In Giornale Italiano di Aritmologia e Cardiostimolazione, Special Issue Proceedings of the 15th Int. Congress "The New Frontiers of Arrhythmias", v. 5, n.1, Mar 2002, 241-245

New WEBGIS technologies for geo-location of epidemiological data: an application for the surveillance of the risk of Lyme borreliosis disease

Cesare Furlanello, Stefano Merler, Stefano Menegon, Sebastiano Mancuso*, Gianni Bertiato*
*ITC-irst, Trento, *ULSS 1 Belluno, Italy*

Abstract

We present a technology for the accurate and fast geo-location of medical data and the creation of central data archives, specifically designed for the development of disease risk maps and of other functions for modern epidemiology and surveillance. A WEBGIS system is configured as an Internet web service integrated with connectivity to a Geographical Information System (GIS). We developed for the ULSS Belluno a WEBGIS for the accurate mapping of tick-borne diseases, with specific attention to Lyme borreliosis, which may cause cardiac manifestations as atrioventricular conduction abnormalities, complete atrioventricular block, myocarditis and dilated cardiomiopathy. A first tree-based predictive model has been developed for risk classification of tick bites from 256 samples (data gathered through the Belluno Lyme WEBGIS), with a descriptive accuracy of 81.9% and a predictive accuracy of 75.1% . An experimental risk GIS map was therefore obtained from the model by considering altitude, week of sampling and vegetation type as predictor variables.

The analysis of spatial patterns in modern epidemiology has been greatly supported by the development of Geographic Information Systems (GIS), software environments whose main functionality is the processing of geo-referenced information. Adding a geographical location to facts and findings of medical interest allows to consider the biotic or abiotic factors that influence the spatial distribution of a disease, as in the famous historical example of John Snow mapping cholera cases in relation to water pumps in London in 1854 [1]. In order to recognize the associations between disease distribution and environmental variables and to advance public health measures, possibly on the base of predictive models, it is however crucial to organize a flow of data, effective and systematic. GIS are still seen as tools for specialists, thus it is rare to find precise geolocation be included in disease notifications within a continuous process of data collection and disease surveillance. On the contrary, the study of infectious ilnesses, but also the epidemiology of all the main health risks, should include elements for an effective monitoring system on a spatial basis in order to implement surveillance of outbreaks of emergent diseases.

An important example is given by Lyme borreliosis, an emergent tick-borne disease, now the most common vector-borne ilness in the United States, and extensively distributed in Europe, Asia and Australia. The causative agent, the bacterial spirochete *Borrelia burgdoferi* has a high transmission potential due to the ability of the compe-

tent vectors, as the *Ixodes* ticks, of feeding on a very wide range of vertebrate host species. The distribution of Lyme disease has been detected by mapping the distribution of its vector [2,3], by considering infection prevalence on vectors [4], and with information from serologic analysis of potential hosts or of patients and retrospective reports of tick bites. Lyme disease may result in a disseminated illness unless treated early. In 8% of patients, it may be cause of Lyme carditis, with cardiac manifestations as atrioventricular (AV) conduction abnormalities at different levels, intra-atrial, atrioventricular, intraventricular [5,6]. The leading manifestation is represented by complete AV block with asystole and cardiac arrest [7,8,9]. Ectopic complex ventricular beats and ventricular tachycardia are documented, and Lyme disease seems to be associated to myocarditis and dilated cardiomiopathy [10,11]. Many of these abnormalites may be transient after careful diagnostic evaluation and treatment. As suggested by several authors [8,9,12,13], in endemic areas borreliosis should be suspected in patients in AV block of unknown origin in young patients, or in atypical profiles for cardiac abnormalities. However, sensitivity of diagnostic serology tests for Lyme disease is unreliable, as it may result negative in earlier (first 4 to 8 weeks) or later stages [14]. It is thus important to recognize the insurgence of Lyme borreliosis from the main symptom of its early stage, the erythema *migrans,* or for any available retrospective information, in particular whether the patient may recall a tick bite in a currently endemic areas. Risk mapping and prompt notification of tick bites and of new cases of Lyme disease are thus fundamental as an appropriate information on possible exposure [12]. The same information is also instrumental for enforcing public awareness and disease control measures.
In this paper we present a new public health resource for risk mapping, surveillance and diagnosis support of Lyme disease and of other tick-borne ilnesses. The system integrates several geo-informatics tools and it has been experimentally applied in the year 2001 for creating a digital risk map for the ULSS 1 Belluno area (a project funded by FEOGA, Italy-Austria INTERREGG II).

Methods and Data

Methodology: the software technology

This study completes the statistical methodology introduced in [12,4] with a technological infrastructure for continuous notification of epidemiological data related to the control of tick-borne diseases. At the same time this infrastructure may be used for many other data collection and management tasks for modern epidemiology.

Our WEBGIS technology is based on an original modification and integration of several of the major open source software systems for data management (PostrgreSQL), geographical data visualization through internet (MapServer), geographical data analysis (GRASS GIS), computational statistics (R), synchronized update of software (CVS), and Internet server management (Apache). Other software functions which are integrated in the system are based on languages such as Perl, Java, Javascript, and PHP to facilitate the interaction between the functionalities of a central server and the authorized users from remote clients. In particular, on-field findings and disease notifications may be inserted via Internet in the central data base using an interface that allows the automatic registration of geographical coordinates by clicking on appropriate digital maps. The maps and all the WEBGIS functions are visualized in a browser window, and thus remain as much as possible independent from the hardware and operative systems of the users. The service may be accessed from any client computer connected to Internet, opening also opportunities for the forthcoming mobile computing hardware. The interface provides raster and vector base maps, including a cover by high resolution (1 m) digital colour ortophoto from the recent Volo Italia 2000. also included are roads, game district borders, digitized 1:25000 cartography, vegetation maps. A digital elevation model (DEM) is also available: at the time of the study we reconstructed an interpolated model at the 40 m cell resolution, currently being updated to a cell resolution of 25 m. Most of the digital raster maps are in GeoTIFF format, while shapefile / DBF and MapInfo formats are used for vectors and site lists; however data in GRASS format are also included that can be directly accessed using the GDAL conversion libraries. The PostGIS extension and the OGR conversion library PostgreSQL allow an integration of the geographical data structures with the geolocated data base of epidemiologic information. Figure 1 summarizes the structure of the system.

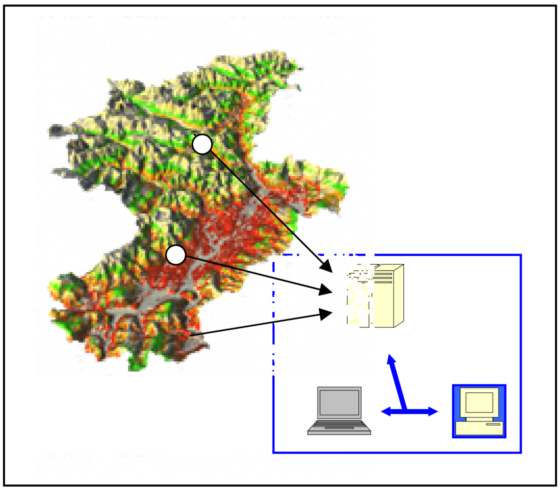

Fig. 1. The WEBGIS system architecture

This system may therefore take advantage from geographical resources available in very different formats from different administrations. In this sector, data interoperability is a serious problem, but the generic user of this WEBGIS does not need to care about integration of data sources nor about formats.

These resources ease the process of accurately locating the sites analysed for tick presence and the areas in which a tick bite may have occurred. All the users share the same cartography, each at the needed definition, and automatic tessellation of maps and aerial imagery is performed. It must be noted that more than 12Gb are needed to store the colour ortophoto cover. It is important to remark that system use is restricted to authorized users of ULSS Belluno and project participants, with access protected by password and session control. The WEBGIS server is currently hosted by ITC-irst in Trento on a biprocessor Pentium with data mirroring (Raid 1).

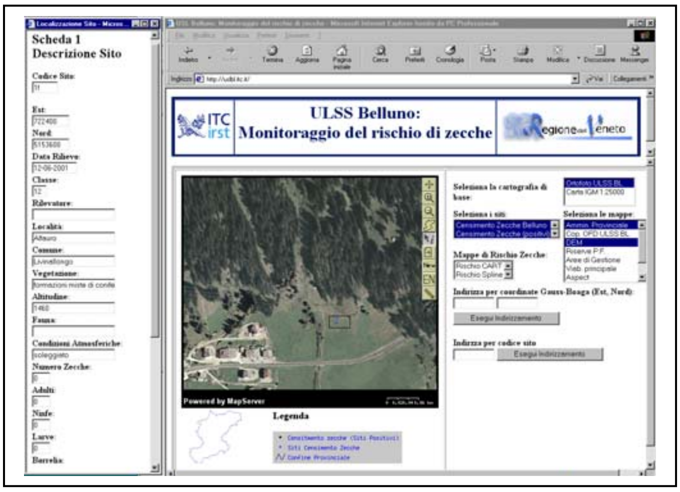

Fig. 2. A data notification example. Site coordinates in the Gauss-Boaga coordinate system are automatically attached to a form (on the left) with on-site measures and results of PCR analysis for the presence of Borrelia.

The main use of this system is the construction of geolocated data bases for epidemiology and surveillance. For a Lyme borreliosis study, the interface may be used for planning in advance the data collection and for accurately locating where the tick specimen were collected in order to build accurate risk maps. It may also be used for a retrospective location of a tick bite, e.g. after the tick has been found and extracted by a public health operator. The interface might also be used for systematic verification of risk exposure to Lyme whenever a possible Lyme carditis manifestation is suspected. The location may be compared with a Lyme risk map, developed as described below.

The interface is also convenient for entering historical data without requiring a complete GIS software: as a fraction of GIS plotting and data analysis functionalities is required, the WEBGIS interface avoids that a GIS specialist has to be hired or trained, and the need for expensive software maintenance. If any specific geostatistical and epidemiological computation is needed, it can be run offline using data exported from the system. If the same computation has to be run several time, or has to be promptly available after a new data or a group of data has been introduced, this functionality may be standardized, implemented on the WEBGIS server and made available to all the authorized users. The visualization of possibly automatically updated maps of the spatial distribution of a disease has a clear advantage over static maps in terms of disease diagnostic and control.

In any case, the system is designed to provide a complete data acquisition process, continuous over time, backed by a data base management system which may be connected to other medical informatics resources. In the next sections, we discuss the experimental use of the system for the collection of research data and the creation of a predictive GIS model.

Experimental design, Data and Modelling techniques

The goal of the 2001 data collection was to understand the spatial aspects of Lyme disease risk distribution in the province of Belluno. Following the experimental results obtained in the similar Alpine landscape of Trentino
[3,12], sampling effort was spatially distributed in order to adequately cover the environmental diversity potentially associated with tick habitats. As it was intended to develop a GIS model, capable to generalize to the entire province the findings of a limited number of samples, the experimental design was based on 30 classes obtained by combining three main digital data set of Belluno: namely we considered the product space generated by 5 vegetation classes, roe deer density (quantized in 2 classes), and 3 elevation classes (14-543 m, 544-1079 m,1079-1589 m, at 250 m cell res. and for cut points defined by 3-quantiles and considering habitat limited below the 1600 m on the base of all the Trentino studies.
Sampling classes were defined and colour marked on each of the 15 IGM (Istituto Geografico Militare) raster maps (1:25000) which cover the ULSS Belluno territory, eliminating classes which globally covered less than 1000 ha.

A total of 256 sampling sessions in 214 distinct sites was performed by standard flagging for 5' – with 5 interval of duration 1'. The location of each sampling site was georeferenced on a map and then finely checked with the WEBGIS system described above, also entering additional site descriptions. Almost 4300 *Ixodes ricinus* were collected during the study in 2001. Molecolar biology tests (PCR) were then performed for *Borrelia burgdoferi, Ehrlichia, Rickettsia, Babesia* and TBEV. Prevalence for *Borrelia* in nymphs (1114 examined) resulted 14% , conferming the previous findings in Trentino [3,4].
Sampling sites resulted distributed between 230 m and 1520 m (m=687,med=935,stdev=318). Altitude resulted significant for tick presence ((t = -6.528, df = 245, p-value = 0, conf.level=0.99), see FIG. 3. Average elevation for tick presence was mu-P = 839 m, while for sampling sites where ticks were not found average elevation was mu-A = 1100 m.

For risk prediction, in this study we considered the recursive partitioning procedure *rpart* within the statistical analysis environment S-PLUS 6.0 and its Open Source version R [15], in order to estimate a classification models for tick-presence. The *rpart* implementation allows the treatment of missing variables through the use of surrogates splits, allowing to perform the modelling exercise without dropping out incomplete cases.

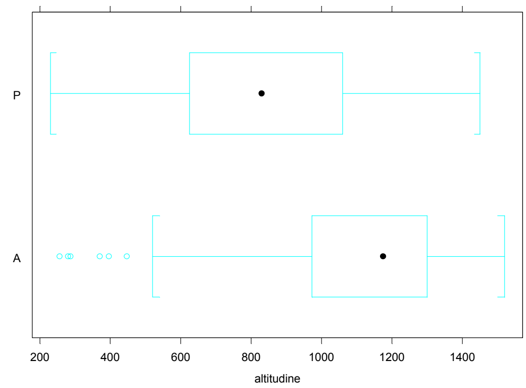

FIG. 3: Tukey boxplots of altitude distribution for tick presence (P) and absence(A)

Results

The rpart procedure automatically selected altitude, experimental design class (altitude x vegetation type from maps x density of roe deer from maps), and vegetation evaluated on site. Weather conditions, and presence of wildlife on site did not resulted significant. According to the model, ticks are likely to be present above 1124 m only in grass highlands with roe deer presence and within the June peak, while below 1124 m ticks are everywhere present with the exclusion of vegetation of medium-high mountain and in absence of roe deer. This model has descriptive accuracy of 83.3% anda predictive accuracy of 71.5%. Ten-fold cross validation was used to select an optimal model for prediction on novel data, obtaining a reduced model with 80.9% descriptive accuracy and 75.1% predictive accuracy. This is an acceptable error which may reduced by considering more advanced classification models as bagging or boosting, as demonstrated in previous computational studies [4]. Finally, a tree based classifier was developed considering only variables available on GIS: altitude and main vegetation types. A GIS raster map was created with cell size of 250 m, resulting in the first experimental map for Belluno, at a 21% of descriptive error.

Conclusion and current development

In this work we have discussed the transferability of our GIS methodology for Lyme risk assessment at an interregional scale. A first risk model has been found for Belluno that it is similar to the one developed as a numerical cartography for Trentino [4]. We are currently developing a new methodology based on geolocated remote sensing input data for a continuous update of predictive models.
These results are now based on an integrated data server system with Internet connection to digital cartography (WEBGIS). The WEBGIS technology may constitute a powerful instrument for surveillance in epidemiology of infectious diseases. In particular, it is expected that the availability of the WEBGIS will allow the continuous monitoring of Lyme disease, now an important emergent infectious illness for the Belluno area (see Tab.1), with a more specific attention and support to the diagnosis of Lyme carditis.

TABLE I. The evolution of Lyme disease in the study area (ULSS1 Belluno)

Year	Residents	Not Residents
1992	5	11
1993	9	11
1994	31	12
1995	6	4
1996	46	12
1997	72	13
1998	53	28
1999	58	35
2000	48	17

Acknowledgments

The authors wish to thank Roberto Flor of ITC for his continuous technical support in the integration of the WEBGIS architecture. Bruno Caprile, Markus Neteler and Gianni Lazzari of ITC-irst contributed with many important suggestions to this project and to its management. The first risk models of tick presence were developed in Trentino with Claudio Chemini and Anna Paola Rizzoli of the Centre of Alpine Ecology. G. Piccolin and colleagues at the ULSS Biology Lab studied prevalence of Lyme in the sampled ticks. Tick data were collected and managed by Pasa and colleagues at the Wildlife Management Service of Belluno and by the Forest Corps. This project received funds by FEOGA, Italy-Austria Veneto-Carinzia INTERREGG II, DGRV 3739/99, cod. AVEN222057.

References

1. Kitron U: Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *J Med Entomol* 1998; 35: 435-445.
2. Guerra M, Walker E, Jones C, et al: Predicting the risk of Lyme disease : Habitat suitability for *Ixodes scapularis* in the North Central United States. *Emerging Infectious Diseases* 2002 March; *8(3)*

3. Merler S, Furlanello C, Chemini C, Nicolini G: Classification tree methods for analysis of mesoscale distribution of *Ixodex ricinus* (Acari: Ixodidae) in Trentino, Italian Alps. *J Med Entomol* 1966; 33: 888-893.
4. Rizzoli A, Merler S, Furlanello C, Genchi C: Geographical information systems and bootstrap aggregation (bagging) of tree-based classifiers for Lyme disease risk assessment in Trentino, Italian Alps. *J. Med. Entomol.* 2002; in press.
5. Steere AC, Batsford WP, Weinberg M, et al: Lyme carditis: cardiac abnormalities of Lyme disease. *Ann Intern Med* 1980; 93: 8-16
6. Klein J, Stanek G, Bitter R, Horvat R, et al: Lyme borreliosis as a cause of myocarditis and heart muscle disease. *Eur Heart J* 1991 Aug; 12 Suppl D:73-5
7. Vasilijevic Z, Dimitrovic R, Naumovic Z, Ostojic M, ET AL. Common form of Lyme borreliosis carditis – complete heart block with sincope: report on 3 cases. *Cardiology* 1996 Jan-Feb; 87(1): 76-8
8. Rosenfeld ME, Beckerman B, Ward MF, Sama A: Lyme carditis: complete AV dissociation with episodic asystole presenting as sincope in the emergency department. *J Emerg Med* 1999 Jul-Aug: 17(4): 661-664
9. Batalla Celorio A, Gonzalez Camino F, Suarez E, Martinez J, et al: Transient complete atrioventricular block in Lyme disease. *Rev Esp Cardiol* 1999; 52: 529-531.
10. Cuffe MS, Lyme disease: an active agent in dilated cardiomiopathy? *Am Hearth J* 1999 Aug; 138 (2).
11. Scheffold N, Sucker C, Berger-Klein J, Kaag N, et al: Acute myocarditis and cardiomiopathy in Lyme borreliosis. *Z. Kardiol* 2000 Nov; 89 (11):1046-52
12. Furlanello C, Merler S, Chemini C: A new bootstrap method for risk assessment of exposure to Lyme disease. *G Ital* Cardiol 1998; 28, 1; 167-171
13. Weissman K, Jagminas L, Shapiro MJ: Frightening dreams and spells: a case of ventricular asystole from Lyme disease. *Eur J Emerg Med* 1999 Dec; 6(4): 397-401
14. Lardieri G, Salvi A, Camerini F, Cinco M, et al: Isolation of *Borrelia burgdoferi* from myocardium [letter]. *Lancet* 1993;342:490
15. Thernau TM, Atkinson EJ: An introduction to recursive partitioning using the RPART routines. *Mayo Foundation Technical report* 1997;