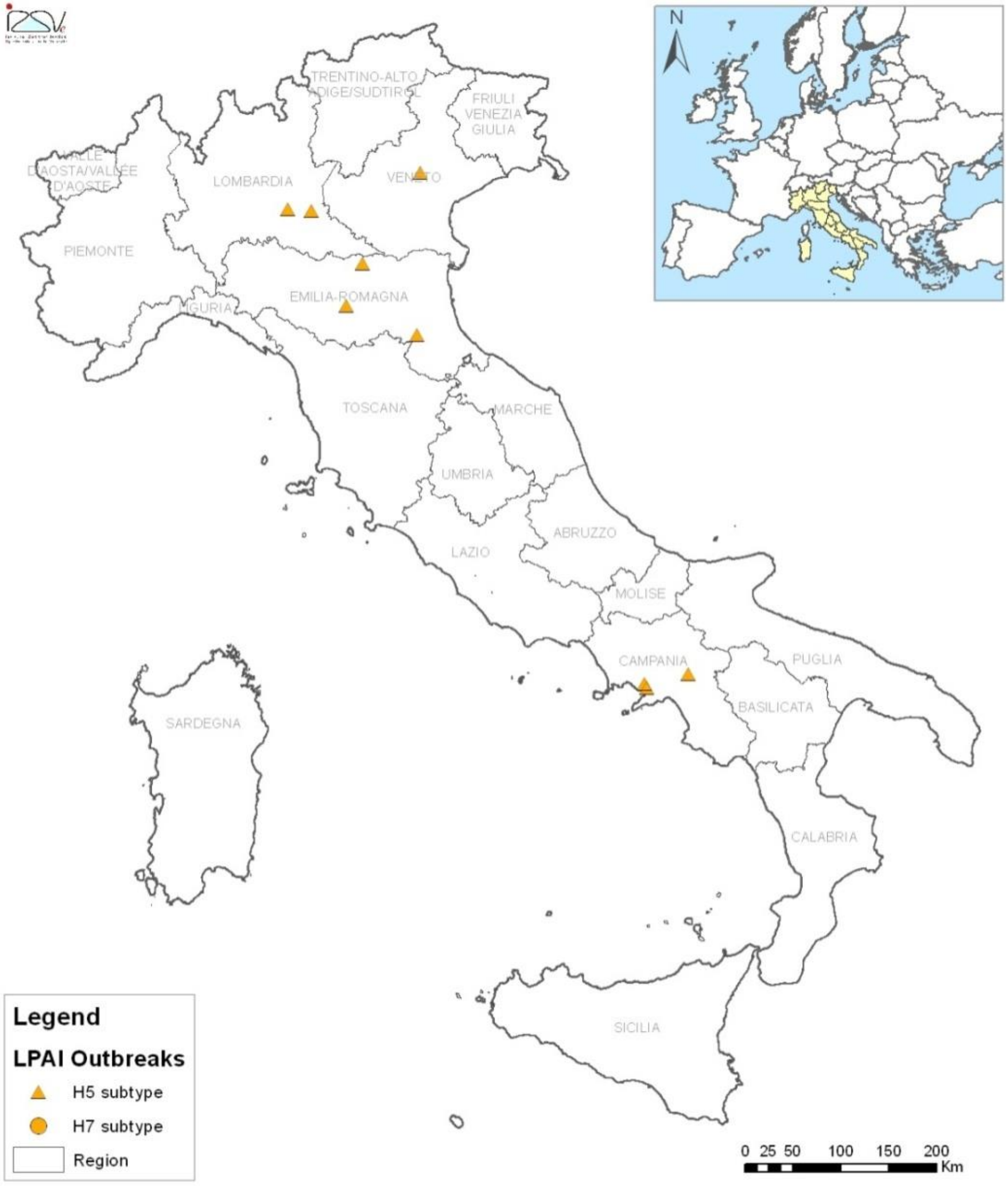
Useful GIS functions

- Selection
- Overlay
- Buffering

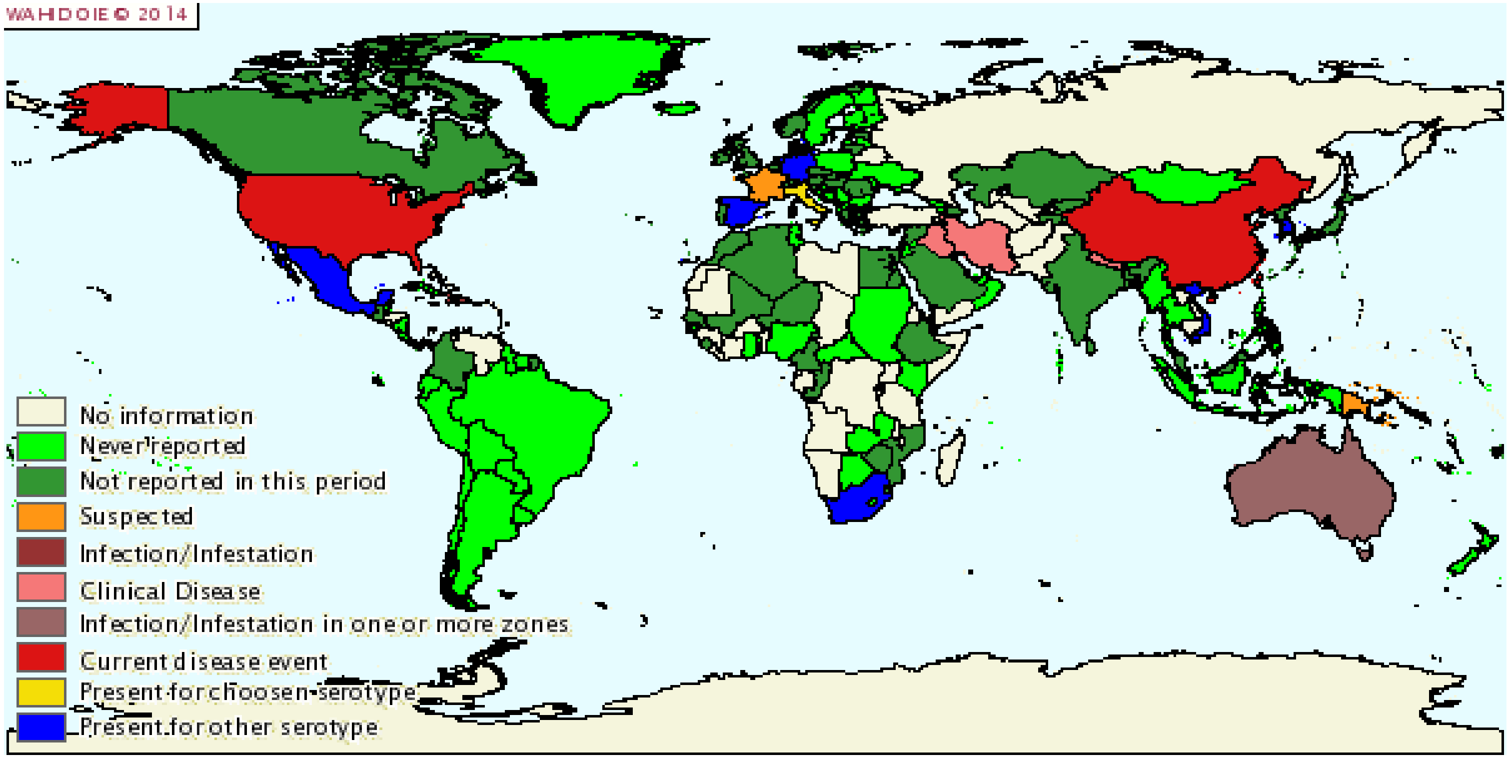


Showing spatial epidemiological data on maps

- The way of representing data depends on the aims of the map
  - Reporting the occurrence of a disease
    - Number of cases
    - Geographical extent
  - Surveillance
  - Control measure management
  - Research outcome
    - Risk Maps



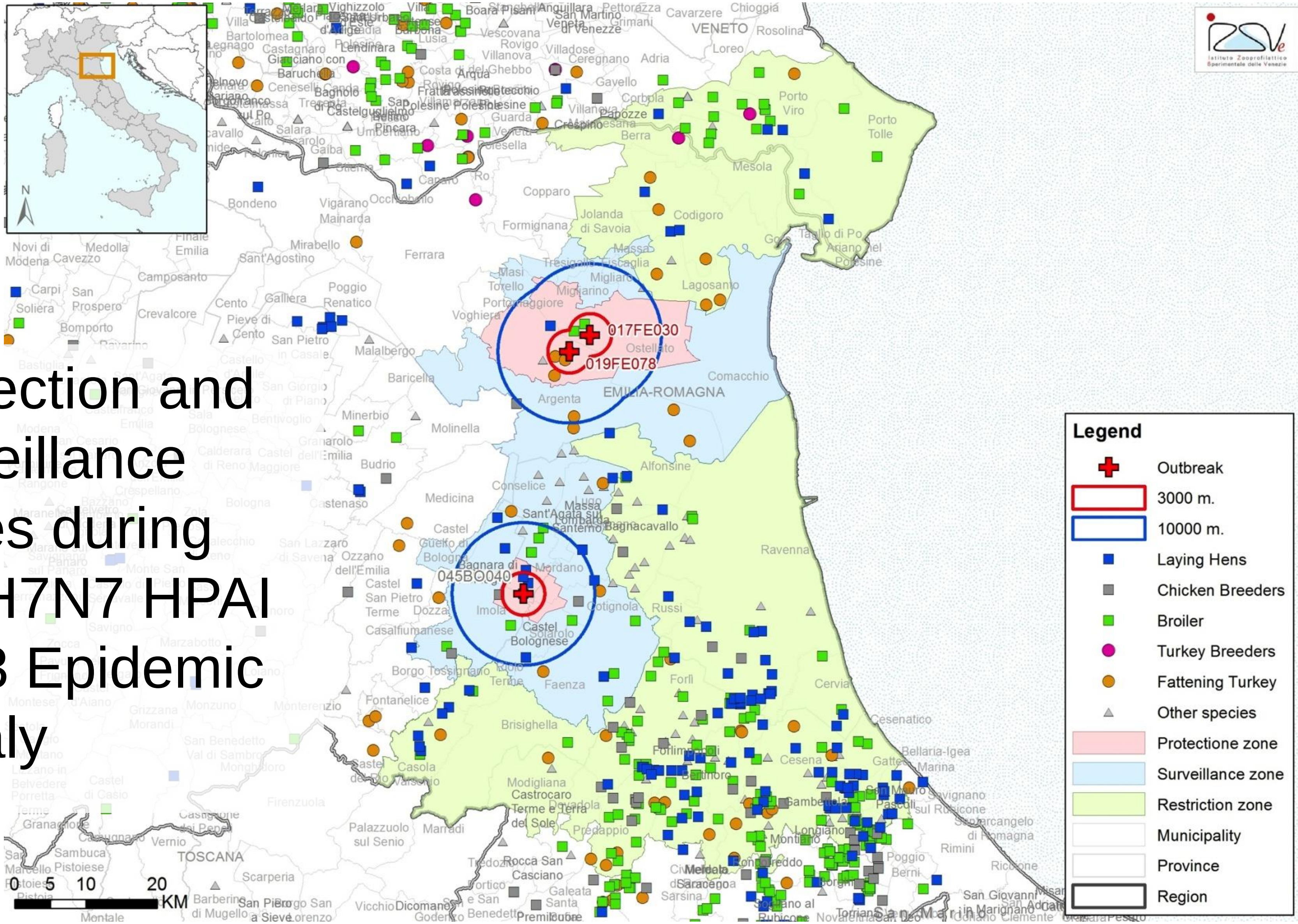
Report of AI outbreaks in Italy, for the Standing Committee on the Food Chain and Animal Health - European Commission

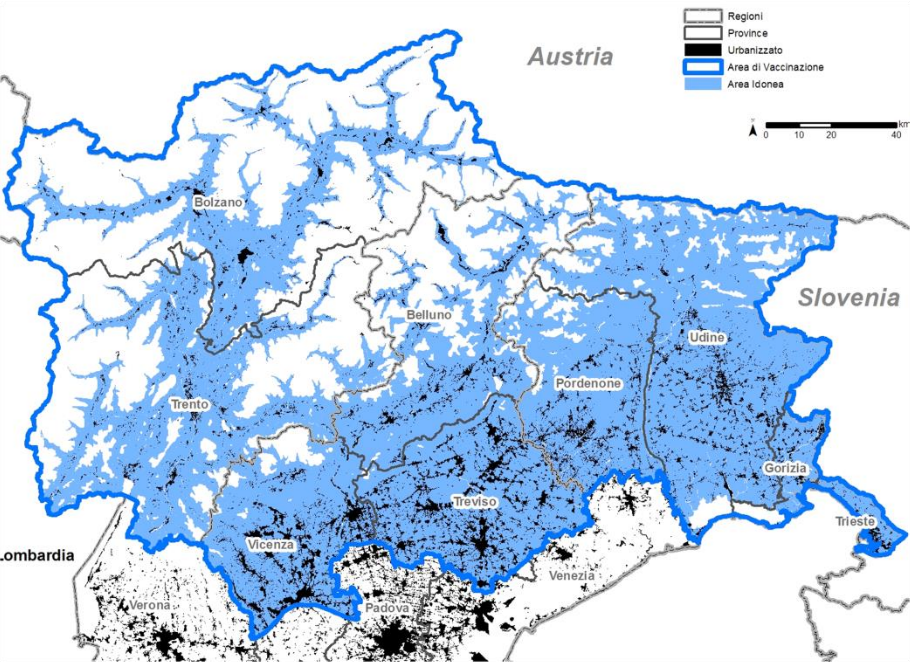


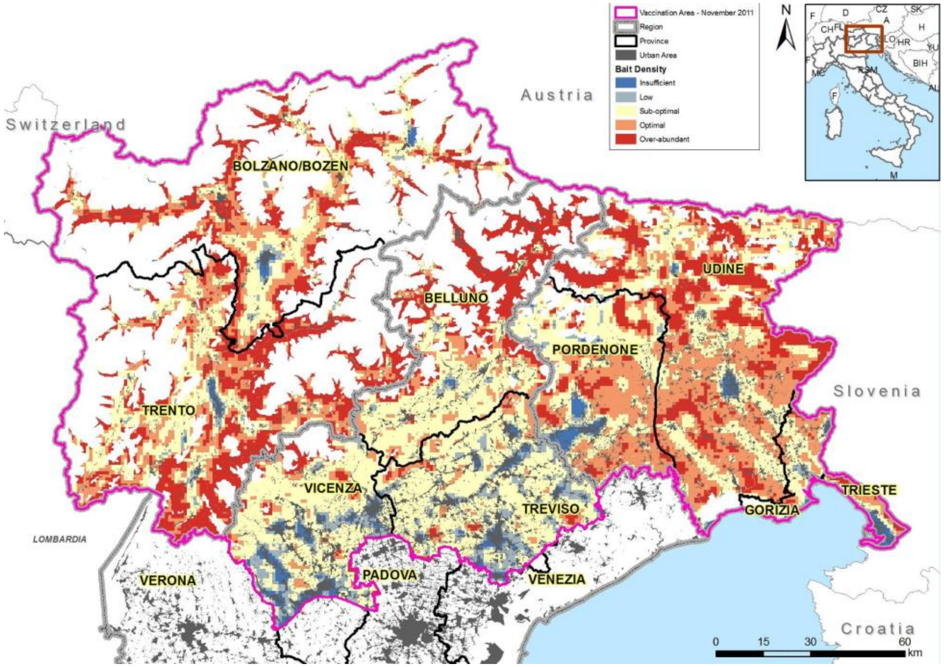
OIE WAHID interface

Distribution of Low Pathogenic AI H5 – Jan-Jun 2013

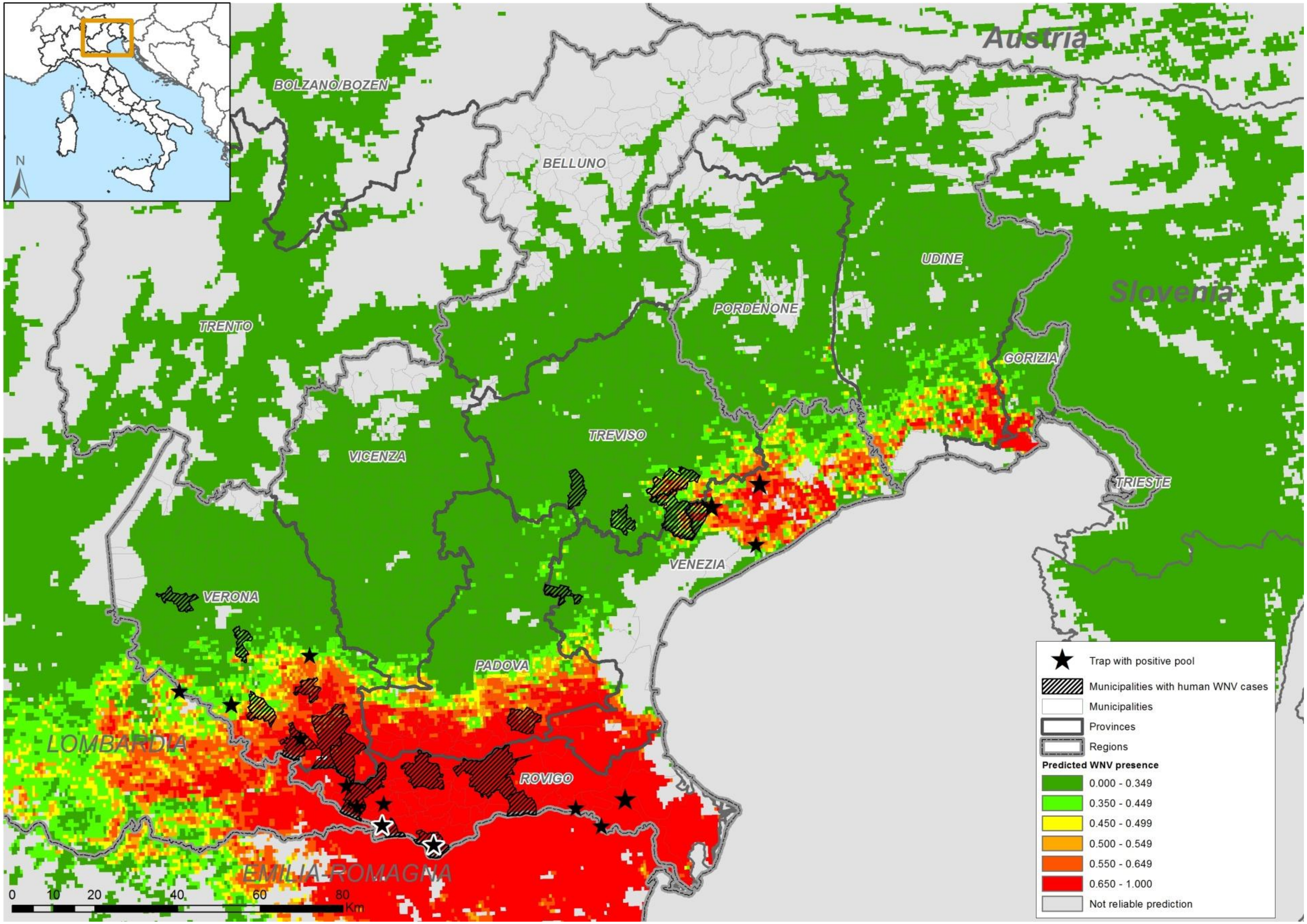
Protection and surveillance zones during the H7N7 HPAI 2013 Epidemic in Italy



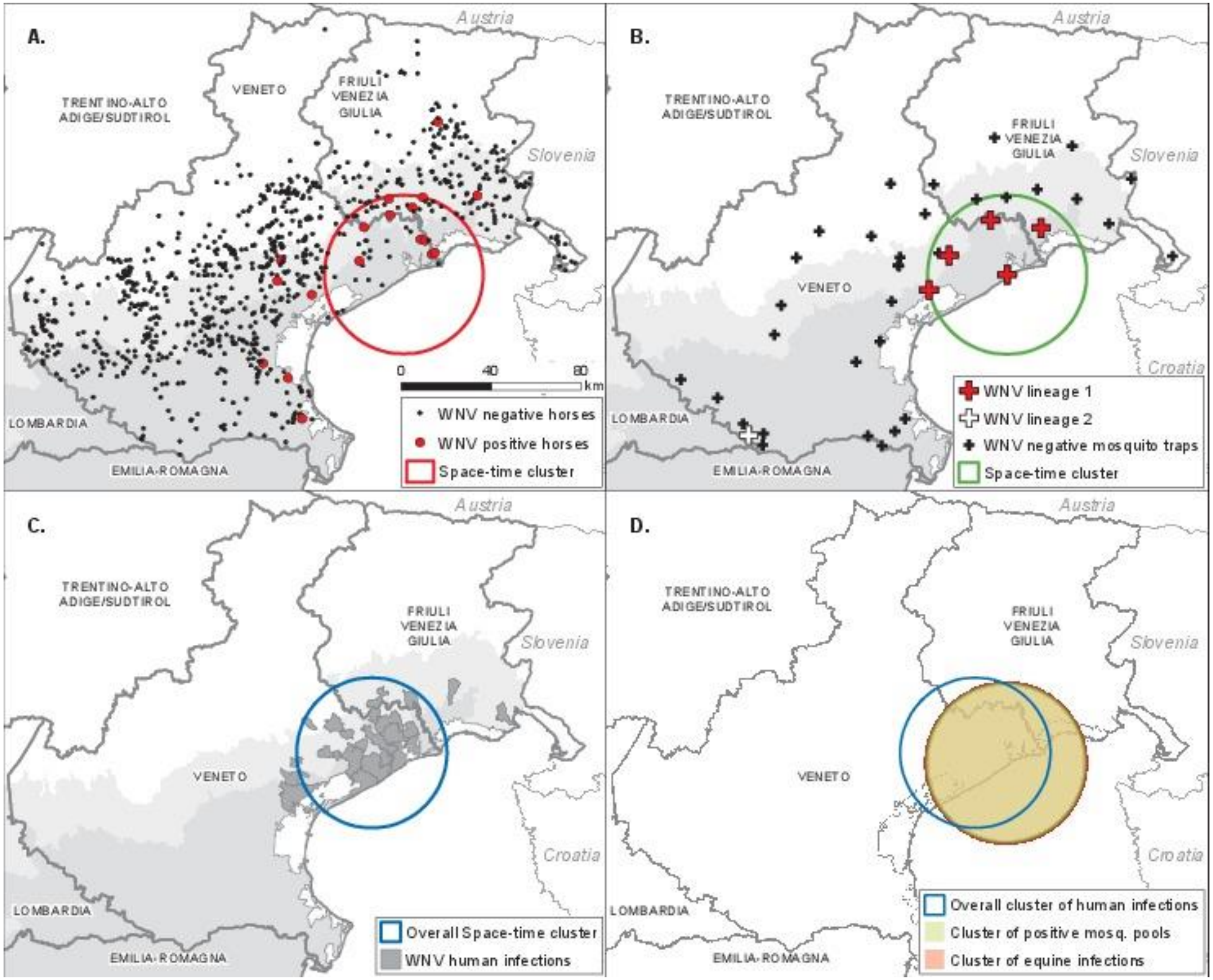




- Management of the Aerial Oral Fox Vaccination against rabies (2009 – 2014)

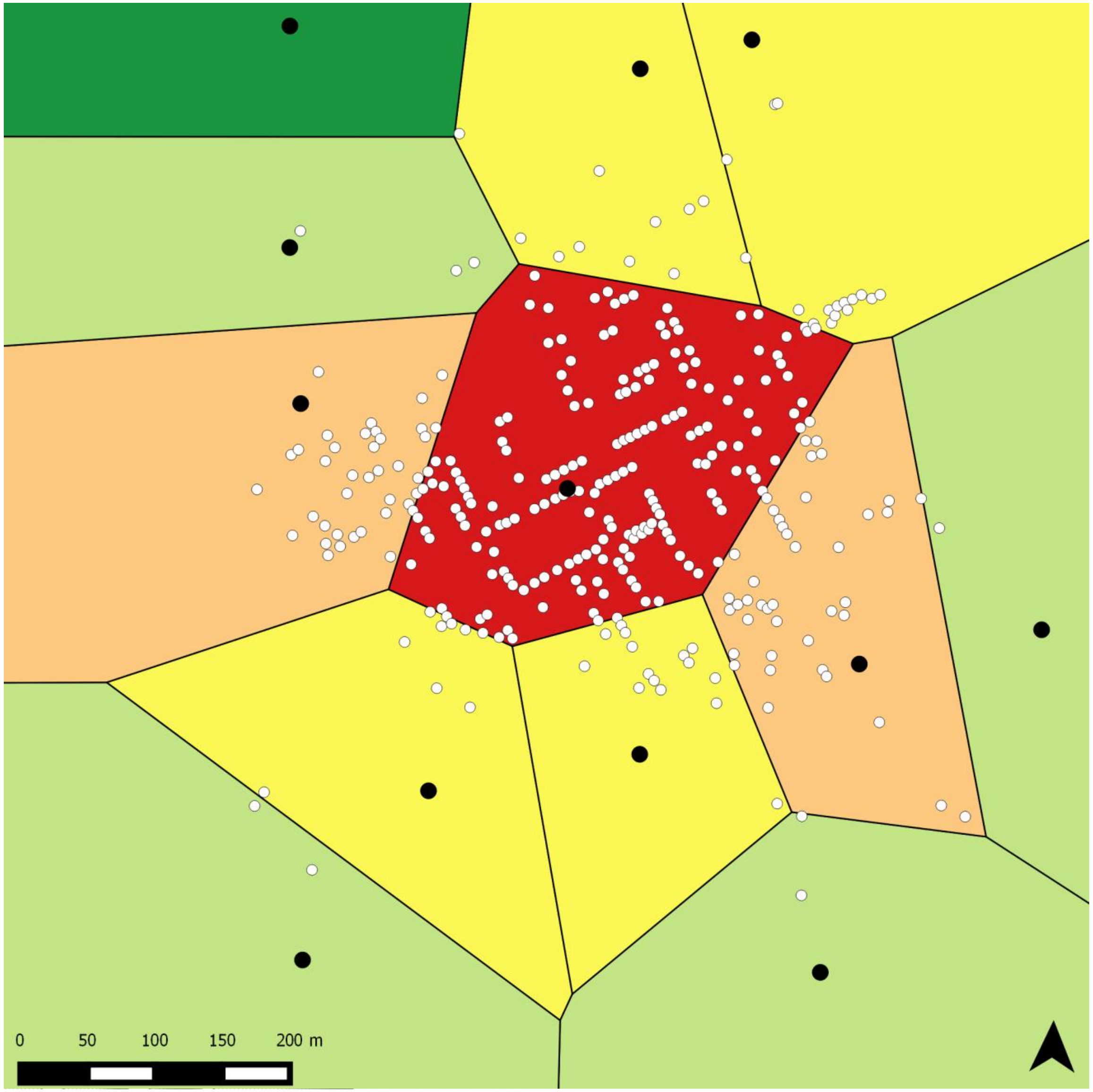


- Risk Maps for WNV in North-eastern Italy

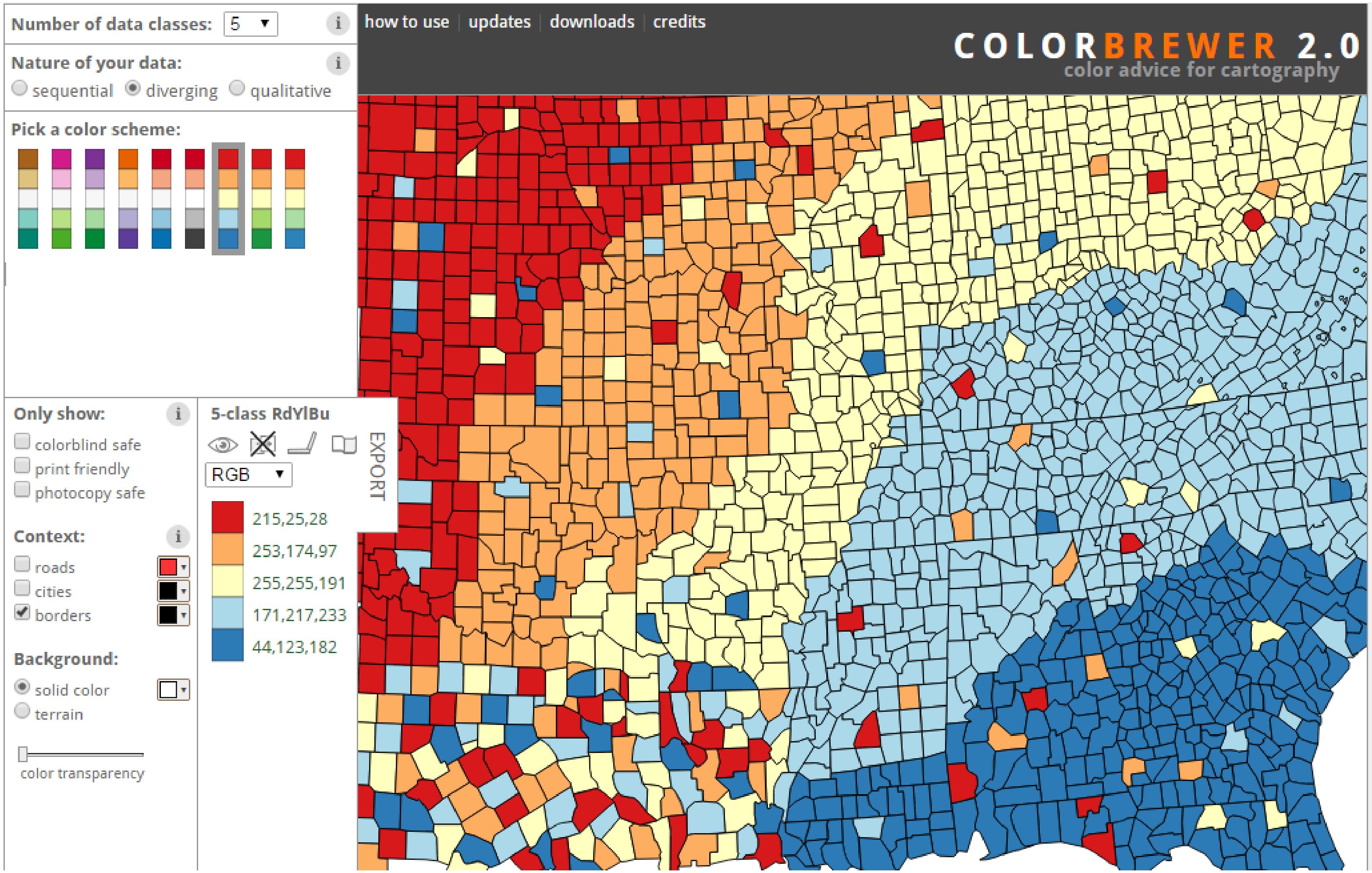


- Risk Maps for refining WNV surveillance

Same data, different maps



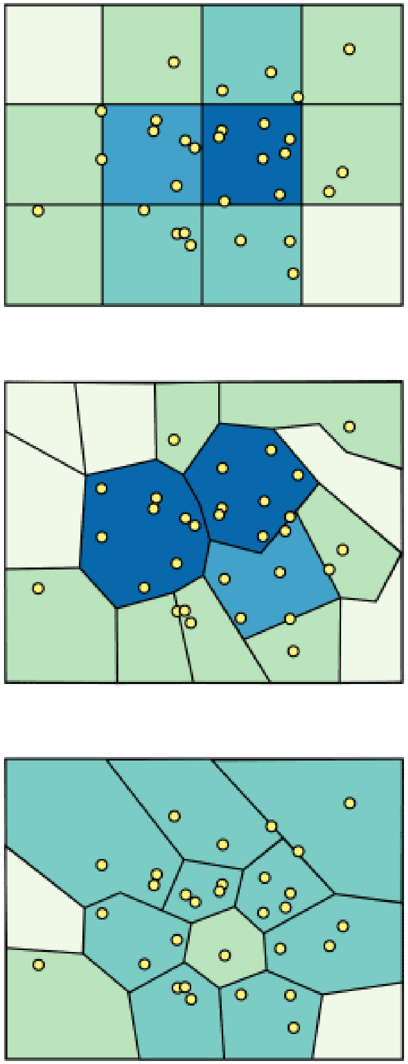
Recommended Web Site



http://colorbrewer2.org/

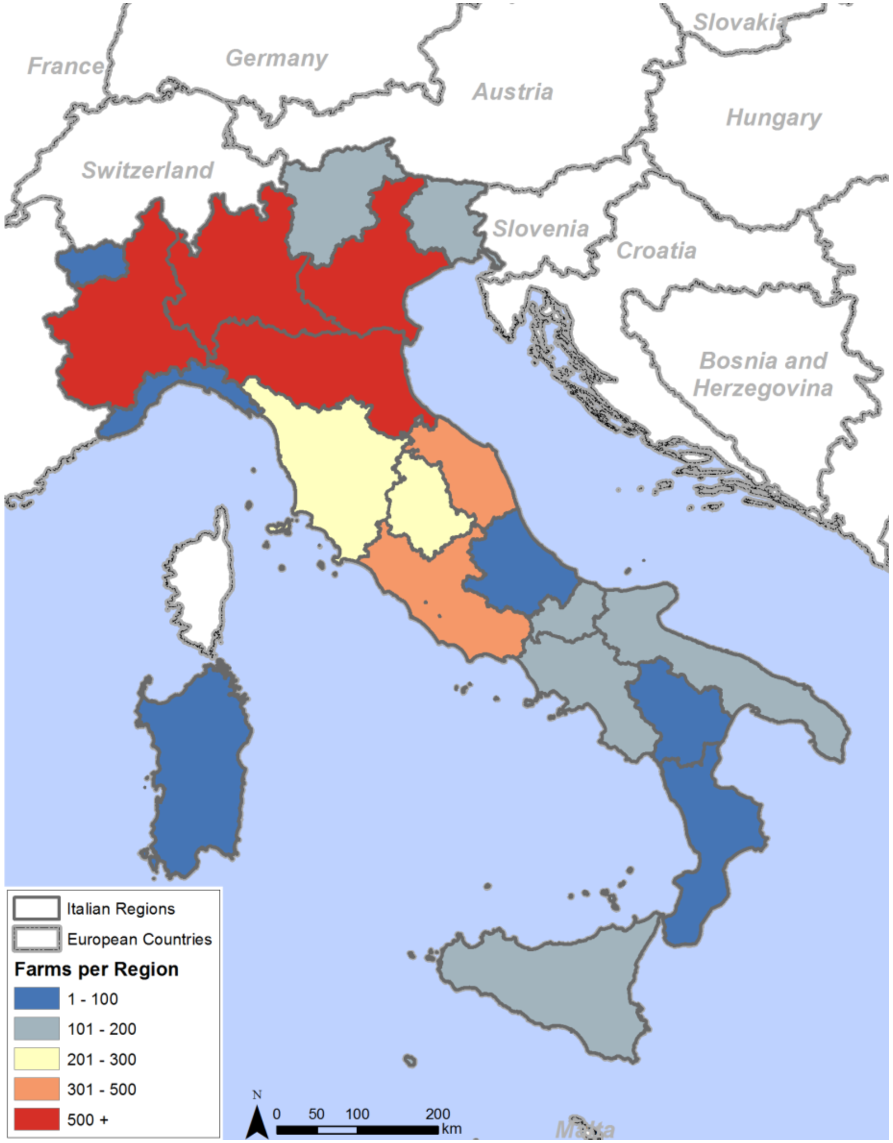
The Modifiable Areal Unit Problem

- The **MAUP** is "*a problem arising from the imposition of artificial units of spatial reporting on continuous geographical phenomena resulting in the generation of artificial spatial patterns*" (Heywood, 1988)



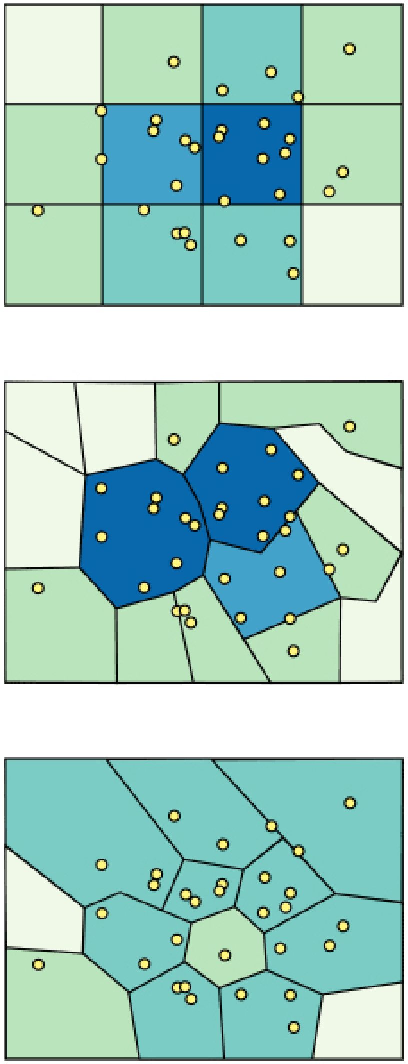
Scale MAUP

- The scale at which one chooses to analyze information can produce different results.
- Choose your scale to match your research question
- If possible, choose a finer scale than you think is necessary.
    - Finer-scale data can be aggregated, while coarser scale data cannot easily be divided.



Zone MAUP

- Related to the grouping schemes used for data analysis (**zones**)
  - May persist even when the considered units are all of the same scale.



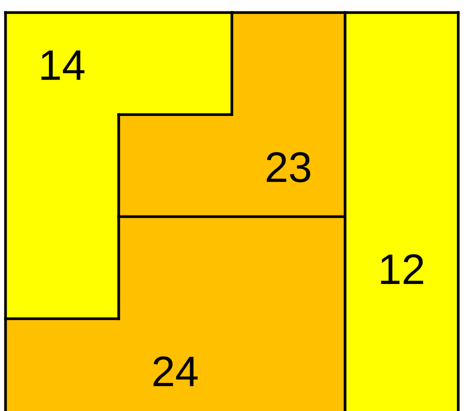
Zone MAUP - Examples

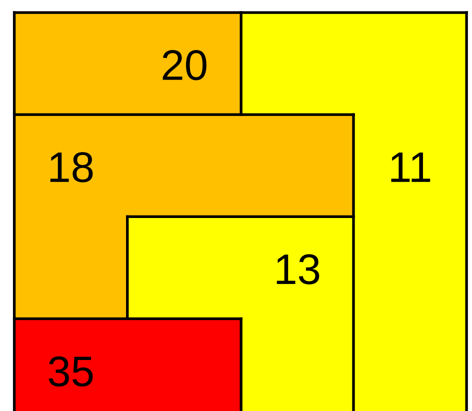
Population per cell

100	200	400	100
100	400	300	200
400	200	200	100
300	100	200	100

Prevalence

20	20	10	10
10	20	30	10
10	10	20	10
40	20	20	20

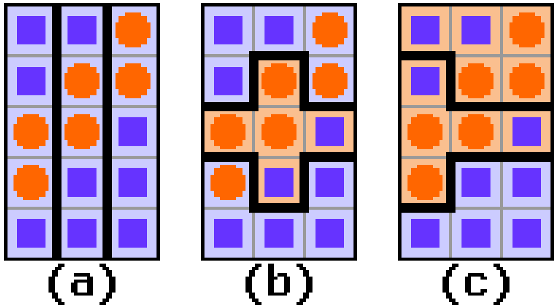




MAUP

- MAUP may add unwanted sources of error or mis-interpretation
  - It may also be used to intentionally manipulate the results (though not advisable).

This occurs in politics so often that it has it's own terms!
'Gerrymandering' and 'Political redistricting'



http://en.wikipedia.org/wiki/Gerrymandering

MAUP - Conclusions

- The MAUP is fundamentally an unsolvable issue
  - Loss of information or bias are inevitable when aggregating/grouping data.

- Information should be grouped only when strictly necessary
  - Additional analyses that one can perform outweigh this effect or and minimize the negative effects of this grouping.

Spatial analysis

What is Spatial Analysis

'Analytical techniques to determine the spatial distribution of a variable, the relationship between the spatial distribution of variables, and the association of the variables of an area…
… It refers to the analysis of phenomena distributed in space and having physical dimensions' (Data West Research Agency)

- Statistical/Mathematical methods accounting for geographical locations

What is Spatial Analysis

- Analyses of phenomena occurring in space, and having a *physical dimension:*
  - Outbreaks
  - Livestock farms
  - Potential risk factors
  - Sources of contaminants
  - ...

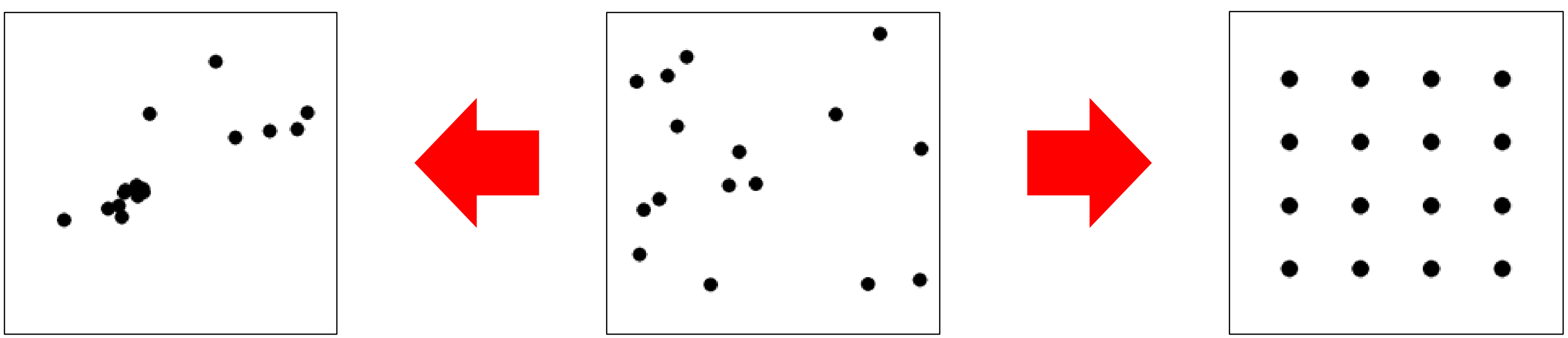
- *How are events of interests distributed in space?*

Spatial analysis in Public health - Scopes

- Descriptive analysis
  - Visualising and preliminary analysing the spatial distribution of events of interest
  - Exploratory Spatial Data Analysis (ESDA)
- Hypotheses generation
  - Unexpected clustering of events might be related to the presence of factors that facilitate the spread of diseases
- Surveillance and early warning
  - Spatial analysis may be integrated in Surveillance plans, to provide information on how a disease spread, and whether there are areas requiring targeted surveillance measures
- Simulation models
  - Mathematical/Statistical models to simulate how a disease might spread

Complete Spatial Randomness (CSR)

- Point process in which all of the events occur in space in a purely random fashion (Poisson distribution)
- Any point in the space has the same chance to *host* an event
- CSR is the null hypothesis for ***Point Pattern Analysis*** (PPA)



Clustering **CSR** ***Over-dispersion***

Type of Spatial Analyses

Global

- Clustering of events on a *large scale*
- No information on **where** in the area the events are aggregated

Local

- Clustering of events on a smaller scale
- **Hot-spots:** zones where events occur more frequently than expected
- **Cold-spots:** zones where events occur less frequently than expected

Focussed

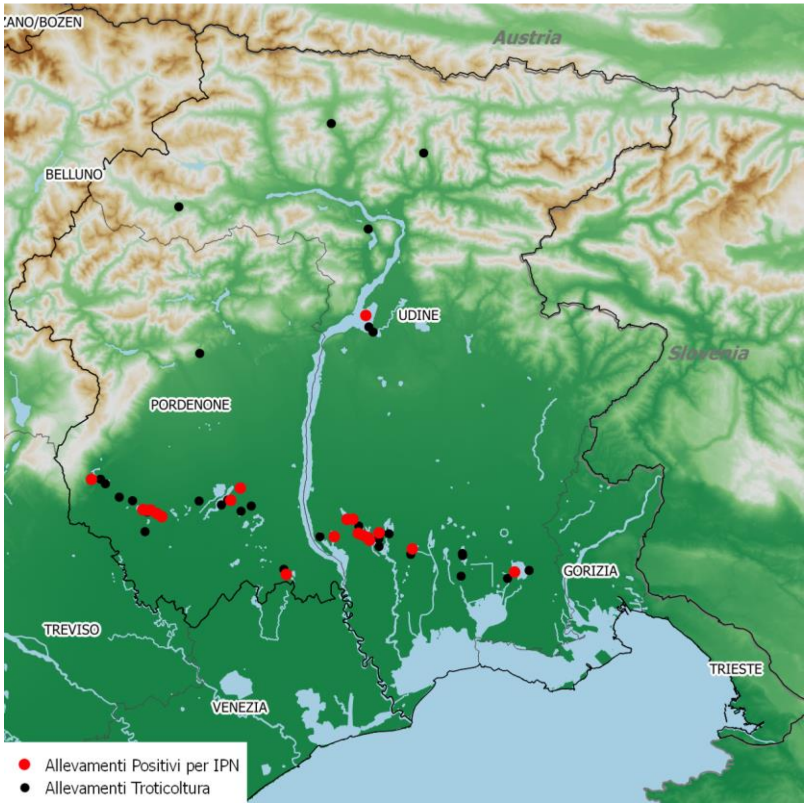
- Clustering of events around pre-defined points of interest (e.g. Surces of contaminants)

Global spatial analysis – Case study

Study of prevalence and risk factors for the introduction of the Infective Pacreatic Necrosis (IPN) in trout farms in North-eastern Italy

- Study Area: North-eastern Region of Friuli Venezia Giulia
- Population of reference: All of the trout farms in the region
- Events of interest: Farms affected by IPN

Global spatial analysis – Case study

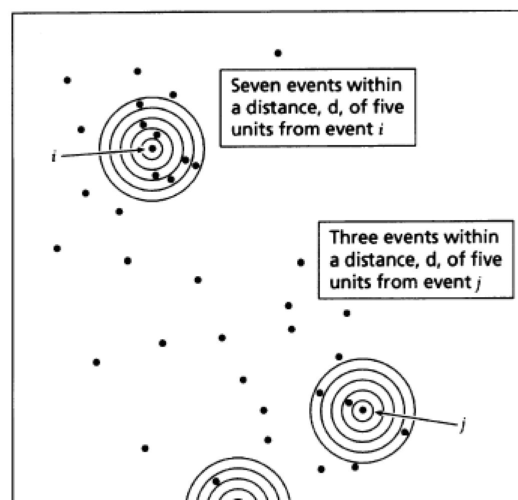


Population: 51 fish farms

Cases: 20

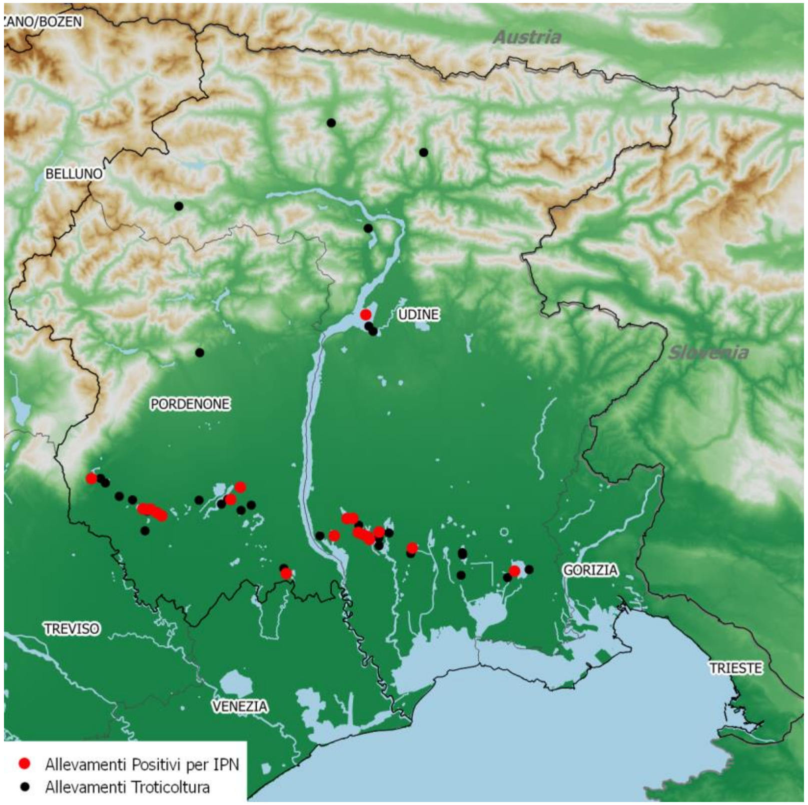
Clustering measure:
Ripley's *K(d)*

Counts of events within increasing search radiuses



Gatrell et al., 1996

Global spatial analysis – Case study

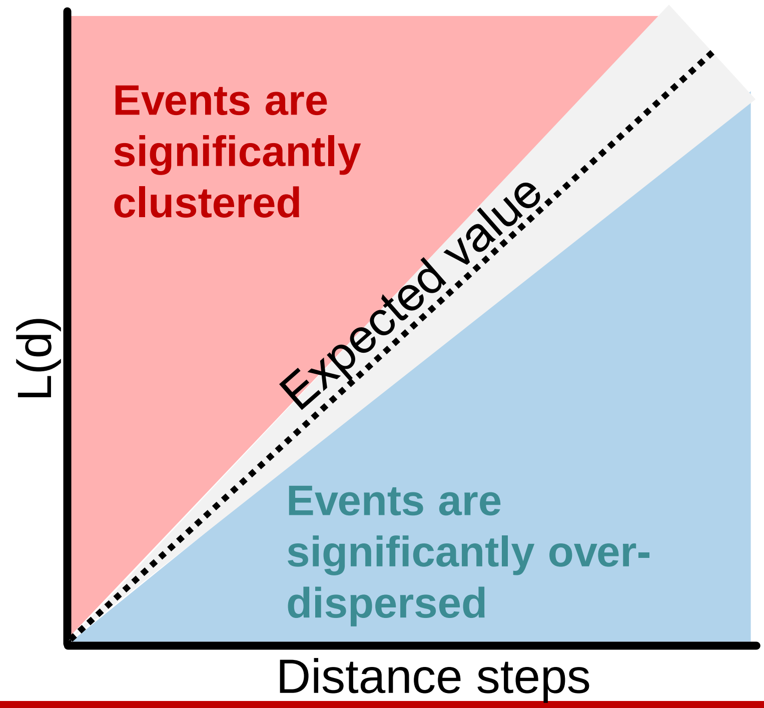


Population: 51 fish farms

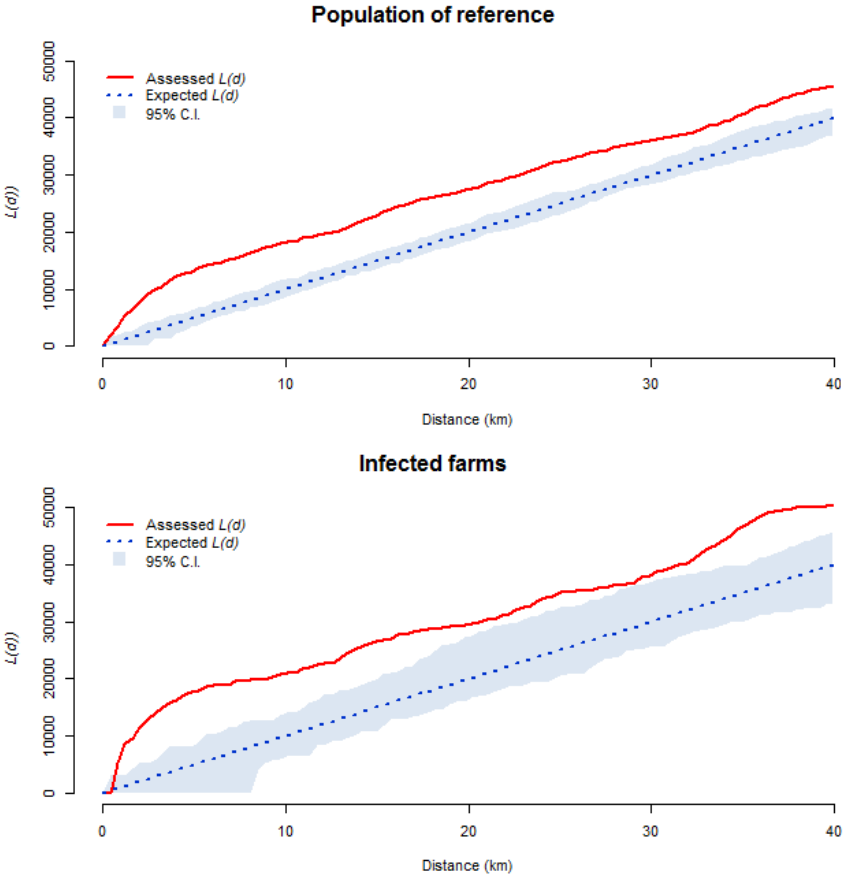
Cases: 20

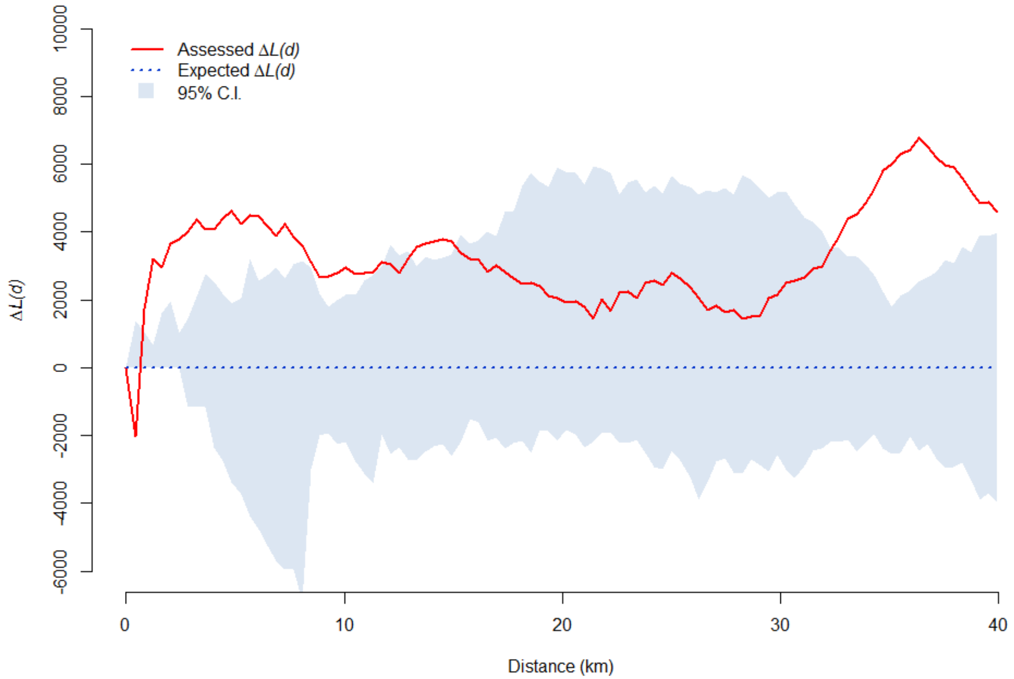
Clustering measure:
Ripley's *K(d)*

Count of events within increasing search radiuses



Global spatial analysis – Case study





Local spatial analysis – Case study

- Identifies where significant spatial aggregtions of events occur (Hot-spots)

- Case study:

Geospatial Health 8(2), 2014, pp. 509-515

Retrospective and spatial analysis tools for integrated surveillance of cystic echinococcosis and bovine cysticercosis in hypo-endemic areas

Rudi Cassini[1], Paolo Mulatti[2], Claudia Zanardello[2], Giulia Simonato[1], Manuela Signorini[1], Stefania Cazzin[2], Pier Giorgio Tambalo[3], Mario Cobianchi[2], Mario Pietrobelli[1], Gioia Capelli[2]

Global spatial analysis – Case study

- Study Area: North-eastern Italian Region of Veneto
- Period of reference: 2006-2010
- Population of refernce: Dairy cattle farms (n=8173)
- Cases of Cystic Echinococcosis in cattle (n=251)

Global spatial analysis – Case study

Cassini et al, 2014

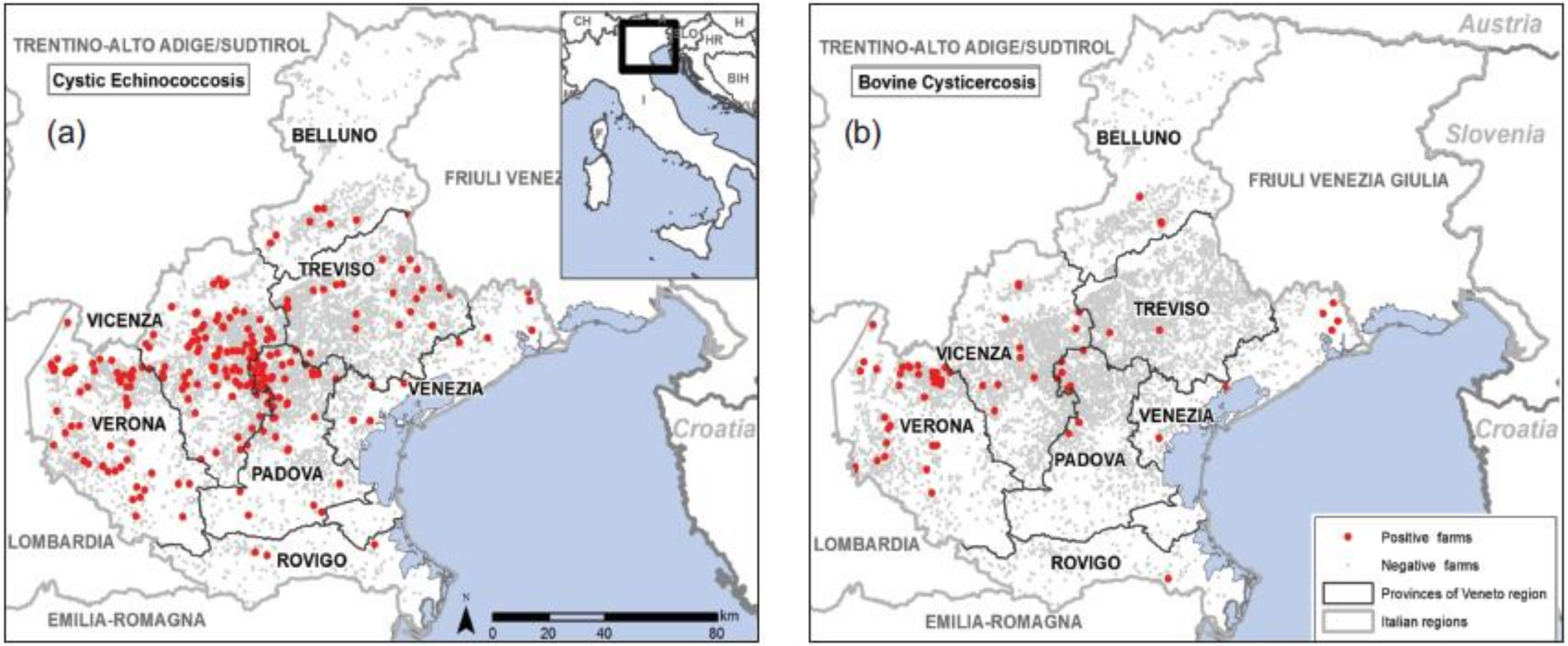


Fig. 1. Map of the Veneto region with dairy farms testing negative (grey dots) for CE (a) and farms testing positive (red dots) for BC (b).

Global spatial analysis – Case study

- (Purely Spatial) Scan Statistic
  - Scans the study area using *Spatial Windows* with increasing radiuses;
  - Counts number of positive and negative farms within the window and compares the number to the distribution of events outside of the spatial window;
  - Statistical significance tested though Monte Carlo
- Also a Space-Time version is available

Global spatial analysis – Case study

Cassini et al, 2014

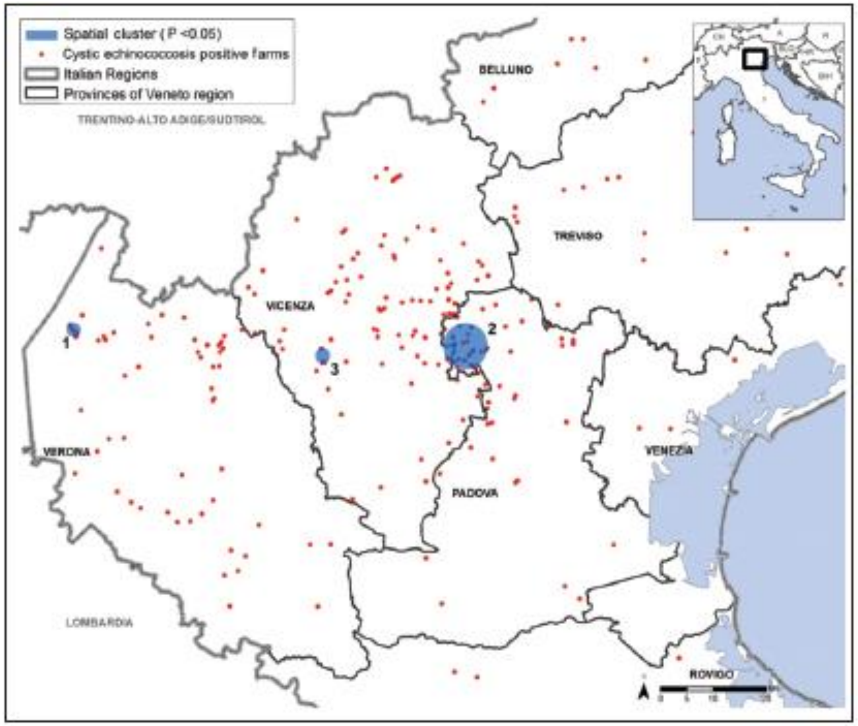


Fig. 2. Geographical localization of CE clusters in the Veneto region.

Cluster ID	Radius (m)	No. of farms included	No. of cases observed	No. of cases expected	P-value
1	1,333	18	10	0.56	<0.001
2	4,219	198	22	6.08	0.01
3	1,384	7	5	0.22	0.012

Spatial Analysis – Spatial Autocorrelation

Everything is related to everything else, but near things are more related than distant things

- Firs Law of Geography (W. Tobler)

- How are variables of interest geographically distributed?

- Autocorrelation:
  - **Positive →** Variables similar to each other are also spatially closer
  - **Negative →** Events that are spatially close have variables with different values

Spatial Analysis – Spatial Autocorrelation

- Global Spatial Autocorrelation
  - Moran's I
    - Similar interpretation of Pearson's coefficient
    - Varies between -1 (max negative autocorrelation) and 1 (max positive autocorrelation)
- Local Spatial Autocorrelation
  - Local Index of Spatial Autocorrelation (LISA)
    - *Local* version of Moran's I
    - More frequently used for aggregated data (polygons)
    - Relies on defining contact matrices
  - Getis-Ord's G e Getis-Ord's G*

Hotspots (positive autocorrelation of high values)

Coldspots (positive autocorrelation of low values)

Spatial Analysis – Spatial Autocorrelation

Epidemiol. Infect., Page 1 of 8.
doi:10.1017/S0950268811001282

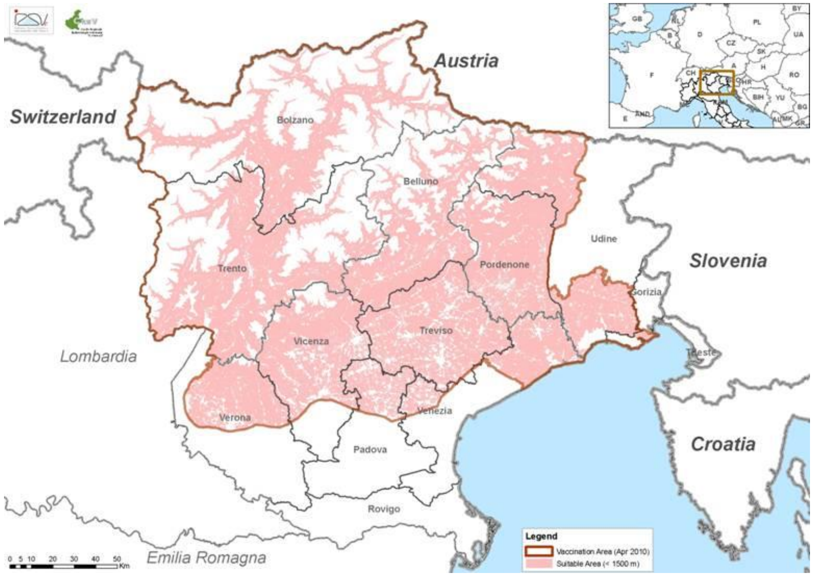
Emergency oral rabies vaccination of foxes in Italy in 2009–2010: identification of residual rabies foci at higher altitudes in the Alps

P. MULATTI[1]*, T. MÜLLER[2], L. BONFANTI[1] AND S. MARANGON[1]

[1] *Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro (Padua), Italy*
[2] *Institute of Epidemiology, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Wusterhausen, Germany*

Spatial Analysis – Spatial Autocorrelation

- Background
  - 2008-2011 Rabies epidemic in North-eastern Italy
  - Second emergency Oral Rabies Vaccinaton (ORV) (April-May 2010)



Suitable area for vaccination

Exclusion of:

- Urban areas
- Rivers and water basins
- Area above the freezing point

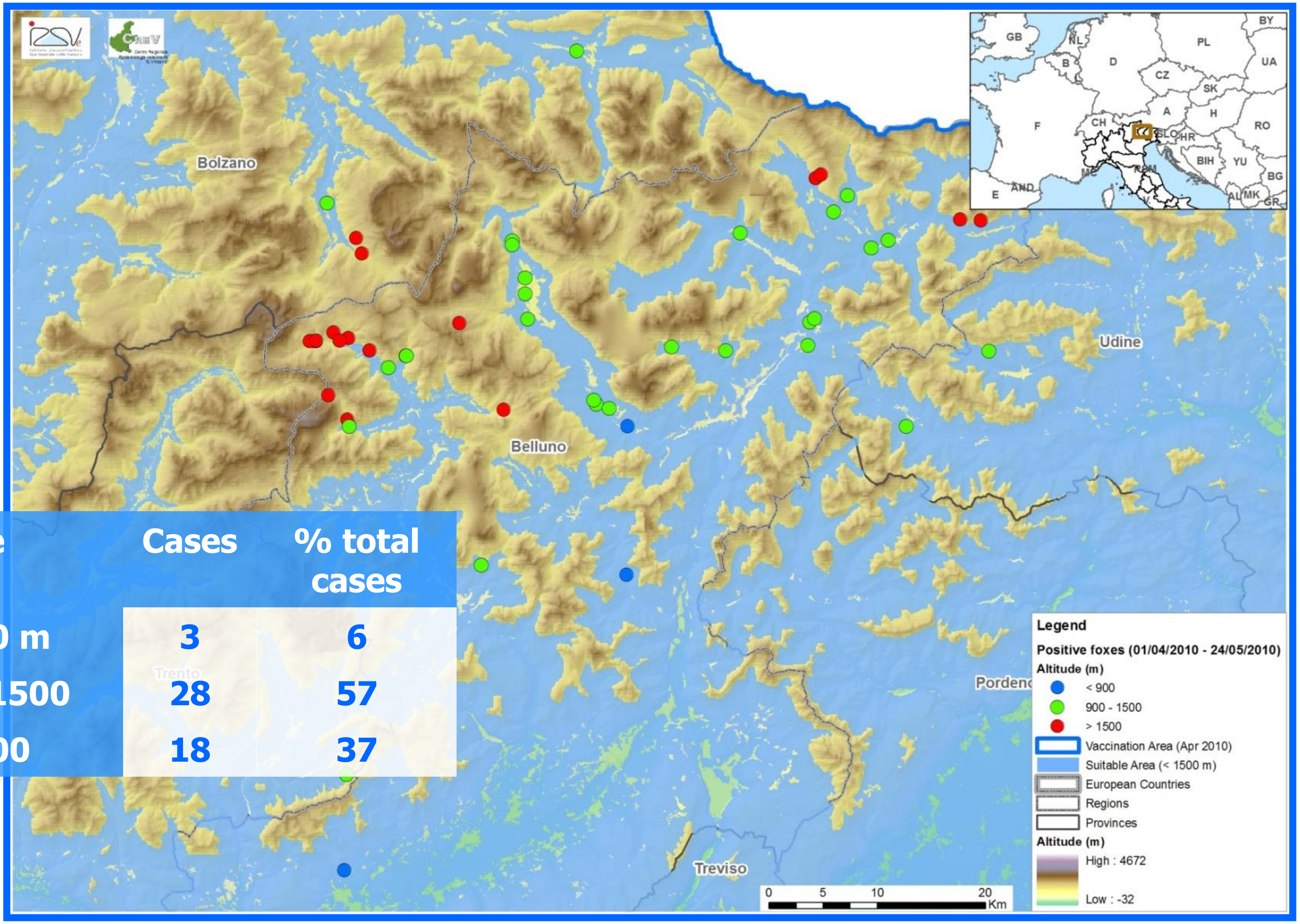
ORV Spring 2010

Freezing point: 1500 m a.s.l.

Spatial Analysis – Spatial Autocorrelation

- Aims
  - Detect clusters of rabies in unvaccinated areas
    - Risk of persisting rabies foci
  - Re-define vaccination strategies
- Explorative visual analysis
  - Rabies cases in foxes classfied as:
    - < 900 m (threshold altitude for the previous winter campaign)
    - 900-1500 m
    - > 1500 m
- Search for Hotspots
  - Getis-Ord’s G*
  - Variable of interest: elevation

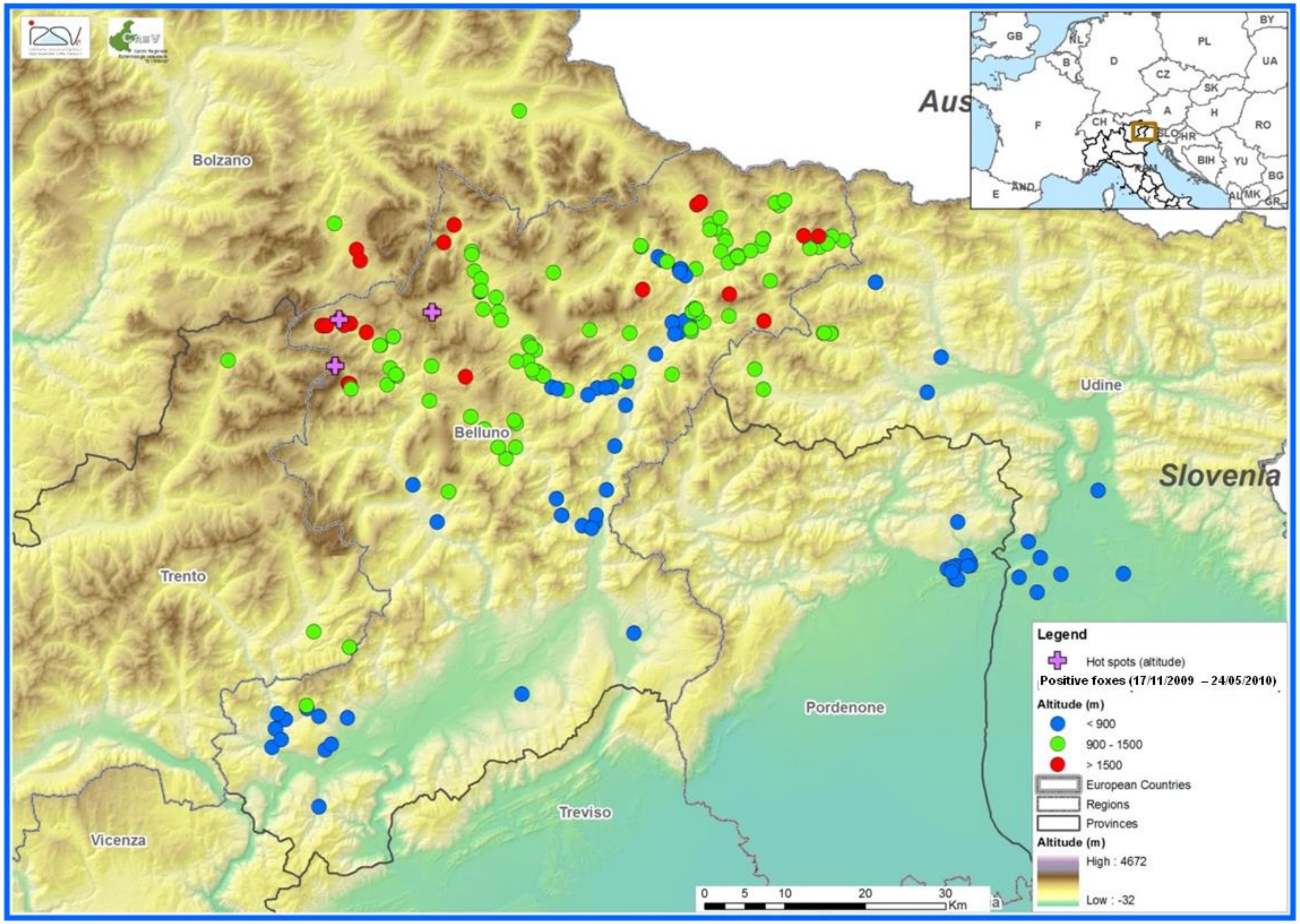
Spatial Analysis – Spatial Autocorrelation



Altitude	Cases	% total cases
# < 900 m	3	6
# 900-1500	28	57
# > 1500	18	37

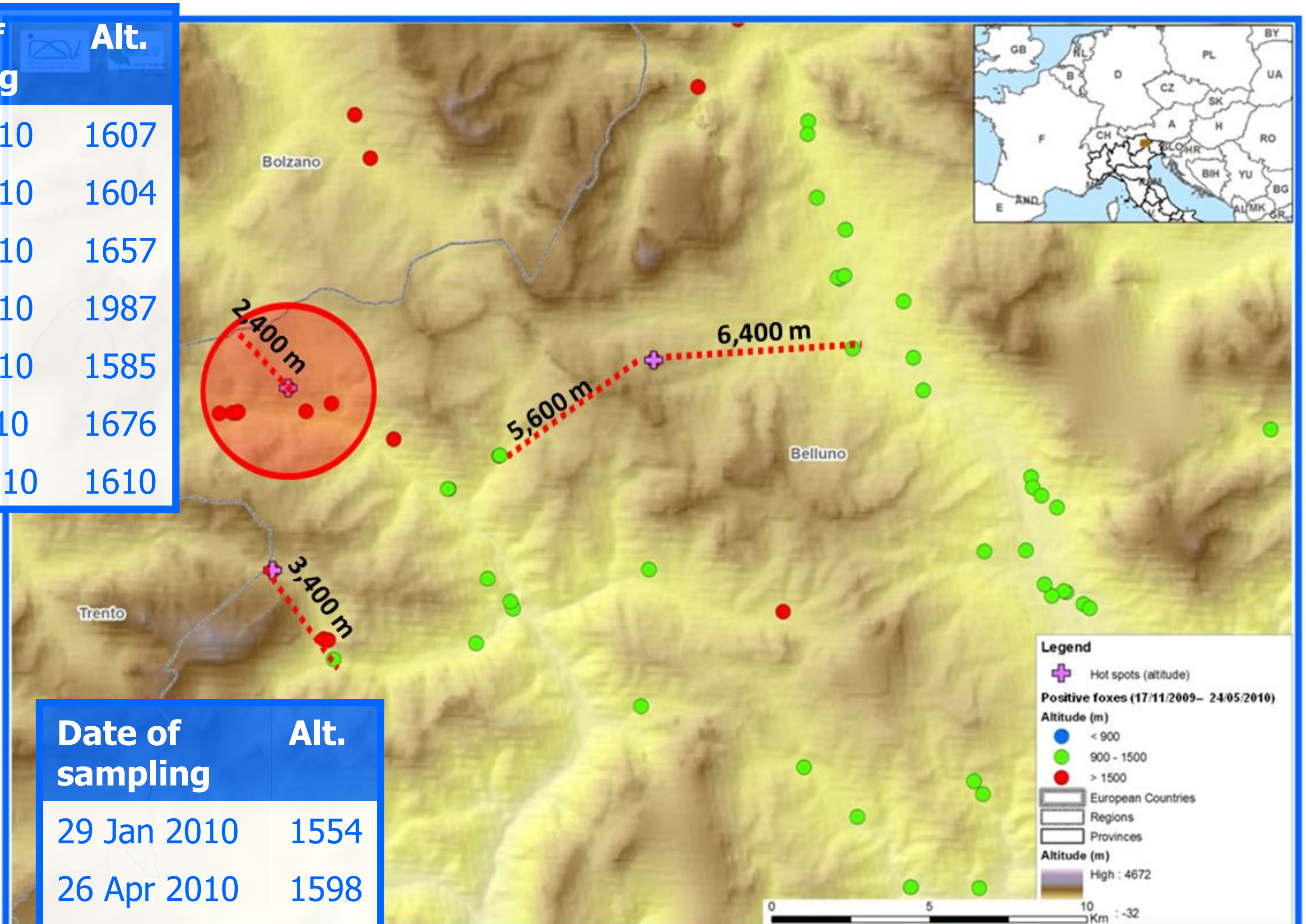
49 total positive foxes
(1 Apr 2010 – 24 May 2010)

Spatial Analysis – Spatial Autocorrelation



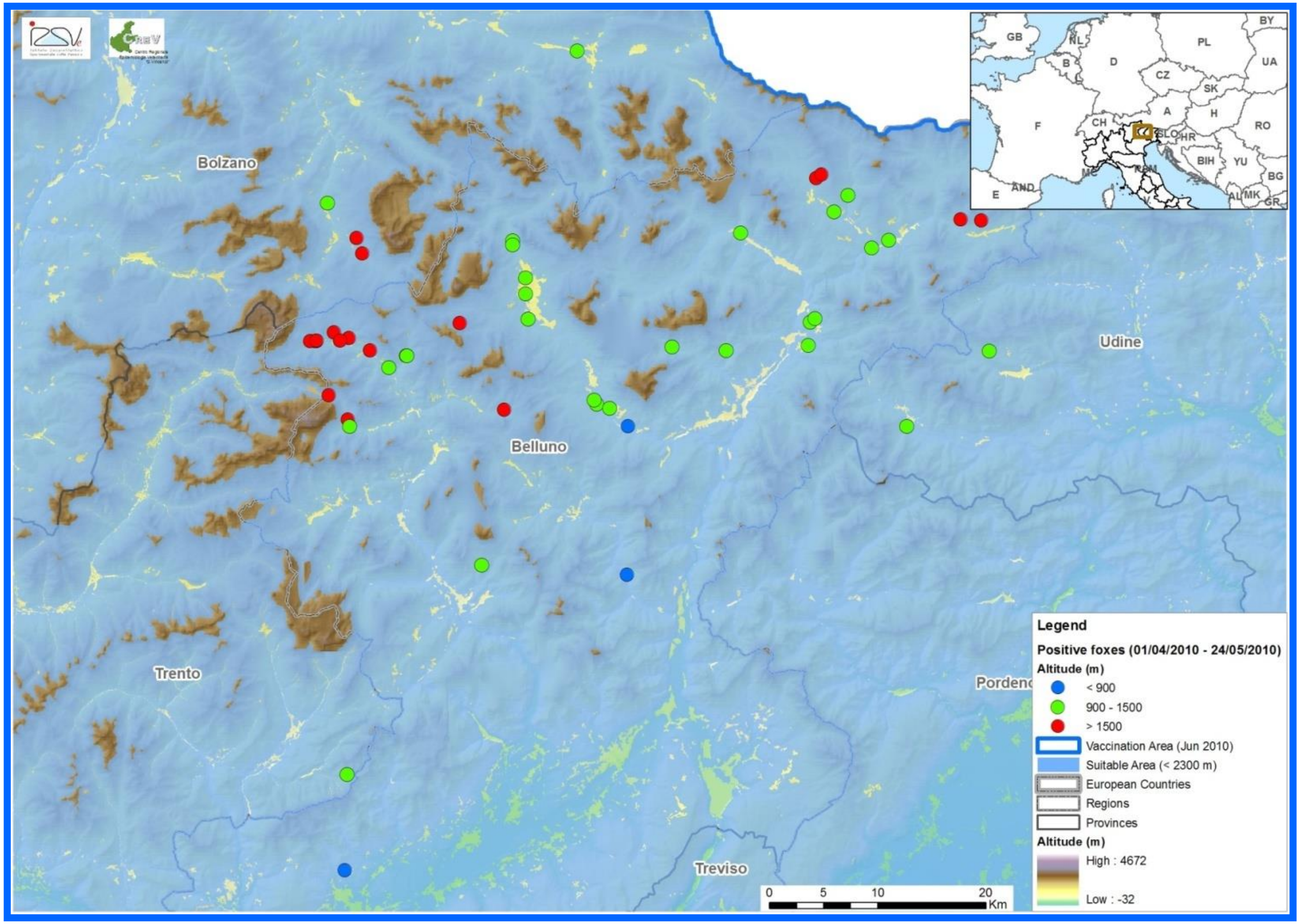
Spatial Analysis – Spatial Autocorrelation

Date of sampling	Alt.
11 Apr 2010	1607
16 Apr 2010	1604
17 Apr 2010	1657
19 Apr 2010	1987
24 Apr 2010	1585
3 May 2010	1676
14 May 2010	1610



Date of sampling	Alt.
29 Jan 2010	1554
26 Apr 2010	1598
10 May 2010	1446
21 May 2010	2061

Spatial Analysis – Spatial Autocorrelation

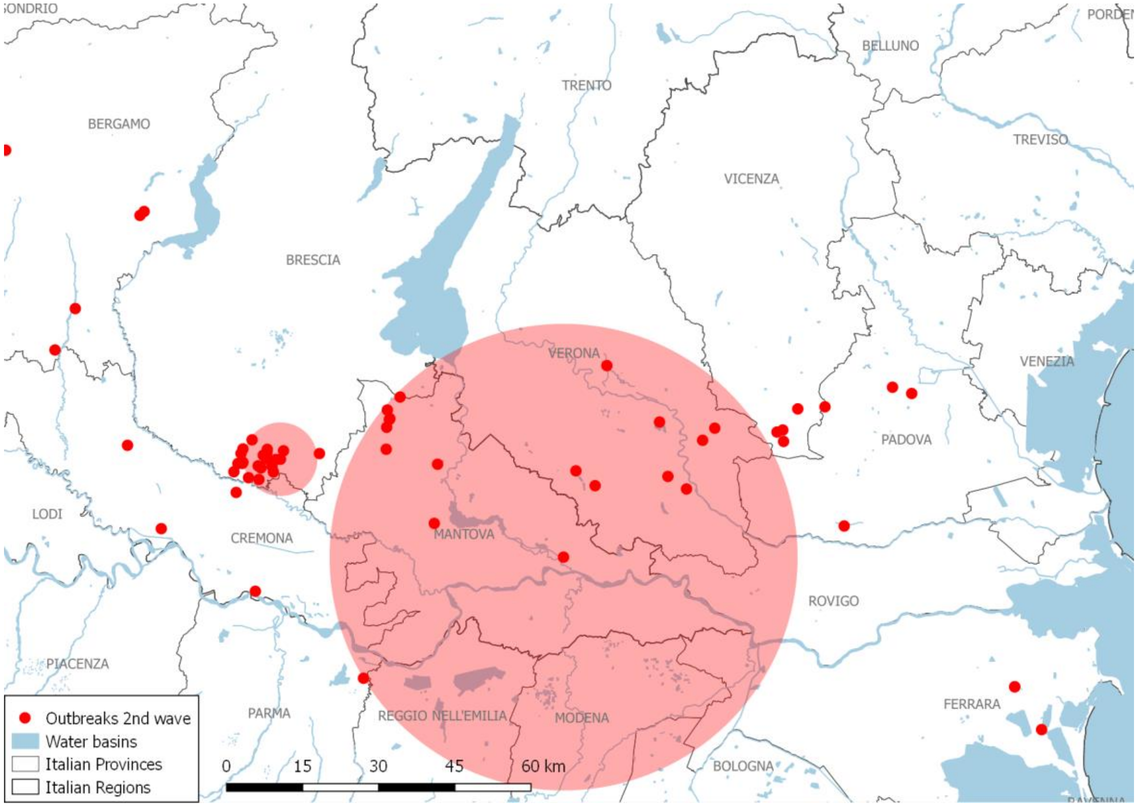


Space-time analysis

- Aims:
  - Assessing whether events of interest that occur in close proximity in space, are also in close proximity in time
- Greater support in hypotheses generation (especially in case of infectious diseases)
- The definition of the time scale and of the time unit is paramount
  - An uncorrect *time unit* might lead to bias when evaluating the space-time clustering
- Global Space-time analyses
  - e.g. Ripley's K(d,t)
- Local Space-time analyses
  - e.g. Space-time Scan Statistic

Space-time analysis

- Case study
  - Second epidemic wave of HPAI H5N8 in Italy (July 2017 – Decemer 2017)
  - 67 outbreaks
  - Assessing occurrence of space-time clusters
- Space-Time Scan statistic
  - Space-time permutation (does not require a population of reference)
    - Random labelling of occurrence dates among events
  - Uses a space-time search window



ID Cluster	Radius (km)	Start	End	No. Outbreaks included	p-value
1	7.25	2017/10/30	2017/11/12	14	< 0,001
2	45.98	2017/05/29	2017/08/27	19	0,001

Final remarks

- Accounting for space in Public Health allow to:
  - Get a better grip on dynamics of phenomena of interest
  - Define Risk-Based approachs on a geographic basis (risk-areas, surveillance)
  - Generate hypotheses on presence of risk factors (i.e. infected farms, contaminants, etc.)

- The application of spatial analysis on Public Health activities requires a multi-disciplinary approach:
  - Statistcs/Mathematics
  - Veternary
  - Epidemiology
  - GIS

Thanks for your attention!

Paolo Mulatti
pmulatti@izsvenezie.it

Istituto Zooprofilattico Sperimentale delle Venezie

OIE Collaborating Centre for epidemiology, training and control of emerging avian diseases

World Organisation for Animal Health

OIE Headquarters

CHINESE ACADEMY OF INSPECTION AND QUARANTINE

中国检验检疫科学研究院