World Development Vol. 78, pp. 188–203, 2016
0305-750X/© 2015 Elsevier Ltd. All rights reserved.
www.elsevier.com/locate/worlddev
http://dx.doi.org/10.1016/j.worlddev.2015.10.031

Understanding the Evidence Base for Poverty–Environment Relationships using Remotely Sensed Satellite Data: An Example from Assam, India

GARY R. WATMOUGH [a], PETER M. ATKINSON [b], ARUPJYOTI SAIKIA [c] and CRAIG W. HUTTON [b,*]

[a] *Columbia University, New York, USA*
[b] *University of Southampton, UK*
[c] *Indian Institute of Technology Guwahati, Assam, India*

Summary. — This article presents results from an investigation of the relationships between welfare and geographic metrics from over 14,000 villages in Assam, India. Geographic metrics accounted for 61% of the variation in the lowest welfare quintile and 57% in the highest welfare quintile. Travel time to market towns, percentage of a village covered with woodland, and percentage of a village covered with winter crop were significantly related to welfare. These results support findings in the literature across a range of different developing countries. Model accuracy is unprecedented considering that the majority of geographic metrics were derived from remotely sensed data.

Key words — livelihoods, poverty, environment, remote sensing, India

1. INTRODUCTION

The relationships between poverty and natural resources are complex and can vary across both time and space (Dasgupta, Deichmann, Meisner, & Wheeler, 2005; De Sherbinin, Carr, Cassels, & Jiang, 2007; De Sherbinin *et al.*, 2008). Despite these complexities, local land cover types, isolation from markets, and agro-ecological conditions often have a significant impact on rural poverty (Barrett *et al.*, 2006; Okwi *et al.*, 2007). A limited set of population characteristics can be measured using censuses which aim for complete population coverage, while more detailed snapshots of population characteristics can be obtained using cross-sectional, nationally representative household sample surveys. Furthermore, environmental variables which may have important contributions to local poverty can be measured synoptically across large areas from space using satellite imagery (Rogan & Chen, 2004). The complex nature of poverty means that it is difficult to quantify poverty–environment relationships and it is often difficult to integrate understanding of poverty–environment relationships into policies targeting development and ecological protection. This paper investigates if remotely sensed satellite data could be used as a tool to increase understanding of the relationships between welfare and local environmental conditions.

The World's poorest communities often rely upon natural resources and ecosystem services for their livelihoods and subsistence. The majority of these communities are ecologically marginalized as they are concentrated in socially and environmentally fragile areas (Sen, 2003). The pressure that natural resources experience from population growth is a significant barrier to sustainable human development and ecological conservation. Complex feedback loops mean that changes in population or environment will have consequences for the other. However, understanding of these links and feedback loops is limited, and greater understanding of the inter-linkages is needed to help foster more effective policies for sustainable human development and ecological conservation. To this end, one area of population–environment research concentrates on estimating the contribution of environmental resources to household incomes. In small case-study examples, environmental resources, and especially forest products such as wood fuel and building materials and non-timber forest products (NTFP) such as food and medicine, have been found to significantly contribute to rural household incomes (Babulo *et al.*, 2009; Cavendish, 2000; Kamanga, Vedeld, & Sjaastad, 2009; Mamo, Sjaastad, & Vedeld, 2007; Vedeld, Angelsen, Bojö, Sjaastad, & Kobugabe Berg, 2007). The Center for International Forestry Research (CIFOR) population–environment network (PEN) is an attempt to study these population–environment relationships using a larger dataset from approximately 8,000 households across 24 developing countries. Using the PEN data, Angelsen *et al.* (2014) found 28% of household incomes were derived from natural resources across the CIFOR study, lending some support to findings from previous smaller case studies. However, natural resources were found to contribute primarily to subsistence in poorer households and income in wealthier households, indicating a more complex pattern than was indicated by small-scale case study examples. Furthermore, Wunder,

* We thank the BRAHMATWINN project for providing data for this study. John Duncan provided advice on the use of MODIS data products. We are grateful for comments from Cheryl Palm, Mark Musumba, Alex de Sherbinin on a previous draft of the paper. Figures 1 and 5 were created by Mark Dover, University of Southampton. Financial support for the project was provided by Earth Institute, Columbia University – New York; Geography and Environment, University of Southampton – United States of America; the Royal Geographical Society – United Kingdom, (RGS-IBG) and the Gilchrist Educational Trust – United Kingdom. Final revision accepted: October 3, 2015.

Börner, Shively, and Wyman (2014) found that forests appeared only to be used as a last resort safety-net when alternatives had been exhausted where only 10% of households identified extraction strategies as the most important response to a shock. This finding somewhat contradicted the results of smaller scale and individual case study examples which indicated that forests and woodland were commonly used as safety nets (Angelsen & Wunder, 2003) especially in areas highly susceptible to epidemiological shocks (De Sherbinin *et al.*, 2008).

Quantifying environmental incomes in poorer regions of the World can have important policy implications for human development and ecological conservation. The majority of previous studies have used household surveys, but while often being nationally representative they do not cover an entire population (Weeks *et al.*, 2012). The way in which data are collected and the coverage of these data can have a significant impact on analysis results. This is illustrated in some of the results from analysis of the global comparative PEN dataset which differed to those found using smaller case-study datasets (Wunder *et al.*, 2014). Therefore, using national census datasets covering an entire population in population-environment research is important because it can give more meaning to the relationships and be more applicable to government policy.

(a) *Population–environment research using remote sensing*

In urban areas, vegetation and welfare are commonly associated positively due to the prominence of gardens, recreation spaces, and tree-lined streets in wealthier residential areas (Rashed, Weeks, Roberts, Rogan, & Powell, 2003). Thus, biophysical metrics such as the normalized difference vegetation index (NDVI) derived from satellite data have been associated with welfare. For example, NDVI has been found to be positively correlated with a range of wealth indicators such as; median household income (Lafary, Gatrell, & Jensen, 2008; Pearsall & Christman, 2012); per capita income (Lo & Faber, 1997; Tooke, Klinkenber, & Coops, 2010); household value (Lafary *et al.*, 2008) and educational achievement (Mennis, 2006). Similar relationships have also been found between leaf area index (LAI) and median household income (Jensen, Gatrell, Boulton, & Harper, 2004). In Accra, Ghana, NDVI and percent vegetation cover were both found to be correlated with wealth (Weeks *et al.*, 2012). The opposite relation was found in Detroit, where high levels of NDVI were associated with high incidences of poverty (Ryznar & Wagner, 2001). This result was explained by urban decay and out-migration which, led to the abandonment of previously developed areas of the city and resulted in vegetation regrowth (increasing greenness).

There has been a strong focus on modeling relationships between poverty and remotely sensed biophysical parameters in urban and sub-urban locations. However, the majority of the World's poorest populations are rural and reliant upon natural resources and ecosystem services for subsistence and income generation (Barbier, 2010; Barrett, 2005). Studies integrating socioeconomic and remotely sensed environmental data have found several socioeconomic and demographic conditions to be significant determinants of land use and land cover change (Liverman & Cuesta, 2008; Liverman, Moran, Rindfuss, & Stern, 1998). Increased development has been related to environmental and socioeconomic change in Nigeria (Twumasi & Merem, 2006) and in rural India female literacy (which is related to poverty) has been found to be strongly correlated with local environmental conditions estimated from satellite data (Watmough, Atkinson, & Hutton, 2013). It appears however, that the choice of dependent variable can be important as biophysical parameters were not significantly related to child malnutrition in Africa (De Sherbinin, 2011) or female health in Accra (Weeks *et al.*, 2012).

Growth in agriculture is a particularly important driver to reduce poverty in developing countries (Diao, Hazell, & Thurlow, 2010). Without improved seed varieties developed during the green revolution agricultural yields in developing countries would be much lower and food prices and poverty rates higher than they are today (Evenson & Gollin, 2003). Improved agricultural technology can reduce poverty by raising farm productivity which leads to more job opportunities, increased farm incomes, and lower local food prices (Asfaw, Kassie, Simtowe, & Lipper, 2012; Kassie, Shiferaw, & Muricho, 2011; Minten & Barrett, 2008). However, if poverty is to be reduced, growth in yield should benefit small-scale farmers, as evidence suggests that the relationship between yield increases and poverty reduction is weaker in places where growth favors large-scale industrial farmers over small holders (Pauw & Thurlow, 2011).

Ecosystem provisioning services such as water, woodland, and biodiversity have a potentially important influence on local poverty and welfare. Water resources, especially for irrigation, tend to have significant positive relationships with rural development (De Janvry & Sadoulet, 2010; Huang, Rozelle, Lohmar, Huang, & Wang, 2006; Hussain, 2007). Non-timber forest products (NTFP) contribute significantly to household incomes (Babulo *et al.*, 2009; Mamo *et al.*, 2007; Vedeld *et al.*, 2007) and biodiversity can be an important factor for rural development and poverty alleviation (Campbell, 1996; Ligon & Sadoulet, 2008). Thus, poverty and welfare could have strong relationships with remotely sensed measures of agricultural prevalence, productivity, and yield and other important ecosystem services such as availability of water, woodland, and biodiversity.

Chronic or long-term poverty and transient more short-term poverty often result from different sets of factors which can have important implications for policy makers. It is not in the scope of this paper to distinguish between different forms of poverty or suggest policy implications. However, it is worth noting that location and environment are often found to have significant impacts on both chronic and transitory poverty. Barrett (2005) describes *chronic* poverty as often being a result of low levels of productive asset ownership, an inability to make full use of owned assets, and shocks that can deplete accumulated assets and wealth such as natural disasters and disease. Krishna (2004, 2006) found that, in Kenya and India, respectively, descents into poverty were associated with drought and the costs of accessing health care, and ascents out of poverty were associated with income diversification. Therefore, poverty could have relationships with data derived from remotely sensed imagery such as land cover change that could be used to indicate areas affected by shocks such as flooding and drought. Relationships may also exist between income diversification opportunities and remotely sensed metrics such as land use and land cover associated with non-agricultural income sources.

Communities in remote locations with limited access to transportation are less able to make full use of productive assets that they own such as land and human capital and they are more susceptible to chronic poverty (Barrett, 2005). Remote regions have lower access to a range of features important for development such as education, information, health, and finance (Blaikie, Cameron, & Seddon, 2002; Khandker, Zaid, & Gayatri, 2006; Sen, 2003). Adoption rates of new seed varieties and improved agricultural technologies are often found to be lower in regions with poor access to information and markets (Minten & Barrett, 2008). Remoteness can also have an effect on access to non-farm employment and financial capital, which is also an important aspect of poverty reduction as it can help income diversification (Barrett *et al.*, 2006). Diversification of income is associated with spatial location as communities with physical access to larger population centers may have fewer barriers to income diversification. Thus, locations of large populated centers identified by satellite data could be used to estimate proxies of community access and be correlated with poverty.

Environmental parameters can have an important impact on poverty. However, the complex nature of population, development, and poverty means that it is difficult to quantify these relationships. This lack of understanding creates limitations for a range of policies including development and ecological protection. Coupling remotely sensed environmental data with socioeconomic datasets could help to increase the understanding of poverty–environment relationships. This paper explores and quantifies the associations between welfare and environmental metrics derived from satellite sensor data across an extensive area of a developing country. The associations revealed in the analysis are used to suggest potential underlying mechanisms within the poverty–environment nexus.

2. STUDY SITE

We selected Assam in Northeast India as a case study (Figure 1). Assam is a good example to test the relationships between poverty and environmental metrics derived from satellite data due to relatively high poverty rates and a large reliance on agricultural practices. In 2001, 36% of the population was below the national poverty line compared to a national average of 26% (World Bank, 2011a). Poverty levels in Assam are persistently high (Mehta, 2003) and poverty increased in absolute terms from 7.8 million to 9.5 million during 1983–2000 (Government of Assam, 2003).

Industrial growth has been slow due the isolation of the Northeast region from the rest of India. Historical growth of industry during the British colonial rule focused on a small number of products such as tea, jute, and oil. Agrarian expansion during the colonial era led to large amounts of forest land being leased for Tea cultivation and Tea continues to be a major commercial crop in Assam in the present day. However, the cost of clearing the forest land meant only a select few were able to benefit (Saikia, 2014). Jute was an important commercial crop in Assam in the early twentieth century. However, the collapse of Jute prices during the Great Depression of the 1930s resulted in many Jute farmers switching to paddy cultivation (Saikia, 2014). After independence, the partition

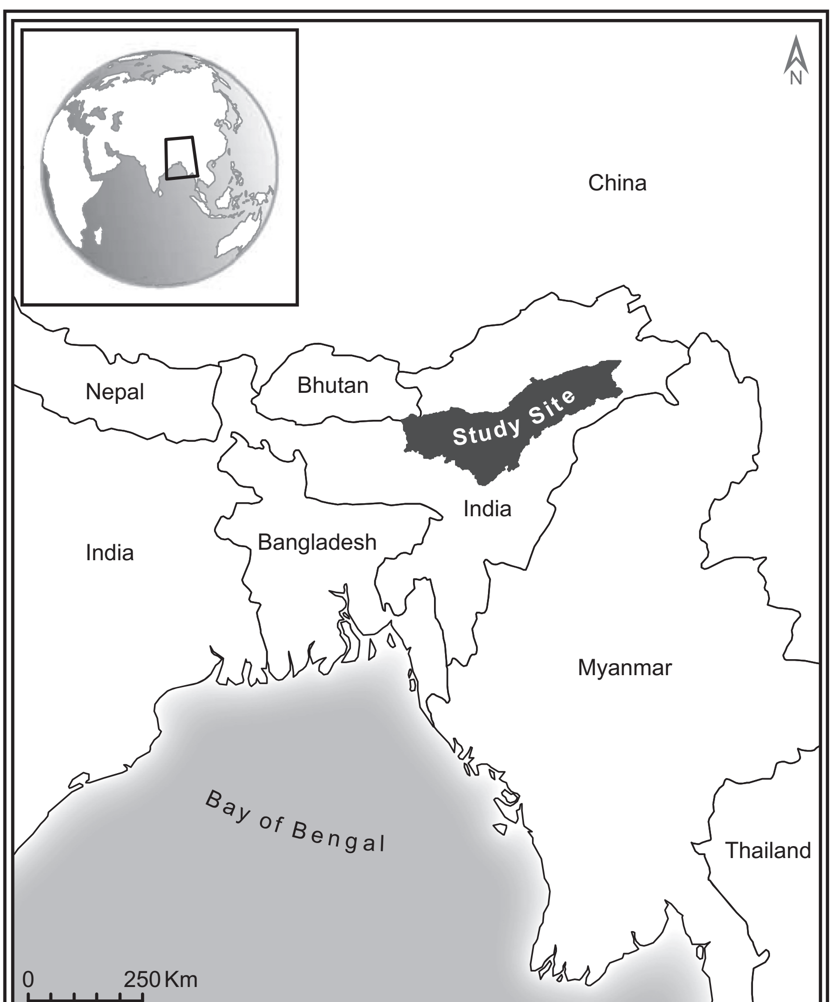

Figure 1. *Location of study site in North-East India. Clearly shows how Assam and the northeast region are isolated from the rest of India.*

of India into East Pakistan and subsequently into Bangladesh left the Northeast region of India cut-off from the rest of India as navigable river channels and railways which once connected Assam to major ports such as Calcutta now cross Bangladesh. Today, the main method of reaching the rest of India is using the poor road and rail infrastructure (Baruah, Das, & Dutta, 2004). This poor connectivity along with the small plot sizes due to cultivable land being absorbed into forest conservation schemes or Tea plantations and more recently due to population pressure (Saikia, 2014) are thought to be responsible for the low levels of agricultural productivity and slower growth rates (Datt & Ravallion, 1998).

Irrigation in Assam has been estimated at 22% (Daimari & Mishra, 2005) which is significantly lower than the national average of 35% (Fan, Hazell, & Haque, 2000), and relative drought conditions during the winter months (November to March) result in low levels of multiple cropping (Baruah *et al.*, 2004). Estimates suggest that 76% of the population is reliant on agriculture or allied industries for incomes and subsistence (Daimari, 2005) and therefore this low productivity will have significant contributions to poverty in the region. Forest and woodland areas are historically linked with poverty and landless communities in Assam. From the early 1920's Forest Villages were established in Assam which allowed poor landless people to inhabit reserved forests. This was in part to deal with lack of labor available to the Forest Department for maintaining and protecting the forests. In 1976 the National Commission of Agriculture adopted a social forestry policy which was the further development of the National Forest policy of 1952 which sought to provide forest products to villages while protecting large forest areas by encouraging tree planting in common areas of villages and along roadsides (Saikia, 2005).

Annual monsoon flooding is vital for the padi-rice crops and replenishing soil nutrients. However, the increasing number of flood events (Jamir, Gadgil, & De, 2008) and associated river bank erosion (Sarma, 2005) cause extensive damage and socioeconomic problems. Land lost through bank erosion in Assam was estimated to be 10.3 km^2 per year from 1912 to 1996 (Sarma, 2005). The deposition of river sediments creates sand banks and islands that often become the site of semi-permanent habitation (Brocklesby & Hobley, 2003). However, this newly formed land is often unstable and prone to flooding meaning that communities located in these areas are some of the poorest in the region.

Fieldwork was conducted in Assam in 2009 to identify associations between welfare and land cover dynamics, and detailed methods are presented in Watmough *et al.* (2013). The results indicated that wealthy communities had extensive cropland cover mostly used for *Kharif crops* (sown during summer monsoon and harvested around October) and very little cultivation during the dry winter season. Cover of grass and bare land was low and communities were either close to larger market towns or had access to paved roads. Poorer communities were often found on land closer to rivers which was more susceptible to flooding and erosion. There was also low woodland cover and few roads resulting in poorer access to major towns. However, the close proximity to water bodies did allow the growth of *Rabi* crops (sown during the dry winter period) often due to small plot sizes and to insure against losses to *Kharif* harvest from flooding, deposition, and erosion. For villages located close the large rivers the erosion and deposition problems appeared to remove many of the benefits of increased soil fertility created from alluvial deposits. This may be different in areas with flood protection but it was not possible to visit these areas. Very poor communities often inhabited three distinct areas; river banks, river islands (Chars), and dense forested areas. Communities on river banks and river islands were characterized by high levels of bare sandy cover, and very small agricultural plots were used intensively to produce several crops per year (Watmough *et al.*, 2013).

3. DATA AND METHODS

Description of the data and analysis is split into three sections; (i) construction of a village level welfare index using socioeconomic data; (ii) estimation of local environmental covariates using remotely sensed satellite sensor data; (iii) modeling the relationships between welfare and the environment.

(a) *Socioeconomic data*

Socioeconomic data were available at the community level from a subset of the 2001 Indian National Population and Household Census (Registrar General, 2011). Communities ranged in size from small hamlets containing only 2 or 3 households to towns containing several thousand households. Urban areas were removed from the data using the census 'Status' variable which indicated large populated towns and cities resulting in 14,227 rural communities each with a population ranging in size from 10 to 15,000 people.

(i) *Creating a community buffer*

Environmental metrics local to each community were required for the statistical modeling of population-environment relationships. The census data were provided with coordinates for each community centroid which meant that the spatial extent of communities was unknown. Radial buffer zones have been used in the past to link population data with the local environment (Entwisle, Walsh, Rindfuss, & Vanwey, 2005) and a previous study in Assam approximated community spatial extents using a range of radial buffer zones and Thiessen polygons (Watmough *et al.*, 2013). This analysis found little difference between the goodness-of-fit statistics for models based on; Thiessen polygons, fixed radial buffer zones of 500 m, 1,000 m and 3,000 m radius and buffer zones that were weighted in size using total community population and total land owners in a community (Watmough *et al.*, 2013). The number of land owners was used as a weighting factor because agriculture was the dominant economic activity in smaller communities which were often surrounded by agricultural land. Larger communities had more people engaged in non-agricultural work, thus, reducing the amount of land surrounding communities with fewer land owners. Limited conclusions about the most appropriate polygon could be drawn from these results. However, with no way of identifying a 'true' community extent a 1,000 m buffer zone was used in this study. The 1,000 m buffer zone was chosen because it reduced the overlap between adjacent polygons and subsequently reduced the amount of environmental data extracted to more than one community. This polygon design may introduce issues associated with the modifiable areal unit problem (Jelinski & Wu, 1996). However, field observations and discussions with local academics suggested that in the absence of other information a 1,000 m buffer zone could be used.

(b) *Constructing a welfare index*

Income and expenditure data were not available in the census and so a relative welfare index was constructed by combining several census variables as described below.

(i) *Education*

Education can determine decision-making (Montgomery, Gragnolati, Burke, & Paredes, 2000) as well as occupation type and income (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006a). Adult literacy has been used to calculate human capital when exploring socioeconomic vulnerability in India (O'brien *et al.*, 2004). Links have also been found between female literacy and rural poverty (Hannum & Buchmann, 2005; Watmough *et al.*, 2013). These relationships mean that education can be a suitable indicator of welfare and, as such, measures of education are used in both the Human Development Index (HDI) and the Multidimensional Poverty Index (MPI) (Alkire & Santos, 2014). Years of schooling or level of educational achievement are often used, but neither was available for Assam. Instead, question 12 of the individual particulars section of the census defined a respondent as literate if able to read and write with understanding in one of several pre-determined languages. No formal educational achievements or years in school were required. A person who could read but not write was classed as illiterate, as were all children six years of age or under (Registrar General & Census Commissioner, 2011). Therefore, female literacy was used as an indicator of education in Assam and was defined as:

$$\text{Female Literacy} = \frac{\text{number of literate adult females}}{\text{total adult female population}} \tag{1}$$

(ii) *Land ownership*

Employment indicators related with welfare include job security and employment type (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006b). Census data for Assam were available on the type of employment, which was split into several categories; cultivators (land owners), agricultural laborers (working on land not owned by the household), household industry workers, and other types of workers. These categories did not indicate job security or salary levels and it was felt that village level employment rates were inappropriate for the present study because they included poorly paid and insecure job types such as rock breaking. Employment in these types of industries is not reflective of wealth and so an employment index was not used here. Instead, land ownership (cultivators) was included as an indicator of welfare. In this case, land ownership referred to households cultivating land owned by the household and not working on another household's land and was defined as:

$$\text{Land owners} = \frac{\text{number of land owners(cultivators)}}{\text{total adult population}} \tag{2}$$

(iii) *Social status*

In India, groups classed as Scheduled Tribe and Scheduled Caste are more likely to be undernourished, poor, and marginalized than the Hindu majority (Jha, 2008). Average incomes for Hindus have been found to be approximately 59% higher than those for Scheduled Tribal groups and 68% higher than for Scheduled Caste groups. Approximately 71% of Hindu's were classed as non-poor compared with 54% Scheduled Caste and 53% Scheduled Tribe (Borooah, 2005). The HDI and Human Poverty Index (HPI) have been found to be approximately 30% lower for Scheduled Tribes than the all-India average (Sarkar, Mishra, Dayal, & Nathan, 2006). Alkire and Seth (2013a) reported that the reduction in the MPI in India varied according to different social groups with the slowest reduction in poverty being 6.3% for Scheduled Tribe groups compared to 12.2% for the general Indian population. These statistics indicate the disparity between different social and cultural groups in India. Therefore, the proportion of the village population registered as a member of a Scheduled Caste or Scheduled Tribe was included as an indicator of deprivation and was defined as:

$$\text{Scheduled Caste(Tribe)} = \frac{\text{number of Scheduled Casteor(Tribe)members}}{\text{total population}} \tag{3}$$

Indicator variables were normalized using a zero to one scale (zero representing a low and one representing a high welfare score). This meant that Scheduled Caste and Scheduled Tribe variables were combined and inverted to create a non-deprived variable (Eq. (4)).

$$\text{Nondeprived group} = 1 - (\text{Scheduled Caste proportion} + \text{Scheduled Tribe Proportion}) \tag{4}$$

(iv) *Access to water*

Access to resources such as water, sanitation, and roads are more common in wealthy communities and the provision of public services can increase wealth and socioeconomic positions (Rutstein & Johnson, 2004). The Demographic and Health Survey (DHS) wealth index includes variables indicating access to community resources such as water (Rutstein, 2004). Open, unprotected, and surface water sources such as springs, lakes, and rivers can be indicative of a low socioeconomic position (Gwatkin, Rutstein, Johnson, Pande, & Wagstaff, 2007; Houweling, Kunst, & Mackenbach, 2003). Consequently, access to safe drinking water has been used in the MPI (Alkire & Seth, 2013b) and a range of asset indices (Booysen, Van Der Berg, Burger, Maltitz, & Rand, 2008). Village-level water source was included in the welfare index for Assam. The census contained nine water source classes which were split into improved and surface water categories. Improved water included water from Tap, Well, Tank, Tube well, and Hand pump sources while surface water included water from River, Canal, Lake, and Spring source. The water source indicator was defined as a binary variable where improved water had a value of 1 and surface water a value of 0. Construction of the welfare index followed the approach used by the Demographic and Health Survey and outlined in Rutstein (2004).

(c) *Weighting the index*

A weighted relative welfare index was created using normalized variables of female literacy, land ownership, deprived class, and water source and was constructed using Eq. (5):

$$\text{Relative Welfare} = V_1W_1 + V_2W_2 + \ldots + V_3W_3 \tag{5}$$

where V are the normalized index variables (female literacy, land ownership, deprived class, and water source) and W are the variable weights estimated using principal components analysis (PCA). PCA is a data reduction technique that transforms a set of correlated variables into a set of orthogonal (un-correlated) principal components (PC) which are linear combinations of the original variables. The loadings for the first PC are typically taken as the variable weights when constructing an asset index as the first PC is assumed to represent income-related wealth (Filmer & Pritchett, 2001). It has been recommended that PCA be used in cases when continuous variables are used and factor analysis be used when there are both continuous and categorical variables (Filmer & Pritchett, 2001). However, the weightings for categorical variables derived from factor analysis and PCA are usually very similar and since PCA is easier to run and interpret it is often used in cases when continuous and categorical variables are used in combination (Howe, Hargreaves, & Huttly, 2008). To allow focus on the poorer and wealthier communities, the relative welfare scores were converted to quintiles by splitting the range into five equal-sized groups named; Lowest; Second; Middle; Fourth; Highest as in the DHS wealth index (Rutstein, 2004) and Angelsen *et al.* (2014).

(d) *Remotely sensed environmental data*

The complete coverage of the census prevents the collection of matching environmental data using household surveys as it would be costly and time consuming. However, environmental variables can be measured synoptically across large areas from space using satellite imagery (Rogan & Chen, 2004). In this study, remotely sensed satellite data were used to estimate proxies for environmental conditions. The 1,000 m buffer zone surrounding community centroids was used to approximate the spatial extent of each community and a Geographic Information System (GIS) was used to extract remotely sensed environmental conditions local to each community. Remotely sensed data included Landsat Enhanced Thematic Mapper Plus (ETM+) which was used to create a land cover map of Assam (see below) and the MODIS Normalized Difference Vegetation Index (NDVI). The NDVI provides an indication of vegetation vigor and when incorporated into a time-series can be used to identify different vegetation types (Rogan & Chen, 2004). All remotely sensed data used were freely available and downloaded from the USGS GloVis database (http://glovis.usgs.gov/). The following sections give details on how the two remotely sensed datasets were utilized in this study.

(i) *Land use and land cover*

An annual NDVI trend was estimated for 2001 (the same year as census enumeration) using 23 MODIS 250 m spatial resolution MOD13Q1 16-day composites (Solano, Didan, Jacobson, & Huete, 2010). The NDVI mean and standard deviation across the year for each community was calculated by extracting the NDVI pixel values in each radial buffer zone. It was hypothesized that large mean NDVI values for a radial buffer would indicate areas of dense forest, while buffer zones with small standard deviations in NDVI would indicate villages with dense forest cover or very small amounts of vegetation cover (river islands, river banks) and buffer zones with large standard deviations in NDVI would indicate villages with multiple cropping patterns across the year.

Nine land use/land cover classes (Table 1) were derived from 30 m spatial resolution Landsat ETM+ data from October 2001 and March/April 2002. The date of the 2001 imagery was chosen to be prior to the main harvest in October which matched closely the time of census enumeration, and 2002 imagery helped to distinguish between different crop cycles (see below). Cloud and semi-transparent cloud were removed from each image using the method detailed in Watmough, Atkinson, and Hutton (2011). Cloud free data were atmospherically corrected and converted to reflectance using ATCOR-2 (Richter, 2007). Object-based analysis using eCognition 5.5 (Definiens, 2006) was used to classify the images. The segmentation step used a scale parameter of 8 (determines the size of the objects), the homogeneity criteria included a 90% emphasis on spectral homogeneity and 10% on shape and an equal weighting of 0.5 was given to compactness (perfectly compact object is a square) and smoothness (optimizes for smoothness of object boundaries). Segmentation was performed using Landsat ETM+ 2001 Bands 2–5 and 7 as well as the ETM+ NDVI values from 2001 to 2002. For the classification, a nearest neighbor classifier was used to classify bare land, grassland, *Kharif* and *Rabi* crops, water and woodland using within-scene samples with a minimum membership of 0.5 (Table 1). The classification accuracy of each class was in excess of 90%, estimated using a confusion matrix and independent random samples collected from within the image.

To classify *Rabi* crops, ETM+ data from 2001 to 2002 were used because the majority of areas were un-vegetated in the October 2001 scenes. Therefore, only using the 2001 imagery would have resulted in land being classified as bare ground. Combining the October 2001 images with images from March and April 2002 enabled the classification algorithms to distinguish between land that was bare ground throughout the period and land that was bare ground only once. An object that had a low or no vegetation signal in October and a strong vegetation signal in March–April was classified as *Rabi* (winter)

Table 1. *Land use and land cover class definitions applied to landsat ETM+ data*

Class	Object-based class definition
Bare ground	NN: μ and σ of Landsat ETM+ B2–B5 and B7 for years 2001 and 2002, μ and σ TC wetness and greenness
Built-up	Manual editing
Grassland	NN: μ and σ of Landsat ETM+ B2–B5 and B7, μ and σ TC wetness and greenness
Kharif crop (*sown in monsoon winter and harvested in October*)	NN: μ and σ of Landsat ETM+ B2–B5 and B7, μ and σ TC wetness and greenness, NDVI October 2001, NDVI April/March 2002
Plantation	NN: μ and σ of Landsat ETM+ B2–B5 and B7, μ and σ TC wetness and greenness, NDVI 2001, NDVI 2002, Area > 500,000 m^2
Rabi crop (*sown in November/ December and harvested April/ March*)	NN: μ and σ of Landsat ETM+ B2–B5 and B7, μ and σ TC wetness and greenness, NDVI October 2001, NDVI April/March 2002
Water	NN: μ Landsat ETM+ B5 and B7, μ TC wetness, NDVI 2001
Wetland	Manual classification
Woodland	NN: μ and σ of Landsat ETM+ B2–B5 and B7, μ and σ TC wetness and greenness, NDVI October 2001, NDVI April/March 2002

NN (Nearest Neighbor); μ (mean); σ (standard deviation), B2–B5 and B7 (Bands of the Landsat satellite data), TC (Tasseled Cap Transformation).

Crop. An object with a strong vegetation signal in October and low vegetation signal in March–April was classified as *Kharif* (summer) crop (Table 1).

(ii) *Estimating local environmental conditions*

The 1,000 m radial buffer zones were used to link the remotely sensed data with the census data. The "*isectpolyrast*" function in the Geospatial Modeling Environment (Beyer, 2010) was used to extract the proportion of land within each buffer zone that was covered by the land cover classes. This resulted in nine variables indicating the proportion of the buffer zone covered in each land cover. The buffer zones were also used to extract the mean and standard deviation of the NDVI for the 12 months of 2001. This resulted in two variables, one indicating the mean NDVI value for the year across the entire buffer zone and the other indicating the standard deviation of the NDVI value for the year across the entire buffer zone.

(iii) *Access to major towns*

A proxy for community remoteness was created by estimating travel time to major market towns for each community. Euclidean distances are inappropriate for this purpose in Assam as substantial water bodies create boundaries to travel. Thus, a least accumulative cost surface was created in ArcMap 9.3 (Esri, 2010) to estimate time (in hours) to travel from each village to the nearest major market town (defined in the 2001 Census). Road data from the Digital Chart of the World (DCW) were available from the DCW 1993 version at 1:1,000,000 scale which was the best data available in the public domain (Nelson, De Sherbinin, & Pozzi, 2006). The data contained the following road types: paved; un-paved; cart track; footpath; track pass. Roads were converted to a raster with 30-m spatial resolution and combined with the land cover classification. Average speeds for Assam were based on previous studies (Black *et al*., 2004; Noor *et al*., 2006) and are presented in Table 2. It was assumed that travel on national highways, paved roads, and un-paved roads would be undertaken using motor vehicles, while travel on paths, tracks, and across other land cover types would be on foot. Water was classified as "no data to ensure that it was treated as a barrier to travel. This had the effect of forcing the cost surface to take a route that involved no travel on water, following closely the patterns observed during field observation. Specifying water as a barrier to travel led to 1,507 villages with no estimated travel time as they were completely surrounded by water. The cost estimation was repeated for these villages using a speed of 1.5 km h^{-1} for water. The environmental metrics extracted from remotely sensed satellite sensor data are summarized in Table 3 along with the hypothesized relationships with village-level welfare.

Table 2. *Estimated travel speeds for different road and land use/cover types used to generate time to travel cost surface based on values in Black et al. (2004) and Noor et al. (2006)*

Cover/road type	Estimated speed (km h^{-1})
Bare land	4
Built-up	30
Grassland	4
Kharif crop (sown in rainy season and harvested October)	4
Plantation	3
Rabi crop (sown winter and harvested spring)	4
Water	0 (1.5 for villages having to use waterways in second pass)
Wetland	1.5
Woodland	3
National highway	60
Paved road	60
Un-paved road	30
Cart track, foot path, track-pass	5

(e) *Modeling welfare using satellite-derived environmental metrics*

Classification trees segment data into a number of regions by asking simple questions of the predictor space. They are simple to implement and interpret and do not assume a normal distribution. A further advantage to classification trees is that they are hierarchical in nature, meaning that each variable can be used multiple times allowing for non-linear relationships to be handled (Ghimire, Rogan, Galiano, Panday, & Neeti, 2012) which can be important for modeling population–environment relationships. However, classification trees tend to over-fit the data by creating complex rules. The random forest is an ensemble computer learning algorithm first developed by Breiman (2001) to extend the predictive power of classification trees (Breiman, 2001). The random forest algorithm constructs a series of classification trees and results

Table 3. *Remotely sensed environmental metrics used in the models and the hypothesized relationships they have with welfare*

Remote sensing variable	Hypothesized relationship with welfare
Time to travel to market	Linear and negative: higher travel associated with low welfare
Proportion of woodland cover in village	Nonlinear: high levels of woodland associated with low welfare; low levels of woodland associated with low welfare; moderate levels of woodland associated with higher welfare
Proportion of *Kharif* crop cover in village	Linear and positive: more land under summer crop cover associated with higher welfare
Proportion of *Rabi* crop cover in village	Linear and negative: more land under winter crop cover associated with low welfare as closer to river and associated hazards
Proportion of water cover in village	Linear and negative: high levels associated with low welfare as indicated by close proximity to rivers and maybe river islands
Proportion of bare and grass cover in village	Linear and negative: more land under these unproductive covers associated with lower welfare
NDVI mean	Linear and negative: higher mean NDVI indicates dense forest which is associated with low welfare.
NDVI Std Dev	Nonlinear: High values indicate multiple cropping which is associated with low welfare. Low values indicate dense forest and small levels of vegetation cover associated with low welfare. Moderate values associated with higher welfare as indicates a mix of vegetation and land covers

in multiple predictions for each data point (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012). The final predictions from the random forest model are derived using a voting system where each tree in the forest is allowed to cast a single vote on how each data point should be classified. The final classification is the most commonly occurring class (welfare quintile in this case) for each individual observation point (Breiman, 2001; Ghimire *et al.*, 2012). To reduce over-fitting, random forests use a bootstrap approach where each tree is constructed using around two thirds of the total available data (Biau, 2012; Rodriguez-Galiano *et al.*, 2012). Rather than considering every predictor variable at each decision or node, a random sample of predictor variables is also used (Ghimire *et al.*, 2012). Without this additional sampling approach of the predictor space the majority of the trees constructed would be similar, resulting in high levels of correlation within the model and reducing the reliability of predictions (Breiman, 2001; Rodriguez-Galiano *et al.*, 2012). Error is estimated using the out-of-bag (OOB) sample which is the error associated with the third of the data withheld from estimating an individual tree. This process is repeated hundreds or thousands of times

Table 4. *Mean values for independent variables in each wealth quintile*

	Market time (mins)	Bare land (%)	Grass (%)	Wood land (%)	Winter crop (%)	Wet land (%)	Summer crop (%)	Water (%)	Plant-ation (%)	NDVI mean
Lowest	133.2	3	4	60	3	1	24	4	1	0.64
Second	93	3	4	43	3	2	38	5	2	0.58
Middle	68.4	2	3	44	3	1	37	4	5	0.57
Fourth	58.8	2	3	45	2	1	40	4	3	0.59
Highest	44.4	1	3	46	1	1	43	3	2	0.58

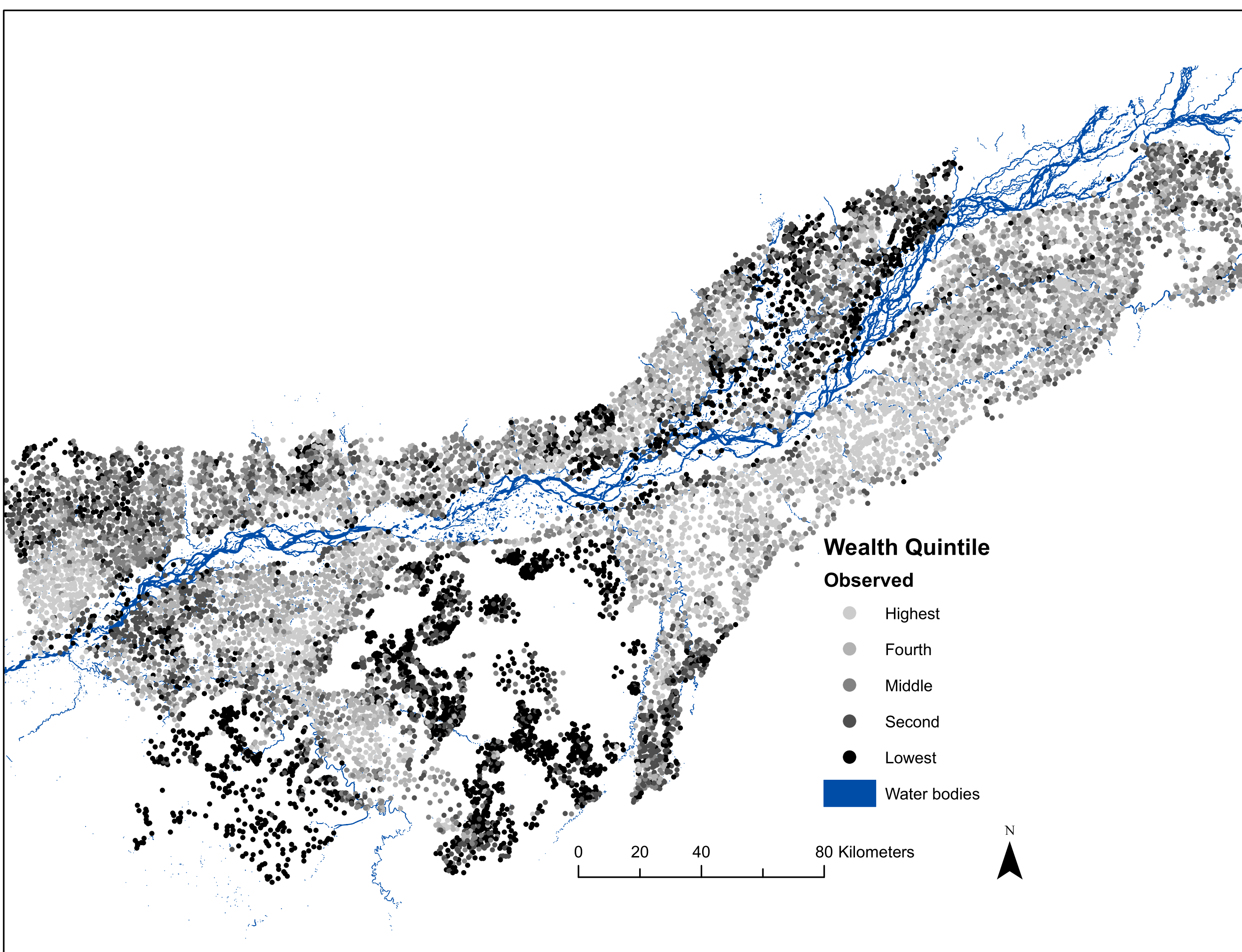

Figure 2. *Relative welfare scores estimated from the 2001 Indian census for Assam using the welfare equation in Section 3. Each point represents one of the 14,227 communities in the study region with darker points representing lower welfare scores.*

and the overall classification accuracy estimated by averaging the OOB error from each of the component trees. Convergence approaches such as random forests, bagging, and boosting have been found to increase model prediction accuracy (Ghimire *et al.*, 2012).

A random forests model was run on the entire Assam dataset of 14,227 samples using the "*randomForest*" package (Liaw & Wiener, 2014) in *R 3.0.3*. The use of withheld test data is not required due to random forests using bootstrapping with replacement. The number of random covariates considered in each node split is determined by *mtry* with the default value $\sqrt{p}$ appropriate for classification problems when the number of observations n is greater than the number of explanatory covariates p (Genuer, Poggi, & Tuleau, 2008; Ghimire *et al.*, 2012). The number of iterations of the model was set to 1,000 ($ntree = 1{,}000$) as increasing the number of trees in the random forests model reduces OOB errors and increases the stability of variable importance metrics (Genuer *et al.*, 2008). Overall accuracy was calculated using an averaged OOB error score and a classification confusion matrix to compare the observed welfare quintile (from the census data) with the quintile predicted using the random forests model for each village.

4. RESULTS

Summary statistics indicated that travel times to major towns are greater in regions with lower welfare scores (Table 4), which supports field observations. Villages in the lowest quintile have more woodland cover and less *Kharif* (summer) crop cover than the other four quintiles (Table 4), which also supports the field observations. There is very little *Rabi* (winter) crop, bare ground, and grassland cover in the data. However, summary statistics indicated that these cover types were highest in those communities from the lowest and second welfare quintiles (Table 4). Figure 2 shows the spatial coverage of welfare scores across the study site estimated from the census data. The lowest welfare quintiles (black) are often clustered in areas of dense forest, along river banks, on river islands, and in wetland regions. The higher welfare quintiles (brighter shades of gray) are often further from major rivers.

The random forests model gave an overall out-of-bag (OOB) error of 0.64, which is a predictive accuracy of 36%. The OOB misclassification error was 39% (accuracy 61%) for the lowest quintile and 43% (57% accuracy) for the highest quintile. However, the misclassification errors were 77%, 78%,

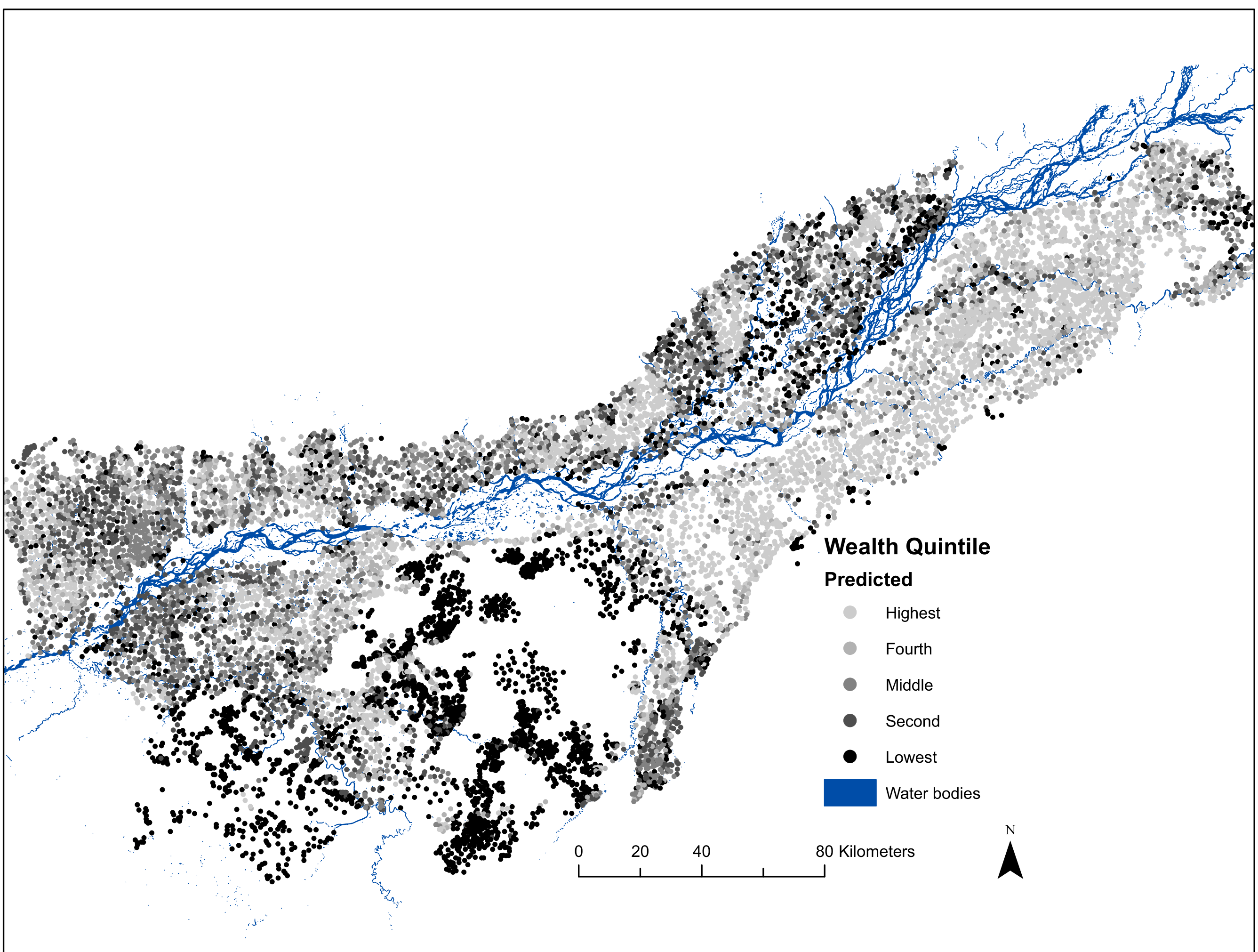

Figure 3. *Predicted welfare quintiles using the random forests model. Each point represents one of the 14,227 communities in the study region with darker points representing lower welfare scores.*

and 80% for the second, middle, and fourth welfare quintiles respectively. The model predicted accurately the low welfare areas in the dense forest region of Karbi Anglong and along the north bank of the Brahmaputra (Figures 3 and 4).

The out-of-bag mean decrease in accuracy and the mean decrease in the Gini Index (which is an average of the total amount by which the Gini Index decreases when a particular predictor variable is used in the model) can be used to indicate variable importance within the random forest model (Figure 5). The time to market town and proportion of woodland cover variables had the most important contribution to model prediction in terms of the OOB mean decrease in accuracy and Gini Index measures (Figure 5). Bare ground, grassland, winter cropland, and the standard deviation of the NDVI had considerably lower contributions to overall model prediction.

Table 5 indicates the importance of the contribution of each variable to the random forest classification for each welfare quintile. In most cases, time to market town and proportion of woodland cover produced the largest contribution to each welfare quintile and bare ground and grassland produced the smallest contribution. The breakdown of variable importance for each wealth quintile indicates which environmental metrics are related with different levels of welfare. Interpreting these results further is almost impossible using ensemble classifiers such as random forests (James, Witten, Hastie, & Tibshirani, 2013). Classification trees segment data into a number of regions by asking simple questions of the predictor space and allow us to interpret more easily how predictor variables are associated with welfare quintiles (Ghimire *et al.*, 2012). Therefore, for pedagogical purposes an exploratory classification tree was created using the '*tree*' function in *R 3.0.3* to examine the relationships between welfare quintiles and environmental predictor variables.

A k-fold cross-validation ($k = 10$) approach was used to identify the most parsimonious tree using the "*cv.tree*" function (Ripley, 2014). This process trains a tree using nine data folds and tests the accuracy of the tree using the remaining tenth fold. The process is repeated so that each data fold is withheld from model training and used to test the model accuracy once. The accuracy over the k repetitions is averaged and the optimal tree size (number of terminal nodes) is calculated by comparing the change in model residual deviance for different numbers of tree nodes. The full classification tree was pruned using the "*prune.misclass*" function to contain only the optimal number of terminal nodes. The pruned or optimal tree indicated that time to market town was the most important predictor of welfare followed by proportion of woodland cover (Figure 6), which supports the random forest model output.

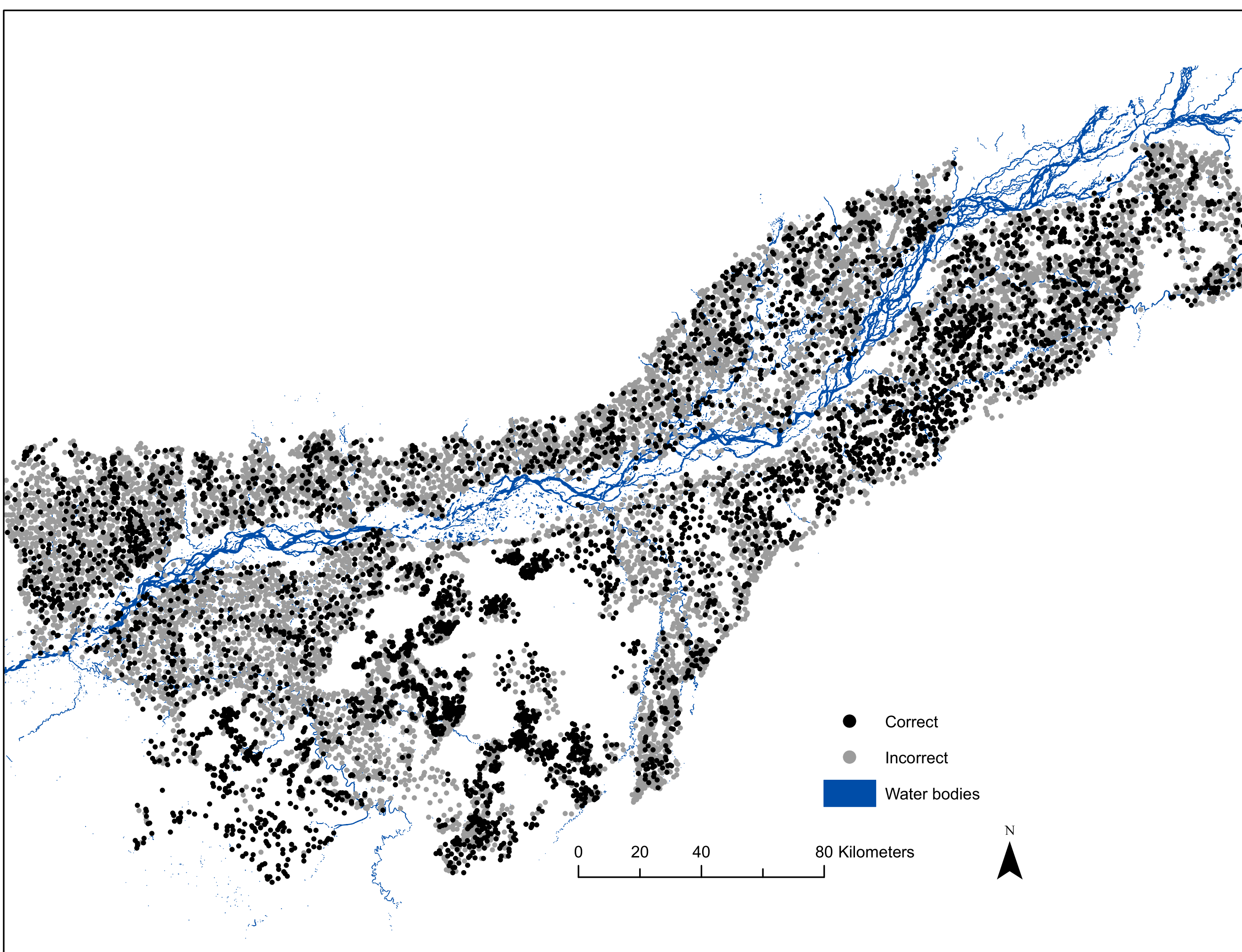

Figure 4. *Comparison of observed versus predicted welfare quintiles across Assam. Each point represents a community in Assam (n = 14,227) the dark points indicate where the random forests model correctly predicted the welfare quintile and lighter shade indicates incorrectly predicted welfare quintiles.*

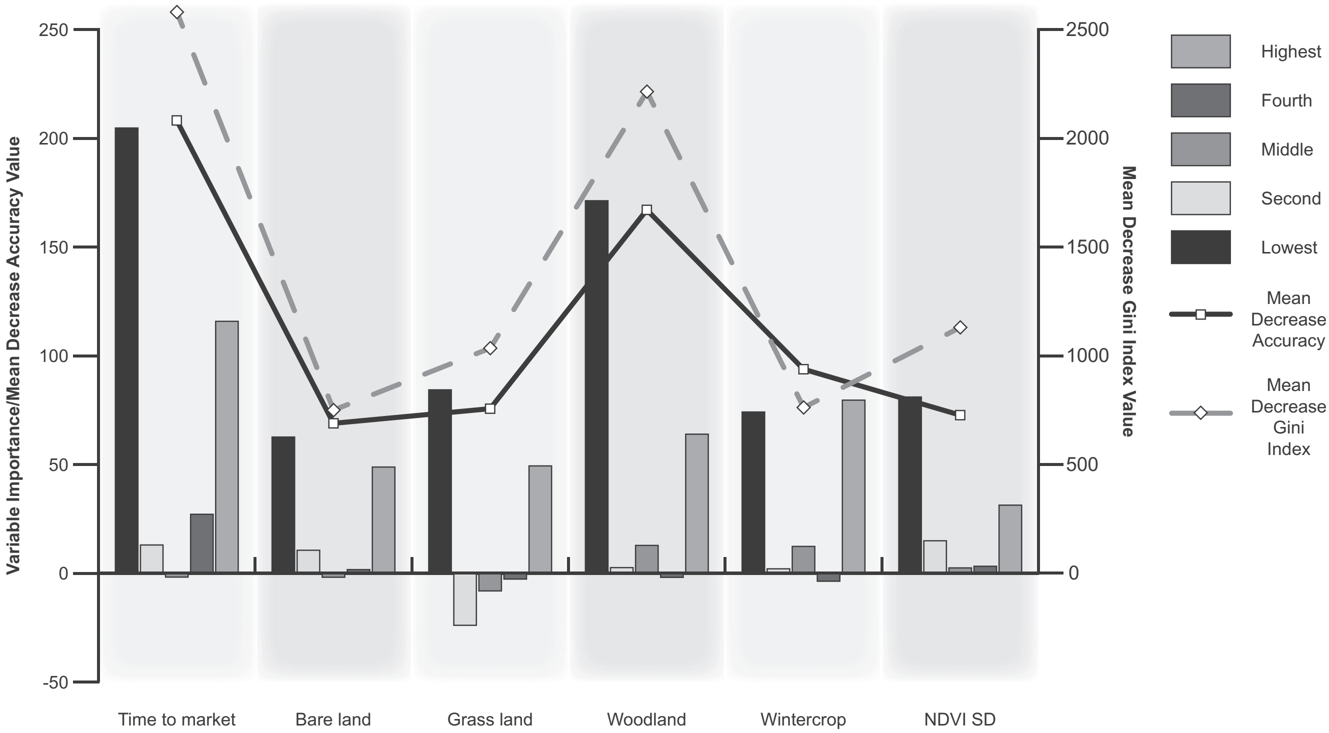

Figure 5. *Results of the random forest model. Variable importance per welfare quintile is displayed in the columns. The solid line shows the mean decrease in accuracy in the overall model associated with removing each of the environmental variables from the model. The dashed line represents the mean decrease in Gini Index of the entire model associated with removing each of the environmental variables from the model.*

Table 5. *Variable importance in terms of out of bag mean decrease in accuracy per welfare quintile. Ranking of importance from top to bottom*

Lowest	Second	Middle	Fourth	Highest
Time	NDVI mean	Woodland	Time	Time
Woodland	Time	Winter cropland	NDVI mean	Winter cropland
Grass land	Bare ground	NDVI mean	Bare ground	Woodland
NDVI mean	Woodland	Time	Woodland	Grass land
Winter cropland	Winter cropland	Bare ground	Grass land	Bare ground
Bare ground	Grassland	Grass land	Winter cropland	NDVI mean

The classification tree also indicates that villages over one hour from a major town were classified as belonging to the two lowest welfare quintiles and all villages within one hour of a major market town belonged to the middle or highest groups. Villages over one hour from a major market town and containing over 80% woodland cover were in the lowest welfare quintile while those with less than 80% woodland cover were in the second welfare quintile. Villages within an hour of a major market town and containing over 35% woodland cover were in the highest welfare quintile. Villages within an hour of a market town that had less than 35% woodland cover and zero *Rabi* crop cover also belonged to the highest welfare quintile while those with less than 35% woodland cover, but some winter crop cover belonged to the middle welfare quintile.

5. DISCUSSION AND CONCLUSIONS

The relationship between poverty and environment was investigated for an extensive area in India. The analysis used a random forests approach applied to a relative welfare index and environmental metrics derived from satellite sensor data. Time to major market town, woodland cover, and to a lesser extent winter crop cover were associated with welfare. Overall prediction accuracy was 36%, but this increased to 61% and 57% when predicting the lowest and highest welfare quintiles, respectively. The results indicated that satellite sensor data are strongly associated with aspects of rural welfare for an extensive region of a developing country. For example, as a predictor of poverty (the lowest welfare quintile), the method achieved an accuracy of 61%, which is very high when it is considered that the environmental information was derived only from non-temporal satellite sensor imagery at a medium spatial resolution of 30 m.

The overall accuracy of the random forests model of 36% was comparable to that achieved in a similar study by Tooke *et al.* (2010) in Canadian cities and is promising considering that welfare was modeled using only environmental metrics derived from satellite sensor data. The high prediction accuracy of the lowest (61%) and highest (57%) welfare quintiles indicated that there are large relationships between the extremes of welfare in Assam and environmental metrics derived from satellite sensor data. The high prediction accura-

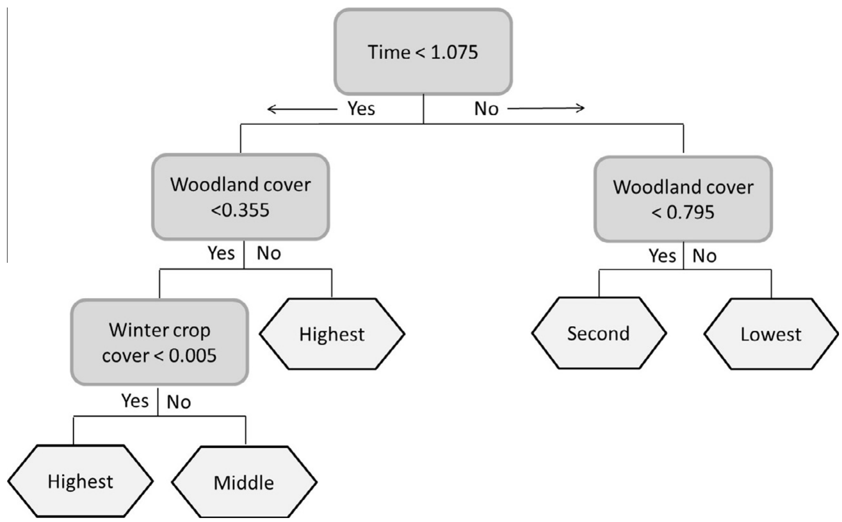

Figure 6. *Optimal tree derived by cross-validation. The first binary split (Time < 1.075) indicates that time to major market town was the most important predictor of wealth. Those observations that were above this threshold (ie were over 1 h from a market town) were sent to the right of the split and those that fell within the threshold (i.e., with 1 h of a market town) were sent to the left of the split. The majority of villages that were over 1 h from a market town belonged to the lowest or second lowest welfare quintiles, those villages within an hour of a market town belonged to higher welfare quintiles.*

cies for low and high welfare groups were also comparable to previous studies using satellite sensor data to predict poverty in North America (Ogneva-Himmelberger, Pearsall, & Rakshit, 2009; Pearsall & Christman, 2012).

The lower accuracies for the three middle welfare quintiles are likely due to the way that they were defined. Splitting the welfare scores into quintiles produced five categories each containing the same number of observations or villages. This supported statistical modeling, but the boundaries between quintiles were drawn arbitrarily and resulted in only small differences between the second and middle categories and the middle and fourth categories. Collapsing the quintiles into three categories (lowest, middle, and highest) where the middle category contained second, middle, and fourth welfare scores caused the model to predict only the middle categories as this optimized accuracy (60% of the observations were correctly classified). Alternative approaches to discretization could be explored in future.

The research would benefit from comparing the welfare index to other measures of poverty and welfare in India to assess its sensitivity, but these data are not currently available at similar spatial resolutions. In future, the welfare index could be combined with variables from the Agricultural Census or the National Living Standards Survey. The index assumed that the water source for every home within a village was the same. Aggregating water source in this way is not ideal as there will be variation within communities. However, access to these community goods is often more common for wealthy communities and households and the provision of public services can increase wealth and socioeconomic position (Rutstein, 2004). Furthermore, the current models only used local environmental conditions within a 1,000 m buffer zone. Poverty–environment relationships are complex (De Sherbinin *et al.*, 2007, 2008) and, therefore, future research should seek to include multi-level environmental data from other spatial extents to represent micro-and meso-scale environmental conditions. It is currently not possible to compare these results with other studies that quantified the contribution of environmental resources to household incomes such as Angelsen *et al.* (2014). Future research should seek to compare census data with income variables from surveys similar to those in CIFOR's PEN study to quantify how census data could be used to upscale the high-resolution household survey datasets to regional scales.

Isolated communities are often found to have higher levels of poverty (Sen, 2003; Stifel & Minten, 2008) and improving access to markets has been associated with increased economic diversity (Dewi, Belcher, & Puntodewo, 2005). The power of roads to reduce travel times and improve access mean that large amounts of international development assistance is directed toward road building schemes (World Bank., 2011b). Results in this study indicated that access to major market towns was the most important variable for predicting welfare and many of the correctly predicted village welfare quintiles are clustered in parts of Assam with poor access (Figure 4). Furthermore, the classification tree indicated that the hypothesis in Table 4 was correct as villages over one hour from a major town belonged to the two lowest welfare categories while villages within an hour belonged to higher welfare categories (Figure 3). Replacing time in the model with straight-line distance to roads and major towns did not result in these metrics being the most important variables (model not shown). This likely indicates that rivers and waterways within Assam act as barriers to travel and significantly increase the effort required to travel. These barriers were accounted for in the cost surface estimation, but not in straight-line distance calculations. This result indicates the importance of estimating access to a particular location over a simple Euclidean distance metric which does not account for the effort required.

Woodland and non-timber forest products can contribute significantly to incomes and subsistence in rural areas (Babulo *et al.*, 2009; Mamo *et al.*, 2007; Sunderlin *et al.*, 2005; Wunder *et al.*, 2014). In Assam forest and woodland have historically been closely related with poverty (Saikia, 2005). We predicted that the relationship between woodland and poverty would be complex and non-linear (Table 4). Past research has indicated that woodland services can be a fallback/safety-net option during shocks such as drought and reduced agricultural yields (Wunder *et al.*, 2014). It can also be a significant and regular source of income (Babulo *et al.*, 2009) or in some cases the only source of income. This complexity is somewhat mirrored in the results from Assam.

In terms of the random forest predictions, woodland had a large contribution to the lowest and middle welfare groups but the contribution was lower for the second, fourth, and highest welfare groups (Table 5). The classification tree indicated that villages over one hour from a major town with the highest levels of woodland cover were in the lowest welfare quintile. In areas that were within an hour of a major town, woodland cover was higher in the highest welfare quintile (Figure 4). This was supported by field observations which suggested that some of the poorest communities were located in densely forested regions. However, wealthier communities often had more land which was less intensively cultivated, increasing the likelihood of some woodland cover in wealthy areas. This illustrates that woodland cover can have a bimodal interpretation in Assam, depending on the context. The results also indicate the importance of using a non-linear approach to modeling population-environment relationships. This reflects, to some extent, the discussion in the introduction about conflicting interpretations of vegetation greenness and poverty relationships in Lafary *et al.* (2008) and Ryznar and Wagner (2001). The results here indicate that the random forests approach was able to handle and give meaning to the dual and opposing association between welfare and woodland which would likely be missed using linear models.

Past studies have found significant positive relationships between wealth and NDVI (Lafary *et al.*, 2008; Tooke *et al.*, 2010). NDVI in Assam had relatively low variable importance. There are two potential explanations for this; (i) NDVI was highly correlated with the woodland cover variable which would potentially mask any effect the NDVI had on the welfare since woodland was the second most important variable, and; (ii) the relationship between a vegetation measure like NDVI and welfare is perhaps different to that in urban and suburban areas of developed countries. Assam has a strong reliance on agriculture and, therefore, the NDVI signal may be more consistent due to higher levels of vegetation cover across space. Perhaps relationships would be different if longer-term NDVI trends were measured. For example, large and unexpected deviations from these longer-term averages may indicate areas that have suffered a shock such as drought or erosion, whereas longer-term changes in the trends may indicate a region that is experiencing income diversification.

Field observations indicated that *Rabi* (winter) crops were rarely grown in wealthy areas of Assam as these areas were further from river irrigation water, had larger amounts of land available per household and had access to additional sources of non-farm income (Watmough *et al.*, 2013). The variable importance values (Table 5) and the classification tree indicated that winter crop cover was associated with communities belonging to the middle welfare quintile which was somewhat unexpected (Table 4). However, poverty is often found to be associated with lacking the resources to take advantage of available productive assets as well as lack of adequate productive assets (Barrett, 2005). The result likely indicated a similar pattern in Assam. Perhaps, communities from the lowest and second quintiles either did not have access to agricultural land or did not have the assets available to make use of the land that was available. Field observations revealed that in most areas *Rabi* crops were vegetables grown for sale in markets. These communities have complex welfare structures as they were located along the banks of rivers in areas that are vulnerable to flooding, deposition, and erosion damage. However, the communities appeared to use the close proximity to rivers to provide irrigation water and concentrated on accumulating cash reserves over human and physical capital such as education and consumer assets. Insufficient information was collected on these regions during fieldwork and more research is needed to investigate how multiple cropping patterns are associated with wealth in Assam.

The cover of bare ground and grassland were intended as proxies for unproductive areas and water as a proxy for river bank and river island locations (Table 4). However, these variables had low relative importance in the random forest model and none of these variables featured in the classification. The intended purpose of these land covers could be problematic. Grassland may not have been unproductive land but rather agricultural land under crop cycles that differed from the main *Kharif* and *Rabi* crops; using just a snap-shot of remote sensing data would have missed this information. Water was intended to indicate areas along river banks and islands that may suffer from shocks due to flooding, erosion, and deposition. However, some regions of Assam have flood defenses which were not considered in this work. Furthermore, the resolution of the imagery used may mean that other water resources such as irrigation canals were missed.

6. POLICY IMPLICATIONS: USING REMOTELY SENSED SATELLITE DATA TO EXPAND THE EVIDENCE BASE OF POVERTY–ENVIRONMENT RELATIONSHIPS

Understanding of poverty–environment relationships is driven by a series of often highly detailed case studies. However, research by Dasgupta *et al.* (2005) indicated that poverty–environment relationships can differ spatially. Furthermore, Wunder *et al.* (2014) demonstrate that, while highly informative, individual case studies can have a limited scope for wider scale policy development. Sustainable poverty alleviation will require multi-dimensional interventions (Frost *et al.*, 2007) of which environmental considerations are but one. Therefore, future research should aim to increase our understanding of poverty–environment relationships across wider spatial extents. The results of the random forests model are coherent and logical which mean that we can begin to investigate and potentially understand poverty–environment relationships on a larger regional scale. Furthermore, classification trees and random forest models are able to deal with the non-linearity present in poverty–environment relationships.

The results also mean that we can begin to think about interventions to managing environmental resources which could contribute to development and to understand what changes are likely in environmental resources when development occurs. Therefore, this paper, by linking poverty to environment variables from space on a regional scale opens new possibilities for exploring population–environment relationships and environmental management for poverty alleviation. More research is required, particularly in; (i) application of the method to explore poverty–environment relationships in rural areas in other developing regions of the world; (ii) coupling the environment data supplied through satellite remote sensing with additional sources of information, for example, from the Agricultural Census and the National Living Standards Survey, and; (iii) extending the method to include temporal remotely sensed information and household panel surveys.

REFERENCES

Alkire, S. & Seth, S. (2013a) Multidimensional Poverty Reduction in India between 1999 and 2006: Where and How? *OPHI Working Paper Series*, 42.

Alkire, S., & Santos, M. E. (2014). Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index. *World Development, 59*, 251–274.

Alkire, S., & Seth, S. (2013b). Selecting a targeting method to identify BPL households in India. *Social Indicators Research, 112*, 417–446.

Angelsen, A. & Wunder, S. (2003) Exploring the forest–poverty link: Key concepts, issues and research implications. *CIFOR Occasional Paper*. Bogor, Indonesia: CIFOR.

Angelsen, A., Jagger, P., Babigumira, R., Belcher, B., Hogarth, N. J., Bauch, S., ... Wunder, S. (2014). Environmental income and rural livelihoods: A global-comparative analysis. *World Development, 64* (Supplement 1), S12–S28.

Asfaw, S., Kassie, M., Simtowe, F., & Lipper, L. (2012). Poverty reduction effects of agricultural technology adoption: A micro-evidence from rural Tanzania. *The Journal of Development Studies, 48*, 1288–1305.

Babulo, B., Muys, B., Nega, F., Tollens, E., Nyssen, J., Deckers, J., *et al.* (2009). The economic contribution of forest resource use to rural livelihoods in Tigray, Northern Ethiopia. *Forest Policy and Economics, 11*, 109–117.

Barbier, E. B. (2010). Poverty, development, and environment. *Environment and Development Economics, 15*, 635–660.

Barrett, C. B. (2005). Rural poverty dynamics: Development policy implications. *Agricultural Economics, 32*, 45–60.

Barrett, C. B., Marenya, P. P., Mcpeak, J., Minten, B., Murithi, F., Oluoch-Kosura, W., ... Wangila, J. (2006). Welfare dynamics in rural Kenya and Madagascar. *The Journal of Development Studies, 42*, 248–277.

Baruah, D. C., Das, P. K., & Dutta, P. K. (2004). Present status and future demand for energy for bullock-operated paddy-farms in Assam (India). *Applied Energy, 79*, 145–157.

Beyer, H. L. (2010) Geospatial Modelling Environment. *Spatial Ecology*. http://www.spatialecology.com/gme/: Beyer, H.L.

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research, 13*, 1063–1095.

Black, M., Ebener, S., Najera Aguilar, P., El Morjani, Z., Ray, N. & Vidaurre, M. (2004) Using GIS to measure Physical Accessibility to Health Care. *International Health Users Conference*. Washington DC: WHO-PAHO.

Blaikie, P., Cameron, J., & Seddon, D. (2002). Understanding 20 years of change in West-Central Nepal: Continuity and change in lives and ideas. *World Development, 30*, 1255–1270.

Booysen, F., Van Der Berg, S., Burger, R., Maltitz, M. V., & Rand, G. D. (2008). Using an asset index to assess trends in poverty in seven sub-saharan African countries. *World Development, 36*, 1113–1130.

Borooah, V. K. (2005). Caste, inequality, and poverty in India. *Review of Development Economics, 9*, 399–414.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brocklesby, M. A., & Hobley, M. (2003). The practice of design: Developing the chars livelihoods programme in Bangladesh. *Journal of International Development, 15*, 893–909.

Campbell, B. (1996) *The Miomboin Transition: Woodlands and Welfare in Africa, Bogor*, Indonesia, Center for International Forestry Research (CIFOR).

Cavendish, W. (2000). Empirical regularities in the poverty–environment relationship of rural households: Evidence from Zimbabwe. *World Development, 28*, 1979–2003.

Daimari, P. & Mishra, S. K. (2005) Poverty and inequality in rural Assam: An indicative study of seven villages in Udalguri sub-division, Assam (India). *North Eastern Hill University Working Paper Department of Economics, NOrth Eastern Hill University*.

Dasgupta, S., Deichmann, U., Meisner, C., & Wheeler, D. (2005). Where is the poverty–environment nexus? Evidence from Cambodia, Lao PDR, and Vietnam. *World Development, 33*, 617–638.

Datt, G., & Ravallion, M. (1998). Farm productivity and rural poverty in India. *The Journal of Development Studies, 34*, 62–85.

De Janvry, A., & Sadoulet, E. (2010). Agricultural growth and poverty reduction: Additional evidence. *The World Bank Research Observer, 25*, 1–20.

De Sherbinin, A. (2011). The biophysical and geographical correlates of child malnutrition in Africa. *Population, Space and Place, 17*, 27–46.

De Sherbinin, A., Carr, D., Cassels, S., & Jiang, L. (2007). Population and environment. *Annual Review of Environment and Resources, 32*, 345–373.

De Sherbinin, A., Vanwey, L. K., Mcsweeney, K., Aggarwal, R., Barbieri, A., Henry, S., ... Walker, R. (2008). Rural household demographics, livelihoods and the environment. *Global Environmental Change, 18*, 38–53.

Definiens (2006) eCognition. http://ecognition.com/products.

Dewi, S., Belcher, B., & Puntodewo, A. (2005). Village economic opportunity, forest dependence, and rural livelihoods in East Kalimantan, Indonesia. *World Development, 33*, 1419–1434.

Diao, X., Hazell, P., & Thurlow, J. (2010). The role of agriculture in African development. *World Development, 38*, 1375–1383.

Entwisle, B., Walsh, S. J., Rindfuss, R. R., & Vanwey, L. K. (2005). Population and upland crop production in Nang Rong, Thailand. *Population and Environment, 26*(6), 449–470.

Esri (2010). *ArcMap: Release 9.3.*. Redlands, CA: Environmental Systems Research Institute.

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the green revolution, 1960 to 2000. *Science, 300*, 758–762.

Fan, S., Hazell, P., & Haque, T. (2000). Targeting public investments by agro-ecological zone to achieve growth and poverty alleviation goals in rural India. *Food Policy, 25*, 411–428.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of India. *Demography, 38*, 115–132.

Frost, P., Campbell, B., Luckert, M., Mutamba, M., Mandondo, A., & Kozanayi, W. (2007). In search of improved rural livelihoods in semi-arid regions through local management of natural resources: Lessons from case studies in Zimbabwe. *World Development, 35*, 1961–1974.

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006a). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health, 60*, 7–12.

Genuer, R., Poggi, J.-M., & Tuleau, C. (2008). Random forests: Some methodological insights. *Institute National De Recherche En Informatique Et En Automatique, 6729*, 35.

Ghimire, B., Rogan, J., Galiano, V. R., Panday, P., & Neeti, N. (2012). An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GIScience & Remote Sensing, 49*, 623–643.

Goalobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006b). Indicators of socioeconomic position (part 2). *Journal of Epidemiology and Community Health, 60*, 95–101.

Government of Assam (2003). Assam Human Development Report. In *Planning and Development Department*. Dispur Guwahati: Government of Assam.

Gwatkin, D. R., Rutstein, D., Johnson, K., Pande, R. P. & Wagstaff, A. (2007) Socio-economic differences in health, nutrition and population. *Publication of the HNP/Poverty Thematic Group of the World Bank*.

Hannum, E., & Buchmann, C. (2005). Global educational expansion and socio-economic development: An assessment of findings from the social sciences. *World Development, 33*, 333–354.

Houweling, T. A. J., Kunst, A. E., & Mackenbach, J. P. (2003). Measuring health inequality among children in developing countries: Does the choice of indicator of economic status matter?. *International Journal of Equity in Health, 2*.

Howe, L. D., Hargreaves, J. R., & Huttly, S. R. (2008). Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging Themes in Epidemiology, 5*.

Huang, Q., Rozelle, S., Lohmar, B., Huang, J., & Wang, J. (2006). Irrigation, agricultural performance and poverty reduction in China. *Food Policy, 31*, 30–52.

Hussain, I. (2007). Poverty-reducing impacts of irrigation: Evidence and lessons. *Irrigation and Drainage, 56*, 147–164.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York: Springer, 426.

Jamir, T., Gadgil, A. S., & De, U. S. (2008). Recent floods related natural hazards over West coast and Northeast India. *Journal of Indian Geophysical Union, 12*, 179–182.

Jelinski, D., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology, 11*, 129–140.
Jensen, R., Gatrell, J. D., Boulton, J., & Harper, B. (2004). Using remote sensing and geographic information systems to study urban quality of life and urban forest amenities. *Ecology and Society, 9*, 5.
Jha, R. (2008). Economic reforms and human development indicators in India. *Asian Economic Policy Review, 3*, 290–310.
Kamanga, P., Vedeld, P., & Sjaastad, E. (2009). Forest incomes and rural livelihoods in Chiradzulu District, Malawi. *Ecological Economics, 68*, 613–624.
Kassie, M., Shiferaw, B., & Muricho, G. (2011). Agricultural technology, crop income, and poverty alleviation in Uganda. *World Development, 39*, 1784–1795.
Khandker, S. R., Zaid, B., & Gayatri, B. (2006). The poverty impact of rural roads: Evidence from Bangladesh. In World Bank (Ed.). Washington, DC: World Bank.
Krishna, A. (2004). Escaping poverty and becoming poor: Who gains, who loses, and why?. *World Development, 32*, 121–136.
Krishna, A. (2006). Pathways out of and into poverty in 36 villages of Andhra Pradesh, India. *World Development, 34*, 271–288.
Lafary, E. W., Gatrell, J. D., & Jensen, R. R. (2008). People, pixels and weights in Vanderburgh County, Indiana: Toward a new urban geography of human–environment interactions. *Geocarto International, 23*, 53–66.
Liaw, A. & Wiener, M. (2014) Package 'randomForest': Breiman and Cutler's random forests for classification and regression. *In:* Team, R. D. C. (ed.) *R package.* 4.6-10 ed. http://cran.r-project.org/web/packages/randomForest/randomForest.pdf: R Development Core Team.
Ligon, E. & Sadoulet, E. (2008) Estimating the effects of aggregate agricultural growth on the distribution of expenditures *Background paper for the World Development Report* 2008 [Online].
Liverman, D. M., & Cuesta, R. M. R. (2008). Human interactions with the Earth system: People and pixels revisited. *Earth Surface Processes and Landforms, 33*, 1458–1471.
Liverman, D., Moran, E. F., Rindfuss, R. R., & Stern, P. C. (1998). *People and pixels: Linking remote sensing and social science*. Washington D.C.: National Academy Press, 256.
Lo, C. P., & Faber, B. J. (1997). Integration of landsat thematic mapper and census data for quality of life assessment. *Remote Sensing of Environment, 62*, 143–157.
Mamo, G., Sjaastad, E., & Vedeld, P. (2007). Economic dependence on forest resources: A case from Dendi District, Ethiopia. *Forest Policy and Economics, 9*, 916–927.
Mehta, A. K. (2003) Multidimensional poverty in India: District level estimates. *CPRC:IIPA Working Paper*, Chronic Poverty Research Centre, Manchester.
Mennis, J. (2006). Socioeconomic-vegetation relationships in urban residential land: The case of Dever, Colorado. *Photogrammetric Engineering and Remote Sensing, 72*, 911–921.
Minten, B., & Barrett, C. B. (2008). Agricultural technology, productivity, and poverty in Madagascar. *World Development, 36*, 797–822.
Montgomery, M. R., Gragnolati, M., Burke, K. A., & Paredes, E. (2000). Measuring living standards with proxy variables. *Demography, 37*, 155–174.
Nelson, A., De Sherbinin, A., & Pozzi, F. (2006). Towards development of a high quality public domain global roads database. *CODATA Science Journal, 5*, 223–265.
Noor, A. M., Amin, A. A., Gething, P. W., Atkinson, P. M., Hay, S. I., & Snow, R. W. (2006). Modelling distances travelled to government health services in Kenya, Modélisation des distances parcourues pour accéder aux services de santé gouvernementaux au Kenya, Modelaje de las distancias viajadas a las unidades sanitarias gubernamentales en Kenya. *Tropical Medicine & International Health, 11*, 188–196.
O'brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., Javed, A., Bhadwal, S., Barg, S., Nygaard, L. & West, J. (2004) Mapping vulnerability to multiple stressors: climate change and globalization in India. Global Environmental Change, 14, 303-313.
Ogneva-Himmelberger, Y., Pearsall, H., & Rakshit, R. (2009). Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts. *Applied Geography, 29*, 478–487.
Okwi, P. O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A., ... Owuor, J. (2007). Spatial determinants of poverty in rural Kenya. *Proceedings of the National Academy of Sciences, 104*, 16769–16774.
Pauw, K., & Thurlow, J. (2011). Agricultural growth, poverty, and nutrition in Tanzania. *Food Policy, 36*, 795–804.
Pearsall, H., & Christman, Z. (2012). Tree-lined lanes or vacant lots? Evaluating non-stationarity between urban greenness and socio-economic conditions in Philadelphia, Pennsylvania, USA at multiple scales. *Applied Geography, 35*, 257–264.
Rashed, T., Weeks, J. R., Roberts, D., Rogan, J., & Powell, R. (2003). Measuring the physical composition of urban morphology using multiple endmember spectral mixture models. *Photogrammetric Engineering & Remote Sensing, 69*, 1011–1020.
Registrar General and Census Commissioner (2011) Census Data 2001 - Metadata *In:* Registrar General and Census Commissioner (ed.). http://www.censusindia.gov.in/Metadata.Metada.htm#2i: Government of India.
Richter, R. (2007) Atmospheric/Topographic Correction for Satellite Imagery: ATCOR 2/3 User Guide. Weissling, Germany: DLR German Aerospace Centre.
Ripley, B. (2014) tree: Classification and Regression Trees *In:* Team, R. D. C. (ed.) 1.0-35 ed. http://CRAN.R-project.org/package=tree: R Development Core Team.
Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104.
Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning, 61*, 301–325.
Rutstein, S. O. & Johnson, K. (2004) The DHS Wealth Index. *DHS Comparative Reports No. 6* [Online].
Ryznar, R. M., & Wagner, T. W. (2001). Using remotely sensed imagery to detect urban change: Viewing detroit from space. *Journal of the American Planning Association, 67*, 327–336.
Saikia, A. (2005) Jungles, Reserves, Wildlife: A History of Forests in Assam, Guwahati, India, Wildlife Areas Development and Welfare Trust Assam 372.
Saikia, A. (2014). *A century of protests: Peasant politics in Assam since 1900*. New Delhi: Routledge, 500.
Sarkar, S., Mishra, S., Dayal, H., & Nathan, D. (2006). Development and deprivation of scheduled tribes. *Economic and Political Weekly, 41*, 4824–4827.
Sarma, J. N. (2005). Fluvial process and morphology of the Brahmaputra River in Assam, India. *Geomorphology, 70*, 226–256.
Sen, B. (2003). Drivers of escape and descent: changing household fortunes in rural Bangladesh. *World Development, 31*, 513–534.
Solano, R., Didan, K., Jacobson, A., & Huete, A. (2010). *MODIS vegetation index user's guide (MOD13 Series)*. Vegetation and Phenology Lab, University of Arizona.
Stifel, D., & Minten, B. (2008). Isolation and agricultural productivity. *Agricultural Economics, 39*, 1–15.
Sunderlin, W. D., Angelsen, A., Belcher, B., Burgers, P., Nasi, R., Santoso, L., *et al.* (2005). Livelihoods, forests, and conservation in developing countries: An overview. *World Development, 33*, 1383–1402.
Tooke, T. R., Klinkenber, B., & Coops, N. C. (2010). A geographical approach to identifying vegetation-related environmental equity in Canadian cities. *Environment and Planning B: Planning and Design, 37*, 1040–1056.
Twumasi, Y., & Merem, E. (2006). GIS and Remote Sensing Applications in the Assessment of Change within a Coastal Environment in the Niger Delta Region of Nigeria. *International Journal of Environmental Research and Public Health, 3*, 98–106.
Vedeld, P., Angelsen, A., Bojö, J., Sjaastad, E., & Kobugabe Berg, G. (2007). Forest environmental incomes and the rural poor. *Forest Policy and Economics, 9*, 869–879.
Watmough, G. R., Atkinson, P. M., & Hutton, C. W. (2011). A combined spectral and object-based approach to transparent cloud removal in an operational setting for Landsat ETM+. *International Journal of Applied Earth Observation and Geoinformation, 13*, 220–227.
Watmough, G. R., Atkinson, P. M., & Hutton, C. W. (2013). Predicting socioeconomic conditions from satellite sensor data in rural developing countries: A case study using female literacy in Assam, India. *Applied Geography, 44*, 192–200.

Weeks, J. R., Getis, A., Stow, D. A., Hill, A. G., Rain, D., Engstrom, R., ... Ofiesh, C. (2012). Connecting the dots between health, poverty and place in Accra, Ghana. *Annals of the Association of American Geographers, 102*, 932–941.

World Bank. (2011a) India: Priorities for Agriculture and Rural Development [Online]. Available: http://go.worldbank.org/8EFXZBL3Y0 [Accessed 8th August 2011.

World Bank. (2011b) *Transport* [Online]. Washington D.C.: The World Bank Group. Available: http://go.worldbank.org/H4FAFOWTU0 [Accessed 20th October 2012].

Wunder, S., Börner, J., Shively, G., & Wyman, M. (2014). Safety nets, gap filling and forests: A global-comparative perspective. *World Development, 64*(Supplement 1), S29–S42.