ARTICLE

Geographical Information Systems and Bootstrap Aggregation (Bagging) of Tree-Based Classifiers for Lyme Disease Risk Prediction in Trentino, Italian Alps

ANNAPAOLA RIZZOLI,[1] STEFANO MERLER,[2] CESARE FURLANELLO,[2, 3] AND CLAUDIO GENCHI[4]

J. Med. Entomol. 39(3): 485–492 (2002)

ABSTRACT The risk of exposure to Lyme disease in the province of Trento, Italian Alps, was predicted through the analysis of the distribution of *Ixodes ricinus* (L.) nymphs infected with *Borrelia burgdorferi* s.l. with a model based on bootstrap aggregation (bagging) of tree-based classifiers within a geographical information system (GIS). Data on *I. ricinus* density assessed by dragging the vegetation in 438 sites during 1996 were cross-correlated with the digital cartography of a GIS, which included the variables altitude, exposure and slope, substratum, vegetation type and roe deer density. Ticks were more abundant at altitudes below 1,300 m a.s.l., in the presence of limestone and vegetation cover with thermophile deciduous forests and high densities of roe deer. A bootstrap aggregation procedure (bagging) was used to produce a model for the prediction of tick occurrence, the accuracy of which was tested on actual tick counts assessed by a further dragging campaign carried out during 1997 to determine infection prevalence and resulted in average 77%. Other tests of the model were made on additional and independent data sets. The prevalence of infection with *Borrelia burgdorferi* s.l, determined by polymerase chain reaction on 2,208 nymphs collected by random dragging in 245 transects selected within eight areas where the model predicted the occurrence of *I. ricinus* during 1997, was 17.5% and was positively correlated to tick abundance and roe deer density. These findings were used to relate the output of the bagged model (probability of tick occurrence) to the density of infected nymphs through a stepwise model selection procedure and thus to produce a GIS digital map of the probability distribution of infected nymphs in the Province of Trento at high resolution scale (50 by 50-m cell resolution). The application of the bagging procedure increased the accuracy of the prediction made by a single classification tree, a well-known classification method for the analysis of epidemiological data.

KEY WORDS *Ixodes ricinus*, Lyme disease, geographical information system, bagging, tree-based classifiers, risk prediction

LYME DISEASE IS the most common bacterial infection transmitted by ticks in the boreal hemisphere (Gray et al. 1998a). In Europe, the tick *Ixodes ricinus* (L.) acts as the main vector of infection that is maintained by a series of competent reservoirs, at least nine species of small mammals, seven species of medium-sized mammals, and 16 species of birds (Gern et al. 1998a, Gray 1998, Humair et al. 1999).

In Italy, Lyme disease was first recorded in 1983. The causative agent *Borrelia burgdorferi* sensu lato (s.l.) and the genospecies recognized as pathogenic for humans, *Borrelia burgdorferi* sensu stricto (s.s.), *Borrelia afzelii* and *Borrelia garinii* (Johnson et al. 1984, Baranton et al. 1992, Canica et al. 1993) were isolated from humans and the tick *I. ricinus* (Burioni et al. 1988, Genchi et al. 1994, Bandi et al. 1997, Ciceroni and Ciarrocchi 1998, Cinco et al. 1998).

Currently, the disease is endemic in at least six regions of central and northeastern Italy (Liguria, Emilia- Romagna, Friuli Venezia Giulia, Veneto, and Trentino-Alto Adige) with 1,171 cases recorded between 1986 and 1997, although this number is considerably underestimated (Pavan et al. 2000). The incidence rate of Lyme disease in the Trentino-Alto Adige region, one of the "hot-spots" of infection, was 8.02 cases/100,000 inhabitants estimated for the period 1986–1997 (Pavan et al. 2000).

The mesoscale distribution of *I. ricinus* in the Province of Trento was determined by dragging the vegetation and by screening roe deer shot by hunters (Chemini et al. 1993, 1997). Environmental variables including altitude, exposure, vegetation type, and geological substratum were used to develop a tree-based

---

The use of wild animals and the collection of data followed a protocol approved by the Government of the Autonomous Province of Trento and is on file in our laboratory.

[1] Centro di Ecologia Alpina, Viote del Monte Bondone, 38040 Sardagna (Trento), Italy.

[2] Istituto per la Ricerca Scientifica e Tecnologica, 38050 Povo (Trento), Italy.

[3] E-mail:furlan@itc.it.

[4] Istituto di Patologia Generale Veterinaria, Facoltà di Medicina Veterinaria, Università degli Studi di Milano, 20100 Milano, Italy.

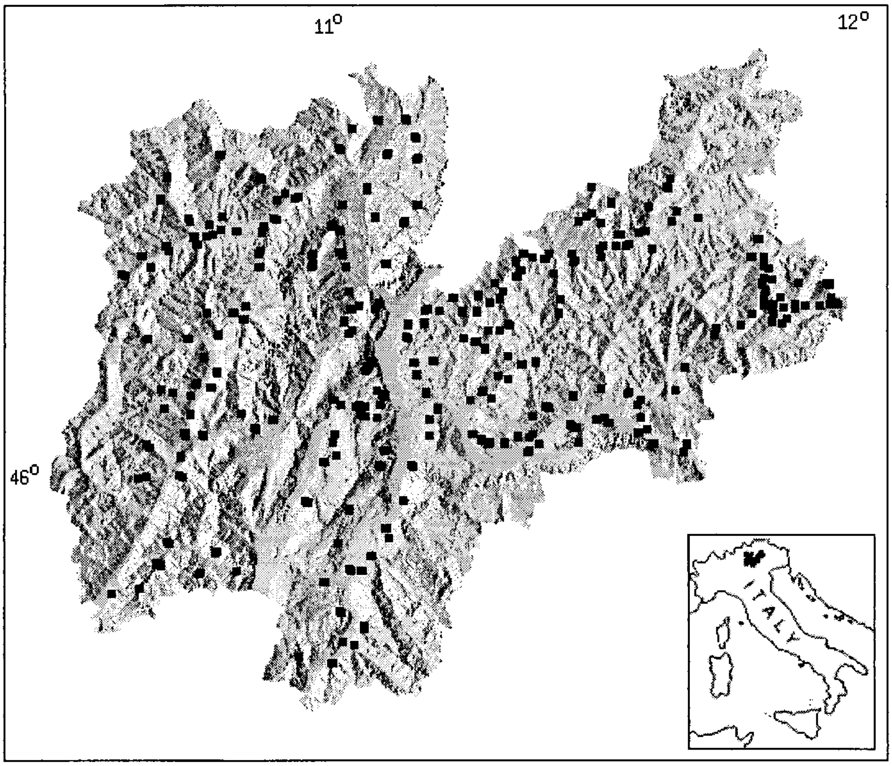

Fig. 1. Study sites sampled for *I. ricinus* in the Province of Trento, Italian Alps.

model for the assessment of the habitat patterns with the highest probability of tick occurrence (Merler et al. 1996). The model showed that altitude and geological substratum were the most important variables affecting *I. ricinus* occurrence, with a decrease of tick presence above 1,100 m and on volcanic rocks. At lower altitudes, ticks occurred more frequently in woods and coppices with a low or medium cover density, in vegetation types linked to human land use change, such as coppices with *Ostrya carpinifolia* Scopoli, and patches of *Corylus avellana* L. and *Populus tremula* L. (Merler et al. 1996).

The objective of this study was to provide local human health authorities with a high degree resolution predictive risk map of *I. ricinus* tick distribution and probability of infection with *Borrelia burgdorferi* sensu lato (s.l.) for the province of Trento. We combined the use of a geographical information system (GIS), already recognized as a powerful tool for the geospatial analysis of the distribution of vector borne diseases (Glass et al. 1995, Nicholson and Mather 1996, Furlanello et al. 1997) with a bootstrap aggregation (bagging) of tree-based models (Breiman 1996). This approach was used to analyze the structure of the relationships among tick density, tick prevalence of infection, and a series of critical ecological variables with the aim to improve the ability of the well-known classification trees methods in analyzing epidemiological data.

Materials and Methods

Study Area. The study was carried out in the Province of Trento, which is a mountainous region (6,200 km^2) located in the south central Alps (Fig. 1). Approximately 70% of its territory is above 1,000 m a.s.l., and 55% is covered by forest; the main forest types are beech forests (*Fagus sylvatica* L.), mountain and subalpine Norway spruce forests [*Picea abies* (L.) Kasten], European silver fir forests (*Abies alba* Miller), thermophile deciduous woods, and coppices with hophornbeam (*Ostrya carpinifolia*) and flowering ash (*Fraxinus ornus* L.).

The temperate oceanic climate is characterized by a strong altitudinal gradient with mean annual temperatures ranging from 12–13°C at the basal level to 4–6°C of upper-montane forests. Mean annual rainfall of the study area ranges from 850 to 1,300 mm. The main lithological substrata include limestone and dolomite, effusive and intrusive acid rocks, metamorphic rocks, and quaternary deposits. The fauna includes most of the species recognized as competent reservoirs for Lyme disease spirochetae (Gern et al. 1998). The roe deer (*Capreolus capreolus* L.) is the most abundant wild ungulate occurring in the study area and its populations have increased steadily over the last 35 yr to a present level of 29,800 individuals, with a mean density of 7 heads/100 ha of suitable habitat and peaks of 85 heads/100 ha.

Although the roe deer is considered reservoir-incompetent for *Borrelia burgdorferi* sensu lato (s.l.), it has an important effect on tick population density because it is one of the main hosts for the adult tick and indirectly effects the density of infected nymphs (Tälleklint and Jaenson 1996, Chemini et al. 1997, Hudson et al. 2001).

Assessment of Tick Occurrence. To determine the distribution and relative abundance of *I. ricinus*, questing ticks were collected in 438 sites of Trento province (Fig. 1) from 10 March to 15 October 1996. Sample sites, with one linear transect each, were chosen randomly within different vegetation types. The density of questing ticks was determined by dragging the vegetation with a 1-m^2 piece of white flannel. Drags were conducted between 1000 and 1800 hours, except on rainy days. On each transect, sampling was carried out for a total of 5 min that corresponded to a distance of ≈100 m of active drag. Ticks were collected from the drag at 1-min intervals, placed in vials, identified at species level, and stored.

Data on distribution and abundance of questing ticks collected in each site were entered and cross-correlated with the digital cartography of a customized GIS, based on the open source software GRASS (Geographic Resources Analysis Support System), including the following environmental variables: altitude, exposure and slope, substratum, vegetation type, and roe deer density (heads/100 ha) (Table 1).

These data were then analyzed with a tree-based classification model with bootstrap aggregation or *bagging* (Breiman 1996, Merler et al. 1996, Furlanello et al. 1997). These procedures allow the improvement of tree-based models: in this study, 100 tree-based classifiers were combined to obtain greater accuracy in the prediction. Single tree-based classifiers are appropriate for the analysis of large multivariate data sets described by a mix of numerical and categorical attributes. The *bagging* reduce the model error in classifiers with high variance, e.g., tree based classifiers, by combining multiple versions of the base model (Breiman 1999, Dietterich 2000).

In our case, the output of a single tree-based classifier T is either 0 (absence of ticks) or 1 (presence of ticks). Multiple versions $\{T_b\}$ b = 1, . . . , B of the model can be obtained by training the model on bootstrap replicates, i.e., random samples with reintroduction $\{L_b\}$ b = 1, . . . , B of the original data L. The bagging model is then defined as the average of the models $\{T_b\}$ b = 1, . . . , B

$$M(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x). \qquad [1]$$

A sample x will be assigned to the class "presence of ticks" whether $M(x) > 0.5$, otherwise it will be assigned to the class "absence of ticks." In general, $M(x)$ can be considered an estimate of the probability of tick occurrence.

Table 1. List of the GIS independent variables considered in this study

Variable	Classification
Lithological substratum	Limestone and dolomite, effusive acid rocks, intrusive acid rocks, metamorphic rocks, quaternary deposits
Vegetation type	Unproductive; Pasture-land; Alpine stone pine (*Pinus cembra* L.) forest, European larch (*Larix decidua* Miller) forest; mountain and subalpine spruce [*Picea abies* (L.) Karsten] forest; European silver fir (*Abies alba* Miller) forest; secondary European larch forest; secondary spruce wood; mesophile pinewood of Scots pine (*Pinus sylvestris* L.); beech (*Fagus sylvatica* L.) forest; thermophile pinewood of Austrian pine (*Pinus nigra* Arnold) and Scots pine; European chestnut (*Castanea sativa* Miller) and black locust (*Robinia pseudoacacia* L.) formation; thermophile deciduous woods and coppices with hophornbeam (*Ostrya carpinifolia* Scop.), flowering ash (*Fraxinus omus* L.), and pubescent oak (*Quercus pubescent* Willd.); Green alder [*Alnus glutinosa* (L.) Gaertner] and mountain pine (*Pinus mugo* Turra) formation; Evergreen oak (*Quercus ilex* L.) formations.
Exposure	E, NE, N, NW, W, SW, S, SE
Altitude	200–1,900 meters a.s.l. (957 ± 340) (mean ± SD)
Deer density	0–32 head/100 ha (6.8 ± 5.6)

The results obtained with the *bagging* procedure were used to predict the distribution of *I. ricinus* in Trentino with a production of a first digital map.

Model Validation. We considered as independent validation set the data relative to an additional dragging campaign carried out during 1997 to assess the rate of infection with *Borrelia burgdorferi* sensu lato (s.l.). A further set of two data base relative to actual tick counts carried out roe-deer harvested by hunters during 1994 and on forestry workers during 1996 were also used to evaluate the accuracy of the prediction obtained with this model. The roe-deer database included the data relative to the number of ticks recorded on 562 roe bucks harvested during the first 2 wk of September 1994 in 56 hunting districts in the Province of Trento (Chemini et al. 1997). These records were split into three data sets accordingly to three level of infestation (level 0: absence of ticks; level 1: 1–10 ticks; level 2: >10 ticks) and analyzed separately. Another database included 98 human cases of tick bites recorded on forestry workers during 1996 and provided by the local Health Office Authority.

Assessment of Density of Infected Nymphs. To determine the density and the distribution of nymphs infected with *B. burgdorferi* s.l., an additional sampling campaign was conducted, following the same methodology described above, from 20 March to 10 October 1997. Sixty sampling areas chosen within eight of the major valleys of the province of Trento were

Table 2. Estimated probability of tick occurrence (mean ± SD) and accuracy of the bagged model in predicting tick occurrence on actual tick counts belonging to six different data sets

Data set	Estimated probability of tick occurrence	Accuracy, %
1. 1997 dragging sites (tick absence)	0.3 ± 0.16	76
2. 1997 dragging sites (tick presence)	0.64 ± 0.31	78
3. Roe deer 0 (no ticks recorded)	0.48 ± 0.30	62
4. Roe deer 1 (number of ticks recorded: 1–10)	0.68 ± 0.31	66
5. Roe deer 2 (number of ticks recorded: >10)	0.79 ± 0.27	81
6. Human tick bites	0.70 ± 0.28	76

selected randomly in habitats where *I. ricinus* was predicted to occur by the model and dragged along a total of 245 transects. The ticks were collected at 1-min intervals from the flannel and placed in small humidified vials. After the identification of the tick species, DNA from each nymph specimen was amplified by polymerase chain reaction (PCR) using primers specific for *B. burgodorferi* rRNA genes. Amplification products were visualized by gel electrophoresis (Bandi et al. 1997). In this study we only analyzed nymphs that were considered at the developmental stage, with the highest epidemiological value (Lane et al. 1991, Gray 1998).

The proportion of infected nymphs collected during 1997 was weighted for the total nymphal density recorded in each site, since prevalence of infection was evaluated on a subsample of the ticks collected as previously described (Tälleklint and Jaenson 1996). The density of questing nymphs was considered as the mean number of nymphs collected during repeated sampling for each of the 60 sampling areas dragged during 1997. To assess the statistical significance of the analysis, only transects with at least eight nymphs examined for spirochetes were considered.

Analysis of variance (ANOVA) was used to test differences among the density of questing nymphs, the proportion of infected nymphs, and the density of infected nymphs. ANOVA was also used to test differences in density of infected nymphs in relation to the lithological substratum and the vegetation types.

The Spearman-rank correlation test was used to assess the relationships between the density of infected nymphs and the density of questing nymphs and for the analysis of the relationships between the density of infected nymphs and the variables altitude and roe deer density.

A stepwise model selection procedure was applied to select a parametric relationship between probability of tick occurrence, as predicted with the bagged model, and density of infected nymphs.

The results were entered in the GIS for production of a digital map of the distribution of infected nymphs in the Province of Trento.

Results

Assessment of Tick Occurrence. A total of 3,422 *I. ricinus* specimens were collected during 1996 along 438 transects for a total surface sampled of 43,800 m^2; 227 out of 438 sites were positive for tick occurrence with a mean tick density of 15.1/100 m^2 (SD = 30.8/100 m^2); 2,341 individuals were larvae (mean = 5.3; SD = 20.2), 996 were nymphs (mean = 2.3; SD = 5.1) and 85 adults (mean = 0.2; SD = 1.4). Density of questing nymphs varied from 0 to 37/100 m^2, with a mean value of 4.4/100 m^2 (SD = 5.1/100 m^2). All the specimens collected belonged to the species *I. ricinus.*

The bagging procedure was applied to develop the predictive model of probability of tick occurrence. The model was based on the bagging of 100 tree-based models. Each model predicted the presence or the absence of ticks as a function of the GIS environmental variable altitude, vegetation type, deer density, exposure, slope, and geological substratum. The structure of each tree-model, a recursive binary split of the prediction variables, easily allowed the transfer of the models within the GIS to produce a family of 100 tick presence maps. The probability of tick occurrence was estimated for each cell of the study area (cell size = 50 m by 50 m) by voting over the family of maps. In accordance with previous results (Merler et al. 1996), a higher frequency of tick occurrence was recorded at an altitude of <1,300 m, on limestone, and vegetation type with thermophile deciduous woods.

Model Validation. The performance of the models is listed in Table 2. For each data set, both the mean predicted probability of tick occurrence and the standard deviation of the estimate are reported. The accuracy in predicting tick presence as tested on the data

Table 3. Sampled areas, number of sites sampled in each area, number (mean ± SD) of questing *I. ricinus* nymphs collected per 100 m^2, prevalence of infection and number (mean ± SD) of infected nymphs per 100 m^2 at eight areas of the Province of Trento sampled during 1997

Area	No. of sites sampled	No. of questing nymphs/100 m^2	Prevalence of infection	No. of infected nymphs/100 m^2
Val d'Adige	9	10.0 ± 8.1	0.17 ± 0.11	2.0 ± 2.5
Valle di Fiemme	8	1.4 ± 2.6	0.08 ± 0.10	0.2 ± 0.2
Giudicarie	7	15.2 ± 19.55	0.14 ± 0.10	3.5 ± 5.5
Valle dei Laghi	6	37.3 ± 35.48	0.11 ± 0.06	5.4 ± 5.9
Valle di Non	8	14.5 ± 15.9	0.11 ± 0.14	1.6 ± 1.9
Pine'	6	5.8 ± 8.3	0.05 ± 0.08	1.0 ± 1.7
Primiero	8	8.6 ± 10.4	0.19 ± 0.08	1.3 ± 1.7
Valsugana	8	3.2 ± 5.5	0.08 ± 0.16	0.2 ± 0.5

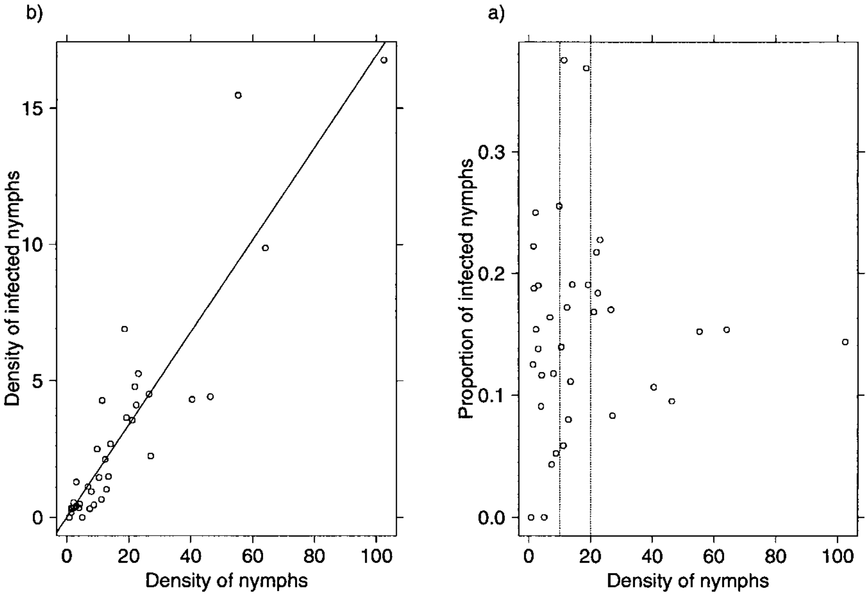

Fig. 2. Proportion of infected questing nymphs/100 m^2 versus number of questing nymphs/100 m^2 (panel a) and number of infected questing nymphs/100 m^2 versus number of questing nymphs/100 m^2 at 60 sites samples during 1997.

set relative to the 1997 dragging campaign was on average 77%. The best fit was obtained with the data set 5 relative to the infestation level two on roe deer, where the presence of ticks was correctly predicted by the model in 81% of the cases. In the other cases the accuracy ranged from 62% (data set 3: roe deer infestation level 0) to 76% (data set 1: tick absence; data set 6: human tick bites).

Assessment of Density of Infected Nymphs and Risk Map. During 1997 a total of 18,482 specimens of *I. ricinus* were collected: 10,196 were larvae, 7,939 nymphs, 180 adult males and 167 adult females; 2,208 nymphs were examined for *B. burgdorferi* (s.l.) infection. Mean prevalence of infection in nymphs was 17.5% (382/2,208) and infected nymphs were recorded in all the sampling areas, which did not significantly differ in the density of nymphs ($F = 2.128$; df = 7, 28; $P = 0.07$) or in the density of infected nymphs ($F = 1.117$; df = 7, 28; $P = 0.35$) (Table 3).

The proportion of infected nymphs increased with the nymphal density (with a peak of 37.5%) until the density of nymphs was <20/100 m^2 and decreased to a value of 14% at higher densities; a positive relationship between density of questing nymphs and density of infected nymphs was also observed (Spearman rank correlation, $r = 0.90$; $P < 0.001$) (Fig. 2).

The relationship between the density of infected nymphs and the probability of tick occurrence as obtained with the stepwise analysis (a cubic polynomial $y = 5.2 \times 10^{-6} x^3$, where x was the probability of tick occurrence and y was the density of infected nymphs, $R^2 = 0.69$), was therefore used to produce a model of the density of infected nymphs.

The altitude resulted negatively correlated with the density of infected nymphs that drastically decreased >1,300 m (Spearman rank correlation, $r = -0.65$; $P < 0.001$) (Fig. 3).

There was a positive correlation between roe deer density and the density of infected nymphs (Spearman rank correlation, $r = 0.25$; $P < 0.001$). The density of infected nymphs doubled over a roe deer threshold

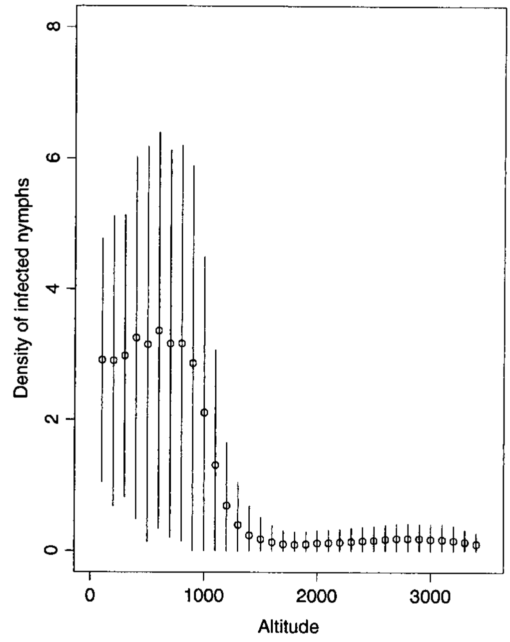

Fig. 3. Number of infected nymphs/100 m^2 (circles) as a function of the altitude. Vertical bars indicate confidence intervals at 95%.

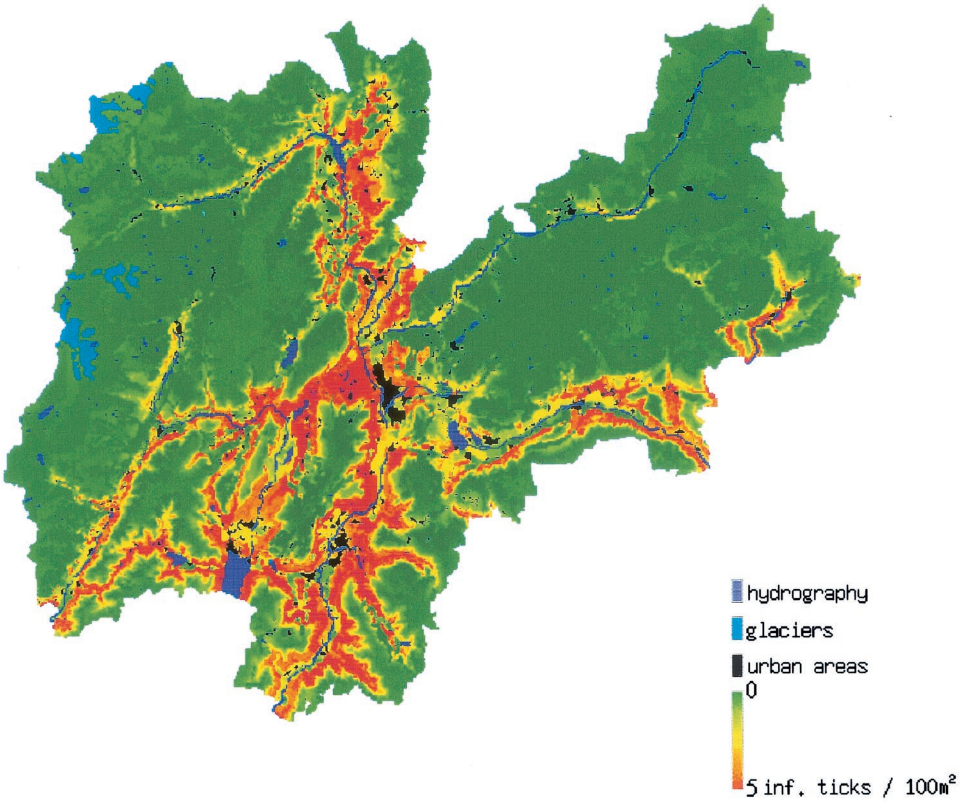

Fig. 4. Map of Lyme disease risk for the Province of Trento (resolution = 50 × 50 m). Colors from green to red indicate increasing density of infected nymphs.

density of 15 heads/100 ha and slightly decreased above densities >60 heads/100 ha.

ANOVA showed a clear relationship between vegetation cover and proportion of infected nymphs ($F = 32.682$; df = 4, 466,059; $P < 0.001$), indicating a higher prevalence of infection in thermophile deciduous forests with hophornbeam, flowering ash, and pubescent oak (*Quercus pubescens* Willd.) (mean density of infected nymphs = 3.8/100 m^2), thermophile pinewoods of Austrian pine (*Pinus nigra* Arnold) and Scots pine (*Pinus sylvestris* L.) (mean density of infected nymphs = 3.2/100 m^2), woods with European chestnut (*Castanea sativa* Miller) and black locust (*Robinia pseudoacacia* L.) (mean density of infected nymphs = 2.6/100 m^2), mesophile pinewoods of Scots pine (mean density of infected nymphs = 2.3/100 m^2) and beech (*Fagus sylvatica* L.) forests (mean density of infected nymphs = 2.7/100 m^2).

ANOVA also showed that the substratum type was related to the density of infected nymphs ($F = 28.783$; df = 4, 592,430; $P < 0.001$), indicating an higher proportion of infected nymphs on limestone (mean density of infected nymphs = 1.6/100 m^2).

The digital map of the density of infected nymphs in the Province of Trento is presented in Fig. 4.

Discussion

In this study, we produced a map of the probability of encountering a questing *I. ricinus* nymph infected with *B. burgdorferi* s.l. in Trentino, a territory with high tourist value and widely used for recreational purposes. This map was obtained by the integration of a bagging procedure with tree-based classifiers within a GIS. This approach was used to analyze the structure of the relationships between tick occurrence and prevalence of infection of ticks with a series of ecological variables (altitude, exposure and slope, substratum, vegetation type, and roe deer density) as classified with the GIS.

We first produced a predictive bagged model of the occurrence of *I. ricinus* in Trentino on the basis of the data collected during a dragging campaign carried out during 1996. The use of tree-based models gave a clear indication of the variables that appeared to have the greatest effect on tick occurrence. They were altitude, vegetation cover and roe deer density. Ticks were more abundant at altitudes <1,300 m, in areas covered with thermophile deciduous woods and coppices with hophornbeam, flowering ash, and pubescent oak characterized by high densities of roe deer.

The accuracy of the prediction obtained with this model was tested on independent databases, relative to actual tick counts carried out by collecting ticks from vegetation during a further dragging campaign, by counting ticks on roe deer shot by hunters and by recording tick bites on forestry workers. The accuracy of the model in predicting tick occurrence was in average 77% in the case of tick collected directly from the vegetation, and ranged from 62 to 81% in the other cases. These results highlighted the limitation of the GIS in describing environmental units especially when high-resolution scale is requested and when environmental variability is high.

A stepwise model selection procedure was applied for the analysis of the relationship between the predicted probability of tick occurrence and the density of infected nymphs as determined by counting and screening with PCR the ticks collected by dragging the vegetation during 1997. The results showed that Lyme disease appears widespread over the territory of the Province of Trento, with a prevalence of infection with the spirochetae *Borrelia burgdorferi* sensu lato (s.l.) of questing nymphs ranging from 5 to 19%, which is consistent with the data recorded in other European countries (Gray et al. 1998b). As observed for tick occurrence, the variables that appeared to exert the greatest influence on prevalence of infection were altitude, vegetation cover and roe deer density. These findings were not surprising because the conditions that favor the occurrence of *I. ricinus* and the spirocheta *Borrelia burgdorferi* s.l. are often overlapping. A significant relationship was recorded between the prevalence of infection and the density of questing nymphs. The general pattern was that prevalence of infection increased with the nymphal density below 20 nymphs/100 m^2, peaked at 20 nymphs/100 m^2 and decreased at higher nymphal densities. This behavior was related to host association in different habitat types and the relative ratio of abundance among different hosts (Tälleklint and Jaenson 1996, Gern et al. 1998). Roe deer density affected both tick abundance and the prevalence of infection. In general, we observed a concurrent increase both in tick density and prevalence of infection below a threshold density of 15 heads/100 ha. Above this value, despite the increased number of ticks, the prevalence of infection decreased. This behavior, as observed elsewhere (Tälleklint and Jaenson 1996), depends on the fact that although this species does not plays a direct role in *Borrelia* transmission and exerts a dilution effect for the borrelicidal activity of its serum (Kurtenbach et al. 1998) it indirectly affects tick prevalence of infection in relation to the number of ticks that are able to feed on competent reservoirs (Tälleklint and Jaenson 1996, Chemini et al. 1997, Hudson et al. 2001).

The relationship observed among vegetation type, tick abundance, and prevalence of infection can be explained by the analysis of the relationship between host and habitat preferences. In fact, vegetation cover, in addition to affecting habitat suitability for ticks, was used to a different extent by the hosts included in the *Borrelia* infection system. The high level of prevalence of infection observed in the thermophile deciduous forests with hophornbeam, flowering, and with pubescent oak may be related to the high density of roe deer usually observed in such habitat in the southern Alps. Furthermore, these associations are usually alternated with meadows, ecotones, and cultivated areas thus providing a suitable habitat for small mammals and other competent reservoirs (Gray 1998, Gray et al.1998b).

The relationships recorded among the density of infected nymphs and the variables considered in this study were used to construct a map of probability distribution of infected nymphs in the territory of the Province of Trento. In our study the application of the bagging procedure increased the accuracy of the prediction made by a single classification tree, reducing the overfitting of the data by averaging over different models.

Acknowledgments

The authors thank Claudio Chemini for continuous scientific support. We also thank Laura Agostini, Stefano De Felici, and Roberto Luise for helping in tick sampling; Damiano Gianelle, Heidi Hauffe, Isabella Cattadori, Peter J. Hudson, and Laura Kramer for the revision of the manuscript; the Forestry, Geology, Parks and Wildlife Services of the Province of Trento for providing GIS data and roe deer counts. This study was partially supported by a grant from Ministry of University and Scientific and Technological Research (MURST) p. 9907151238.

References Cited

Bandi, C., M. Sironi, V. Sambri, M. Fabbi, F. Solari-Basano, and C. Genchi. 1997. Molecular characterisation of Lyme disease borreliae using RAPD analysis and 16S rDNA sequencing. Ann. Ist. Super. Sanit. 33: 225–229.

Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J. C. Piffaretti, M. Assous, and P.A.D. Grimont. 1992. Delineation of *Borrelia burgdorferi sensu strictu, Borrelia garinii* sp. nov. and group VS461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. 42: 378–383.

Breiman, L. 1996. Bagging predictors. Mach. Learn. 26: 123–140.

Breiman, L. 1999. Combining predictors, pp. 31–50. *In* J. C. Sharkey (ed), Combining artificial neural nets. Springer, London.

Burioni, R., R. Grillo, M. Magaro, and G. Dettori. 1988. Lyme disease in Italy: isolation of *Borrelia burgdorferi* from a patient. Eur. J. Epidemiol. 4: 506–510.

Canica, M. M., F. Nato, L. Du Merle, J. C. Mazie, G. Baranton, and D. Postic. 1993. Monoclonal antibodies for identification of *Borrelia afzelii* sp.nov. associated with late cutaneous manifestation of Lyme borreliosis. Scand. J. Infect. Dis. 25: 441–448.

Chemini, C., G. Nicolini, A. Rizzoli, and C. Genchi. 1993. Alcuni dati preliminari sull'ecologia e ruolo patogeno della zecca *Ixodes ricinus* in Trentino. Riv. Med. Trentina 31: 51–56.

Chemini, C., A. Rizzoli, S. Merler, C. Furlanello, and C. Genchi. 1997. *Ixodes ricinus* (Acari: Ixodidae) infestation on roe deer (*Capreolus capreolus*) in Trentino, Italian Alps. Parassitologia 39: 59–63.

Ciceroni, L., and S. Ciarrocchi. 1998. Lyme disease in Italy, 1983–1996. New Microbiol. 21: 407–418.

Cinco, M., D. Padovan, R. Murgia, L. Poldini, L. Frusteri, I. van de Pol, N. Verbeek-De Kruif, S. Rijpkema, and M.

Maroli. 1998. Rate of infection of *Ixodes ricinus* ticks with *Borrelia burgdorferi* sensu stricto, *Borrelia garinii*, *Borrelia afzelii* and group VS116 in an endemic focus of Lyme disease in Italy. Eur. J. Clin. Microbiol. Infect. Dis. 17: 90–4.

Dietterich, T. G. 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. 40: 139–157.

Furlanello, C., S. Merler, and C. Chemini. 1997. Tree-based classifiers and GIS for biological risk forecasting, pp. 316–323. *In* F. C. Morabito (ed.), Advances in intelligent systems. IOS, Amsterdam.

Genchi, C., A. Rizzoli, M. Fabbi, M. Manfredi, V. Sambri, S. Magnino, A. Moroni, F. Massaria, and A. Marangoni. 1994. Ecology of *Borrelia burgdorferi* s.l. in some areas of northern Italy, pp. 232–235. *In* R. Cervini, V. Sambri, and M. La Placa (eds.), Proceedings, Advances in Lyme borreliosis research. VI International Conference Lyme Borreliosis, 19–22 June, Bologna, Italy. Soc. Ed. Esculapio, Bologna, Italy.

Gern, L., A. Estrada-Pena, F. Frandsen, J. S. Gray, T.G.T. Jaenson, F. Jongejan, O. Kahl, E. Korenberg, R. Mehl, and P. A. Nuttall. 1998. European reservoir hosts of *Borrelia burgdorferi* sensu lato. Zentralbl. Bakteriol. 287: 196–204.

Glass, G. E., B. S. Schwartz, J. M. Morgan, D. T., III Johnson, P. M. Noy, and E. Israel. 1995. Environmental risk factors for Lyme disease identified with geographic information systems. Am. J. Public Health 85: 944–948.

Gray, J. S. 1998. The ecology of ticks transmitting Lyme borreliosis. Exp. Appl. Acarol. 22: 249–258.

Gray, J. S., M. Granstrom, M. Cimmino, M. Daniel, G. Gettinby, O. Kahl, T.G.T. Jaenson, F. Jongejan, E. Korenberg, and S. O'Connell. 1998a. Lyme borreliosis awareness. Zentralbl. Bakteriol. 287: 253–265.

Gray, J. S., O. Kahl, J. N. Robertson, M. Daniel, A. Estrada-pena, G. Gettinby, T.G.T. Jaenson, P. Jensen, F. Jongejan, E. Korenberg, K. Kurtenbach, and P. Zeman. 1998b. Lyme borreliosis habitat assessment. Zentralbl. Bakteriol. 287: 211–228.

Hudson, P. J., A. Rizzoli, R. Rosà, C. Chemini, L. D. Jones, and E. A. Gould. 2001. Tick borne encephalitis virus in northern Italy: molecular analysis, relationships with density and seasonal dynamics of *Ixodes ricinus.* J. Med. Vet. Entomol. 15: 304–313.

Humair, P. F., O. Rais, and L. Gern. 1999. Transmission of *Borrelia afzelii* from *Apodemus* mice and *Clethrionomys* voles to *Ixodes ricinus* ticks: differential transmission pattern and overwintering maintenance. Parasitology 118: 33–42.

Johnson, R. C., G. P. Schmind, F. W. Hyde, A. G. Steigerwalt, and D. J. Brenner. 1984. *Borrelia burgdorferi* sp. nov.: etiologic agent of Lyme disease. Int. J. Syst. Bacteriol. 34: 496–497.

Kurtenbach, K., H. S. Sewell, N. H. Ogden, S. E. Randolph, and P. A. Nuttall. 1998. Serum complement sensitivity as a key factor in Lyme disease ecology. Infect. Immunol. 66: 1248–1251.

Lane, R. S., J. Piesman, and W. Burgdorfer. 1991. Lyme Borreliosis: relation of its causative agent to its vector and hosts in North America and Europe. Annu. Rev. Entomol. 36: 587–609.

Merler, S., C. Furlanello, C. Chemini, and G. Nicolini. 1996. Classification tree methods for analysis of mesoscale distribution of *Ixodex ricinus* (Acari: Ixodidae) in Trentino, Italian Alps. J. Med. Entomol. 33: 888–893.

Nicholson, M. C., and T. N. Mather. 1996. Methods for evaluating Lyme disease risks using geographic information systems and geospatial analysis. J. Med. Entomol. 33: 711–720.

Pavan, W. O., P. Ragni, N. Coppola, G. Gallo, E. Nava, J. Simeoni, G. Rovetta, G. Trevisan, and L. Vellucci. 2000. Borreliosi di Lyme: sorveglianza epidemiologica nazionale, pp. 54–56. *In* Proceedings, Malattie Infettive dell'Arco Alpino, V Conv. Int., Siusi allo Sciliar, 23–25 Marzo 2000. PAB, Bolzano, Italy.

Tälleklint, L., and T.G.T. Jaenson. 1996. Relationship between *Ixodes ricinus* density and prevalence of infection with Borrelia-like spirochetes and density of infected ticks. J. Med. Entomol. 33: 805–811.

Received for publication 21 September 2000; accepted 20 December 2001.