

# **Information system for the management of Foot-And-Mouth Disease: validation of geocoding process accuracy in field epidemiology through the comparison of three georeferencing methodologies.**

**Bellini, S. 1, Cantoni, R. 2, Stefini, G. 1, Bonazza, V. 1, Zanardi, G. 1**

1 CERVES: Centro Nazionale di Referenza per le Malattie Vescolari - Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna. Brescia, ITALY.

2 Topotek, centro di competenza in geomatica e rilevamento del Consorzio di Innovazione Tecnologica (InnTec). Brescia, ITALY.

## **Abstract**

The Italian Reference Centre for Vesicular Diseases has realized a system to support decisions and operations in case of epidemic emergencies due to the introduction of FMD, a component of the system is GIS (Geographical Information System). The validity of epidemiological studies using GIS and Geocoding methods depends on the proportion of addresses that can be geocoded as well as the accuracy of the Geocoding process. Several studies have assessed the address matching rate of commercial Geocoding companies and found that matching rates are typically 60–80%. Verification of each subject's address location by other methods, like GPS, would require considerable time and resources, especially for a large study. Anyhow in epidemiology, positional inaccuracy of geocoded addresses may be an important source of misclassification. In this study we have compared the location of farm addresses measured by GPS, Geocoding and GeoZoo.

## **Introduction**

The introduction of Foot and Mouth Disease (FMD) in a Country would be particularly damaging due to its high spreading potential. Outbreaks of this disease could have severe consequences for animal welfare, livestock production, export of animal products and for the environment. The main line of defence when facing this type of danger, is the ability of organizing actions plans to fight the epidemic emergency with efficacy, efficiency and timeliness. CERVES has realized an information system for FMD and a component of the system is GIS.

Geographic information systems (GIS) are increasingly used by epidemiologists to screen and test hypotheses about environmental exposures and diseases. An important first step, in all GIS applications, is represented by the georeferencing operations that could be very long and expensive depending on the adopted method.

Geocoding, also referred to as “address matching”, is a method successfully used in other fields of application and resulted being fast and less expensive than the ones mainly used in epidemiology (GPS and desktop mapping software).

In this study we compared the location of farm addresses measured by three different methodologies: 1) Global positioning system (GPS) receivers, 2) Positions geocoded with a commercially available reference theme 3) Positions achieved by operators, through a home made desktop mapping software (GeoZoo).

Differences in positional accuracy between urban and non-urban areas were also verified. This urban–rural differential is very important since the available reference themes were designed for non epidemiologic purposes and are generally thought to be more accurate and complete in urban areas than in the non-urban ones.

## **Material and Methods**

Lombardy, with an extension of 23834 kmq, is a region of the Northern part of Italy with a densely populated area for cattle (23138 holdings) and pigs (3800 holdings). In Lombardia, a sample of 300 bovine farms, previously georeferenced by GPS, was selected for the comparison and to assess reliability of the other methodologies.

Urban addresses were defined as addresses within the urban region limits (Corine Land Cover definition). The other addresses were considered non-urban.

The methodologies used to collect sites were:

- Global Positioning System: a satellite system with global and continuous cover of the Earth. It allows us to determine the ground position and altimetry of a point with a precision that varies from a few millimetres to several meters in function of the type of equipment and the operating procedures of the survey. The GPS receivers utilized in the study was a Lowrance GlobalNav200, the producer reports a positional accuracy between 1 and 5 m. However, field tests indicate for GPS receivers an accuracy within 15–40 m. The sites have been collected directly by personnel recording latitude and longitude in front of the stable.
- Geocoding: is the process whereby the relative position of an address is linked to a reference theme (database containing both address and locational information: ie, latitude and longitude). Geocoding: works matching address giving back a couple of coordinates. This system is based on a road graph continuously updated, and needs as input a complete address (ie, “via Italia 51, Brescia”). In case the address is not complete, a mean value of latitude and longitude of the street segment is assigned. To determine latitude and longitude of each address Navteq dataset was used.
- GeoZoo: is a home made desktop mapping software, it allows us to register the position of a site with some simple tools over the digital map (orthoimages at scale 1:5.000). This instrument was used by operators of the regional veterinary services and about 30.000 farm positions were collected.

Both, Geocoding and GPS receivers were set to collect coordinates in decimal degrees using the World Geodetic System 1984 map datum. East and North coordinates provided by Geozoo are in Gauss Boaga (Datum Roma40) projection. For the comparison among the three different methods, Latitude and longitude obtained with GPS and Geocoding were converted to East and North coordinates (values measured in meters units representing a point on a plane instead of a geoid) referred to Gauss Boaga (Datum Roma40) projection. To increase the conversion accuracy seven parameters of roto-translation were applied.

GPS was considered the gold standard method to use as comparison. For each address, the distance between the two points was calculated by determining the Euclidean length of the hypotenuse of the right triangle formed by the two points:  $[(x1 - x2)^2 + (y1 - y2)^2]^{1/2}$ , where  $x1$  is the GPS latitude,  $x2$  is the geocoded latitude,  $y1$  is the GPS longitude, and  $y2$  is the geocoded longitude. This formula does not correct for the Earth curvature but does not introduce sizable error in the distance calculation between the two points, since such distances were relatively small.

For the purpose of this study an error of 200 meters was considered acceptable.

## Results

Geocoding address matching was achieved for 209 of the 300 (70%) sites under study. The results obtained by comparing Geocoding and GeoZoo with GPS methodology are shown in Table 1. In addition, a differentiation was also made according to “urban” and “non urban” location of the sites. For geocoding, 76% was the percentage of “errors” within 200 mt (84% for urban area and 75% for the non-urban). Only a small proportion of sites (1 %) was over 1000 m.

For GeoZoo, 97% was the percentage of “errors” within 200 mt (92% for urban area and 97% for the non-urban). Only a small proportion of sites, less than 1%, was over 1000 m.

**Table 1:**

urban sites: 25  
 non-urban sites: 184

		GPS - Geocoding			GPS - GeoZoo		
		tot	urban	non-urban	tot	urban	non-urban
min	meters	4	9	4	8	38	8
max	meters	4261	279	4261	1926	281	1926
mean	meters	180	122	188	102	110	101
< 25mt	%	4	12	3	3	0	3
< 50mt	%	11	16	10	13	8	14
< 100mt	%	31	36	30	64	48	66
< 200mt	%	76	84	75	97	92	97
< 500mt	%	97	100	97	99	100	99
< 1000mt	%	99	100	98	100	100	99
25 th percentile	meters	84	70	85	63	73	62
50 th percentile	meters	133	110	135	81	100	80
75 th percentile	meters	198	176	199	111	129	107
90 th percentile	meters	270	223	273	143	159	140

## Conclusion

According to the results obtained, GeoZoo is more accurate than geocoding. GeoZoo was more accurate in non-urban area, while geocoding works better in an urban context.

The geocoding slightly decreased accuracy in non urban area is due to the Geocoding process itself. Geocoding uses interpolation to estimate the relative position of an address on a line segment in the reference theme. In Geocoding process, the likelihood of inaccuracy is higher for an address located in areas with long street segments compared to the one with short ones, characteristic of urban area. The problems observed using GeoZoo were mainly in urban context, where the correct site position could be not easy to recognize. For the purpose of this study, the 2 tested methodologies were sufficiently accurate. Worth to be considered that also GPS receivers are prone to error in positional accuracy which mainly arise from three sources: satellite related errors, signal propagation errors, and receiver errors. Lowrance reports that the Lowrance GlobalNav200 has a positional accuracy between 1 and 5m. Field tests indicate that civilian GPS receivers have a range of accuracy within 15–40 m.

The efficiency of the matching rate obtained (70%, typically 60–80%) needs to be evaluated also in contest where the methodology is applied. In fact, the trial was conducted in Lombardia, a region densely populated for farm animals, where to use exclusively GPS receivers would have required time and a lot of dedicated personnel. In conclusion, there is non clear best georeferencing method, it is necessary to apply the suitable methodology referred to the proper context and the choice is also strongly conditioned by the available resources (working time needed, costs).

## References

- Krieger N, Waterman P, Lemieux K, et al. (2001) On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health*; 91:1114–1116.
- Moore DA, Carpenter TE. (1999) Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev.*; 21:143–161.
- Vine MF, Degnan D, Hanchette C. (1997) Geographic information systems: their use in environmental epidemiologic research. *Environ Health Perspect*; 105:598–605.
- Drummond WJ. (1995) Address matching:GIS technology for mapping human activity patterns. *J Am Planning Assoc.*; 61:240–251.
- Hofmann-Wellenhof B, Lichtenegger H, Collins J. (1997) Global Positioning Systems: Theory and Practice. Springer-Verlag, New York.
- Eric A. Whitsel, Kathryn M. Rose, Joy L. Wood, Amanda C. Henley, Duanping Liao, Gerardo Heiss (2004) Accuracy and Repeatability of Commercial Geocoding. *Am J Epidemiol*; 160:1023–1029.