SPATIO-TEMPORAL MODELLING OF SMALL MAMMAL DISTRIBUTIONS USING MODIS NDVI TIME-SERIES DATA

C G Marston[a], R P Armitage[a], F M Danson[a], P Giraudoux[b], A Ramirez[a] and P S Craig[a]

[a] School of Environment and Life Sciences, University of Salford, Salford, M5 4WT, UK
[b] University of Franche-Comte, Besancon, France
c.g.marston@pgr.salford.ac.uk

KEY WORDS: Echinococcus, Rodents, Landscape, MODIS, Time-series, Spatio-temporal

ABSTRACT:

This work modelled the spatial distribution of the rodent species that act as hosts in the transmission cycle of the parasitic tapeworm *Echinococcus multilocularis*. The rodent distribution was modelled in relation to landscape characteristics in four ways, using (1) a Landsat ETM+ derived hard classification, (2) single-image Landsat ETM+ derived NDVI, (3) single-image MODIS 16-day composite NDVI, and (4) time-series MODIS 16-day composite NDVI imagery. The MODIS time-series imagery produced the strongest relationships and explained the highest percentage deviance of the relationships present (up to 41.4%), whereas the hard classification method only explained up to 21.2% of deviance. Single-image NDVI datasets produced poor results, with Landsat ETM+ derived NDVI explaining only up to 11.9% of deviance, and MODIS derived NDVI up to only 8.78%. These results confirm that using time-series NDVI data to model rodent distributions is a valid method, and can offer improved results over single date NDVI and hard classification methods.

1. INTRODUCTION AND BACKGROUND

The fox tapeworm *Echinococcus multilocularis* (*Em*) is one of the World's most dangerous human parasites causing fatalities in 95% of infected patients. The tapeworm exists in a transmission cycle between small mammals and foxes, and the parasite is increasingly infecting domestic dogs that predate infected rodents, leading to increased transmission to dog owners via dog faeces contaminated with *Em* eggs. In central China, where the parasite is endemic, there is currently an urgent need to develop spatial models to predict the location and extent of *Em* transmission foci so that limited medical resources can be effectively targeted and appropriate control programmes implemented.

This study aims to assess the relationship between small mammal distributions and spatio-temporal landscape variables, and develop a model to accurately predict the spatial distribution of locations where the parasite hosts are likely to be present. Field observations have indicated that there are spatial relationships between heavily grazed areas with low biomass that are the optimal habitat of rodent species (Ochotona sp.) known to be involved in the *Em* transmission cycle, and small mammal distributions in our study region in Shiqu county, western Sichuan Province, China.

Knowledge of land cover and land cover change is an important input in modelling ecological processes from the regional to the global scale (Schwarz *et al*, 2005), and Geographical Information Systems now facilitate the incorporation of these spatio-temporal land-cover relationships into epidemiological investigations of wildlife diseases (Pfeiffer *et al*, 2002). Several previous investigations have related satellite-derived land cover and Digital Elevation Model data to rodent populations that act as disease transmission vectors of Hantavirus, with studies by Goodin *et al* (2006), Suzan *et al* (2006) and Glass *et al* (2000) showing that direct relationships exist between landscape structure and topographical variables, and the distribution of rodents. Studies by Boone *et al* (2000) also successfully utilised satellite-derived vegetation and topographical data to predict the infection status of Deer mice with the Sin Nombre Virus with up to 80% accuracy in the USA.

These successful studies show that modelling rodent population distributions through landscape characteristics is a viable methodology, however, previous work using remotely sensed data to assess landscape-rodent relationships have concentrated on hard classification methods using single date medium spatial resolution imagery. Limitations result from attempting to classify continuous variables such as vegetation gradients into discrete classes. This may result in information on sub-class variation being lost, when this variation may be integral in the observed relationships. Continuous measurements such as a vegetation index, for example the Normalised Difference Vegetation Index (NDVI), the most widely used vegetation index for retrieval of vegetation canopy biophysical characteristics (Jiang *et al*, 2006), may be better suited to define, characterise and quantify landscape characteristics. This NDVI data could then be modelled with rodent presence; as rodent distribution is often related to the spatial distribution of areas of differing vegetation characteristics, this could be a viable alternative to using a hard classification method.

Furthermore, as vegetation experiences a seasonal growth and senescence cycle, resulting in seasonal landscape change, time-series NDVI data offers the potential to better quantify the spatio-temporal characteristics of this seasonal vegetation cycle rather than measuring NDVI at only one point in time. Time-series NDVI data has previously been used successfully for various applications, including quantifying vegetation cycles, assessing land cover change and to map the spatial distribution of habitats by Beck *et al* (2006), Xiao *et al* (2006), Lunetta *et al* 2006) and Jin *et al* (2005), and could offer significant advantages over single-date vegetation index data.

The research described in this paper develops a multi-temporal model of landscape change over large areas using time-series MODIS NDVI 16-day composite 250m resolution imagery. The main advantage of these datasets is that the regular repeat coverage of the MODIS NDVI data product allows assessment of temporal variability in the landscape which is expected to better describe small mammal distributions.

2. STUDY SITE AND DATA COLLECTION

The study area for this investigation is located near the town of Serxu, Shiqu county, Sichuan Province, China (Figure 1). This site is located on the eastern Tibetan plateau at high altitude between 4000 and 4300m above sea level. Although this area is above the tree line, variation in the herb and shrub layers of vegetation produces a variety of habitats across the study area.

Rodent transects totalling approximately 35km in total length were surveyed in summer 2001, with presence or absence of three species and groups of rodents being recorded at 10 metre intervals, giving a total of 3485 transect points. Visual sightings of rodents, presence of rodent droppings or rodent holes (both identifiable to species or group of species level) were used to determine rodent presence. These rodent groups were *Ochotona curzoniae* or black-lipped pika (Occu), *Ochotona cansus* or gansu pika (Occa), and a generically described group of rodents consisting of *Arvicola terrestris* and *Microtus arvalis* (Smallsm). Each of these has different habitat requirements, and will therefore exhibit different relationships with the spatial landscape and habitat arrangement.

Map and habitat data was not available for this area, therefore it was necessary to generate the required datasets. Shuttle Radar Topography Mission Digital Elevation Model data was acquired and used to produce a 3D topographical model of the study area, enabling topographical variables such as altitude, slope and aspect to be calculated. A Landsat ETM+ image (acquisition date 3rd July 2001) was also obtained and subjected to a supervised classification using training areas of known habitats visited during fieldwork. This Landsat ETM+ image was atmospherically corrected and used to derive a NDVI dataset for the study area. Also, multiple MODIS 16-day composite vegetation index data at 250m spatial resolution were acquired for the period between 6th April 2000 and 24th May 2006, 138 images in total, to allow the spatio-temporal modelling of vegetation change.

Figure 1. Study site location, Shiqu county, China.

3. METHODOLOGY

The rodent transect data were overlaid on the supervised classification derived from the Landsat ETM+ image using ArcView 9.1. A buffer of 500m was created around each of the 3485 rodent transect points. Within these buffers the proportion (%) of each habitat type was calculated. This proportional land cover data was combined with the rodent transect data was analysed using a Generalised Additive Model (GAM).

To explore whether a continuous NDVI dataset better characterised the landscape and its relationship with rodent distribution, a NDVI dataset was generated from the Landsat ETM+ image using Erdas Imagine 8.6. This NDVI image was then overlaid with the rodent index data, and a spatial join was performed to extract the NDVI value for the location of each rodent transect point.

Analysis of the 250m resolution MODIS 16-day composite vegetation index data imagery that had been acquired was also performed. The NDVI datasets were extracted from this data product, and were then reprojected to the UTM WGS84 projection to ensure compatibility with the other datasets utilised in this investigation, and subset to the region of the study site to reduce storage demands and reduce processing time. Initially a single NDVI image corresponding to the same period as the Landsat ETM+ image was acquired (3-11 July 2001), was overlaid with the rodent transect data, and again, a spatial join was performed extracting a MODIS NDVI value for each rodent transect point.

Next, the relationship between rodent distribution and time-series NDVI data was analysed. All 138 time-series NDVI composite images were stacked in acquisition date order which enabled the extraction of seasonal NDVI profiles over a six-year period for any location within the extent of the image. These seasonal NDVI profiles were extracted for each MODIS pixel containing rodent transect points. As MODIS 16-day composite images are produced for the same calendar periods each year (for instance the first image in each year covers the period 1-16 January), it is possible to collapse the six-year NDVI profile down into a single 'standardised' annual NDVI profile for that location. This was initially performed using the mean NDVI value for each 16-day period, and also the median and maximum NDVI values. This should have the effect of 'smoothing' the profile and removing noise. On examination, the 'standardised' profiles using the mean and maximum NDVI values were susceptible to the effects of outlying values, and were therefore disregarded.

Data quantifying the characteristics of the median standardised annual NDVI profile was then extracted, including maximum NDVI, minimum NDVI, mean NDVI, NDVI range, growing season (length of period when NDVI > 0.3), and greening period (when NDVI value rises above 0.3) allowing quantitative analysis methods to characterise spatial and temporal variation in the NDVI to be performed. This data was then combined with the rodent index data, and Landsat NDVI, MODIS single date NDVI and standardised time-series profile data were all analysed using a GAM.

4. RESULTS

The rodent index data, each landscape variable, and the topographical variables of slope, altitude and aspect were entered individually into a Generalized Additive Model. The results, showing % deviance explained, are displayed in Table 1. Although many of the land cover class and topographical variables gave poor results, the disturbed class for Occu (21.2%), Wet Grassland for Occa (18.1%) and Bog for Smallsm (18.1%) showed the highest single class degree of explanation, and indicated that relationships between the rodent presence and proportion of these land cover classes. The strongest relationship existed between Occu and the disturbed land cover class, and when plotted graphically it was obvious that the probability of Occu being present as the proportion of disturbed land increased within the buffered areas (Figure 2), which was supported by field observations.

Variable	Occu	Occa	Smallsm
Village	2.86%	2.86%	2.86%
Road	10.20%	14.00%	8.65%
Bog	3.62%	3.52%	18.10%
Water	1.54%	10.80%	6.24%
Grass	7.24%	5.64%	9.84%
Broadleaf	5.18%	11.30%	11.60%
Bare	4.79%	5.64%	8.89%
Disturbed	21.20%	14.10%	13.30%
Yellowbrush	1.34%	1.58%	4.66%
Wet Grassland	16.10%	18.10%	4.34%
Slope	4.50%	3.00%	3.82%
Altitude	2.92%	8.26%	12.70%
Aspect	6.69%	9.70%	5.09%

Table.1. Deviance explained (%) of rodent distribution related to proportion of each land cover class and topographical variables.

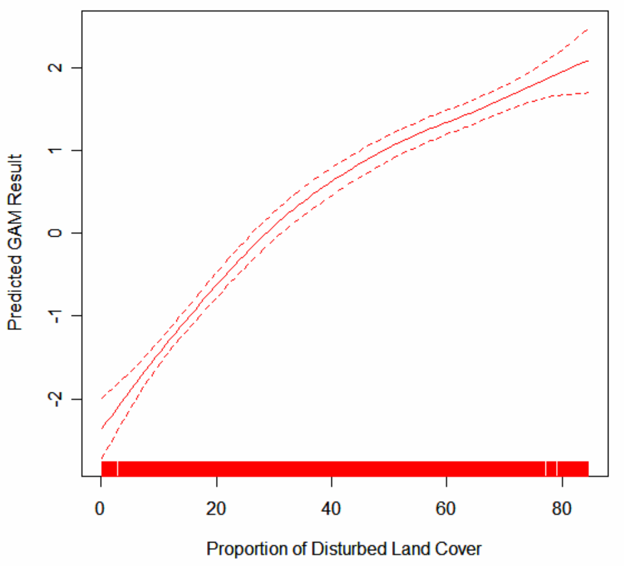

Figure 2. Occu distribution related to proportion of 'Disturbed' land cover class.

The characteristics of this disturbed class are very low biomass levels, and often bare soil, therefore it could be expected that a similar relationship may be observed between rodent distribution and NDVI as was observed with the 'Disturbed' land cover class. However, as Table 2 shows, attempts to model the relationships between rodent distribution and NDVI datasets derived from both the Landsat ETM+ and MODIS imagery did not explain as much deviance as the hard classification model.

The results of the single-date GAM analysis of rodent distribution and the Landsat ETM+ and MODIS NDVI datasets produced poor results for all three rodent groups, with the maximum deviance explained being 11.9% for Occu related to Landsat ETM+ NDVI. This was disappointing as the relationships between the disturbed (low biomass) land cover class and rodent distribution that had previously been confirmed using the hard classification were not repeated when related to a continuous vegetation dataset. The relationship between rodent distribution and MODIS NDVI was even lower, although this may be a result of the coarse 250m spatial resolution of the MODIS imagery.

Variable	Occu	Occa	Smallsm
Landsat ETM+ NDVI	11.9%	3.17%	4.86%
MODIS NDVI	5.03%	4.83%	8.78%

Table.2. Deviance explained (%) of rodent distribution related to single date Landsat ETM+ and MODIS derived NDVI data.

The results of the GAM analysis of rodent distribution and time-series MODIS NDVI data produced improved results over the single-data NDVI data. When each individual variable of the NDVI cycle was analysed in relation to the rodent distribution data the results were poor, accounting for between 3.47% and 15.6% of deviance explained. However, when a combination of all these variables were modelled with rodent distribution results improved significantly, with 41.4% deviance explained for Occu, 37.6% for Occa and 34.7% for Smallsm. Therefore, it can be assumed that modelling several variables relating to the seasonal spatio-temporal NDVI profile produces improved results, and is a more appropriate methodology than using single-date NDVI imagery. The differences in deviance explained values between Occu, Occa and Smallsm may result from Occu being present at a larger number of transect points than Occa or Smallsm, and therefore having more data items to model.

Variable	Occu	Occa	Smallsm
Maximum NDVI	13.7%	9.06%	15.6%
Minimum NDVI	7.06%	9.87%	4.55%
Mean NDVI	13.0%	6.91%	10.2%
NDVI range	3.47%	6.92%	9.04%
Greening period	12.3%	7.45%	9.84%
Growing season length	9.59%	9.54%	6.88%
All variables combined	41.4%	37.6%	34.7%

Table.3. Deviance explained (%) of rodent distribution related to quantified measurements of the standardised annual NDVI profile.

Another limitation of the MODIS time-series data is the 250m spatial resolution. It is likely that sub-pixel landscape and ecological features exist which influence rodent distributions, and cannot be identified using this coarse-resolution imagery. Higher spatial resolution time-series data could overcome this limitation, improving the rodent-landscape models. Other unidentified ecological processes could also influence the accuracy of the time-series model that were not quantified in this investigation. Also fieldwork observations identified areas of ideal rodent habitat that had no rodents present, further complicating the situation. Although the results of this study have shown the advantages and potential of using time-series NDVI datasets to model rodent distributions over single-time NDVI datasets, further research is required to better understand this relationship.

5. CONCLUSION

This investigation modelled rodent distribution in relation to landscape characteristics using four separate landscape-quantifying datasets, (1) a Landsat ETM+ hard classification, (2) single-image Landsat ETM+ derived NDVI, (3) single-image MODIS 16-day composite NDVI, and (4) time-series MODIS 16-day composite NDVI imagery. The success of these methods was variable, with the hard classification method displaying a relationship between Occu distribution and the 'Disturbed' land cover class, but showing poor relationships between the other landscape variables and rodent distribution. Both the Landsat ETM+ and MODIS derived NDVI single-image datasets gave poor results when modelled against rodent distribution, however the MODIS time-series NDVI data gave much improved results over the hard classification and the single-date NDVI datasets: for Occu, time series data explained 41.4% of deviance as opposed to 11.9% (single-image Landsat ETM+ NDVI) or 5.03% (single-image MODIS NDVI). For Occa it was 37.6% as opposed to 3.17% and 4.83%, and for Smallsm it explained 34.7% of deviation as opposed to 4.86% and 8.78%. It is likely that these figures could be improved further should higher spatial resolution time-series imagery become available, if additional ecological data is introduced into the models, and if methods of dealing with the effects of spatial autocorrelation are developed. Even with these limitations, these results show that rodent distributions can be successfully modelled using time-series NDVI data, and that this method is a viable alternative to using both single-image NDVI and hard classification data methods. It is hoped, in turn, that this method will significantly contribute to the regional scale prediction of *Em* rodent host distribution, and therefore also to the regional scale prediction of *Em* transmission foci.

REFERENCES

Beck P S A., Atzberger C., Hogda K A., Johansen B. and Skidmore A K., 2006, Improved monitoring of vegetation dynamics at very high latitudes: A new method for using MODIS NDVI, *Remote Sensing of Environment,* 100(3), pp. 321-334.

Boone J D., McGwire K C., Otteson E W., DeBaca R S., Kuhn E A., Villard P., Brussard P F. and St Jeor S C., 2000, Remote sensing and geographic information systems: charting sin nombre virus infections in deer mice. *Emerging Infectious Diseases,* 6(3), pp. 248-258.

Glass G E., Cheek J E., Patz J A., Shields T M., Doyle T J., Thoroughman D A., Hunt D K., Ensore R E., Gage K L., Irland C., Peters C J. and Bryan R., 2000, Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome, *Emerging Infectious Diseases,* 6(3), pp. 238:247.

Goodin D G., Koch D E., Owen R D., Chu Y K., Hutchinson J M S. and Jonsson C B., 2006, Land cover associated with hantavirus present in Paraguay, *Global Ecology and Biogeography,* 15(5), pp. 519-527.

Jiang Z., Huete A R., Chen J., Chen Y., Yan G. and Zhang X., 2006, Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction, *Remote Sensing of Environment,* 101(3), pp. 366-378.

Jin S. and Sader S A., 2005, MODIS time-series imagery for forest disturbance detection and quantification of patch size effects, *Remote Sensing of Environment,* 99(4), pp. 462-470.

Lunetta R S., Knight J F., Ediriwickrema J., Lyon J G. and Worthy L D., 2006, Land-cover change detection using multi-temporal MODIS NDVI data, *Remote Sensing of Environment,* 105(2), pp. 142-154.

Pfeiffer D U. and Hugh-Jones M., 2002, Geographical information systems as a tool in epidemiological assessment and wildlife disease management, *Revue Scientifique et technique De L Office International Des Epizooties,* 21(1), pp. 91-102.

Schwarz M. and Zimmermann N E., 2005, A new GLM-based method for mapping tree cover continuous fields using regional MODIS reflectance data, *Remote Sensing of Environment,* 95(4), pp. 428-443.

Suzan G., Giermakowski J T., Marce E., Suzan-Azpiri H., Armien B. and Yates T L., 2006, Modelling hantavirus reservoir species dominance in high seroprevalence areas on the Azuero peninsula of Panama, *American journal of Tropical Medicine and Hygiene,* 74(6), pp. 1103-1110.

Xiao X., Boles S., Frolking S., Li C., Babu J Y., Salas W. and Moore III B., 2006, Mapping paddy rice agriculture in south and southeast Asia using multi-temporal MODIS images, *Remote Sensing of Environment,* 100(1), pp. 95-113.

ACKNOWLEDGEMENTS

Acknowledgements must also be made of the National Institute of Health, National Science Foundation, USA, through which the funding for this research was provided, and also the contribution of our Chinese co-workers during fieldwork periods who provided invaluable support and assistance.