

Spatio-Temporal Modeling Of TB In Cattle Herds

Gabrielle E. Kelly

School of Mathematical Sciences, UCD

Abstract

We examine spatial association of bovine TB in cattle herds using data from Ireland. Badgers (*Meles meles*), a protected species under the Wildlife Act 1976 (OAG 2012), have been implicated in the spread of the disease in cattle. Current disease control policies include reactive culling (in response to TB outbreaks) of badgers in the index and neighbouring farms. Kelly and More (2011) using generalized linear geostatistical models, established that TB clusters in cattle herds and estimated the practical spatial ranges at which this occurs. Here this work is extended by taking into account possible anisotropy. Changes in spatial association over two time periods are also examined. The results have direct implications for establishing scale and direction in reactive culling. They are also of import regarding the evaluation of vaccines for badgers and cattle.

Keywords: spatio-temporal association, anisotropy, bovine TB, semi-variograms, linear geostatistical models.

1. Introduction

Bovine tuberculosis (TB) is a disease that affects cattle and badgers (*Meles meles*). The annual herd incidence is around 4% in Ireland (DAFF 2011) while incidence in the UK is roughly double this (DEFRA 2011). Two large-scale field trials, the Four Area Project (FAP) (Griffin, Williams, Kelly, Clegg, and Collins 2005) in Ireland and the Randomized Badger Culling Trial (RBCT) (Bourne, Donnelly, Cox, Gettinby, McInerney, and Morrison 2007) in England presented strong evidence that badgers infect cattle. Both trials involved comparisons between large areas where badgers were pro-actively culled to reference areas where little or no culling was carried out. The RBCT also included comparisons between areas with no culling and areas where reactive culling (in response to TB outbreaks) was carried out and concluded that such culling leads to an increase in bovine TB due to a perturbation of badger habitats (Bourne *et al.* 2007). In Ireland reactive culling typically occurs in the index and neighboring farms - a distance of a few kilometres, while proactive culling in the FAP and

RBCT was in areas of well over 100 km². [Kelly and More \(2011\)](#) using data gathered in two 5-year periods, before (Sept. 1992-Aug. 1997, period 0) and during (Sept. 1997-Aug. 2002, period 1) proactive culling in lands of the FAP established that TB is spatially clustered in cattle herds. They also found the practical spatial correlation ranges at which this clustering takes place and showed spatial clustering persisted following badger removal. In their methods isotropy was assumed but this was based merely on visual inspection of the data.

Here we re-examine the issue of spatial association using the same data by more formally investigating the question of anisotropy. The two time periods are also examined for changes in spatial association. This has direct implications for badger culling policy, since selecting for reactive culling all farms surrounding an index farm may not be appropriate, and it may be more effective if reactive culling took place in a particular direction only. This is also of import regarding the current development of vaccines for badgers and cattle. We note badgers are a protected species in Ireland since the Wildlife Act of 1976 ([OAG 2012](#)).

2. Materials and Methods

2.1. Data

We examine data on TB incidence in cattle herds, from matched removal and reference areas in county Cork in Ireland, used in the Four Area Project (FAP). The study design of the FAP and its results are published in detail in [Griffin *et al.* \(2005\)](#). Two time periods, the 5-year period prior to (Sept. 1992 - Aug. 1997, period 0) and the 5-year period of proactive culling (Sept. 1997 - Aug. 2002, period 1) are studied. Badger removal was intensive and proactive during period 1 in the removal areas, but reactive during period 0 and in the reference areas for both periods. Reactive culling typically took place only in response to severe outbreaks of TB where badgers were implicated as described in [Griffin *et al.* \(2005\)](#). The removal area (area approx 300² km²) has roughly 300 herds while the reference area (area approx 200² km²) has 270 herds. 27% of herds in the removal and 21% in the reference were infected with TB in period 0, and 15% in the removal and 29% in the reference in period 1. The geographical location associated with a cattle herd is based on the centroid of the largest land fragment owned by the farmer and recorded in a GIS database. All cattle herds in Ireland are tested at least annually. A herd was designated infected in a year if it contained at least two cattle that tested positive that year.

Although the FAP included four counties, we restrict attention here to county Cork as the strongest evidence for spatial clustering of TB was found in county Cork in [Kelly and More \(2011\)](#). Also the scale at which spatial correlation occurs in each county varies with infection rate and thus counties need to be considered separately, although analytic methods are similar. A detailed analysis here of each county would detract from the main issues and thus only results from county Cork are presented.

2.2. Statistical Methods

The primary purpose of the analysis is to demonstrate if there is spatial association of herd TB in a geographical area and if this spatial association varies with direction i.e. is anisotropic. A secondary purpose is to examine if spatial correlation changes over time.

Using TB infection status of the herd as the outcome variable, a logistic glm is fitted using

only the covariates log of herd size, and previous history of infection, found to be significant in previous studies (Griffin *et al.* 2005; Kelly and More 2011). The Pearson residuals from the model were obtained. Thus, all or part of the broad-scale (first order) variation in the mean response was identified and removed.

These residuals were then examined for directional effects by constructing directional empirical semi-variograms (Diggle and Ribeiro 2007, Section 5.2). This was done by partitioning the space of herd location vector differences into grid cells and displaying the corresponding average semi-variogram ordinates as a contour plot.

More formally, the Pearson residuals were examined for spatial structure via model fitting. The residuals were assumed to be normally distributed. Two groups of residuals were identified corresponding to: TB positive/negative cattle herds.

Firstly, trend surface models as described in Cressie (1991, chap.4,6); Diggle and Ribeiro (2007) were fitted. Such models include powers and cross-products of GIS co-ordinates as fixed effects in logistic regression models relating TB incidence to covariates.

Then linear spatial geostatistical models (LGM's) with exponential type covariance structure were fitted (Schabenberger and Gotway 2005). A fixed effect of group is included in all models since the sets of residuals are of different order of magnitude by group, but spatial covariance structure may differ among the groups. The LGM model is as follows:

Let $r(s_i)$ be the Pearson residual for the herd at the i th location s_i . Let $u(s_i)$ be a spatial random effect at location s_i . We assume the $u(s_i)$ follow an exponential anisotropic covariance model F with (i,j) th element given by

$$F[i, j] = Cov[u(s_i), u(s_j)] = \sigma^2 \exp[-\theta_1 |x_i - x_j|^{p_1} - \theta_2 |y_i - y_j|^{p_2}]$$

where x and y denote the easting and northing of GIS co-ordinates respectively. The variance-covariance matrix of the data is modeled as

$$Var[r(s)] = c_0 I + F \quad (1)$$

The parameters σ^2 and c_0 , refer to the geostatistical parameters "partial sill" and "nugget" respectively (Schabenberger and Gotway 2005, Section 6.3). Since we have repeated measures on cattle herds (data over 5 years in each time period), the nugget term represents yearly variation within herds. Differing values of θ_1 and θ_2 in the model indicate anisotropy. A partial sill of zero indicates no spatial structure.

We consider the removal and reference areas of county Cork separately since the spatial correlation pattern in each area varies considerably. Moreover, as the 5-year time period, period 1, corresponds to a period of proactive badger culling in the removal area only, these data need to be considered separately. Also, as infection rates are relatively low, computational difficulties associated with sparse data arise when data is considered yearly. Therefore, separate spatial models are fitted by area and 5-year period (four in total; removal area period 0 and 1 and reference area period 0 and 1). The effect of time is also examined, by combining data over both time periods in the reference (removal) area and comparing models with a single covariance structure to those with a different covariance structure for each time period. A measure of goodness-of-fit of a model is taken as: -2 residual log(likelihood). Models with the same mean structure but nested covariance structures are compared by taking the difference of -2 residual log(likelihood) and referring it to a χ^2 distribution with degrees of freedom the difference in number of covariance parameters in the models i.e. a likelihood ratio test (LRT). Since we are testing whether certain variance components are zero, the parameters lie

on the boundary of the parameter space and thus we divide the p-value obtained from the χ^2 by 2 (Self and Liang 1987). The test will be referred to as a modified LRT.

Non-nested models are compared using AIC.

Isotropic models are also fitted i.e.

$$F[i, j] = Cov[u(s_i), u(s_j)] = \sigma^2[\exp(-d_{ij}/\rho)]$$

where d_{ij} is the distance between the locations s_i and s_j and ρ refers to the geostatistical parameter "range". Since covariance in this model reaches zero only asymptotically, the practical range is defined as the distance at which covariances are reduced to 5% of the sill i.e. 3ρ . Isotropic models are compared to the anisotropic models again using modified LRT's. The Gaussian and spherical covariance structures are also fitted in all models to see if fit is improved (Schabenberger and Gotway 2005).

We computed a common measure for the degree of spatial structure via the relative structured variability (RSV) given by

$$\frac{\sigma^2}{\sigma^2 + c_0} \times 100\%$$

(Schabenberger and Gotway 2005, Section 4.2).

The scale in all models is km.

Effects with p-values < 0.05 are regarded as statistically significant.

Models were fitted using the Mixed and Glimmix procedure in the software package SAS version 9.1.3 (SAS Institute Inc., Cary, North Carolina, USA).

3. Results

Initial plots were made of the response variables infected/not infected, as well as the Pearson residuals from the logistic models, versus the locations. Figure 1 displaying 3-d scatterplots of the Pearson residuals of infected herds versus their locations shows some spatial structure particularly in the easting direction.

Directional semi-variograms based on the Pearson residuals of positive herds were then constructed and are shown in Figure 2. Approximately, these show more spatial variation along the east-west axis than north-south apart from the reference area period 0.

In the first model-fitting step, trend surface models were fitted to the Pearson residuals for each area/period and the results are shown in Table 1. Note that a group (infected/not-infected) term together with interaction terms with the GIS co-ordinates were included in the initial models but only effects relating to infected herds were statistically significant. A decreasing trend east-west was found in all area/periods and in the reference area period 0 an increasing trend south-north was also found.

Then LGM's, given by equation (1), were fitted to the Pearson residuals again by area/period. Pearson residuals associated with non-infected herds showed no spatial structure and thus analysis was restricted to the Pearson residuals from infected herds only. Estimates and standard errors of partial sills and nugget were determined. Using modified LRT's there was significant spatial structure in both areas for both periods ($p < 0.001$ in all cases, except the reference area period 0 p-value=0.0162), and the relative structured variability was high (the nugget effect was small and significant in all cases $p < 0.001$). The results are shown in Table 2. The Table shows that in three areas/periods there is spatial correlation only in the east-west direction while in the reference area period 0 there is also spatial correlation in the

Figure 1: 3d scatterplots of standardized Pearson residuals of infected herds from a logistic model fitted to the herd TB incidence, in the badger removal and reference areas of County Cork, (a)removal pre-cull period:Sept.'92-'97, (b)removal cull period:Sept.'97-Aug.'02, (c)reference pre-cull period:Sept.'92-'97, (d)reference cull period:Sept.'97-Aug.'02

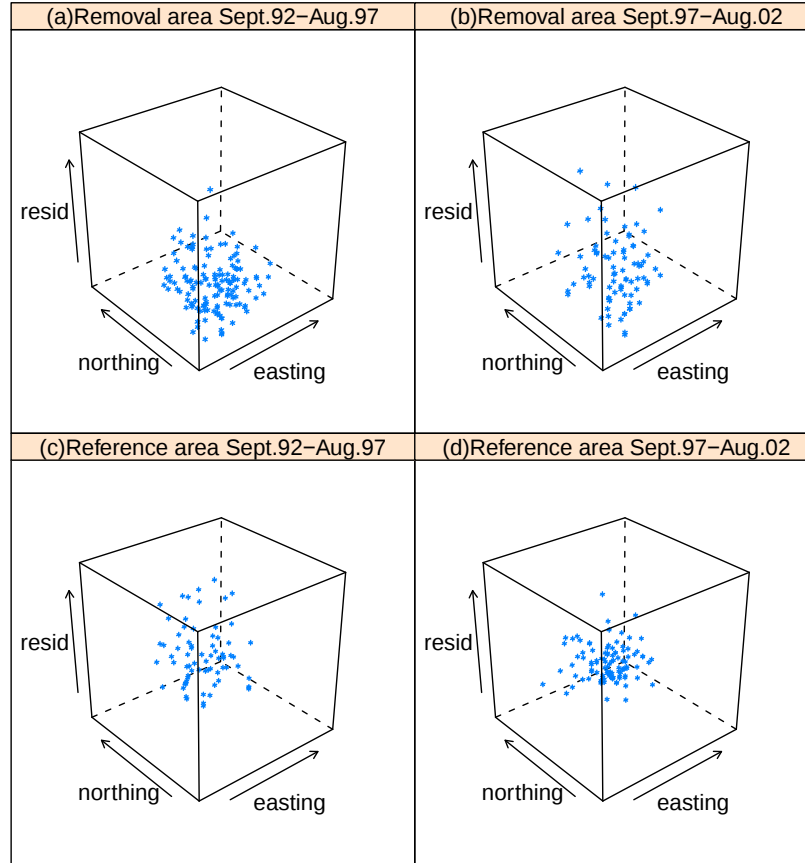


Table 1: Parameter estimates and standard errors (est (s.e.)) and AIC values of the best fitting trend surface models fitted to Pearson residuals of TB positive herds from logistic models. The logistic models were fitted to TB incidence of herds separately for the removal (rem) and reference (ref) areas for the pre- (0) and proactive (1) badger culling periods of county Cork in the Four Area Project.

	Removal period 0		Removal period 1		Ref period 0		Ref period 1	
Variable	est (s.e.)	p-value	est (s.e.)	p-value	est (s.e.)	p-value	est (s.e.)	p-value
easting	-8.49 (1.66)	< 0.0001	-7.10 (3.13)	0.026	-6.65 (3.36)	0.052	-4.37 (1.87)	0.021
northing	-	-	-	-	18.15 (6.33)	0.0055	-	-
AIC	424.5		300.6		259.7		262.9	

Figure 2: Directional empirical semi-variograms of standardized Pearson residuals of infected herds from a logistic model fitted to herd TB incidence, in the badger removal and reference areas of County Cork, pre-cull period:Sept.'92-'97, cull period:Sept.'97-Aug.'02. The contour plot for each area/period was constructed from the sample means of the empirical semi-variogram ordinates in square bins of equal size.

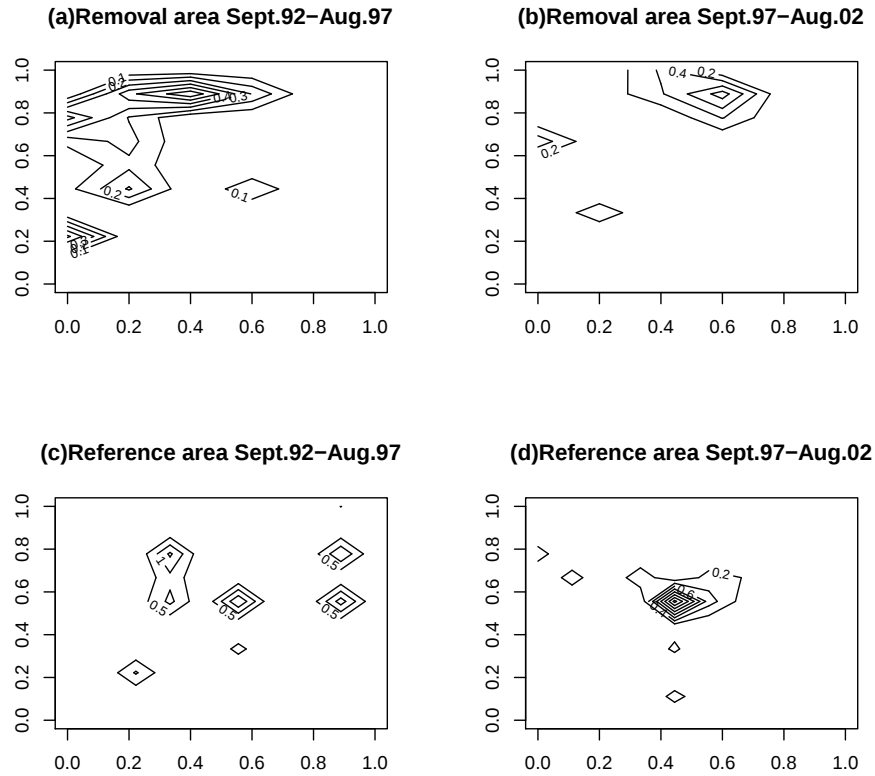


Table 2: Covariance parameter estimates and standard errors (est (s.e.)) of the best fitting anisotropic models, given by equation (1), fitted to the Pearson residuals associated with TB positive herds from logistic models. The logistic models were fitted to TB incidence of herds separately for the removal and reference (ref) areas for the pre- (0) and proactive (1) badger culling periods of county Cork in the Four Area Project. A '-' indicates the effect was not significant in the model by a modified likelihood ratio test. AIC values and the p-value of the LRT test for spatial structure are also shown.

Parameter	Removal period 0		Removal period 1		Ref period 0		Ref period 1	
	est (s.e.)	p-value	est (s.e.)	p-value	est (s.e.)	p-value	est (s.e.)	p-value
partial sill	2.89 (3.61)	0.212	2.51 (1.12)	0.013	8.63 (20.25)	0.335	0.84 (0.73)	0.1228
θ_1	0.92 (1.26)	0.231	2.38 (4.01)	0.277	0.25 (0.65)	0.352	2.71 (11.46)	0.4065
θ_2	-	-	-	-	2.10 (11.23)	0.426	-	-
p_1	1.15 (0.69)	0.049	0.04 (0.08)	0.311	0.09 (0.11)	0.185	0.02 (0.09)	0.4342
p_2	-	-	-	-	1.23 (1.19)	0.151	-	-
nugget	1.19 (0.15)	< 0.0001	0.30 (0.10)	0.001	0.92 (0.35)	0.0039	0.06 (0.02)	0.0003
AIC	437.7		290.5		276.6		239.0	
LRT								
best vs null model		0.0002		< 0.0001		0.0162		< 0.0001

south-north direction.

When the two time periods for the reference area were combined and model (1) fitted, a model with $\theta_2 = 0.0$ fitted best with -2 residual log likelihood = 549.6 and AIC= 557.6. A model fitted to the same data with different covariance parameters (given by model (1)) for each time period had a best fitting model -2 residual log likelihood = 537.8 and AIC= 551.8. By a modified LRT test the difference in goodness-of-fit of these two models indicating that covariance parameters differ for the two time periods. A similar analysis of the removal area also indicated a significant difference for the two time periods.

4. Discussion

The exploratory data plots of Figures 1 and 2 indicate some spatial structure in these data. With respect to Figure 2, as noted in Diggle and Ribeiro (2007), this type of three-dimensional plot requires more data than its two-dimensional counterpart (the ordinary empirical semi-variogram) and yields only gross directional effects. It serves mainly as a tool for displaying the data prior to formal inference.

The results of the model fitting both in Tables 1 and 2 show spatial association of TB varies both over time, between areas and with direction within an area. The trend-surface model results of Table 1 provide useful descriptions of simple, unexplained spatial trends and give the direction of the trend. The LGM's results of Table 2 however, can describe smaller-scale variation and can estimate the scale and direction of this variation. We note the covariance parameter estimates displayed in Table 2 are not significant using Wald type tests. However, as covariance parameter estimates are not typically normally distributed, unless some transformation is applied, such tests are not appropriate here and in all cases modified LRT tests were carried out to test the significance of covariance parameters.

We note Tables 1 and 2 give results that are generally consistent with each other. Diggle and Ribeiro (2007) chose to interpret spatial effects which vary on a scale comparable to or greater than the dimensions of the study region as variation in the mean of the process and smaller-scale effects as variation in the actual process. Cressie (1991) also argues that spatial covariance functions which do not decay essentially to zero at distances shorter than the dimensions of the study region will be poorly identified, and in practice indistinguishable from spatial trends. However, the results of the spatial models also indicate the range of correlation. For example, in the removal area period 0, the estimate of the range parameter is approximately 1.0 in the east-west direction indicating a practical range of approximately 3 km.

It is difficult to examine time-varying spatial dependence mainly due to sparsity of data. There are also other confounding factors - the incubation period of the disease, time difference between annual TB test and disease onset. The time of disease onset is unknown and only TB status at time of annual test is available. Thus, when examining yearly data we chose to group the data into two five year periods based on differing TB-control programs during those periods in the removal area and spatial dependence was examined separately in each area/period. Thus consistent estimates are obtained for each 5-year period. These models were then compared to models fitted to the combined time periods. To this extent, time-varying spatial dependence was examined.

The differing estimates of covariance parameters in different regions and time periods has implications for badger culling policy indicating a single range or direction for culling is not appropriate. Moreover, in evaluating the results of field trials relating to vaccine development the question of anisotropy needs to be incorporated in any comparisons.

We note spatial association persists following complete badger removal (removal area period 1 in Table 2) most likely due to factors other than badgers ?. Further work will explore efficient methods for detecting the direction in which spatial clustering of TB occurs in cattle herds. This may take two forms - firstly through statistical detection of spatial clusters and secondly association with important risk factors. In relation to the latter factor, badgers live in social groups in a system of interconnecting underground burrows called setts and can travel distances of at least 6 km from these setts (Kelly, McGrath, and More 2010). Since, as noted in the Introduction, it is known that the source of some TB infection in cattle is the badger then it may be reasonable to look for spatial clustering in the direction of the nearest badger sett for example.

Acknowledgements

I would like to thank an anonymous referee for helpful comments in relation to this paper.

References

- Bourne F, Donnelly C, Cox D, Gettinby G, McInerney J, Morrison W (2007). *Bovine TB: the scientific evidence. Final report of the Independent Scientific Group on Cattle TB*. Defra: London.
- Cressie NAC (1991). *Statistics for spatial data*. New York: Wiley.

- DAFF (ed.) (2011). *Bovine TB and Brucellosis Eradication Schemes*. Department of Agriculture, Fisheries and Food: Ireland. URL <http://www.agriculture.gov.ie/animalhealthwelfare/diseasecontrol>.
- DEFRA (ed.) (2011). *Monthly publication of national statistics on the incidence of TB in cattle for Great Britain*. Department for Environment, Food and Rural Affairs : UK. URL <http://www.defra.gov.uk/statistics/foodfarm/landuselivestock/cattletb>.
- Diggle P, Ribeiro P (2007). *Model-based Geostatistics*. New York: Springer.
- Griffin J, Williams D, Kelly G, Clegg T, Collins J (2005). “The impact of badger removal on the control of tuberculosis in cattle herds in Ireland.” *Preventive Veterinary Medicine*, **67**, 237–266.
- Kelly G, McGrath G, More SJ (2010). “Estimating the extent of spatial association of *Mycobacterium bovis* infection in badgers in Ireland. Epidemiology and Infection 2010.” *Epidemiology and Infection*, **38**, 270–279.
- Kelly G, More SJ (2011). “Spatial clustering of TB-infected cattle herds prior to and following proactive badger removal.” *Epidemiology and Infection*, **139**, 1220–1229.
- OAG (ed.) (2012). *Wildlife Act 1976, Irish Statute Book*. Office of the Attorney General: Ireland. URL <http://www.irishstatutebook.ie/1976/en/act/pub/0039>.
- Schabenberger O, Gotway C (2005). *Statistical methods for spatial data analysis*. Boca Raton Florida: Chapman and Hall/CRC.
- Self S, Liang K (1987). “Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions.” *Journal of the American Statistical Association*, **82**, 605–610.

Affiliation:

Gabrielle E. Kelly
School of Mathematical Sciences, UCD
Dublin, Ireland
E-mail: gabrielle.kelly@ucd.ie
URL: <http://maths@ucd.ie/gkelly>