Applied Geography 44 (2013) 192–200

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Predicting socioeconomic conditions from satellite sensor data in rural developing countries: A case study using female literacy in Assam, India

Gary R. Watmough [a,b,*], Peter M. Atkinson [b], Craig W. Hutton [c]

[a] Agriculture and Food Security Center, The Earth Institute, Columbia University, Lamont Campus, 61 Route 9W, Lamont Hall 2G, P.O. Box 1000, Palisades, NY 10964, USA
[b] Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK
[c] GeoData Institute, University of Southampton, Southampton SO17 1BJ, UK

Keywords:
Population–environment
Regression
Wealth
Assam
India
Generalised autoregressive error model
Remote sensing
Poverty

A B S T R A C T

Social data from census and household surveys provide key information for monitoring the status of populations, but the data utility can be limited by temporal gaps between surveys. Recent studies have pointed to the potential for remotely sensed satellite sensor data to be used as proxies for social data. Such an approach could provide valuable information for the monitoring of populations between enumeration periods. Field observations in Assam, north-east India suggested that socioeconomic conditions could be related to patterns in the type and abundance of local land cover dynamics prompting the development of a more formal approach. This research tested if environmental data derived from remotely sensed satellite sensor data could be used to predict a socioeconomic outcome using a generalised autoregressive error (GAR_{err}) model. The proportion of female literacy from the 2001 Indian National Census was used as an indicator of socioeconomic conditions. A significant positive correlation was found with woodland and a significant negative correlation with winter cropland (i.e., additional cropping beyond the normal cropping season). The dependence of female literacy on distance to nearest road was very small. The GAR_{err} model reduced residual spatial autocorrelation and revealed that the logistic regression model over-estimated the significance of the explanatory covariates. The results are promising, while also revealing the complexities of population–environment interactions in rural, developing world contexts. Further research should explore the prediction of socioeconomic conditions using fine spatial resolution satellite sensor data and methods that can account for such complexities.

Introduction

The determinants of land use land cover (LULC) change include multi-temporal and multi-level factors relating to biophysical, environmental, economic, socioeconomic variables and government policy (Liverman, Moran, Rindfuss, & Stern, 1998; Walsh, Crawford, Welsh, & Crews-Meyer, 2001). Socioeconomic variables correlated with LULC change have included; the density of rural migration (Wood & Skole, 1998); village population growth (Entwisle, Walsh, Rindfuss, & Vanwey, 2005); lack of financial capital (Perz & Skole, 2003); age structure of the population (Moran, Brondizio, & Vanwey, 2005) and; household assets (Walsh, Rindfuss, Prasartkul, Entwisle, & Chamratrithirong, 2005). Studies in urban and suburban areas of the United States have found significant relationships between socioeconomic variables from national census data and environmental variables (Lafary, Gatrell, & Jensen, 2008; Li & Weng, 2007; Mennis, 2006; Ogneva-Himmelberger, Pearsall, & Rakshit, 2009). The relationships between socioeconomic and environmental metrics suggest that there may be potential for using remotely sensed satellite sensor data as proxies for social data from census and household survey enumeration. Ogneva-Himmelberger et al. (2009) stated that; "… if the relationships between socioeconomic and environmental conditions were understood, practitioners and policy makers could develop an understanding of the geographic distribution of social well-being and quality of life based on a map of vegetation cover" (p. 479) which could significantly add to social data in census and household surveys.

* Corresponding author. Agriculture and Food Security Center, The Earth Institute, Columbia University, Lamont Campus, 61 Route 9W, Lamont Hall 2G, P.O. Box 1000, Palisades, NY 10964, USA. Tel.: +1 (845) 365 8732.
E-mail addresses: gw@ldeo.columbia.edu (G.R. Watmough), P.M.Atkinson@soton.ac.uk (P.M. Atkinson), cwh@geodata.soton.ac.uk (C.W. Hutton).

0143-6228/$ – see front matter
http://dx.doi.org/10.1016/j.apgeog.2013.07.023

Social data provide key information to track changes in populations (Shearmur, 2010), allocate government resources and evaluate policies (UNSD, 2008). The repeatability and universality of census data is vital to chart a country's development both temporally and spatially and allow for focused policy creation (Rindfuss & Stern, 1998). However the fine spatial resolution of census data means that a census is an expensive process and is typically conducted only once every 10 years. This coarse temporal resolution limits the relevance of census data for planning resource allocation by governments and targeting development assistance, especially in rapidly changing economies. For example, a study in 2008 of flood vulnerability in Assam, north-east India used 2001 National Census data (Amoako Johnson & Hutton, 2012; Hutton et al., 2011) which was the most up-to-date data available.

Gaps between census enumerations are typically between five and 10 years but political and social instability have resulted in larger gaps such as 15 years (1993–2008) in Sudan (UNFPA Sudan, 2007) and over 40 years (1970–2013) in Angola. This can mean that countries most in need of reliable social data lack baseline studies from which to plan effectively. The fine spatial resolution temporal resolution that remotely sensed imagery offers over traditional ground survey methods may provide a way of increasing the information available to researchers and policy makers for monitoring socioeconomic conditions. If relationships between socioeconomic and environmental conditions can be found, that are similar to those found in previous studies (Lafary et al., 2008; Li & Weng, 2007; Ogneva-Himmelberger et al., 2009), satellite sensor data could provide proxies for social data for use between census enumeration periods. This study explored and quantified the relationships between a census-derived socioeconomic variable and land cover dynamics derived from remotely sensed satellite sensor data vegetation maps in a rural region of a developing country.

Theory

Relationships between population and environment can be complex (de Sherbinin, Carr, Cassels, & Jiang, 2007) and context-specific, and it is difficult to assign causality between the two. Relationships vary due to a range of mediating and confounding factors such as government policies, international markets, household experiences, demographic, societal and cultural factors (Lambin, Geist, & Lepers, 2003; de Sherbinin et al., 2007). For example, woodland and non-timber forest products (NTFP) can be an important income source for rural households, but the contribution can be influenced by local access to road infrastructure and markets (Babulo et al., 2009). Reflecting the complexity of population–environment relationships, several theoretical approaches have been developed to explore the interactions (Hummel et al., 2013). One method is the Sustainable Livelihoods Approach (SLA) which identifies five capital asset groups that are important for rural community development; financial capital, natural capital, physical capital, social capital and human capital. Livelihoods of households can be assessed by estimating the endowment of each capital group (Chambers & Conway, 1991, p. 33). As an example, communities in remote locations often have larger concentrations of poverty (Sen, 2003) and an SLA assessment may show that these communities have relatively large endowments of natural capital (forest, NTFP and land), but small endowments of physical (poor infrastructure and agricultural development), social, human and financial capital, which make it difficult to take advantage of natural capital assets.

Natural capital assets can provide a significant component of rural livelihoods through the use of natural resources such as water, firewood, agriculture, fishing and NTFP. In 2004 it was estimated that over 1.6 billion people depended on forests for subsistence or income generation (World Bank, 2004, p. 81). The contribution of NTFP to household incomes has been found to vary from 33% in Tanzania (Monela, Kajembe, Kaoneka, & Kowero, 2001), 17 to 45% in Bolivia and Honduras (Godoy et al., 2002), 39% in South-west Ethiopia (Mamo, Sjaastad, & Vedeld, 2007) and 27% in northern Ethiopia (Babulo et al., 2009). The presence of water resources and the ability to irrigate agricultural fields can have significant relationships with rural development (Hussain, 2007). Often the impact is related to yield increases (Huang, Rozelle, Lohmar, Huang, & Wang, 2006) which result in subsequent increases in rural employment and growth in local economic activity (de Janvry & Sadoulet, 2010). Different regions, countries, communities and households will have different experiences depending on the stocks of assets available to them and demographic, social, environmental and climatic conditions. Remotely sensed data can provide information on biophysical conditions in and around habited areas (Rindfuss & Stern, 1998). Considering the potential importance of natural and physical capital to rural populations and the spatial and temporal variation in livelihood conditions it may be that remotely sensed data could significantly contribute to monitoring population change. Remotely sensed data could provide estimates of natural and physical capital endowments and integrate these with census and household survey data to provide more detailed SLA assessments. Furthermore, if links between socioeconomic conditions and remotely sensed land cover dynamics can be established, remotely sensed data could be used to provide proxies for socioeconomic conditions or livelihoods during the temporal gaps between census enumerations.

Several studies have explored the links between socioeconomic conditions and environmental metrics often derived from remotely sensed data, such as fraction of vegetation cover, normalised difference vegetation index (NDVI) and impervious surfaces. NDVI has been found to be positively associated with; median household income (Lafary et al., 2008; Pearsall & Christman, 2012); per capita income (Lo & Faber, 1997; Tooke, Klinkenber, & Coops, 2010); household value (Lafary et al., 2008); educational achievement (Mennis, 2006). Relationships have also been found between median household income and leaf area index (Jensen, Gatrell, Boulton, & Harper, 2004) and between household income and fraction of green vegetation (Li & Weng, 2007). Li and Weng (2007) and Ogneva-Himmelberger et al. (2009) used measures of impervious surface instead of vegetation and found that the relationships of income and poverty with imperviousness were often the inverse of those with vegetation. This is to be expected as urban and suburban vegetation cover is often negatively related to building density. For example, correlations in Massachusetts were largely negative between median income and percent impervious cover and median housing value and percent impervious cover and largely positive between percent minority and percent impervious cover and percent poverty and percent impervious cover (Ogneva-Himmelberger et al., 2009). The use of geographically weighted regression (GWR) models in Lafary et al. (2008), Ogneva-Himmelberger et al. (2009) and Pearsall and Christman (2012) revealed that the relationship size and direction was spatially heterogeneous. All of these environmental metrics can be estimated from remotely sensed data. Thus, it is worth examining if, and how, remotely sensed environmental data could be used to estimate socioeconomic outcomes. It is unlikely that socioeconomic outcomes would have such a significant relationship with NDVI in rural areas of developing countries because vegetation is much more abundant than in urban areas. However, the abundance of land cover classes such as woodland, cropland, water, unproductive land covers such as bare land and physical capital