Veterinary Parasitology 205 (2014) 158–168

Contents lists available at ScienceDirect

Veterinary Parasitology

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

On the use of posterior predictive probabilities and prediction uncertainty to tailor informative sampling for parasitological surveillance in livestock

Vincenzo Musella[a], Laura Rinaldi[b], Corrado Lagazio[c], Giuseppe Cringoli[b], Annibale Biggeri[d,e], Dolores Catelan[d,e,*]

[a] *Department of Health Sciences, University of Catanzaro Magna Graecia, Catanzaro, Italy*
[b] *Department of Veterinary Medicine and Animal Productions, University of Naples "Federico II", Naples, Italy*
[c] *Department of Economics, University of Genoa, Genoa, Italy*
[d] *Department of Statistics, Informatics and Applications "G. Parenti", University of Florence, Florence, Italy*
[e] *Biostatistics Unit, ISPO, Florence, Italy*

ARTICLE INFO

Article history:
Received 31 August 2013
Received in revised form 6 May 2014
Accepted 4 July 2014

Keywords:
Parasitological surveillance
Bayesian kriging model
Posterior predictive probability
Sampling strategies
Livestock

ABSTRACT

Model-based geostatistics and Bayesian approaches are appropriate in the context of Veterinary Epidemiology when point data have been collected by valid study designs. The aim is to predict a continuous infection risk surface. Little work has been done on the use of predictive infection probabilities at farm unit level. In this paper we show how to use predictive infection probability and related uncertainty from a Bayesian kriging model to draw a informative samples from the 8794 geo-referenced sheep farms of the Campania region (southern Italy). Parasitological data come from a first cross-sectional survey carried out to study the spatial distribution of selected helminths in sheep farms. A grid sampling was performed to select the farms for coprological examinations. Faecal samples were collected for 121 sheep farms and the presence of 21 different helminths were investigated using the FLOTAC technique. The 21 responses are very different in terms of geographical distribution and prevalence of infection. The observed prevalence range is from 0.83% to 96.69%. The distributions of the posterior predictive probabilities for all the 21 parasites are very heterogeneous. We show how the results of the Bayesian kriging model can be used to plan a second wave survey. Several alternatives can be chosen depending on the purposes of the second survey: weight by posterior predictive probabilities, their uncertainty or combining both information. The proposed Bayesian kriging model is simple, and the proposed samping strategy represents a useful tool to address targeted infection control treatments and surbveillance campaigns. It is easily extendable to other fields of research.

---

[*] Corresponding author at: Department of Statistics, Informatics and Applications "G. Parenti", University of Florence, Viale Morgagni, 59, 50134 Florence, Italy. Tel.: +39 320 4314303; fax: +39 055 4223560.

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

http://dx.doi.org/10.1016/j.vetpar.2014.07.004
0304-4017/© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The use of geostatistical models are not new in parasitology. They are appropriate every time point data are collected by valid study design (Diggle, 1983; Cressie, 1991). The main goal is usually to predict a continuous risk surface on the whole study area, to explain the spatial pattern of the phenomenon under study, to investigate spatial autocorrelation. Geostatistical models have also been widely used in human epidemiology to obtain smoothed maps of disease prevalence (Schur et al., 2011a, 2011b).

Although most of the literature which applied geostatistical approaches focused on human diseases, these approaches appear to be natural in the context of veterinary parasitology, particularly for livestock. The unit of interest is usually the farm, the animals belonging to the same farm have usually an enclosed pasture area, share the same risk factors and parasites abundance is strongly determined by ecological characteristics of the pasture area (Musella et al., 2011). Therefore mapping farms location, studying the spatial distribution of infections and estimating a continuous risk surface for specific parasites enables focused intervention strategies and addresses important scientific clues.

Bayesian geostatistics are widely used for their flexibility to model uncertainty (Banerjee et al., 2004) and to introduce covariates which can explain the spatial pattern of risk (e.g. Basáñez et al., 2004; Noor et al., 2008; Gosoniu et al., 2010; Magalhães et al., 2011; Forrer et al., 2012). In Gosoniu et al. (2010) a Bayesian geostatistical model has been used both to map Malaria prevalence in Angola and to address the role of different covariates in explaining the spatial heterogeneity of disease risk. Bayesian geostatistical models have been also proposed in the analysis of areal data, in the context of human disease mapping with data at administrative level (e.g. county) (Kelsall and Wakefield, 2002). The aggregate data are collapsed at the geographical coordinate of the centroid of each unit area and are then treated as point data.

Hierarchical spatial Bayesian models were used to investigate the long range spatial trend of four selected parasite infections (Biggeri et al., 2007) and the role of ecological characteristics of the pasture environment in determining the spatial distribution of the helminthes (Musella et al., 2011). In the present paper, we explore a different task: to predict the probability of infection at the location of each single unit of the entire finite farm population and studying its utility for subsequent surveillance and survey planning.

We take advantage of a veterinary parasitological survey conducted in a region of southern Italy. The MAPZOO project (Cringoli, 2006) started in 2005 to study the geographical distribution of a number of helminth infections in the Campania region (southern Italy). As part of the MAPZOO project a Geographical Information System (GIS) was constructed using the administrative boundaries (at provincial and municipal levels) of the Campania region and all the sheep farms of the region were geo-referenced (Cringoli et al., 2007). Here, we are not interested in the description and characterization of the spatial distribution of parasites infection (as in Biggeri et al., 2007; Musella et al., 2011) but in the prediction of the posterior probability of infection for the entire finite population of 8794 farms of the Campania Region, southern Italy. The novelty is not in the geostatistical models used but it resides in showing how informative sampling schemes for the study area can be designed using the outputs (probability maps and uncertainty maps) of the geostatistical models fitted to the first wave sample survey data.

In general, data of a first survey can be based, for example, on a systematic grid sampling scheme (Gyapong and Remme, 2001), on administrative areas as health system units (Gyapong et al., 2002) or based on known ecological and topographical features of the study region (Onapa et al., 2005).

Now, when we are interested in performing a second subsequent survey, an informative sampling scheme – based on the results of the first survey – will permit to better address the researchers' goals: for example, improve the number of positive farms visited in surveillance campaigns by sample proportional to the probability of infection; improve precision in the estimation of the risk surface by sample proportional to the prediction standard error; optimize sample size by sample taking into account spatial autocorrelation as in several spatial sampling approaches.

To summarize, in this work we developed a simple Bayesian kriging model, fitted to point referenced data on 121 farms, to obtain a prediction of helminth infection probability and related uncertainty for each one of the 8794 geo-referenced sheep farms of the region under study. We then describe how these farm predicted risk probabilities and related uncertainty can be used in planning future surveys for veterinary surveillance.

2. Materials and methods

2.1. Study area and data collection

The survey was conducted in the Campania region (southern Italy) which extends over an area of 13,590 km^2. A Geographical Information System (Gauss-Boaga projection Geodetic datum Rome40) was developed using Arc-GIS 9.3 GIS software (ESRI, Redlands, CA, USA). All the sheep farms present in the year 2005 (and then updated till the year 2010) in the region were geo-referenced.

Since this was the first survey the farms were uniformly sampled throughout the entire region by a systematic grid sampling approach (Rinaldi et al., 2006). Specifically, a grid representing quadrants of 10 × 10 km was overlaid on the region map within a GIS. As a result, the territory of the Campania region was divided in 135 equal quadrants, the centroid of each quadrant was identified and among all the farms present into the GIS database the farm closest to the centroid in each quadrant was selected (Fig. 1). Only pastured farms with more than 50 animals were included in the sample. Out of the total 135 quadrants, 121 were investigated (89.6%). In the other 14 quadrants, no farms having the required characteristics were present. Individual faecal samples were collected from animals on the selected ovine farms. On each ovine farm, the animals were divided into two age groups: lambs (4–18 months) and adult sheep (older than 18 months). Five individual faecal samples were