Determining the reliable scale when aggregating data to model risk of disease spread through mapping of animals movements. Combined use of Ripley K® and Besag L® functions applied on farm points and distance distribution study on animal exchange function.

P. Bonnet1*, S. Jean Baptiste2, M. Lesnoff4, A. Workalemu3. 1CIRAD-EMVT Montpellier, Dept. of Medical Geography GEOS, University of Montpellier III Paul Valery France, seconded to ILRI, P.O. BOX 5689 , Addis Ababa, Ethiopia. 2University of Paris France, Faculty of Medicine Kremlin Bicêtre, 3ILRI Ethiopia. 4CIRAD-EMVT seconded to ILRI Ethiopia.

a) Summary

The study highlights the rationale for proper choice of cell size when planning the use of a regular lattice (with aggregation of point data attributes) towards disease spread modeling. Purpose intends to particularly fit with the use of MAS Multi Agent System modeling to study disease spread, as it is based on social and spatial behavior of local stakeholders [Ferber J., 2000]. Study provides illustration from data originated from Western Ethiopian highlands farming systems [Bonnet P., 2002], with calculation of indicators on point aggregation patterns computed with Ripley statistics and the study of some social distance using exchange of animals between farmers as a proxy.

b) Introduction

The study is part of a larger research protocol to study local determinants of re-emergence of CBPP in Africa with particular focus in Western Ethiopian highlands. Determinants of CBPP herd status are firstly factors which favor spread of causative agent [Bonnet P., 2002] like animal movements as important factors for agent circulation and spread. In the particular area of the study, exchange of animals between farmers are provoked by socio-economic drivers such as the common use animals (oxen and bulls) for ploughing. Therefore the study on between farmers distance when animals exchange is undertaken was of primary importance. Moreover since exchange of animals is also driven by social and spatial proximity between farmers the study of neighborhood patterns through nearest neighbor statistics is of primary importance. Ripley function was chosen as a first try.

c) Objectives

MAS simulation is now considered as an interesting tool to study spatial behavior of stakeholders in the context of natural resources management [CIRAD-TERA , 2002]. In the context of animal health the study of disease spread through animal introduction and movements i.e. when disease is considered as a social marker would benefit using such approach [Estival J.-B., 2001]. MAS models of disease spread using spatial cells behavior patterns mimicking exchange of animals between farmers requires the use of square or hexagonal cell grid to simulate spread which results would depend on cell size. Therefore in order to study appropriate size of lattice cells and drive future sensitivity analysis we have studied aggregation function and some social distance in a particular benchmark area using farmers point dataset and results from a survey.

d) Materials & Method

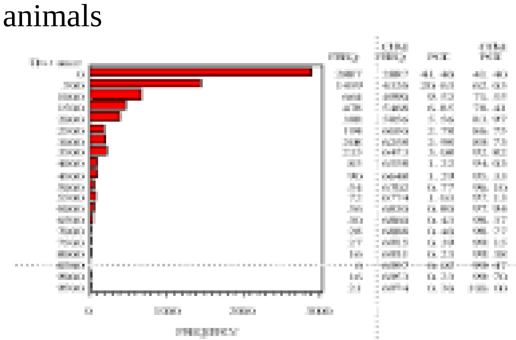
The study was carried out in sedentary crop livestock mixed farming systems in Ethiopian highlands. Data originated from a survey carried out on 6110 farmers. Firstly we analyzed the distribution function of distance between farmers exchanging cattle as observed in a dataset of 13351 exchanges documented. When linking farmers from survey data & within the spatial domain we were able to connect farmers for a total of 6974 exchanges following declaration of farmers. The mode and mean distance (and confidence intervals) when exchanging animal between farms and the distance decay function were studied. Cumulative distribution function of distance was used to assess the % of exchange that would be represented as going out any lattice cell when simulating exchange between cells which is the aim of MAS models.

Secondly we used computations of Ripley K(r) and Besag L(r) functions [Ripley B.D., 1977], [Besag J.E., 1977] computed with Crimestat© statistical package [Levine N., 2002] to look at patterns of spatial structure of farm points (and identify level of clustering if any) and to discuss possible levels of model aggregation for points (into villages or grid cells). We used dual computation with 1000 simulations and use of border correction (rectangular) [Goreaud F., 1999]. We computed Ripley statistics with and without use of an intensity variable z made of number of animals exchanged.

e) Results

First distance function study gives various thresholds representing various cell sizes for lattice in MAS modeling. When taking a threshold of 2000 m as the edge of the cell in a regular square lattice grid we allow 80% of exchange to remain internal to cell and 20% to spread out the cell into neighboring cells in average. When taking a threshold of 4000 m as the edge of the cell in a regular square lattice grid we allow 94% of exchange to remain internal and only 6% to spread out the cell in average. The choice of a threshold is therefore dependent on particular objectives and would serve for sensitivity analysis when simulating. Results from Besag L() suggest that there is high level of aggregation of breeding farms point at low distance bin less than 5000 meters which confirms the first assessment and give a superior limit value for maximum edge to be considered. Distribution of inter point distance as observed is compared with distribution under complete randomness.

Figure 1(left): Distance decay function from distance between farmers exchanging animals



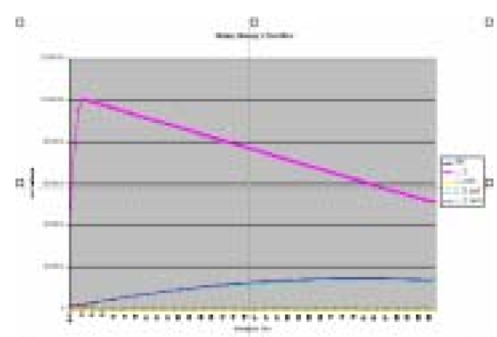


Figure 2 (right): Besag L (from K Ripley) function displaying aggregation at low values of distance bin. Ripley's Besag's L() function plotted as provided in Figure 2 suggests low distance (less than 5000 m) aggregation since observed values display an L function out of randomness envelope curves for short distance.

f) Discussion

Second order statistics and particularly Ripley's K function is used to describe complete distribution of all distance in the point pattern as it is a test on the cumulative distribution function of the full set of inter-point distance [Anselin L., 2003]. Besag L() function is used instead of K() as a re-scaled K function where reference for spatial randomness is linear and put in an horizontal line at zero. On the other hand distance function from survey represents real behavior of farmers and a good proxy for their spatial behavior and disease spread risk. It mimics social distance between farmer i.e. a proxy distance within social networks and therefore seems appropriate to serve into social and behavioral models.

These two methods provide ranges of distance for possible aggregations trials and representations in MAS models. Finally appropriate size of pixel when aggregating farm data into pixel or for the use into socio-spatial modeling is documented. The choice of cell size depends on final objectives of the modeling method.

g) References

1. Anselin L. An introduction to Point pattern analysis using Crimestat. Department of Agricultural and Consumers Economics. Urbana Champaign: University of Illinois, 2003:19.
2. Besag J.E. Comments on Ripley's paper. Journal of the Royal Statistical Society 1977;B39:193-195.
3. Bonnet P. Etude préliminaire des marqueurs de risque socio-spatiaux de la diffusion de la péripneumonie contagieuse bovine (PPCB) dans les hauts plateaux d'Ethiopie. Exploration de données de recensement agricole et d'échanges d'animaux. GEOS Geography of Health. Montpellier France: Université Paul Valéry Montpellier 3, 2002: 65 p. plus annexes.
4. CIRAD-TERA . Systèmes multi agent et gestion des ressources naturelles. In: TERA C, ed. MAS training with CORMAS software. Montpellier: CIRAD, 2002:99.
5. Estival J.-B. Animal movement computer simulation with multi-agent system MAS. Implementation of an application simualtion prototype within the context of CBPP spread in Ethiopia. ISIMA. Clermont-Ferrand: Université de Clermont-Ferrand France, 2001.
6. Ferber J. Les systèmes multi agents, vers une intelligence collective. Paris, 2000.
7. Goreaud F., Pelissier R. On explicit formulas of edge effect correction for Ripley's K function. Journal of Vegetation Science 1999:433-438.
8. Levine N. Crimestat II: a spatial statistics program for the analysis of crime incidents locations (version2.0). Washington DC, Houston USA: National Institute of Justice, 2002.
9. Ripley B.D. Modelling spatial patterns. Journal of the Royal Statistical

Society 1977;B39:172-212.