



Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Covariate selection in multivariate spatial analysis of ovine parasitic infection

V. Musella^a, D. Catelan^{b,c}, L. Rinaldi^d, C. Lagazio^e, G. Cringoli^{d,*}, A. Biggeri^{b,c}

^a Department of Experimental and Clinical Medicine, University Magna Graecia of Catanzaro, Italy

^b Department of Statistics "G. Parenti", University of Florence, Italy

^c Biostatistics Unit, ISPO Cancer Prevention and Research Institute, Florence, Italy

^d Department of Pathology and Animal Health, University of Naples Federico II, Via della Veterinaria, 1, 80137 Naples, Italy

^e Department of Statistical Sciences, University of Udine, Italy

ARTICLE INFO

Article history:

Received 12 October 2009

Received in revised form

15 November 2010

Accepted 15 November 2010

Keywords:

Ovine parasitic diseases

GIS

Land use

Geolithological features

Bayesian covariate selection

Multivariate geographical models

ABSTRACT

Gastrointestinal (GI) strongyle and fluke infections remain one of the main constraints on health and productivity in sheep dairy production. A cross-sectional survey was conducted in 2004–2005 on ovine farms in the Campania region of southern Italy in order to evaluate the prevalence of *Haemonchus contortus*, *Fasciola hepatica*, *Dicrocoelium dendriticum* and *Calicophoron daubneyi* from among other parasitic infections. In the present work, we focused on the role of the ecological characteristics of the pasture environment while accounting for the underlying long range geographical risk pattern. Bayesian multivariate spatial statistical analysis was used.

A systematic grid (10 km × 10 km) sampling approach was used. Laboratory procedures were based on the FLOTAC technique to detect and count eggs of helminths. A *Geographical Information System* (GIS) was constructed by using environmental data layers. Data on each of these layers were then extracted for pasturing areas that were previously digitalized aerial images of the ovine farms. Bayesian multivariate statistical analyses, including improper multivariate conditional autoregressive models, were used to select covariates on a multivariate spatially structured risk surface.

Out of the 121 tested farms, 109 were positive for *H. contortus*, 81 for *D. dendriticum*, 17 for *C. daubneyi* and 15 for *F. hepatica*. The statistical analysis highlighted a north–south long range spatially structured pattern. This geographical pattern is treated here as a confounder, because the main interest was in the causal role of ecological covariates at the level of each pasturing area.

A high percentage of pasture and impermeable soil were strong predictors of *F. hepatica* risk and a high percentage of wood was a strong predictor of *C. daubneyi*. A high percentage of wood, rocks and arable soil with sparse trees explained the spatial distribution of *D. dendriticum*. Sparse vegetation, river, mixed soil and permeable soil explained the spatial distribution of the *H. contortus*. Bayesian multivariate spatial analysis of parasitic infections with covariates from remote sensing at a very small geographical level allowed us to identify relevant risk predictors. All the covariates selected are consistent with the life cycles of the helminths investigated. This research showed the utility of appropriate GIS-driven surveillance systems. Moreover, spatial features can be used to tailor sampling design where the sampling fraction can be a function of remote sensing covariables.

© 2010 Published by Elsevier B.V.

* Corresponding author. Tel.: +39 081 2536283; fax: +39 081 2536282.

E-mail addresses: musella@unicz.it (V. Musella), catelan@ds.unifi.it (D. Catelan), lrinaldi@unina.it (L. Rinaldi), lagazio@dss.uniud.it (C. Lagazio), cringoli@unina.it (G. Cringoli), abiggeri@ds.unifi.it (A. Biggeri).

0167-5877/\$ – see front matter © 2010 Published by Elsevier B.V.

doi:10.1016/j.prevetmed.2010.11.012

1. Introduction

Sheep dairy production has a prominent role in the economy of several countries, in particular those in the Mediterranean Basin. However, it is influenced by several factors such as breeding methods, infectious diseases and parasitic diseases (Park and Haenlein, 2006).

Among the latter, gastrointestinal (GI) strongyle and fluke infections remain one of the main constraints to sheep production both in temperate and tropical countries (Rinaldi et al., 2007). The GI strongyle species of primary concern is *Haemonchus contortus*, a highly pathogenic blood-feeder helminth that causes anaemia and reduced productivity and can lead to death in heavily infected animals (Burke et al., 2007). *H. contortus* has a direct life cycle in the environment involving three larval stages (L₁–L₃). The most common flukes found in sheep in Italy are the liver flukes *Fasciola hepatica*, *Dicrocoelium dendriticum* and the rumen fluke *Calicophoron (Paramphistomum) daubneyi* (Cringoli et al., 2002, 2004a).

In the adult stage, *F. hepatica* is found in the bile ducts and in the immature stage in the liver parenchyma; whereas *D. dendriticum* lives in the bile ducts and gall bladder. The economic and health significance of liver fluke infection is due to the direct losses occasioned by the confiscation of altered livers and also to the indirect ones caused by the digestive disorders derived from the hepatobiliary alterations, such as decreased animal weight, growth delay, and reduced milk production (Cringoli et al., 2006).

Adult rumen flukes (*C. daubneyi*) inhabit the rumen and/or reticulum of the host. Immature flukes, which can cause serious morbidity and even death, are found in the upper small intestine. Both adult and immature rumen flukes can lower nutrient conversion; can cause weight loss and/or decrease milk production; all of which have an important economic impact (Rinaldi et al., 2005b).

Prevalence of *F. hepatica* and *C. daubneyi* would be related to the presence of water, which represents a favourable environment for their intermediate hosts, i.e., amphibious snails.

Favourable biotopes for *D. dendriticum* are connected with dry and calcareous or alkaline soils. In fact, it has a very complex life cycle, because it involves numerous species of land mollusks and ants as first and second intermediate hosts, respectively (for a review, see Manga-González et al., 2001).

Because the above described helminth infections have important significance from an animal health and economic point of view, it is necessary to have accurate and reliable data about their distribution in areas where there is a high concentration of pastured sheep.

The spatial distribution of helminth infections depends on abiotic and biotic environmental factors. Climate, vegetation and soil characteristics interact with the parasite, host and vector biology.

In the present paper, we report the results of a cross-sectional survey aimed at investigating the spatial distribution of *H. contortus*, *F. hepatica*, *D. dendriticum* and *C. daubneyi* in ovine farms from the Campania region in southern Italy.

In a previous article, we investigated the long range spatial trend of risk of the four parasitic diseases (Biggeri et al., 2007). In the present work, while accounting for the underlying long range geographical risk pattern, we focus on the role of the ecological characteristics of the pasture environment. We take advantage of a Geographical Information System (GIS) on the Campania region that includes several appropriate environmental data layers (e.g., land use and geolithological features). Bayesian multivariate spatial statistical analysis was used. We aim to contribute a new methodological approach that makes use of remote sensing covariates for small scale epidemiological surveillance.

2. Methods

2.1. Study area and farm sampling

The survey was conducted in the Campania region in southern Italy (Latitude = 39°59'15"–41°30'25"; Longitude = 13°45'25"–15°48'23") which extends over an area of 13,590 km². The region is mainly hilly and extends from 0 to 1890 m above sea level. The climate is Mediterranean with dry summers and rainy winters.

A complete list of ovine farms, georeferenced within a GIS, was available from the regional project MAP-ZOO (Cringoli, 2006). Since we were interested in the geographical disease pattern, we uniformly sampled the farms throughout the entire region with a systematic grid sampling approach (Cressie, 1991; Rinaldi et al., 2006). Specifically, within the GIS, a grid of quadrants of 10 km × 10 km was overlaid on the region map. While maintaining a good spatial resolution, the size of the grid cells was determined in order to limit the number of empty quadrants (i.e., quadrants with no farm). As a result, the territory of the Campania region was divided into 135 equal quadrants. The centroid of each quadrant was identified and, among all the farms present in the GIS database, the farm closest to the centroid in each quadrant was selected.

No information was collected outside the study area, because the coverage of the farms database was limited to the Campania region.

Only pastured farms with more than 50 animals were included in the sample, because smaller farms were not considered in this kind of parasitological surveillance (Cringoli et al., 2002). Out of the total of 135 quadrants, 121 (89.6%) have at least one farm. All GIS databases were developed using Arc-GIS 9.2 GIS software (ESRI, Redlands, CA, USA). The parasitological survey was done in 2004–2005.

2.2. Animal sampling and laboratory procedures

Rectal faecal samples were collected from animals on the selected ovine farms.

In each ovine farm, the animals were divided into two age groups: lambs (4–18 months) and adult sheep (older than 18 months). The choice of the cut-off age was based on zoo-technical conventional definitions (Cringoli et al., 2004b). Five individual faecal samples were collected from lambs, and 15 from adults. This was done to get the same level of precision of the prevalence estimates. The sample size was then equal for all the farms. The total number

of animals investigated from the 121 ovine farms was 2420.

Five private veterinarians working in the study area were trained for sampling and received a list of ovine farms from which to collect the study samples. Each veterinarian was provided with a copy of the study protocol, and uniform materials for faeces collection and transport.

Once at the laboratory, faecal samples were pooled into four groups of composite samples for each farm (3 composites from adults and 1 composite from lambs) (Nicholls and Obendorf, 1994). Each composite sample was formed from five equal parts by weight of individual faecal samples. The total number of ovine composite samples was 484.

Copromicroscopic examinations were performed using the FLOTAC dual technique (Cringoli et al., 2010), which is based upon the use of two flotation solutions that have complementary specific gravity, and are used in parallel on the same faecal composite. In particular, a sucrose based solution (specific gravity = 1.250) was used in order to detect nematode and eggs, and a zinc sulphate based solution (s.g. = 1.450) was used in order to detect trematoda eggs (Cringoli et al., 2004b).

The analytic sensitivity of the FLOTAC dual technique was 2 eggs per gram (EPG) of faeces. For further details, see Cringoli et al. (2010). In addition, cultures from the pooled composite were made and third stage larvae of *H. contortus* were identified using the morphological keys by van Wyk et al. (2004).

2.3. Geographical information system

A GIS of the Campania region, southern Italy, was constructed utilizing:

- I. a layer of administrative provincial and municipal boundaries at a scale of 1:25,000 provided by the Italian Military Geographical Institute (IGM) as the topographic base map; and digital aerial images collected in the year 2001 and obtained from the Cartographic Office of the Campania region, at a 1.0 m spatial resolution. The images were geo-referenced and then they were merged and the resulting mosaic corrected geometrically when necessary;
- II. the monthly and annual normalized difference vegetation index (NDVI), obtained from Landsat TM 5 images (spatial resolution = 30 m × 30 m);
- III. the elevation, slope and aspect of the study area obtained from the digital elevation model (DEM) (spatial resolution = 40 m) (source: Cartographic Office of the Region Campania);
- IV. the land use, obtained from the photo-interpretation of the digital aerial images described before (point I) and divided into 34 classes;
- V. the geolithological map of the study area (spatial resolution = 10 m) (source: Cartographic Office of the Region Campania), divided into 27 classes, and then grouped into three factors. These data were published in the year 2001.

For data analysis, on the basis of biological considerations briefly reported in the introductory section, we

Table 1

Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005. Land use variables.

Land use variable	Kind of variable
Drinking trough	Presence/absence
Arable	Percentage
Sparse vegetation	Presence/absence
Wood	Percentage
Sparse trees	Percentage
Pasture	Percentage
Woody farming	Percentage
Permanent meadows	Percentage
Rocks	Presence/absence
Riparian vegetation	Presence/absence
Thin wood	Presence/absence
River	Presence/absence
Wooded pasture	Percentage
Arable with sparse trees	Presence/absence
Lake	Presence/absence
Torrent canal	Presence/absence
Pond	Presence/absence
Permanent meadows with sparse trees	Percentage
NDVI vegetation index	Continuous

considered *a priori* 18 covariates that described land use. Two different measurement scales were used. Some of the covariates (e.g., sparse vegetation) were expressed in terms of presence/absence and corresponding binary variables with 1 indicating presence that were included in the statistical models being fitted. Others (e.g., arables) were expressed as the percentage of image pixels with the characteristic of interest. In this last case, variables were then standardized for the statistical analysis. The only continuous variable used was NDVI (Table 1).

As regards the geolithological variables, which are all expressed as percentages, we defined three new factors, labelled impermeable soil, mixed soil and permeable soil and computed a weighted average of the observed variables. First of all, each geological feature was assigned to a single factor (e.g., clayey to impermeable soil) and the classification is described in Table 2. The factor score was then computed defining an appropriate factor model embedded in the more general model used to analyse data (see below).

Data on all variables were extracted from the GIS and refer only to the pasturing area of the sheep farms. The pasturing area was identified with the help of the farmer by inspecting the digital aerial images.

2.4. Statistical analysis

Data analysis was based on the square root transformation of total eggs counted per farm, Y_{ik} , where $i = 1, \dots, I = 121$ indexes farms and $k = 1, \dots, 4$ indexes the parasites. Let's assume that for each i -th farm the response vector $\mathbf{Y}_i = \{Y_{i,k=1}, Y_{i,k=2}, Y_{i,k=3}, Y_{i,k=4}\}$ follows a multivariate normal distribution with vector mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$.

2.4.1. MCAR model for underlying spatial pattern

We first specified a random effects model for $\boldsymbol{\mu}_i$:

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha} + \boldsymbol{\psi}_i \quad (1)$$

Table 2

Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005. Geolithological variables with indication of the classification into factors.

Geolithological variables	Factor
Clay–limestone–arenaceous	Impermeable soil
Arenaceous–clay–limestone–conglomerate	Impermeable soil
Clay	Impermeable soil
Marl–clay–limestone	Impermeable soil
Alluvial	Impermeable soil
Pilitico–gravel–mining lake	Impermeable soil
Arenaceous–conglomerate	Mixed soil
Quarzarenitico–limestone–marble	Mixed soil
Limestone–marl–silica	Mixed soil
Limestone–marl–arenaceous	Mixed soil
Gravelly–sandy–silty alluvial	Mixed soil
Limosa–sandy–pebble	Mixed soil
Pyroclastic–detrital	Mixed soil
Tafacea	Mixed soil
Lime	Permeable soil
Dolomitic	Permeable soil
Conglomerate	Permeable soil
Sandy	Permeable soil
Sands of coastal current	Permeable soil
Travertine	Permeable soil
Detrital–eterometrica	Permeable soil
Pyroclastic–inconsistent	Permeable soil
Lava	Permeable soil

where $\alpha = \{\alpha_{k=1}, \alpha_{k=2}, \alpha_{k=3}, \alpha_{k=4}\}$ is a parasite specific intercepts vector whose elements are *a priori* independent improper uniformly distributed and ψ_i is a random terms vector modelled as multivariate conditionally autoregressive (MCAR; Jin et al., 2005).

This specification is a generalization of the univariate CAR model (Besag, 1974) and can be viewed as a flexible nonparametric way to take into account hidden long range spatially structured confounders (Clayton et al., 1993).

In detail, the multivariate specification assumes that

$$\psi_i | \psi_{j \in S_i} \text{ dist } \text{MVN}(\bar{\psi}_{j \in S_i}, \frac{\Sigma_i}{n_{j \in S_i}})$$

that is the conditional distribution of ψ_i given $\psi_{j \in S_i}$, where S_i represents the set of adjacent areas to the i -th one, is multivariate normal with a vector mean $\bar{\psi}_{j \in S_i} = \Sigma_{j \in S_i} \psi_j / n_{j \in S_i}$, and precision matrix $\Sigma_i / n_{j \in S_i}$, proportional to the number n of areas in S_i . In words, for each k -th parasite, we are assuming that its random terms conditional mean is the average of the random terms of the adjacent quadrants and that these conditional means for all four parasites are correlated.

2.4.2. The selection step: modelling covariate effects

Model Eq. (1) is completed by modelling the vector mean as a function of random effects and geolithological and land use variables. Let's assume we have information on a large number p of covariates and we want to flexibly select the subset of predictors that better describe the variability of faecal egg counts among parasites and farms. This is necessary, because the number of covariates is large. The model becomes:

$$\mu_i = \alpha + \psi_i + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$

where the vector $\beta_j = \{\beta_{j,k=1}, \beta_{j,k=2}, \beta_{j,k=3}, \beta_{j,k=4}\}$ represents the effect of the j -th covariate ($j = 1, \dots, p$) on each of the k -th parasite, and $X_{ij} = \{X_{ij,k=1}, X_{ij,k=2}, X_{ij,k=3}, X_{ij,k=4}\}$ the vector of j -th covariate values for i -th farm. Let γ_{jk} be an indicator variable which takes the value 1 if the j -th covariates for the k -th parasite is in the model and 0 otherwise. The vector of indicators $\gamma = 1$ identifies the subset of predictors in the model and X_γ the design matrix of the covariates for which $\gamma_{jk} = 1$.

The set of $\{\gamma_{jk}\}$ and $\{\beta_{jk}\}$ coefficients are unknown and so we followed the Bayesian approach of George and McCulloch (1993) to make inference about them. In particular, we assumed that β_{jk} are *a priori* independent and distributed as:

$$\beta_{jk} \sim (1 - \gamma_{jk})N(0, \sigma^2) + \gamma_{jk}N(0, c^2\sigma^2)$$

That, *a priori*, is a mixture of two normal distributions, when γ_{jk} is 1 or when γ_{jk} is 0. The variance parameter σ^2 is chosen to be sufficiently small (0.001) in such a way that if $\gamma_{jk} = 0$, then the corresponding regression coefficients are *a priori* constrained to be nearly equal to 0. The constant c^2 is fixed sufficiently large, so that if $\gamma_{jk} = 1$ the prior distribution is almost flat and the corresponding regression coefficient is unconstrained and the variable will be included in the final model if it contributes to better fit of the data. The model is completed by assuming a hyperprior distribution for γ_{jk} as Bernoulli, with parameter $\pi_{\gamma_{jk}}$ equal to 0.5.

2.4.3. The multivariate CAR model with latent factors

In order to improve the model flexibility and parsimony when considering the geolithological characteristics, we included a factor analysis step in the previous model. Geolithological features were modelled as a function of latent traits that were then treated as covariates in the selection step. To clarify the notation, we assumed that the p covariates are divided into two groups: the first one $\{X_{jk}\}$ with index j from 1 to J refers to land use covariates, the second $\{X_{j+1,k}\}$ with index from $J+1$ to p refers to geolithological features. The covariates for geolithological features are then summarized by three latent factors. We *a priori* fixed the number of factors and the geolithological variables that belong to each of them (as explained before, see Table 2 for the classification).

A “measurement model” for the observed geolithological variables was then defined following the Bayesian specification of Congdon (2003):

$$X_{ij} \sim N(\eta_{ij}, \tau_j^2)$$

where $j = J+1, \dots, p$ indexes geolithological covariates. We assume that the expected value η_{ij} is modelled as ($l = 1, 2, 3$ the latent factors):

$$\eta_{ij} = \alpha_j + q_{j,l=1}\lambda_{il=1} + \theta_{j,l=2}\lambda_{il=2} + \theta_{j,l=3}\lambda_{il=3}$$

where θ_{ji} indicates the loading coefficient of the factor analysis, λ_{il} is the common factor to be used as covariates in the full response model and α_j an intercept. The prior distribution of the latent factor λ_{il} is normal $N(0,1)$ and we put also flat normal $N(\cdot, \cdot)$ prior distributions on α_j and θ_{ji} .

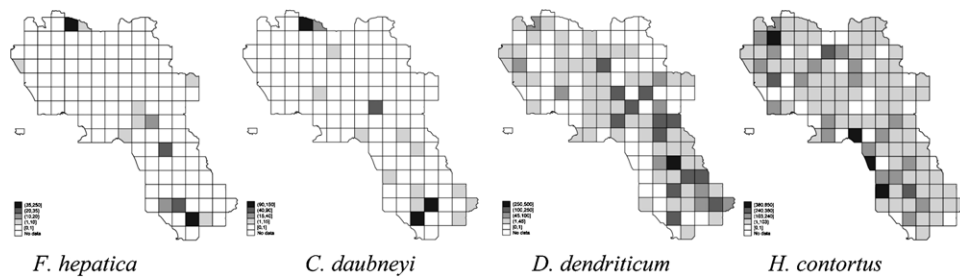


Fig. 1. Observed spatial distributions of four helminths infection (EPG quartiles see text). Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005.

The “full response model” then becomes:

$$Y_i \sim \text{MVN}(\mu_i, \Sigma_i)$$

$$\mu_i = \alpha + \psi_i + [\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_J X_{iJ}] + [\beta_{J+1} \lambda_{i1} + \beta_{J+2} \lambda_{i2} + \beta_{J+3} \lambda_{i3}]$$

where X_{ij} $j=1, \dots, J$ are the vectors of observed land use covariates λ_{il} are the latent factors that summarize the geolithological covariates. As in Section 2.4.2 β_{jk} are *a priori* modelled as:

$$b_{jk} \sim (1 - \gamma_{jk})N(0, \sigma^2) + \gamma_{jk}N(0, c^2\sigma^2)$$

All the models were estimated by Markov Chain Monte Carlo methods using the WinBugs software (Lunn et al., 2000). For each model, we have run two independent chains and the convergence of the algorithm was then evaluated according to Gelman and Rubin (1992). We discarded the first 100,000 iterations (burn-in) and stored 50,000 iterations for estimation of parameters.

Table 3

Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005. Posterior probability of inclusion for each land use and geolithological soil features obtained as latent factors.

	<i>F. hepatica</i>	<i>C. daubneyi</i>	<i>D. dendriticum</i>	<i>H. contortus</i>
<i>Land use variables</i>				
Drinking trough	0.24	0.30	0.42	0.43
Arable	0.02	0.05	0.14	0.15
Sparse vegetation	0.07	0.15	0.28	0.90
Wood	0.19	0.70	0.77	0.28
Sparse trees	0.07	0.05	0.10	0.12
Pasture	0.99	0.04	0.11	0.40
Woody farming	0.08	0.13	0.10	0.23
Permanent meadows	0.04	0.08	0.21	0.41
Rocks	0.10	0.11	0.70	0.34
Riparian vegetation	0.12	0.19	0.22	0.48
Thin wood	0.16	0.15	0.26	0.34
River	0.14	0.19	0.48	0.68
Wooded pasture	0.03	0.14	0.18	0.28
Arable with sparse trees	0.05	0.14	0.79	0.23
Lake	0.55	0.24	0.51	0.54
Torrent canal	0.23	0.21	0.49	0.44
Pond	0.23	0.15	0.48	0.34
Permanent meadows with sparse trees	0.07	0.05	0.07	0.15
NDVI vegetation index	0.01	0.01	0.03	0.01
<i>Geolithological features</i>				
Impermeable soil	0.99	0.53	0.15	0.14
Mixed soil	0.07	0.14	0.26	0.85
Permeable soil	0.03	0.22	0.15	0.99

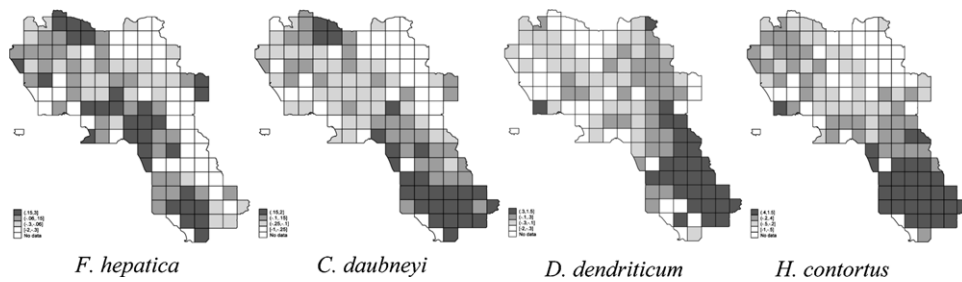


Fig. 2. Spatial distribution of the spatial random terms ψ_{ik} (tertiles, see text) for four helminths infection. Multivariate CAR model. Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005.

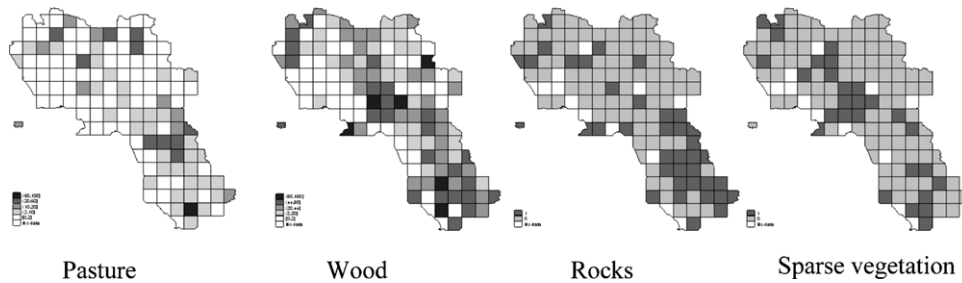


Fig. 3. Spatial distribution of the land use covariates affecting the prevalence of four helminths infection. Multivariate CAR model. Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005.

covariates showed a residual correlation among responses of 0.88 (95% CrI=0.52–0.98) between *F. hepatica* and *C. daubneyi* and almost zero for the other combinations of parasites. The MCAR model, which included the factor analysis step to better summarize the role of geolithological soil features, completely explained the correlation among parasites EPG; the residual correlation among responses was around zero. The importance of each covariate and latent factor in predicting parasite-specific EPG was evaluated inspecting the posterior probability of inclusion $\pi_{\gamma_{ijk}|Y}$ (pp) (Table 3).

Pasture and impermeable soil have a posterior probability higher than 0.99 to be included in the model for *F. hepatica*; in fact the ratio of average EPG (RR) for farms with large pasture areas vs. farms with small pasture areas (categorized as lesser or greater of 20% of the pasturing area of the flock) is 11.3 (95% CI=9.5–13.4) and extremely high for those with impermeable soil vs. those without (RR=156; 95% CI=114–215). Here impermeable soil is the latent factor which was dichotomized by setting a cut-

off at one theoretical standard deviation from the mean, and 7% of the farm was classified at a high risk. Wood (Yes/No RR=4.9; 95% CI=4.1–6.0) is selected for *C. daubneyi* (pp of 0.70). Wood (Yes/No RR=2.8; 95% CI=2.7–3.0), rocks (Yes/No RR=2.7; 95% CI=2.5–2.9) and arable with sparse trees (Yes/No RR=2.0; 95% CI=1.9–2.1) for *D. dendriticum* (posterior probabilities were 77%, 70% and 79% respectively) and sparse vegetation (Yes/No RR=1.4; 95% CI=1.47–1.50), river (RR=2.6; 95% CI=2.5–2.7), mixed soil (having dichotomized the latent factor as less than or greater than zero, 33% of farms were classified in this category: RR=4.8; 95% CI=4.6–5.0) and permeable soil (we dichotomize the latent factor lesser or greater than zero; 47% of farm was classified in the category less than zero: RR 5.9; 5.6–6.3) are chosen for *H. contortus* (posterior probabilities: 90%, 68%, 85%, 99%, respectively).

Fig. 2 shows the posterior means of the spatial random terms ψ_{ik} for the four helminths. This represents the residual long range spatial pattern. The maps highlight a spatially structured variability with a long range

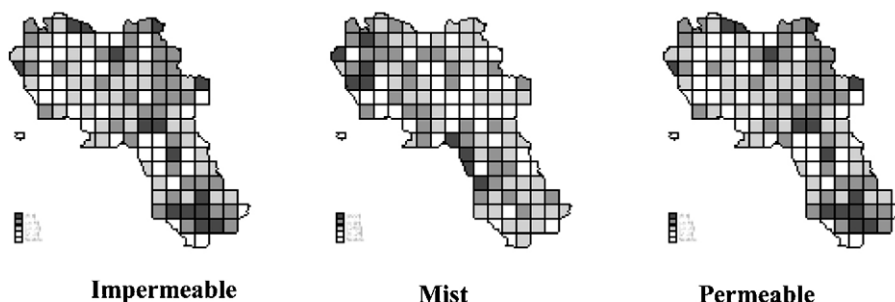


Fig. 4. Spatial distribution of the latent factors for geolithological soil features. Multivariate CAR model. Cross sectional survey on ovine parasitic diseases. Campania region 2004–2005.

north–south gradient common to all four of the parasites considered (see Biggeri et al., 2007 for further details). Fig. 3 shows the spatial distribution of the selected land use covariates. The spatial distributions of the latent variables are reported in Fig. 4.

4. Conclusion and discussion

We studied the multivariate spatial distribution of the GI strongyle *H. contortus*, the liver flukes *F. hepatica* and *D. dendriticum* and the rumen fluke *C. daubneyi* in pastured sheep farms with more than 50 animals from the Campania region in southern Italy.

In ecological studies which utilize GIS and remote sensing (RS) tools, environmental data are usually extracted from buffer zones centered on the farm (Cringoli et al., 2004a) or centroids of the main pasture (Rinaldi et al., 2005a). However, the smaller the area in which parasitological and environmental data were collected, the greater the possibility for making accurate inferences, because averages over large areas may introduce strong ecological bias in correlation studies with survey-based data. For this reason, in the present study, ecological analysis was made by including only the values of environmental data that fell into the pasturing areas of sheep digitalized on aerial images within the GIS.

In a previous study, GIS and RS were used to identify environmental features that influenced the distribution of *C. daubneyi* in sheep from the southern Italian Apennines (Cringoli et al., 2004a). Specifically, a multivariate stepwise discriminant analysis model was developed that included moors and heathland, sclerophyllous and coniferous forest vegetation, autumn–winter NDVI and presence of streams, springs and brooks on pastures. However, it should be noted that the data-layers utilized in that study were characterized by a lower resolution compared to those used in the present study. For example, the DEM in Cringoli et al. (2004a) had a spatial resolution of 100 m compared to the 40 m utilized in the present study. In addition, land cover data were obtained from the Corine land cover map (1:100,000), which described land cover (and partly land use) according to a nomenclature of 44 classes; on the contrary, in the present study we used more accurate landscape data based on the photo-interpretation of aerial digital images.

Another difference between the two studies is that in Cringoli et al. (2004a) the environmental variables were extracted from 3 km buffer zones centered on the sheep farms. In the present study, for each of the 121 ovine farms, the environmental variables were extracted, with the help of the farmer, from the sheep pasturing area digitalized on the aerial images.

With regard to statistical issues, we used a Gaussian likelihood on square root transformation of egg counts and an improper MCAR prior for the area specific spatial random terms (Jin et al., 2005). The Bayesian MCAR model highlighted a long range spatial trend, covariates, and latent factors that summarized geolithological soil features and explained the residual conditional correlation between the helminths. If they are spatially structured and provided the model is correctly specified,

spatial random terms may also account for farm-level factors.

Data were very ill-defined with a high percentage of zero counts for *F. hepatica* and *C. daubneyi* and posterior inference was expected to be sensitive to prior specifications. A zero-inflated spatial Poisson regression model on egg count is an alternative (Lambert, 1999; Biggeri et al., 2006), although it is at the cost of a very heavy computational burden. A sensitivity analysis showed that the square root transformation was robust and gave results close to those obtained by zero inflated models (Biggeri et al., 2006).

We did not apply any correction for the edge effect, i.e., the bias due to missing information at the border of the study area. Among the possible solutions, the use of an external buffer area was not feasible, because of the coverage of the list of farms, while the use of an internal buffer area would have strongly limited our inferences, which in this case would have been restricted to only the inner region. An edge effect cannot be excluded, even if we do not expect it to be strong on covariates' effects, which was the main focus of the present study.

From a parasitological point of view it is interesting to note that the distribution of the four studied helminths in sheep farms was strongly spatially structured. The explanation of this long range geographical pattern, modelled by the multivariate CAR random terms, was not the main purpose of the present analysis, which is indeed the role of ecological covariates at the level of each pasturing area. However, in a previous paper we investigated the presence of a common pattern among the four parasites. This spatial structure was driven by the two more severe parasitic infections and may be attributable to the coverage of veterinary preventive or prophylactic treatments (Biggeri et al., 2007). Our model, which includes spatially structured random effects (ψ_{ik}), is therefore adjusting for hidden confounders such as management practice and demographic factors. Effect estimates of environmental covariates should then not be biased by the exclusion of that information.

Pasture and impermeable soil (at a lesser degree lake, $pp=0.55$) were strong predictors of *F. hepatica* risk and wood (at a lesser degree impermeable soil $pp=0.53$) of *C. daubneyi*. These findings are highly consistent with the life cycle of these two trematoda which involves amphibious snails (e.g., *Lymnaea truncatula*) as intermediate hosts, and thus has strong environmental determinants and strong needs of water. In particular, the land use (i.e., pasture and wood) and geolithological (impermeable soil) types entered in the model are indicators of zones where typically there is a presence of water (permanently or temporarily). These results are also in agreement with those previously obtained in ruminants by Cringoli et al. (2004a) for *C. daubneyi* and by Malone et al. (1998) for *F. hepatica*.

The risk of infection by these two flukes is thus influenced by environmental determinants as well as the number and distribution of animals, the presence of infected snails, and grazing management, which allows animals to access herbage or water containing metacercariae (Tum et al., 2004, 2007).

Variables wood, rocks and arable with sparse trees explained the spatial distribution of the lancet fluke *D.*

dendriticum. These findings are again consistent with the complex life cycle of this fluke. Indeed, it is usually found where dry and calcareous or alkaline soils are present, because they represent a favourable biotope for its intermediate hosts, i.e., several species of land mollusks and various ants (Manga-González et al., 2001; Otranto and Traversa, 2003; Díaz et al., 2007).

Finally, sparse vegetation and river, mixed and permeable soil explained the spatial distribution of the GI strongyle *H. contortus*. The life cycle of this parasite involves a series of larval stages (L1–L3) in the environment. The susceptibility of *H. contortus* pre-infective stages to desiccation (Rose, 1963; Waller and Donald, 1970; Rossanigo and Gruner, 1995) is highly characteristic of this GI strongyle species, limiting its distribution to areas with warm, moist summers and creating a natural barrier to development that results in sporadic development of the free-living stages (O'Connor et al., 2007).

A priori, our results using remote sensing covariates were not obvious. The retrieval of known associations, obtained by different study designs and covariates, confirms the validity of our approach. We expect that our research stimulates the set-up of appropriate GIS-driven surveillance systems. Moreover, spatial features can be used to tailor sampling design where the sampling fraction can be a function of remote sensing covariables (see for example Theobald et al., 2007).

The use of multivariate analysis of spatial distribution of veterinary parasitic infections together with a covariate from RS on a very small geographical level proved to be a useful tool to investigate the biology of parasites and for epidemiological surveillance. Land use variables were useful to evaluate characteristics of intermediate hosts in parasitic diseases and geolithological features provided useful information to investigate environmental attributes for monitoring purposes.

Acknowledgements

The work of Vincenzo Musella was partially supported by Master of Epidemiology, University of Turin (Italy).

References

- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B* 36, 192–236.
- Biggeri, A., Catelan, D., Dreassi, E., Lagazio, C., Rinaldi, L., Cringoli, G., 2006. Multivariate spatially structured variability of ovine parasitic infections. In: Proceedings of the International Workshop On Spatio-Temporal Modelling (Metma3), Pamplona, Spain, 27th, 28th, and 29th September 2006.
- Biggeri, A., Catelan, D., Dreassi, E., Rinaldi, L., Musella, V., Veneziano, V., Cringoli, G., 2007. Multivariate spatially structured variability of ovine parasitic infections. *Geospat. Health* 2 (1), 97–104.
- Burke, J.M., Kaplan, R.M., Miller, J.E., Terrill, T.H., Getz, W.R., Mobini, S., Valencia, E., Williams, M.J., Williamson, L.H., Vatta, A.F., 2007. Accuracy of the FAMACHA system for on-farm use by sheep and goat producers in the southeastern United States. *Vet. Parasitol.* 147, 89–95.
- Clayton, D.G., Bernardinelli, L., Montomoli, C., 1993. Spatial correlation in ecological analysis. *Int. J. Epidemiol.* 22, 1193–1202.
- Congdon, P., 2003. Applied Bayesian Modelling. Wiley, Chichester, UK.
- Cressie, N., 1991. Statistics for Spatial Data. Wiley, New York.
- Cringoli, G., Rinaldi, L., Veneziano, V., Capelli, G., Malone, J.B., 2002. A cross-sectional coprological survey of liver flukes in cattle and sheep from an area of the southern Italian Apennines. *Vet. Parasitol.* 108, 137–143.

- Cringoli, G., Taddei, R., Rinaldi, L., Veneziano, V., Musella, V., Cascone, C., Sibilio, G., Malone, J.B., 2004a. Use of remote sensing and geographical information systems to identify environmental features that influence the distribution of paramphistomosis in sheep from the southern Italian Apennines. *Vet. Parasitol.* 122, 15–26.
- Cringoli, G., Rinaldi, L., Veneziano, V., Capelli, G., Scala, A., 2004b. The influence of flotation solution, sample dilution and the choice of McMaster slide area (volume) on the reliability of the McMaster technique in estimating the faecal egg counts of gastrointestinal strongyles and *Dicrocoelium dendriticum* in sheep. *Vet. Parasitol.* 123, 121–131.
- Cringoli, G., Rinaldi, L., Veneziano, V., Genchi, C., 2006. A dual anthelmintic treatment strategic scheme for the control of fasciolosis in dairy sheep farms. *Parasitol. Res.* 99, 700–705.
- Cringoli, G., 2006. Mappa Parassitologica 7. MAPZoo. In: Cringoli, G., Cavallo, S., Rinaldi, L. (Eds.), *Un Sistema Informativo zootecnico-sanitario in Campania. Primo anno di Attività*, ISBN 88-89132-10-8, p. 236.
- Cringoli, G., Rinaldi, L., Maurelli, M.P., Utzinger, J., 2010. FLOTAC: new multivalent techniques for qualitative and quantitative coproscopic diagnosis of parasites in animals and humans. *Nat. Protoc.* 5, 503–515.
- Díaz, P., Paz-Silva, A., Sánchez-Andrade, R., Suárez, J.L., Pedreira, J., Arias, M., Díez-Baños, P., Morrodo, P., 2007. Assessment of climatic and orographic conditions on the infection by *Calicophoron daubneyi* and *Dicrocoelium dendriticum* in grazing beef cattle (NW Spain). *Vet. Parasitol.* 149, 285–289.
- Gelman, A., Rubin, D.R., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* 7, 457–511.
- George, E.I., McCulloch, R., 1993. Variable selection via gibbs sampling. *J. Am. Stat. Assoc.* 88, 881–889.
- Jin, X., Carlin, B., Banerjee, S., 2005. Generalized hierarchical multivariate car models for areal data. *Biometrics* 61, 950–961.
- Lambert, D., 1999. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1), 1–14.
- Manga-González, M.Y., González-Lanza, C., Cabanas, E., Campo, R., 2001. Contributions to and review of dicrocoeliosis, with special reference to the intermediate hosts of *Dicrocoelium dendriticum*. *Parasitology* 123, S91–S114.
- Malone, J.B., Gommers, R., Hansen, J., Yilma, J.M., Slingerberg, J., Snijders, F., Nachtergaele, F., Ataman, E., 1998. A geographic information system on the potential distribution and abundance of *Fasciola hepatica* and *F. gigantica* in East Africa based on Food and Agriculture Organization databases. *Vet. Parasitol.* 78, 87–101.
- Nicholls, J., Obendorf, D.L., 1994. Application of a composite faecal egg count procedure in diagnostic parasitology. *Vet. Parasitol.* 52, 337–342.
- O'Connor, L.J., Kahn, L.P., Walkden-Brown, S.W., 2007. The effects of amount, timing and distribution of simulated rainfall on the development of *Haemonchus contortus* to the infective larval stage. *Vet. Parasitol.* 146, 90–101.
- Otranto, D., Traversa, D., 2003. Dicrocoeliosis of ruminants: a little known fluke disease. *Trends Parasitol.* 19, 12–15.
- Park, Y.W., Haenlein, G.F.W. (Eds.), 2006. Handbook of Milk of Non-Bovine Mammals. Blackwell Publishing, Ames, Iowa, USA/Oxford, UK, p. 449.
- Rinaldi, L., Fusco, G., Musella, V., Veneziano, V., Guarino, A., Taddei, R., Cringoli, G., 2005a. *Neospora caninum* in pastured cattle: determination of climatic, environmental, farm management and individual animal risk factors using remote sensing and geographical information systems. *Vet. Parasitol.* 128, 219–230.
- Rinaldi, L., Perugini, A.G., Capuano, F., Fenizia, D., Musella, V., Veneziano, V., Cringoli, G., 2005b. Characterization of the second internal transcribed spacer of ribosomal DNA of *Calicophoron daubneyi* from various hosts and locations in southern Italy. *Vet. Parasitol.* 131, 247–253.
- Rinaldi, L., Musella, V., Biggeri, A., Cringoli, G., 2006. New insights into the application of geographical information systems and remote sensing in veterinary parasitology. *Geospat. Health* 1, 33–47.
- Rinaldi, L., Veneziano, V., Cringoli, G., 2007. Dairy goat production and the importance of gastrointestinal strongyle parasitism. *Trans. R. Soc. Trop. Med. Hyg.* 101, 745–746.
- Rose, J.H., 1963. Observations on the free-living stages of the stomach worm *Haemonchus contortus*. *Parasitology* 53, 469–481.
- Rossanigo, C.E., Gruner, L., 1995. Moisture and temperature requirements in faeces for the development of free-living stages of gastrointestinal nematodes of sheep, cattle and deer. *J. Helminthol.* 69, 357–362.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* 10, 325–337.
- Theobald, D.M., Stevens Jr., D.L., White, D., Urquhart, N.S., Olsen, A.R., Norman, J.B., 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environ. Manage.* 40, 134–146.

- Tum, S., Puotinen, M.L., Copeman, D.B., 2004. A geographic information systems model for mapping risk of fasciolosis in cattle and buffaloes in Cambodia. *Vet. Parasitol.* 122, 141–149.
- Tum, S., Puotinen, M.L., Skerratt, L.F., Chan, B., Sothoeun, S., 2007. Validation of a geographic information system model for mapping the risk of fasciolosis in cattle and buffaloes in Cambodia. *Vet. Parasitol.* 143, 364–367.
- van Wyk, J.A., Cabaret, J., Michael, L.M., 2004. Morphological identification of nematode larvae of small ruminants and cattle simplified. *Vet. Parasitol.* 119, 277–306.
- Waller, P.J., Donald, A.D., 1970. The response to desiccation of eggs of *Trichostrongylus colubriformis* and *Haemonchus contortus*. *Parasitology* 61, 195–204.