Reviewer’s report

Title: Spatiotemporal CART model for predicting the outbreak of cryptosporidiosis in Queensland, Australia

Version: 1 Date: 14 November 2009

Reviewer: Xiao-Nong Zhou

Reviewer’s report:

General Comments
It is interested to the readers that this paper is the first time to explore the relationship between geospatial distribution and social-ecological impact on the cryptosporidiosis that is a water-borne disease caused by infection with a protozoan parasite. Some outbreaks have been reported in some countries, the early warning system for this disease transmission is warrant to investigation. Therefore, this manuscript is well developed the system by using spatiotemporal CART models. The written is acceptable to be published.

Major compulsory revisions

The main problem of this manuscript is that the author has not quite clearly defined the function of each method, and made confused to readers by some conflict results from various methods which caused the results out of the original objectives of the manuscript.

Minor Essential Revisions:

1) Based on the objectives described in the end of introduction: “The aims of this paper are to examine the potential impact of socio-ecological factors on the incidence of cryptosporidiosis using spatiotemporal CART models and explore their potential as a predictive model for cryptosporidiosis in Queensland, Australia”, the main results have to focus on the (1) general analysis of the spatial temporal distribution of cryptosporidiosis; (2) potential impact of socio-ecological factors, by using the regression CART models; (3) outbreak prediction model with a risk mapping. My comments on those three stages are: (1) In the general analysis, the spatiotemporal Bayesian model are enough to explain all issues already, but the authors only double check the spatial issue by two methods (spatial autocorrelation analysis and spatial empirical Bayes rates smoothing ) which did not take the advantage of Bayesian model to explore both geographical and time series issues. (2) In correlation analysis part, the authors using three stage spatiotemporal CART models, but did not clearly define the logical issues among these three stages. It recommended that in the first stage, it only explore the non-present and present, then the present results could be got into the second stage for the regression CART models, finally the results from regression CART models with over outbreak’s criteria could be used in the third stage. (3) It is quite good to produce the risk map by outbreak CART model.
2) The term of “social-ecological impact” is not quite suitable used in this manuscript, since the results of this manuscript showed the incidence or outbreak are only related to temperature and SEIFA. SEIFA was been used for standing for social factors, but has not been discussed very clear in the manuscript, it is need to further explain well. And the temperature standing for ecological factors by the authors are not quite well defined, and it is suggested to use temperature directly since single parameter could not be standing as ecological factors. Therefore both title and objectives need to be amended.

Discretionary revisions:

1) The original datasets was not clearly described, for example, the covering periods of SEIFA datasets did not been given.
2) The temperature data set was only got from one year, why the results from CART analysis come out the more than one year.
3) It is suggest providing a supplement document on the statistical programme both on CART analysis or Bayes analysis.
4) It is suggest that drawing a general design scheme for the whole prediction system to make clear function of each model.
5) The main results from the different models need to discuss well, particularly on how to explain the each results in terms of social and biological meanings.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.