Emerging Methods in using GIS to Analyse Barmah Forest Virus Disease in Queensland, Australia

Naish, S.,[1] Mengersen, K.[2] and Tong, S.[1]
[1]School of Public Health, Queensland University of Technology, Kelvin Grove campus, Herston Road Kelvin Grove - 4059, Brisbane, Queensland, Australia, E-mail: s.naish@qut.edu.au
[2]Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

Abstract
Barmah Forest virus (BFV) disease is an emerging mosquito-borne disease in Australia. We aimed to outline some recent methods in using GIS for the analysis of BFV disease in Queensland, Australia. A large database of geocoded BFV cases has been established in conjunction with population data. The database has been used in recently published studies conducted by the authors to determine spatio-temporal BFV disease hotspots and spatial patterns using spatial autocorrelation and semi-variogram analysis in conjunction with the development of interpolated BFV disease standardised incidence maps. This paper briefly outlines spatial analysis methodologies using GIS tools used in those studies. This paper summarises methods and results from previous studies by the authors, and presents a GIS methodology to be used in future spatial analytical studies in attempt to enhance the understanding of BFV disease in Queensland. The methodology developed is useful in improving the analysis of BFV disease data and will enhance the understanding of the BFV disease distribution in Queensland, Australia

1. Introduction

This paper follows a series of studies conducted by the authors and others on the epidemiology of Barmah Forest virus (BFV) disease in Queensland, Australia. Quinn et al (2005) have conducted geographical information system (GIS)-based study on BFV disease at local government area (LGA) level and revealed that there were large variations in the distribution of Queensland. Tong et al., (2005) has evidenced that the disease has geographically expanded over Queensland. Naish et al., (2006) have conducted time series analysis on a BFV database where data was grouped into large regional localities centred around a large city. Followed by this study, Naish et al., (2009) conducted another study that considered all the coastal regions around the large cities in Queensland and assessed the risk determinants for the transmission of BFV disease. Recently, Naish et al., (2011) conducted their first GIS-based study at the statistical local area (SLA, the smallest geographic unit in Australian census) level based on the largest available dataset (1993-2008, 16 years) and quantified the risk estimates of BFV disease incidence rates over time and space in Queensland. They have also conducted several spatial analysis including spatial autocorrelation, semi-variogram modeling and interpolation of incidence rates.

They found that there were significant spatio-temporal differences in the BFV disease across Queensland. Naish et al., (2011) have more recently conducted GIS-based studies into BFV distribution using spatial scanning statistical tools (SatScan) at the SLA level and determined that there were significant spatio-temporal clusters of BFV disease along coastal areas in Queensland. Although these previous studies have provided much insight into BFV diseases transmission, more can be achieved. All the above studies have investigated the disease burden using census boundary groupings which are arbitrarily located geographic boundaries. Naturally, the transmission of BFV disease does not confine itself to arbitrary boundaries. The relatively low incidence rates of the disease and the very large study area compound the issues imposed by such boundaries. Certain issues such as data reliability (Russell, 1994), residential address and location of infection site (Muhar et al., 2000) have already been addressed. Other perennial concerns in areal-based surveillance data analysis are the ecological fallacy, the geographic unit of spatial analysis and smaller disease counts (Muhar et al., 2000). To overcome some of these issues, we have recently developed a grid-based GIS database. This is acknowledged to be usually the most appropriate system for analysis of epidemiological data with environmental data (Fotheringham and Rogerson, 2009, Lloyd, 2010 and Thomson and Connor, 2000). It is of interest to the authors to investigate the spatial relationships of the disease with highly localised wetland features and urban density indicators in an attempt to assess the likely habitats of the BFV mosquito. This technique requires a much finer geographic resolution (Jerrett et al., 2003, Kitron, 1996, Moore and Carpenter, 1999 and Thomson and Connor, 2000) than that used in the previous GIS studies of BFV disease (Quinn et al., 2005 and Tong et al., 2005). Also of interest is, spatial factors such as temperature, rainfall, socio-economic indicators, population size and tides.

2. Methodology

2.1 Study Area

The overall study area is the Australian State of Queensland shown in Figure 1. As observed from the figure, Queensland is a very large area, occupying 1 723 936 km^2 with approximately four million people (Australian Bureau of Statistics, 2009).

In order to provide clear example maps, some sections of this paper will zoom into a small area around an arbitrarily selected city, Mackay, which is approximately mid way along the eastern seaside coast of Queensland.

2.2 Spatial Data Collection

2.2.1 BFV disease cases

De-identified BFV disease notification data (n = 9,431) from January 1992 to December 2008 (17 years) for the study area was obtained in spreadsheet format from the Queensland Department of Health (2008). These data included date of notification, age, gender and residential address. Of the total 9,431 case data supplied, 6,718 (72%) cases were successfully geocoded using GIS tools. Geocoding at the street level allows the analysis to proceed to fine geographic resolution yet maintains the privacy of the cases by not being able to identify the person. The finer resolution geocoding allows detailed investigation of the local environmental conditions surrounding each case, that grouping case to arbitrary governmental or statistical boundaries does not permit.

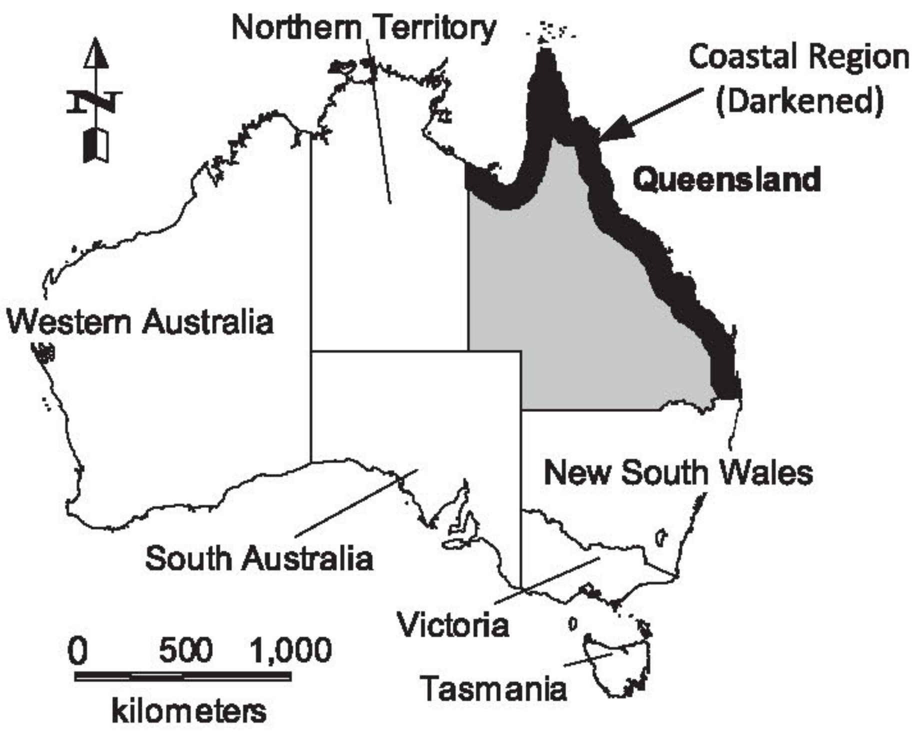

Figure 1: Study area, Queensland, Australia.

2.2.2 Wetlands

Wetland location and type data were obtained from the Queensland Department of Environmental Protection Agency (EPA) in polygon format (2009). Examples of the data obtained are shown in Figure 2 surrounding the city of Mackay. Also shown in the figure are the locations of the geocoded BFV cases.

2.2.3 Meteorology

Monthly gridded surface climate data (maximum and minimum temperature and rainfall) were obtained from 1992 to 2008 from Australian Bureau of Meteorology (2009).

2.2.4 Tides

Tidal data for the years 1992-2008 were obtained from Queensland Department of Transport and Main Roads (2009). The data supplied was in spreadsheet format with two low and high tide readings daily for each tidal monitoring station.

2.2.5 Climate zones

Koeppen classifications of climate zone were obtained from Australian Bureau of Meteorology (2009). It comprise a) equatorial; b) tropical; c) subtropical; d) desert; e) grassland; and f) temperate zones.

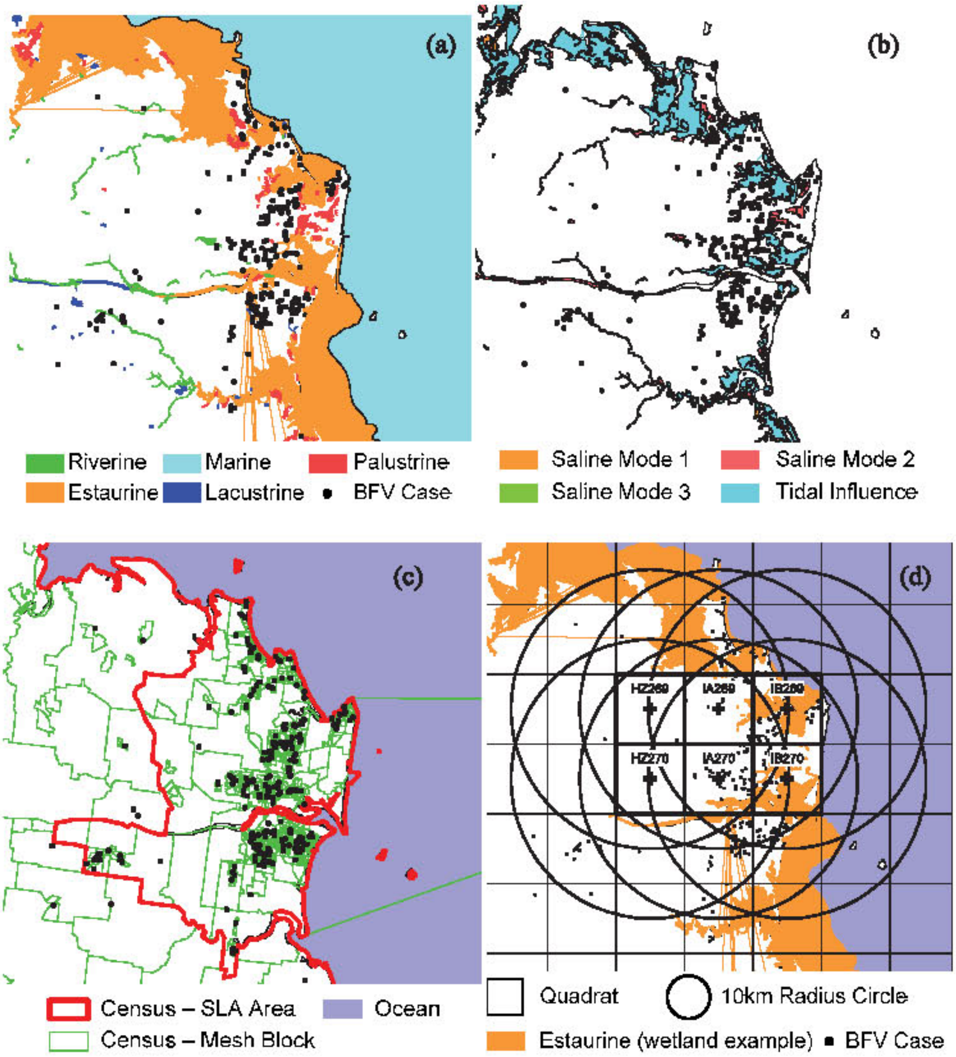

Figure 2: a & b) Example of wetlands near Mackay, Queensland; c) Example of SLA and mesh blocks in Mackay and d) Example of quadrats with their unique identifier codes, 10 km radius circle and centroid points.

2.2.6 Socio-economic and demographic data

Socio-economic indices and demographic data were obtained from the Australian Bureau of Statistics (2009) for the study period. Socio-economic Indexes for Area (SEIFA) and demographic data at the SLA level were based on CDATA 2006 of ABS, a database which provides information of 2006 Australian Census of Population and Housing, digital statistical boundaries, base map data and socio-economic data. We aimed to directly insert SEIFA and demographic data from CDATA into the database. SEIFA included four indices: the Index of Relative Socio-economic Advantage and Disadvantage (IRSAD, the higher IRSAD index, the higher the socioeconomic position), the Index of Relative Socio-economic Disadvantage (IRSD, reflecting disadvantage such as low income and education level, high unemployment and unskilled occupations), the Index of Economic Resources (IER, reflecting the general level of availability to economic resources of residents and households) and the Index of Education and Occupation (IEO, reflecting general educational level and occupational skills).

2.7 Census

The SLA boundaries are arbitrary in terms of BFV disease transmission, and cover a large area. The use of mesh block data from Australian Bureau of Statistics (shown as green boundaries in Figure 2) (2009) are available for the year 2006 census and contain data on overall population and number of dwellings. These data are also in arbitrary boundaries, but the scale is significantly more attributable to the geocoded BFV case data so it is considered a valuable source of information.

2.3 Spatial Data Sampling Design

The data sampling technique used is presented here. The sampling has been designed to improve the ability to determine the local environmental, ecological and climatic influences on the transmission of BFV disease. A 5km square grid lattice was created as suggested in the literature (Daniel et al., 2004, Moore and Carpenter, 1999 and Thomson and Connor, 2000) within the bounds of 9°S to 30°S and 140°E to 154°E. These bounds cover the entire land mass of Queensland. To remove grids outside of the Queensland land mass and prominent island areas, only the grids which were located over land were retained and the extraneous grids discarded. Data from the BFV cases database, population, wetland types and climate variables (rainfall and temperature) were extracted for each quadrat (grid square). Each quadrat was allocated a unique identifier code. An example of the quadrats is shown in Figure 2. The centroid of each quadrat was generated (Rushton, 1998). This was used to extract the climate data, the SLA census data, and the nearest tide station and establish other location data such as local government area. Epidemiologic applications of this approach include West Nile virus disease rates (Wniters et al., 2008), dengue fever (Wu et al., 2009) and malaria incidence rates (Olson et al., 2009). To overcome the problem of small counts, a 10km radius circle was generated around the centroid of each quadrat (Rushton, 1998). This circle was then used to sample the BFV case data, wetland data, and census mesh block data. The number of BFV cases falling within each circle was calculated and attributed to the quadrat. The radius size was chosen to be 10km as this encompasses the environmental conditions which may contribute to the transmission of BFV diseases, namely mosquito breeding sites and BFV mosquito flight range (Whelan, 1997). The 10km radius circle was also used to calculate the proportion of area of each wetland type falling within the circle. The same method was applied to calculate the proportion of population and number of dwellings obtained from the mesh blocks. It is noted that the 10km circles have considerable overlaps with the adjacent quadrats, so that potentially important data are not excluded from the attributes on any individual quadrat. If the quadrat square polygon was only used to extract data from the wetland polygons, then there would be little difference to the arbitrary boundaries such as SLAs. It is important to consider the location of the grid in relation to the Queensland coastline. In order to assess this, the distance from the centre of the grid to the Queensland coastline was calculated. The quadrats within 100km from coastline are shown in Figure 2. Tidal influenced wetlands were observed to extend up to 50km inland from the coast. Since more BFV cases are coastal and most of the Queensland population lies near the coast, it was decided to limit this study to coastal areas. As tidal influences are specific interest it was decided to extend the study to 100km so that the areas without tidal influence could be compared with the tidal influences. Selecting these quadrats for the entire Queensland coastline removes potential discrepancies in the results if only a single city area was analysed. This also complies with statement by Russell et al (2009) that BFV distribution was highly concentrated along coastal lines in Australia. In this way, the database and its consequent analysis should provide the ability to produce robust risk assessment methods and potentially determine the effects of environmental and climatic changes on BFV disease transmission in Queensland.

2.4 Software

Vertical Mapper (v 3.7), a GIS tool was used to transfer all the data into the grid data. Vertical Mapper was incorporated into MapInfo Professional (v 10.5), which was then used as a platform to perform the data link, data transfer, data integration, data geocoding and spatial display. Spatial data quality was maintained throughout the database processing.

3. Final Database and Results

Based on the above spatial sampling design, the resultant quadrat database includes 14,000 grids. We aim to use this grid-database in our future study to determine the spatial risk factors for the spatial distribution of BFV disease in Queensland, Australia by employing spatial regression modelling methods.

4. Conclusions

This paper has discussed GIS and spatial methodology intended to be used in future spatial analytical studies on BFV disease transmission in Queensland, Australia.

Acknowledgements

The authors would like to acknowledge data providers and financial supporters.

References

Australian Bureau of Meteorology, 2009, Data Obtained from National Climate Centre, Queensland, Commonwealth of Australia. Retrieved September 20, 2009, from www.bom.gov.au/climate/

Australian Bureau of Statistics, 2009, Census of Population and Housing - Basic community profiles, Commonwealth of Australia, Canberra, ABS. Retrieved from wwwabscomau

Daniel, M., Kolar, J. and Zeman, P., 2004. GIS Tools for Tick and Tick-Borne Disease Occurences. *Parasitology, 129*, S329-S352.

Fotheringham, A. S., and Rogerson, P. A., 2009, *The Sage Handbook of Spatial Analysis*. London: Sage Publications.

Jerrett, M., Burnett, R., Goldberg, M. S., Sears, M., Krewski, D. and Catalan, R., 2003, Spatial Analysis for Environmental Health Research: Concepts, Methods And Examples. *Journal of Toxicology and Environmental Health, 66*, 1783-1810.

Kitron, U., 1996, Landscape Ecology and Epidemiology of Vector-Borne Diseases: Tools for Spatial Analysis. *Journal of Medical Entomology, 35*, 435-445.

Lloyd, C., 2010, *Spatial Data Analysis - An Introduction for GIS Users*. Oxford: Oxford University Press.

Moore, D. A. and Carpenter, T. E., 1999, Spatial Analytical Methods and Geographic Information Systems: use in Health Research And Epidemiology. *Epidemiologic Reviews, 21*, 143-161.

Muhar, A., Thalib, L., Dale, P. and Arito, E., 2000, The Spatial Distribution of Ross River Virus Infections in Brisbane: Significance of Residential Location and Relationships with Vegetation Types. *Environmental Health and Preventive Medicine, 4*, 184-189.

Naish, S., Hu, W., Mengersen, K. and Tong, S., 2011, Spatial and Temporal Clusters of Barmah Forest Virus Disease in Queensland, Australia. *Tropical Medicine and International Health, 16*, 884-893.

Naish, S., Hu, W., Mengersen, K. and Tong, S., 2011, Spatio-Temporal Patterns of Barmah Forest Virus Disease in Queensland, Australia. *PLoS ONE, 6*, e25688.

Naish, S., Hu, W., Nicholls, N., Mackenzie, J. S., Dale, P. and McMichael, A. J., 2009, Socio-Environmental Predictors of Barmah Forest Virus Transmission in Coastal Areas, Queensland, Australia. *Tropical Medicine and International Health, 14*, 247-256.

Naish, S., Hu, W., Nicholls, N., Mackenzie, J. S., McMichael, A. J. and Dale, P., 2006, Weather Variability, Tides, and Barmah Forest Virus Disease in the Gladstone Region, Australia. *Environmental Health Perspectives, 114*, 678-683.

Olson, S. H., Gangnon, R., Elguero, E., Durieux, L., Foley, J. A. and Patz, J., 2009, Links Between Climate, Malaria, and Wetlands in the Amazon Basin. *Emerg Infect Dis, 15*, 659-662.

Queensland Department of Health, 2008, National Notifiable Diseases Surveillance System, Communicable Diseaes Australia, Department of Health and Ageing, Australian Government. Retrieved September 26, 2004, from http://www.health.gov.au/internet/main/Publishing.nsf/Content/cda-cdna-index.htm

Queensland Department of Transport and Main Roads, 2009, Maritime Safety Queensland:Tides. Queensland official tidal tables and boating safety guide: tidal notes and definitions. Retrieved September 24, 2004, from www.msq.qld.gov.au

Queensland Environmental Protection Agency, 2009, Queensland Wetlands Data, Queensland Wetland Program. Retrieved November 29, 2009, from www.wetlands@epa.qld.gov.au

Quinn, H. E., Gatton, M. L., Hall, G., Young, M. and Ryan, P. A. 2005, Analysis of Barmah Forest Virus Disease Activity in Queensland, Australia, 1993-2003: Identification of a Large, Isolated Outbreak of Disease. *Journal of Medical Entomology, 42* 882-890.

Rushton, G., 1998, Improving the Geographic Basis of Health Surveillance using GIS. In A. C. Gatrell and M. Loytonen (Eds.), *GIS and Health* (63-79). London: Taylor and Francis.

Russell, R. C., 1994, Ross River Virus: Disease Trends and Vector Ecology in Australia. *Bulletin of Social and Vector Ecology, 19*, 73-81.

Russell, R. C., 2009, Mosquito-Borne Disease and Climate Change in Australia: Time for a Reality Check. *Australian Journal of Entomology, 48*, 1-7.

Thomson, M. C. and Connor, S. J., 2000, Environmental Information Systems for the control of Arthropod Vectors of Disease. *Medical and Veterinary Entomology, 14*, 227-244.

Tong, S., Hayes, J. F. and Dale, P., 2005, Spatiotemporal Variation of Notified Barmah Forest Virus Infections in Queensland, Australia, 1993-2001. *International Journal of Environmental Health Research, 15*, 89-98.

Whelan, P. I., 1997, *Problem Mosquito Species in the Top End of the NT Pest and Vector Status Habitats and Breeding Sites Northern Territory*: Department of Health and Community Services.

Wniters, A., Eisen, R. J., Fuentes, S. L., Moore, C. G., Pape, W. J. and Eisen, L., 2008, Predicitve Spatial Models for Risk of West Nile Virus Exposure in Eastern and Western Colorado. *Am J Trop Med Hyg*, 79.

Wu, P. C., Lay, J. G., Guo, H. R., Lin, C. Y., Lung, S. C. and Su, H. J, 2009, Higher Temperature and Urbanization Affect the Spatial Patterns of Dengue Fever Transmission in Subtropical Taiwan. *The Science of The Total Environment,* 407, 2224-2233.