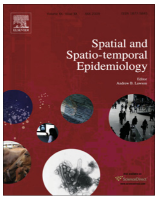
Spatial and Spatio-temporal Epidemiology 11 (2014) 125–133

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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



Using exceedance probabilities to detect anomalies in routinely recorded animal health data, with particular reference to foot-and-mouth disease in Viet Nam

K.K. Richards [a], M.L. Hazelton [a,*], M.A. Stevenson [b,1], C.Y. Lockhart [c], J. Pinto [c], L. Nguyen [d]

[a] Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North 4412, New Zealand
[b] EpiCentre, Institute of Veterinary, Animal, and Biomedical Sciences, Massey University, Private Bag 11-222, Palmerston North 4412, New Zealand
[c] Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, Rome 00153, Italy
[d] Department of Animal Health, Phuong Mai, Dong Da, Ha Noi, Viet Nam

ARTICLE INFO

Article history:
Available online 28 September 2014

Keywords:
Bayesian
Exceedance probabilities
Logistic regression
Random effects
Under-reporting
Veterinary epidemiology

ABSTRACT

The widespread availability of computer hardware and software for recording and storing disease event information means that, in theory, we have the necessary information to carry out detailed analyses of factors influencing the spatial distribution of disease in animal populations. However, the reliability of such analyses depends on data quality, with anomalous records having the potential to introduce significant bias and lead to inappropriate decision making. In this paper we promote the use of exceedance probabilities as a tool for detecting anomalies when applying hierarchical spatio-temporal models to animal health data. We illustrate this methodology through a case study data on outbreaks of foot-and-mouth disease (FMD) in Viet Nam for the period 2006–2008. A flexible binomial logistic regression was employed to model the number of FMD infected communes within each province of the country. Standard analyses of the residuals from this model failed to identify problems, but exceedance probabilities identified provinces in which the number of reported FMD outbreaks was unexpectedly low. This finding is interesting given that these provinces are on major cattle movement pathways through Viet Nam.

1. Introduction

Technological advances are providing ever better opportunities for the recording, processing and storing of disease event information. This, in theory, should provide the responsible authorities with the necessary information to carry out detailed analyses of the factors influencing the spatio-temporal distribution of disease in animal populations. However, the reliability of such analyses depends on the quality of the data. Anomalies in data records have the potential to introduce significant bias in the results of descriptive analyses and fitted statistical models, skewing results and potentially leading to inappropriate decision making.

The usual statistical methodology for identifying anomalous data is residual analysis (e.g. Gail, 1991). A large residual indicates a data record that is poorly predicted by the model in use, indicating either a problem with the datum in question or significant shortcomings for the model as a whole (or both). For simple models, such as linear regressions with fixed effects, this type of analysis is straightforward to implement and interpret. However, residual analysis becomes more challenging as the underlying data types and models become more complicated. For example,

---

[*] Corresponding author. Tel.: +64 6 356 9099; fax: +64 6 350 5682.
E-mail address: m.hazelton@massey.ac.nz (M.L. Hazelton).

[1] Current address: Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria 3010, Australia.

http://dx.doi.org/10.1016/j.sste.2014.08.002
1877-5845/© 2014 Elsevier Ltd. All rights reserved.

until quite recently there was no generally accepted definition of a residual for a spatial point process (Baddeley et al., 2005).

Hierarchical statistical models for spatio-temporal epidemiological modelling present particular problems. Such models seek to describe the variation in disease incidence or prevalence by incorporating random variables at multiple levels (see, for example, Lawson, 2013). In principle this means that we can define residuals at the corresponding levels of the hierarchy. Moreover, these models often provide such flexibility in their representation of the data that records which are truly 'odd' appear, superficially at least, to fit adequately.

Our goal in this paper is to promote the use of *exceedance probabilities* as a tool for identifying and assessing anomalous data in spatio-temporal models for areal disease count data. Exceedance probabilities can be applied at any level of the model to describe the extent to which an individual random term, or combination of random terms, is unusual, in the sense of lying in the extreme tails of the specified distribution. Exceedance probabilities have been used previously to detect anomalous clusters of cases in point process data (e.g. Diggle et al., 2005; Davies and Hazelton, 2013), and have also been used to identity regions with unusually high relative risk when modelling areal count data (e.g. Best et al., 2005;Lawson, 2010). We aim to showcase wider uses of exceedance probabilities, demonstrating their application to individual stochastic terms in areal count models and also noting their utility for highlighting localities with unexpectedly low (as well as high) reported rates of disease.

We illustrate the use of exceedance probabilities as a diagnostic tool for anomalous data through a case study involving outbreaks of foot-and-mouth disease (FMD) in Viet Nam for the period 2006–2008. We describe the data in detail in Section 2. We develop Bayesian models for the data in Section 3. These are fitted using Markov chain Monte Carlo methods. Our preferred model (chosen using standard model selection techniques) is flexible, incorporating random interaction terms to account for otherwise unexplained spatio-temporal heterogeneity. In Section 4 we define exceedance probabilities for our model, and describe their computation within a Bayesian framework. These exceedance probabilities highlight some anomalous data that would otherwise have remained hidden. In Section 5 we discuss these findings, and relate the geographical distribution of data anomalies to patterns of cattle movement across Viet Nam.

2. The Viet Namese foot-and-mouth disease dataset

2.1. FMD in Viet Nam

FMD is a highly infectious disease of cloven-hoofed animals (principally cattle, sheep and pigs, Bachrach, 1968; Bachrach, 1968) spread by direct and indirect animal contact as well as airborne spread. The FMD virus causes fever and affects the epithelial tissue causing vesicular lesions on hard wearing body parts such as the mouth, snout, feet and occasionally teats. These lesions normally cause a loss of appetite, lameness, and a reduction in productivity. FMD is fatal in approximately 2% of adult and 20% of neonatal cases (Radostits et al., 2007). Seven known serotypes of FMD have been recorded: A, O, C, Asia-1, and South African Territories (SAT) 1, 2, and 3 (Davies, 2002). Recent studies of FMD in Asia include Ding et al., 2013; Wang et al., 2013; Muroga et al., 2012.

FMD is endemic in Viet Nam, with three out of the seven serotypes (A, O, Asia-1) isolated from identified cases on a regular basis. It is observed primarily in cattle, buffalo and pigs. FMD has significant effects at a number of levels (Cocks et al., 2009). For livestock owners, FMD reduces productivity, can cause death in young cattle, and imposes a cost for treatment of sick animals. Cattle and buffalo in Viet Nam are used for draught power, so FMD infection can impact other areas of farming. It also affects the country as a whole through the cost of intervention strategies (such as vaccination) (Cocks et al., 2009) and its impact on international trade.

There is a significant amount of trade in cattle and buffalo between Viet Nam and other FMD-endemic countries, such as the Lao People's Democratic Republic (Laos), Thailand, and the People's Republic of China (Gleeson, 2002). By no means all of this is regulated, with unmonitored movement of cattle across international borders believed to be quite prevalent. This may have an important impact on the spatial epidemiology of FMD in Viet Nam.

2.2. The data

We have available data from the Vietnamese Department of Animal Health, consisting of 2734 reported commune-level outbreaks of FMD for the period 1 January 2006 to 31 December 2008. For every reported case we have an estimated date of onset, geographical area (described below), and the serotype of the outbreak. Any other cases from the same commune during the outbreak were assumed to be the same strain as the first case. All three endemic serotypes (O, A, and Asia-1) were observed, although the number of occurrences of type A was low compared with serotypes O and Asia-1.

Viet Nam is divided administratively into provinces, and then further subdivided into districts and communes. There are 64 provinces, with an average size of just over 5000 km^2, and 11,052 communes with an average area close to 30 km^2. These geographical divisions are shown in Fig. 1, which displays the provincial structure of Viet Nam (Fig. 1, and the division of a single province into communes (Fig. 1b).

For these analyses, FMD cases were identified spatially by their commune location. As an illustration, Fig. 2 shows the geographical pattern of disease for each of the years 2006, 2007 and 2008. Specifically, each point represents the centroid of a commune in which a least one case of FMD in cattle was observed during the respective year.

In addition to these disease data, we have information on covariates cattle density, enhanced vegetation index (EVI) and elevation (in metres). Cattle density is specified for each year at the provincial level only, and is measured as the number of cattle per square kilometre. It is based on