

SPATIAL AND TEMPORAL MODELLING FOR PARASITE TRANSMISSION STUDIES AND RISK ASSESSMENT

DANSON F.M.*, ARMITAGE R.P.* & MARSTON C.G.*

Summary:

Spatial and temporal modelling of parasite transmission and risk assessment require relevant spatial information at appropriate spatial and temporal scales. There is now a large literature that demonstrates the utility of satellite remote sensing and spatial modelling within geographical information systems (GIS) and firmly establishes these technologies as the key tools for spatial epidemiology. This review outlines the strength of satellite remotely sensed data for spatial mapping of landscape characteristics in relation to disease reservoirs, host distributions and human disease. It is suggested that current satellite technology can fulfill the spatial mapping needs of disease transmission and risk modelling, but that temporal resolution, which is a function of the satellite data acquisition characteristics, may be a limiting factor for applications requiring information about landscape or ecosystem dynamics. The potential of the Modis sensor for spatial epidemiology is illustrated with reference to mapping spatial and temporal vegetation dynamics and small mammal parasite hosts on the Tibetan plateau. Future research directions and priorities for landscape epidemiology are considered.

KEY WORDS : remote sensing, landscape, epidemiology, modelling.

It is widely acknowledged that the processes determining the transmission risk of parasites between vectors and host populations operate at a range of spatial and temporal scales. Many of the pertinent features of transmission studies are captured by the term 'spatial epidemiology' which aims to determine the spatial nature of the factors that drive disease incidence, prevalence or risk. Spatial epidemiology builds on the fundamental principle that transmission risk will, in general terms, decrease with distance from an infected host (Ostfeld *et al.*, 2005). For example, the distance of villages from mosquito breeding grounds have been shown to affect local scale malaria prevalence (Thomson *et al.*, 1999) and the proximity to forest and pasture land cover is a risk factor for visceral leishmaniasis (Werneck *et al.*, 2002).

The spatial epidemiology of many zoonoses shows close links between human populations, vector distributions and landscape, and the term 'landscape epi-

miology', first coined by Pavlovsky (1966), indicates the importance that the composition and physical characteristics of the landscape may play in determining transmission and risk. Landscape epidemiology requires data to develop and test models of transmission risk and these data are often now derived from satellite remote sensing. Herbreteau *et al.* (2006) found 86 research articles using remote sensing data for spatial epidemiological studies of human infectious disease, published between 1975 and 2004, and found that 73 % of studies related to parasitic diseases, including those transmitted by mosquitoes or ticks. When Lyme disease, Rift valley fever and Hantavirus Pulmonary syndrome were included this proportion increased to 88 %. Most of these studies used NOAA Advanced Very High Resolution Radiometer (AVHRR) or Landsat data and the spatial variable mapped was either the Normalized Difference Vegetation Index (NDVI), or land cover or land use. The aim of this review is to provide an overview of recent developments in spatial and temporal modeling of parasite vector distribution and disease incidence, with a focus on spatial and temporal information that may be extracted from remotely sensed data to provide landscape information to drive landscape epidemiology models.

REMOTE SENSING AND LANDSCAPE EPIDEMIOLOGY

Comprehensive reviews of the use of remote sensing and geographical information systems (GIS) for epidemiological applications are provided in Thomson & Connor (2000), Hay *et al.* (2000), Curran *et al.* (1998) and Ostfeld *et al.* (2005). Ostfeld *et al.* (2005) usefully sub-divided examples of spatial epidemiology using remote sensing into those which attempted to map vector distributions, those that mapped pathogen reservoirs, and those that mapped human cases. Furthermore they identified local, regional and continental scales of study, all making use of data from satellite imagery.

The two approaches used for mapping disease transmission are correlation of host distributions with cli-

* Centre for Environmental Systems Research, School of Environment and Life Sciences, University of Salford, Manchester M5 4WT, UK. Correspondence: F.M. Danson. Tel.: +44 161 295 4038 – Fax: +44 161 295 5015. E-mail: f.m.danson@salford.ac.uk

matic data, or correlations with landscape. At regional scales correlation with climatic data from the AVHRR sensor have been successfully used to predict the spatial distribution of tsetse, African trypanosomiasis and malaria (Rogers, 2000; Hay *et al.*, 1998). Such models depend on limiting environmental factors determining the survival of pathogens, although these limits may not be known *a priori*. In contrast, local to regional studies classify habitat areas for disease vectors, such as the mosquito (*Anopheles* sp.), which provides a transmission pathway for malaria (Wood *et al.*, 1992) and tsetse flies (*Glossina* sp.), which are linked to trypanosomiasis (Kitron *et al.*, 1996; Rogers, 2000). Landscape-related methods have also been used to map encephalitis (Randolph, 2000), tick habitats in relation to Lyme disease (Nicholson & Mather, 1996) and dengue transmission risk maps (Carbajo *et al.*, 2001). *Loa loa* (Cobbold, 1864) habitats have been mapped to explore the epidemiology of eye-worm river blindness – onchocerciasis (Thomson *et al.*, 2000). Spatial models for snail vector-borne diseases of human fasciolosis transmission in the Andes have used satellite-derived vegetation indices (Fuentes *et al.*, 2001) in addition to spatial models of schistosomiasis distribution in sub-Saharan Africa (Brooker, 2002).

SATELLITE IMAGERY FOR LANDSCAPE EPIDEMIOLOGY

The primary source of landscape maps for epidemiology to date has been data from the Multispectral Scanning System and Thematic Mapper instruments on board the Landsat series of satellites. Landsat 1 was launched in 1972, and was followed by a series of similar satellites with the most recent, Landsat 7, carrying the Enhanced Thematic Mapper instrument, launched in 1999. A unique and important feature of data from the Landsat system is the 35 year global archive of spatially consistent data at resolutions of either 80 m (Landsat 1-3) or 30 m (Landsat 4, 5, 7). This data set represents the only globally comprehensive (apart from the poles) record of land surface changes over the past 35 years and represents an irreplaceable resource for epidemiology where changes in land cover or land use in the past are important. The characteristics of the Landsat ETM sensor that make it particularly suitable for landscape epidemiology are wide area coverage (185 × 185 km), medium spatial resolution (15-30 m), wide spectral coverage (visible, shortwave infrared and thermal infrared wavelengths) and repeat coverage (16 days). In fact in general it is the spatial, temporal and spectral characteristics of satellite sensors that together determine their usefulness for a particular application. Spatial resolution repre-

sents, in general terms, the amount of spatial detail that may be extracted from an image, and it may be argued that for epidemiological modeling the spatial resolution of remotely sensed data is no longer a limiting factor since data may be obtained for any part of the world at spatial resolution between 60 cm and 1.1 km. Temporal resolution describes the minimum time interval between regularly imaging the same point on the Earth's surface. Some satellite sensors, like the Landsat ETM have a fixed view angle so that the satellite repeat cycle is the same as the satellite imaging revisit time. Other sensors, like SPOT HRVIR and Ikonos have an off-nadir viewing capability that means that they may be programmed to point to locations to the side, front, or behind the sub-satellite position. This means that the satellite imaging revisit time may be increased for a given point on the Earth but, by virtue of this fact, the revisit times for other locations that may be imaged from the same point are decreased. Latitude also controls satellite imaging revisit time since the overlap of image swaths increases with latitude, with minimum overlap and longest revisit time at the equator. There is widespread misunderstanding of the nature of satellite imaging revisit time for the new generation of very fine spatial resolution sensors. The Ikonos sensor, for example has a nominal satellite repeat cycle of 140 days, but exploiting the pointing capability of the instrument the satellite imaging revisit time may be increased to 3-4 days. Crucially, imagery are only recorded when there is user demand so that there is no regular repeated coverage for any part of the Earth's surface.

There is a trade-off between spatial and temporal resolution of satellite data (Fig. 1) with finer spatial resolution data available less frequently than coarser spatial resolution data. The importance of these concepts for disease transmission studies depends on the fre-

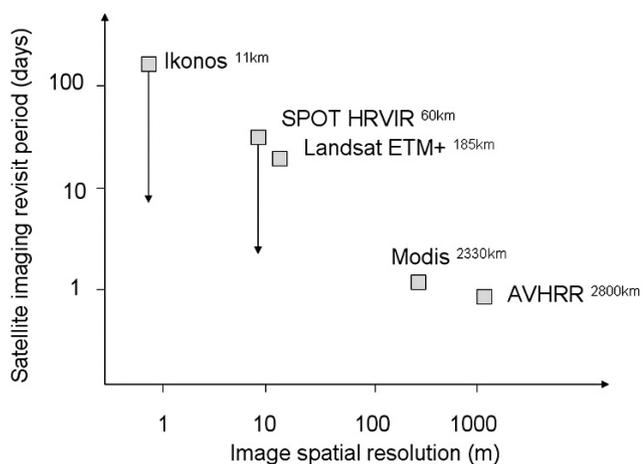


Fig. 1. – Relationship between image spatial resolution and satellite imaging revisit period for selected sensors relevant to spatial epidemiology. Superscripts indicate image swath width and arrows show range of satellite imaging revisit period using sensor off-nadir pointing capability.

quency with which spatial information on landscape properties is required. Transmission models based on land cover change may require data every few years; studies requiring data on the seasonal dynamics of landscape may require data on a monthly basis, or more frequently if the study area is frequently cloud covered. The relevant spectral characteristics of remotely sensed data are the number of spectral wavebands measured and the spectral dispersion and width of individual wavebands. Larger numbers of wavebands potentially allow better discrimination of objects on the Earth's surface, based on their spectral reflectance "signature". Most satellite based remote sensing systems collect data in less than ten wavebands, although a few now provide data in 30 or more wavebands. The review of Ostfeld *et al.* (2005) indicates that most landscape epidemiology studies using remotely sensed data use simple image classification or vegetation indices to build empirical transmission models. The full range of spectral information in remotely sensed imagery is rarely used. A new generation of medium spatial resolution satellite sensors, notably Terra and Aqua Modis, and ERS Meris, now provide high temporal resolution in a larger number of wavebands than the Landsat sensors (Tatem *et al.*, 2004; Gobron *et al.*, 1999). In addition to more spectral information the wide swath width of these sensors means that the satellite imaging revisit time is significantly greater, with Modis imaging areas on a daily basis, with a spatial resolution of 250 m, north and south of 30 degree latitude. The implications for landscape epidemiology is that, for the first time, local to regional scale landscape information may be acquired frequently enough to study seasonal changes in vegetation characteristics which may in turn be related to vector or host population dynamics

A further advantage of the Modis sensor in particular is the range of higher level data products that are derived from the data and made freely available to the scientific community. It is now possible for users to download atmospherically, radiometrically and geometrically corrected reflectance data and a range of derived products like vegetation index images, leaf area index images or land cover maps. In this way researchers may choose to directly use these products in landscape epidemiology investigations without a detailed knowledge of the algorithms used to derive the products.

MODELLING APPROACHES

The most common approach to build transmission and risk models is to measure the spatial distribution of vectors, pathogen reservoirs or human cases and relate these distributions to a range of environmental variables that are hypothesized to

cause the observed spatial distributions. Where landscape factors are thought to be relevant it is most commonly land cover maps, derived from remotely sensed data, that are used as the explanatory independent variable. Land cover maps may be used to directly define the spatial distribution of vectors or reservoirs, or in the case of human disease, landscape may be used as a proxy for the location of vectors or pathogen reservoirs to determine spatial relationships with infected individuals or disease foci. Most landscape epidemiology applications use statistical methods to search for relationships between vectors, reservoirs or human cases and given the binary nature of the observations (presence or absence) logistic regression methods are widely applied. Such methods may be used to explore the explanatory power of different environmental variables in order to develop robust statistical models which may be spatially extrapolated to new areas, based on measuring the same environmental variables. The key limitation of this approach to build parasite transmission models is that risk is likely to more closely related to parasite abundance than the presence or absence of a host. Furthermore, statistical models do not necessarily provide ecological insights into the transmission mechanisms at work and of the 29 papers cited by Ostfeld *et al.* (2005) which used spatial data to predict disease, thirteen were found not to identify the disease transmission mechanism.

Most studies using land cover maps for transmission and risk mapping employ pixel-based classification where each pixel is uniquely assigned to one land cover class. A development of this approach is to derive landscape metrics from the pixel-based classification data in order to better describe the spatial composition and arrangement of landscape elements (Graham *et al.*, 2004a). This approach provides a potentially richer description of the vector or host habitats, but the results of statistical models based on correlations with landscape metrics are complex, and ecological interpretation often very difficult. A further alternative is to employ a range of "soft-classification" techniques where each pixel has a probability of membership to each of the classes defined. Such approaches may find application in landscape epidemiology models in the future, but they are most likely to be successful when underpinned by good ecological understanding of the transmission mechanisms.

Statistically-based transmission and risk models must also account for spatial autocorrelation and stationarity in the dependent variable (host or disease data). Autocorrelation occurs because observations that are close in space are more likely to be similar than those that are further apart (Legendre, 1993; Graham *et al.*, 2004b; Osborne *et al.*, 2007). When spatial autocorrelation is present and unaccounted for in the data, estimates of model fits may be optimistic, although it is

currently very difficult to account for spatial autocorrelation when building such models. Stationarity refers to variation in the modelled relationships over space and may be problematic if locally derived models are extrapolated beyond the area where they were developed. Ideally spatial autocorrelation will be accounted for through optimal sampling of host or infected humans. However the widely observed overdispersion of parasites in a few hosts, and spatial clusters of human disease, may be expected to lead to spatial autocorrelation in such data. It appears to be a paradox of spatial epidemiology that transmission risk decreases with distance from an infected host, but in spatial sampling of infected hosts, samples that are closer together will be spatially autocorrelated.

A range of alternative modelling techniques are now being developed to determine relationships between wildlife distribution and environmental data including geographically weighted regression, variable coefficient modelling, multiple criteria evaluation (MCE) and Dempster-Shafer (D-S) analysis, a variant of Bayesian analysis (Brunsdon *et al.*, 1998; Hastie & Tibshirani, 2005; Osborne *et al.*, 2007). Some of these methods may be used to develop spatial model of disease transmission but have yet to be tested in this context.

True spatial-temporal modelling of vectors, hosts or human disease is rare for most zoonoses. The main obstacle to the development of spatio-temporal transmission models to date has been a lack of data on the temporal dynamics of hosts, reservoirs and disease, coupled with a lack of local scale environmental data at appropriate temporal resolution. The development of medium resolution sensors like Modis may lead to the development of more local scale spatial-temporal transmission models but to date there has been little progress. The importance of understanding the relationships between climate change and disease transmission at both local and regional scales points to the need to develop such methods. The following section presents some early results of our work relating spatial-temporal landscape dynamics at local scale to the distribution of rodent hosts for a zoonotic parasite.

SPATIAL AND TEMPORAL CHARACTERIZATION OF LANDSCAPE FOR EPIDEMIOLOGY

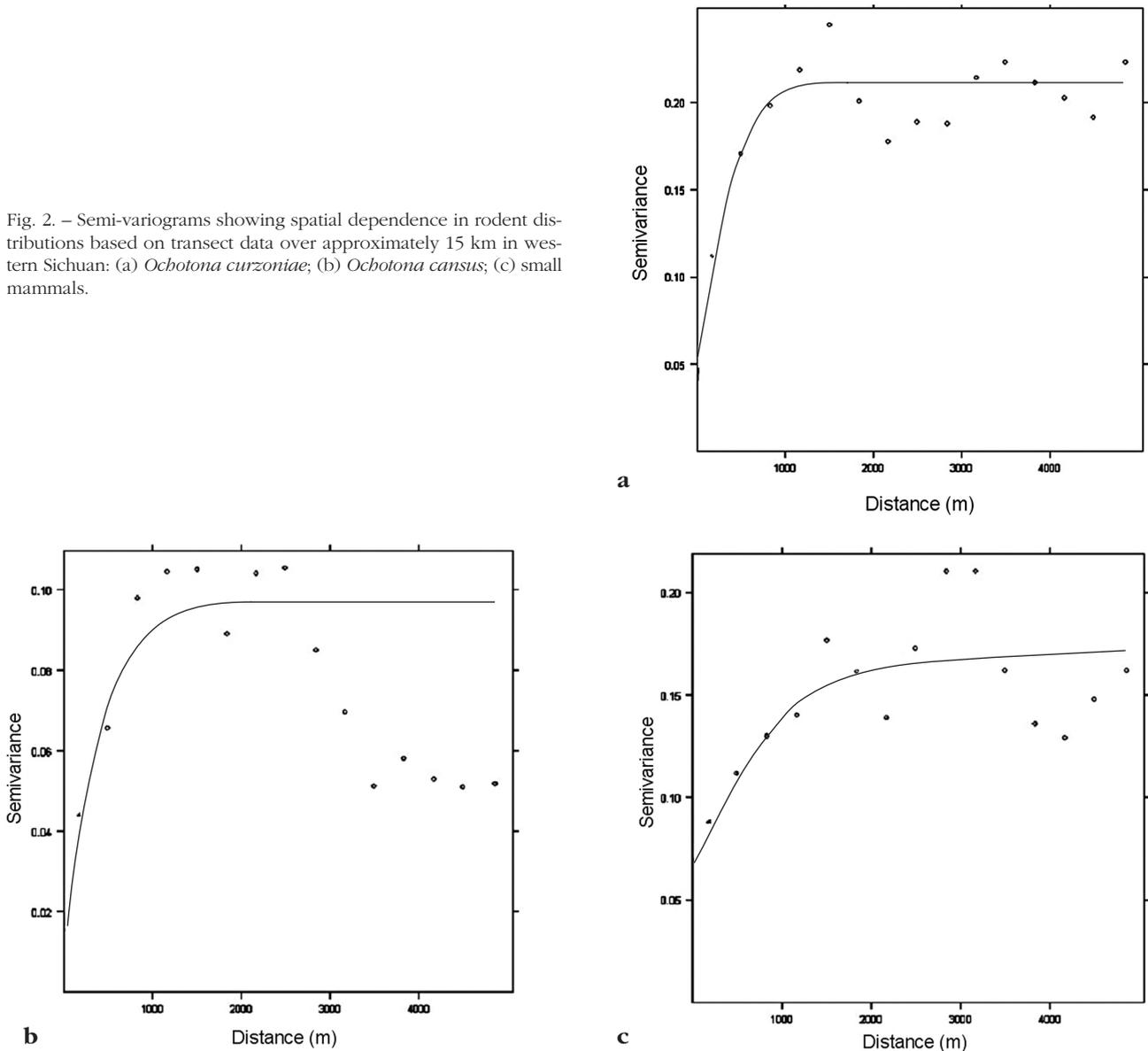
The fox tapeworm *Echinococcus multilocularis* (Em) has a wildlife cycle involving a definitive canid host (fox, dog, wolf) and a small mammal intermediate host. Several regions in central and western China have been identified as important foci for the human disease alveolar echinococcus, caused by infection with the larval stage of the Em tapeworm (Craig *et al.*, 2000; Graham *et al.*, 2005). Research efforts in

the landscape epidemiology of Em transmission have successfully related human AE prevalence to landscape around villages in south Gansu Province (Danson *et al.*, 2004) and are now focusing on relating small mammal distributions to landscape characteristics in endemic areas in Serxu County, western Sichuan, China. Here, rodent transects 35 km in length and covering a range of landscape types were surveyed in summer 2001, with presence or absence of three species or groups of rodents recorded at 10 m intervals (3,485 transect points). Visual sightings of rodents, presence of rodent faeces or rodent holes were used to determine rodent presence. These rodent groups were *Ochotona curzoniae* or black-lipped pika (OC), *Ochotona cansus* or gansu pika (OA), and a group of smaller rodents including *Arvicola terrestris* and *Microtus arvalis* (SM). Shuttle Radar Topography Mission Digital Elevation Model data was acquired to derive topographical variables such as altitude, slope and aspect for all transect points. 138 Modis 16-day composite vegetation index data at 250 m spatial resolution were acquired for the period 6th April 2000 to 24th May 2006. The rodent index data were related to a range of environmental variables using the time-series Modis data using a generalised additive model

These images were stacked in acquisition date order enabling for any location within the extent of the Modis image a temporal signature of vegetation seasonal growth and senescence cycle to be acquired. As the Modis 16-day composite periods are the same for each year of image collection, a median value of the composite images for each composite period was taken and used to collapse the six years of data into a single annual vegetation cycle profile. This profile was then quantified in terms of maximum, minimum and mean NDVI throughout the year, NDVI range, time of greening period and length of growing season. These variables were then related to rodent distribution using a Generalised Additive Model.

As the seasonal dynamics of a vegetation cycle can be indicative of specific habitat types, modelling of these spatio-temporal vegetation dynamics can also be useful in the determination of the spatial distribution of the Em transmission vector hosts that are dependent on those specific habitats. The results of this analysis display this, with 41.4 % deviance explained when relating OC to the quantified vegetation profile, 37.6 % for OA, and also 34.7 % deviance explained for SM which has a different set of habitat preferences. The need for caution is illustrated by calculating spatial dependence in the rodent index data. The results (Fig. 2) show that the GAM models are likely to be over optimistic estimates of the explained deviance since the rodent presence/absence data show spatial dependence up to a lag of approximately 1,000 m. This sort of information should inform optimal sampling schemes to capture

Fig. 2. – Semi-variograms showing spatial dependence in rodent distributions based on transect data over approximately 15 km in western Sichuan: (a) *Ochotona curzoniae*; (b) *Ochotona cansus*; (c) small mammals.



spatial variation in host distributions whilst avoiding the dilemma of spatial autocorrelation. The early potential of using time-series Modis vegetation index data shown here would no doubt be improved with further development and wider application.

FUTURE PROSPECTS

This review has outlined the prospects for the development of spatial and temporal transmission models where landscape characteristics are hypothesized to affect the distribution of hosts, parasites or human cases. Some significant advances have been made to relate regional scale climate and land cover data to disease transmission dynamics and the potential effects of climate change drive the need for fur-

ther developments. At local to regional scale this review suggests that progress is being made for some zoonoses where the transmission mechanism are well understood, but that these are the exception rather than the rule. Further progress towards the development of true spatio-temporal transmission models requires collaborative research involving multi-disciplinary teams to build process based transmission models that are accurate, robust and useful for disease mapping and control.

ACKNOWLEDGEMENTS

Financial support from the US National Institutes of Health from the Fogarty International Center (EID TW001565-05). The content is solely the res-

possibility of the authors and does not necessarily represent the official views of the Fogarty International Center or the National Institutes of Health.

REFERENCES

- BROOKER S. Schistosomes, snails and satellites. *Acta Tropica*, 2002, 82, 207-214.
- BRUNSDON C., FOTHERINGHAM S. & CHARLTON M. Geographically weighted regression - modelling spatial non-stationarity. *The Statistician*, 1998, 47, 431-443.
- CARBAJO A.E., SCHEIGMANN N., CURTO S.I., DE GARIN A. & BEJARAN R. Dengue transmission risk maps of Argentina. *Tropical Medicine and International Health*, 2001, 6, 170-183.
- CRAIG P.S., GIRAUDOUX P., SHI D., BARTHOLOMOT B., BARNISH G., DELATTRE P., QUERE J.P., HARRAGA S., BAO G., WANG Y., LU F. & VUITTON D.A. An epidemiological and ecological study of human alveolar echinococcosis transmission in South Gansu, China. *Acta Tropica*, 2000, 77, 167-177.
- CURRAN P.J., ATKINSON P.M., FOODY G.M. & MILTON E.J. Linking remote sensing, land cover and disease. *Advances in Parasitology*, 2000, 47, 37-80.
- DANSON F.M., CRAIG P.S., MAN W., SHI D. & GIRAUDOUX P. Landscape dynamics and risk modelling of human alveolar echinococcosis. *Photogrammetric Engineering and Remote Sensing*, 2004, 70, 359-166.
- FUENTES M.V., MALONE J.B. & MAS-COMA S. Validation of a mapping and prediction model for human fasciolosis transmission in Andean very high altitude endemic areas using remote sensing data. *Acta Tropica*, 2001, 79, 87-95.
- GOBRON N., PINTY B., VERSTRAETE M. & GOVAERTS Y. The MERIS Global Vegetation Index (MGVI): description and preliminary application. *International Journal of Remote Sensing*, 1999, 20, 1917-1927.
- GRAHAM A.J., DANSON F.M., GIRAUDOUX P. & CRAIG P.S. Ecological epidemiology: landscape metrics and human alveolar echinococcosis. *Acta Tropica*, 2004a, 91, 267-278.
- GRAHAM A.J., ATKINSON P.M. & DANSON F.M. Spatial analysis for epidemiology. *Acta Tropica*, 2004b, 91, 219-225.
- GRAHAM A.J., DANSON F.M. & CRAIG P.S. Ecological epidemiology: the role of landscape structure in the transmission risk of the fox tapeworm *Echinococcus multilocularis* (Leukart 1863) (Cestoda: Cyclophyllidae: Taeniidae). *Progress in Physical Geography*, 2005, 29, 77-91.
- HASTIE T. & TIBSHIRANI R. Varying-coefficient models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 1993, 55, 757-796.
- HAY S.I., SNOW R.W. & ROGERS D.J. Prediction of malaria seasons in Kenya using multi-temporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 1998, 92, 12-20.
- HAY S.I., RANDOLPH S.E. & ROGERS D.J. Remote sensing and geographical information systems in epidemiology. *Advances in Parasitology*, 2000, 47.
- HERBRETEAU V., SALEM G., SOURIS M., HUGOT J.P. & GONZALEZ J.P. Thirty years of use and improvement of remote sensing, applied to epidemiology: from early promises to lasting frustration. *Health and Place*, 2006, 13 (2), 400-403.
- KITRON U., OTIENO L.H., HUNGERFORD L.L., ODULOJA A., BRIFHAM W.U., OKELLO O.O., JOSELYN M., MOHAMED AHMED M.M. & COOK E. Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *Journal of Animal Ecology*, 1996, 65, 371-380.
- LEGENDRE P. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 1993, 74, 1615-1673.
- NICHOLSON M.C. & MATHER T.N. Methods for evaluating Lyme disease risks using geographic information systems and geospatial analysis. *Journal of Medical Entomology*, 1996, 33, 711-720.
- OSBORNE S.E., FOODY G.M. & SUÁREZ-SEOANE S. Non-stationarity and local approaches to modelling the distributions of wildlife. *Diversity and Distributions*, 2007, 13, 313-323.
- OSTFELD R.S., GLASS G.E. & KEESING F. Spatial epidemiology: an emerging (or re-emerging) discipline. *TRENDS in Ecology and Evolution*, 2005, 20 (6), 328-336.
- PAVLOVSKY E.N. Natural nidality of transmissible diseases: with special reference to the landscape epidemiology of zoonoses. *The Challenge of Epidemiology: Issues and Selected Readings*, 1966, 401-405.
- RANDOLPH S.E. Ticks and tick-borne disease systems in space and from space. *Advances in Parasitology*, 2000, 47, 217-243.
- ROGERS D.J. Satellites, space, time and the African trypanosomiasis. *Advances in Parasitology*, 2000, 47, 129-171.
- TATEM A.J., GOETZ S.J. & HAY S.I. Terra and aqua: new data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation*, 2004, 6, 33-46.
- THOMSON M.C. *et al.* Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. *American Journal of Tropical Medicine and Hygiene*, 1999, 61, 2-8.
- THOMSON M.C. & CONNOR S.J. Environmental information systems for the control of arthropod vectors of disease. *Medical and Veterinary Entomology*, 2000, 14, 227-244.
- THOMSON M.C., OBSOMER V., DUNNE M., CONNOR S.J. & MOLYNEUX D.H. Satellite mapping of *Loa loa* prevalence in relation to ivermectin use in west and central Africa. *The Lancet*, 2000, 356, 1077-1078.
- WERNECK G.L. *et al.* The urban spread of visceral leishmaniasis: clues from spatial analysis. *Epidemiology*, 2002, 13, 364-369.
- WOOD B.L., BECK L.R., WASHINO R.K., HIBBARD K.A. & SALUTE J.S. Estimating high mosquito producing rice fields using spectral and spatial data. *International Journal of Remote Sensing*, 1992, 13, 2813-2826.