Journal of Environmental Management 313 (2022) 114950

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Earth observation and geospatial data can predict the relative distribution of village level poverty in the Sundarban Biosphere Reserve, India

Charlotte L.J. Marcinko [a,*], Sourav Samanta [b], Oindrila Basu [b], Andy Harfoot [c], Duncan D. Hornby [c], Craig W. Hutton [c], Sudipa Pal [b], Gary R. Watmough [d,e,**]

[a] School of Engineering, University of Southampton, Southampton, SO17 1BJ, UK
[b] School of Oceanographic Studies, Jadavpur University, 188, Raja S.C. Mallik Road, Jadavpur, Kolkata, 700032, India
[c] School of Geography and Environmental Sciences, University of Southampton, Southampton, SO17 1BJ, UK
[d] School of Geosciences, Institute of Geography, University of Edinburgh, Drummond Street, Edinburgh, Scotland, United Kingdom
[e] Global Academy of Agriculture and Food Security, University of Edinburgh, Easter Bush Campus, Edinburgh, Scotland, United Kingdom

ARTICLE INFO

Keywords:
SDGs
Random forest
Poverty
Population environment
Remote sensing
Socio-ecological systems

ABSTRACT

There is increasing interest in leveraging Earth Observation (EO) and geospatial data to predict and map aspects of socioeconomic conditions to support survey and census activities. This is particularly relevant for the frequent monitoring required to assess progress towards the UNs' Sustainable Development Goals (SDGs). The Sundarban Biosphere Reserve (SBR) is a region of international ecological importance, containing the Indian portion of the world's largest mangrove forest. The region is densely populated and home to over 4.4 million people, many living in chronic poverty with a strong dependence on nature-based rural livelihoods. Such livelihoods are vulnerable to frequent natural hazards including cyclone landfall and storm surges. In this study we examine associations between environmental variables derived from EO and geospatial data with a village level multi-dimensional poverty metric using random forest machine learning, to provide evidence in support of policy formulation in the field of poverty reduction. We find that environmental variables can predict up to 78% of the relative distribution of the poorest villages within the SBR. Exposure to cyclone hazard was the most important variable for prediction of poverty. The poorest villages were associated with relatively small areas of rural settlement (<~30%), large areas of agricultural land (>~50%) and moderate to high cyclone hazard. The poorest villages were also associated with less productive agricultural land than the wealthiest. Analysis suggests villages with access to more diverse livelihood options, and a smaller dependence on agriculture may be more resilient to cyclone hazard. This study contributes to the understanding of poverty-environment dynamics within Low-and middle-income countries and the associations found can inform policy linked to socio-environmental scenarios within the SBR and potentially support monitoring of work towards SDG1 (No Poverty) across the region.

Credit author statement

Charlotte Marcinko: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization. **Sourav Samanta**: Data Curation, Investigation and Writing - Original Draft. **Oindrila Basu**: Formal analysis, Methodology, Writing – Original Draft. **Andy Harfoot**: Methodology, Formal analysis and visualisation. **Duncan Hornby:** Data Curation, Investigation, Writing - Original Draft. **Craig Hutton**: Conceptualisation, Writing - Review & Editing, Supervision, Funding acquisition. **Sudipa Pal**: Formal analysis, Methodology, Writing – Original Draft. **Gary Watmough:** Methodology, Conceptualisation, Writing - Original Draft, Writing - Review & Editing.

1. Introduction

Measuring, monitoring and understanding the dynamics of poverty is vital to achieving the United Nations' Sustainable Development Goals (SDGs), particularly to SDG 1 which aims to end poverty in all its forms

* Corresponding author.
** Corresponding author.
E-mail addresses: cljm1g08@soton.ac.uk, C.L.J.Marcinko@gmail.com (C.L.J. Marcinko), sourav.samanta@gmail.com (S. Samanta), ajph@soton.ac.uk (A. Harfoot), ddh@geodata.soton.ac.uk (D.D. Hornby), ch9@soton.ac.uk (C.W. Hutton), gary.watmough@ed.ac.uk (G.R. Watmough).

https://doi.org/10.1016/j.jenvman.2022.114950
Received 26 September 2021; Received in revised form 26 February 2022; Accepted 20 March 2022
Available online 1 April 2022

everywhere. Census enumeration and household surveys are the standard methods used to measure progress towards development and provide governments with vital statistics for planning resource allocation. However, there are often relatively large time lags between enumeration. For example, Census enumeration is once every 10 years for most countries and household sample surveys can be every 3–5 years depending on the country (Blumenstock, 2016). Given the rapid socioeconomic and demographic changes occurring in many countries these time lags limit the ability to monitor progress (Njuguna and McSharry, 2017; Devarajan, 2013) and have led the UN to call for a data revolution (IEAG, 2014). This data revolution requires significantly more data collection at higher frequency, which is often infeasible using traditional survey methods due to the high costs involved (Jerven, 2017; Demombynes and Sandefur, 2014). This has led to studies examining the associations between socioeconomic conditions and environmental proxy indicators, specifically looking at how non-traditional geospatial and earth observation (EO) data could be used to support the monitoring of SDG progress (Anderson et al., 2017; Kavvada et al., 2020; Hargreaves and Watmough, 2021).

The assertion that environmental data can inform the monitoring of socioeconomic conditions comes from the fact that there are often close links between aspects of people's wellbeing and their environment (Okwi et al., 2007; Engstrom et al., 2017; Watmough et al., 2019), and that regions of relative wealth or poverty can be associated with particular land cover characteristics (e.g., Emmanuel, 1997; Lafary et al., 2008). In rural regions of low- and middle-income countries (LMIC), where there is high reliance on agriculture and other natural resources for livelihoods and food production, broad environmental proxies such as access to market, proximity to forest, presence of vegetation, agriculture, built environment and elevation have previously been found to provide useful insights into socioeconomic conditions. For example, Amoako Johnson (2016) identified statistical associations between the distribution of increasing soil salinity and poverty outcomes within the Ganges Brahmaputra Meghna Delta of Coastal Bangladesh. Watmough et al. (2016) identified that the time to travel to market along with percentage of a community area covered with woodland and winter crop were significantly associated with poverty within the state of Assam, India. The length of growing season, household size and percentage of bare land were associated with poverty in Sauri, rural Kenya (Watmough et al., 2019). Furthermore, Berchoux et al. (2019) found strong spatial associations between measurements of physical, natural, financial and social capitals derived from household surveys and environmental proxy indicators in the Mahanadi delta.

Poverty and wellbeing are complex and, within rural regions of LMIC countries, often explicitly linked to livelihoods which are known to be multi-dimensional, combining different elements between people and across time and space (Scoones, 2015). The associations between poverty and environmental characteristics are complex and can vary depending upon the region being examined and the scale of analysis (Watmough and Marcinko, 2021). Environmental proxies alone cannot predict all aspects of socio-economic conditions. However, the associations found in previous studies indicate the potential value of using environmental proxies obtained from EO and other non-traditional geospatial data to help map and monitor socioeconomic conditions between census enumeration periods (Watmough et al., 2019; Engstrom et al., 2017; Steele et al., 2017). Higher frequency estimates of socioeconomic conditions could be extremely informative to decision makers, promoting evidence-based decision making and accurate targeting of resources. Furthermore, examining the associations between socioeconomic conditions and environmental proxies also provides greater insight into the complex relationships between social and ecological systems. Such knowledge is vital for identifying areas of potential conflict and trade-off between socioeconomic and environment focused SDGs and therefore successfully achieving the UN's 2030 development agenda.

As part of a wider study of interactions between Sustainable Development Goals in deltas, we examine associations between poverty and a range of population and environmental proxies within the Sundarban Biosphere Reserve (SBR), a complex socio-ecological system within the Ganges-Brahmaputra-Meghna (GMB) delta. A random forest machine learning method is used to predict a village level measure of multidimensional poverty from variables related to Land Use Land Cover (LULC), agricultural yield, natural hazard risk, access to urban areas and population density gained from satellite, transport and population data. We identify the importance of these data in predicting village level poverty and use model results to gain a greater understanding of the dynamics of poverty across the region.

2. Study area

The Sundarban Biosphere Reserve (SBR, 1989) (21° 32' N - 22° 40' N and 88° 05' N - 89° 51' E) is located within the state of West Bengal, India within the Ganges-Brahmaputra-Meghna delta (Fig. 1). The region comprises an area of 9630 km^2 including approximately 4200 km^2 of globally important mangrove ecosystem. A network of interconnecting channels, canals and creeks divides the land into over 100 islands, approximately half of which (54/102) are inhabited with a total population of over 4.4 million people (PCA, 2011). The inhabited areas consist of 19 administrative blocks which form the entire South 24 Parganas district and the southern part of the North 24 Parganas district of West Bengal.

The SBR region is one of the most impoverished and vulnerable in India with an average per capita income of approximately USD 0.5 per day (Nishat, 2019) and about half of the population living below the poverty line (Sánchez-Triana et al., 2014). Over 87% of people living in the Southern 24 Parganas district lack food security, with household surveys indicating that many suffer from malnutrition (Sánchez-Triana et al., 2014). Poor health is a problem for many, due to the effects of inadequate freshwater supply and sanitation combined with limited access to a public health system. Educational attainment is low with only 16% of the population attending secondary education and there are limited employment opportunities. This endemic poverty is compounded by exposure to the impacts of climate change, rising sea-levels and coastal erosion (Marcinko et al., 2021 and references therein). The region is especially vulnerable to cyclone hazard and storm surges, having experienced multiple cyclones over recent decades including extremely severe cyclone Sidr in 2007, extreme cyclone Aila in 2009 (Chakraborty, 2015) and more recently super cyclonic storm Amphan in 2020. These events threaten, not only life, but can also devastate local infrastructure and natural resources depended upon for livelihoods in the region. Despite these conditions, the population in the region is increasing with many people migrating into the area from within West Bengal and neighbouring Bangladesh (Sánchez-Triana et al., 2014).

Livelihoods within the SBR predominantly depend on agriculture (Banerjee A, 1998, Marcinko et al., 2021), the alluvial peat providing fertile conditions for paddy fields (FAO, 2005). However, the impact of frequent natural hazards has seen the majority of the population adapt to multiple livelihood activities throughout the year to earn their living (HDR, 2009). The tidal creeks and inland natural wetlands create livelihood opportunities in fishing (riverine, tidal-brackish water, marine) and aquaculture (freshwater or brackish) (Banerjee A, 1998). Prior to the formation of the National Park in 1984, mangrove forests provided livelihoods in the form of logging and harvesting timber. However, the subsequent ban on logging in protected areas has restricted forest-based livelihoods to the collection of Non-Timber Forest Products (NTFP) in the Buffer zones of the National Park (DasGupta and Shaw, 2013). Honey and wax collection along with other NTFP provide livelihoods to considerable numbers of people in fringe forest areas (Nishat, 2019) such as Gosaba and Kultali blocks. A growing tourist industry relating to access to the mangrove provides some direct and indirect support to local farmers, fisherfolks, local transport vehicle owners and youths who are employed by the tourism companies and the Forest Department in a

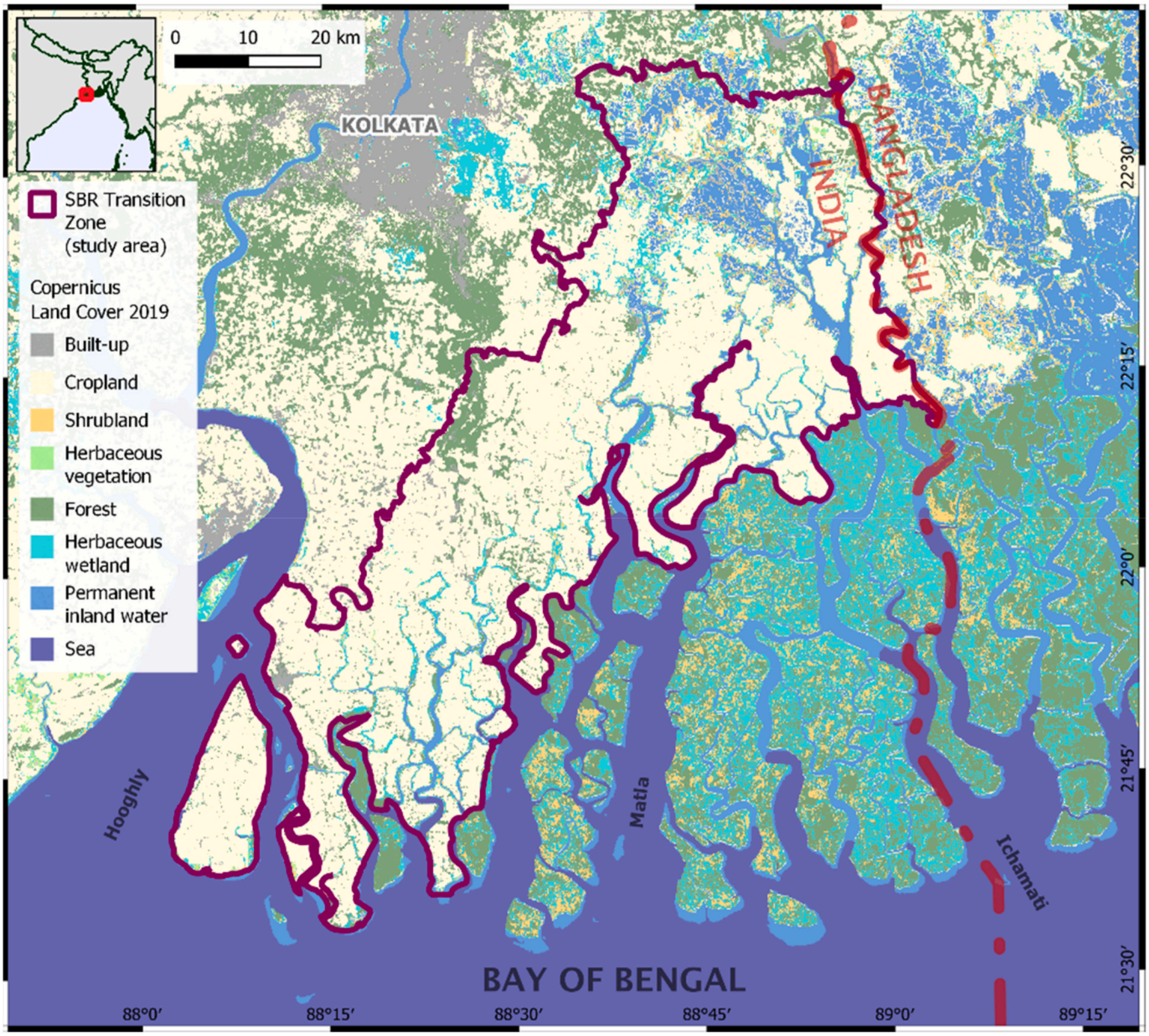

Fig. 1. Sundarban Biosphere Reserve (SBR), West Bengal, India. The SBR Transition Zone, which comprises the populated study area, is outlined in purple. The protected mangrove forest is shown in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

variety of roles including as guides, helpers and maintenance workers. Although tourism from both national and international visitors is increasing, the benefits from this sector are slow to trickle down to the majority of the SBR's inhabitants (Guha and Ghosh, 2007). Therefore, rural household incomes significantly rely on natural resource-dependent activities (Sánchez-Triana et al., 2014).

3. Data and methods

3.1. Village level multi-dimensional poverty index

A Multi-dimensional Poverty Index (MPI) is designed to assess acute poverty across the three equally weighted dimensions of Health, Education and Living standards. MPIs are used by the United Nations Development Programme and by many governments to gain an understanding of the dynamics of poverty (OPHI-UNDP Handbook, 2019). We constructed a village level MPI (vMPI) following the Alkire-Foster methodology (Alkire et al., 2011) and using 2011 Indian Census Data (PCA, 2011), to allow the comparison of poverty levels across villages within the SBR. The index was calculated using information from eight indicators which are grouped into the three equally weighted dimensions of Health, Education and Living standards (Table 1) following the approach taken in Santos and Alkire (2011). The equal weighting approach assumes that the three dimensions are contributing equally to overall poverty which is a simplification. However, weightings have been tested in the past and yielded high correlations, indicating that varying dimension weights did not have a large impact on overall MPI results (Alkire et al., 2010; Santos and Alkire, 2011). Due to lack of data availability, several indicators typically used in the construction of an MPI were replaced by the nearest proxy. Justification for indicator choices is provided within Table 1.

A deprivation score ranging between 0 and 1 was calculated for each household by taking the weighted sum of the different indicators. The score increases as the number of deprivations of the household increases, reaching a maximum of 1 when the household is deprived in all aspects. A household not deprived in any indicator receives a score equal to 0 (Alkire et al., 2011). A poverty cut-off ratio of 1/3 was applied in line with UNDP (2015) standards, ensuring that a household was identified as poor if it had a deprivation score higher or equal to 1/3. Households with a poverty score below 1/3 had their deprivation score set to zero. The poverty information is subsequently aggregated into two indices; (1) the incidence of poverty (*H*), representing the proportion/headcount ratio of households that experience multiple deprivations, and, (2) the intensity of deprivation (*A*), representing the

Table 1
Description of indicators used to construct the village level MPI score weighted to give an overall 1/3 weighting to Health, Education and Standard of Living. Indicators have been derived from the 2011 Indian census.

Name	Description	Weight	Indicator Justification
Health			
Unsafe drinking water	% households without access to safe drinking water	1/9	Safe drinking water is vital for good health and positive correlations exist between low-income households and the occurrence disease from unsafe drinking water (Bedi et al., 2015)
No clean fuel	% households using unclean fuel (dung, crop residue, firewood or charcoal)	1/9	The use of solid fuel expose people to smoke and pollutants longer and can lead to occurrence of respiratory infections (such as asthma, pulmonary tuberculosis, Acute Lower Respiratory Infection) and perinatal mortality (World health statistics, 2018)
No proper sanitation	% households without access to adequate sanitation/ latrine	1/9	High occurrence of diarrhoea in children under 5 have been linked to unimproved sanitation (Kumar and Das, 2014).
Education			
Illiteracy	% illiteracy	1/3	Illiteracy was considered as a proxy for years of schooling and school attendance.
Standard of Living			
Electricity	% households without electricity	1/12	Standard MPI indicator
Household environment	% households that cook indoors without a kitchen	1/12	A household is considered as deprived if it cooks inside the house but doesn't have a separate kitchen. Studies show that it is an indicator for lack of wealth (Mohanty, 2011) and lack of availability of separate kitchen plays vital role in determining the indoor air quality (Duflo et al., 2008).
Poor housing	Floor made of mud	1/12	Standard MPI indicator
Assets	% households that have not more than one radio, tv, mobile phone, bicycle, moped or refrigerator and not have a car	1/12	Standard MPI indicator

average proportion of (weighted) deprivations poor households' experience/deprivation share. The vMPI was computed as the product of these two components following Equation (1).

$$vMPI = H \ x \ A \quad (1)$$

This method identifies both poor households as well as the acuteness or intensity of the poverty situation (Alkire et al., 2011). As the asset data within the 2011 Census data was only available as a proportion at the village level, and household values were unavailable, random allocation of assets down to the household level was undertaken using a bootstrapping approach (Heinrich, 1998) over 1000 replications. The final household asset score was taken as the mean allocation value.

3.2. *Explanatory environmental and population variables*

The environmental and population variables used in this study were primarily derived from data collected in 2011. This allowed the landscape characteristics around villages to be captured during the same period as the census data used to calculate the poverty index. When deriving recent changes in landscape, data from 2011 were compared with data from 2001. Where data were not available for 2011 directly, such as the travel time estimates, these data were adjusted as described in section 3.2.1. Summary statistics were performed on 19 potential explanatory environmental and population variables to explore data characteristics and identify potential issues including multicollinearity. Several bivariate correlations had relatively large correlation coefficients and as such several indicators were dropped from the analysis. A large correlation was defined as a spearman correlation coefficient >0.7 significant at the $p < 0.001$ level. Several variables were also removed due to lack of variance across the dataset. The percentage of single, double and triple crop as independent variables were removed due to high correlations with the percent of agricultural land and the percent of land used for multiple crops. Variables indicating the percent of bare land, brick kiln, mud flat and land converted from agriculture to aquaculture were removed due to lack of variance. Fourteen explanatory variables were taken forward for analysis and are detailed in Table 2.

Table 2
List of environmental and population explanatory variables used in modelling associations with village level poverty.

Explanatory Variables	Description and Source
Travel time to Kolkata (comprising 1 variable)	Google maps API - rolled back to 2011 by removing bridges not in place (method described in section 3.2.1).
Travel time to nearest town (comprising 1 variable)	Google maps API (method described in section 3.2.1).
Land cover % within village boundary for rural settlement (built up); multiple crop; aquaculture; waterbody; mangrove; river/stream and total agricultural land. (comprising 7 variables)	2011 Landsat-TM land cover map at 30 m resolution (method described in section 3.2.2).
NDVI – Greenness index (comprising 1 variable)	MODIS NDVI for agricultural pixels within village (method described in section 3.2.3).
Land converted from agriculture to rural settlement (comprising 1 variable)	Based on Landsat-TM land cover maps for 2001–2011 at 30 m resolution.
Population density (comprising 1 variable)	Indian Census 2011 (PCA, 2011).
Cyclone hazard index (comprising 1 variable)	Based on Tropical cyclone risk model (TCRM) (method described in section 3.2.4).
Flood hazard index (comprising 1 variable)	Based on Landsat-TM data at 30 m resolution (May 26, 2009) after the severe cyclone Aila (method described in section 3.2.4).

3.2.1. *Time to travel to urban centres*

Travel routes from each village centroid within the study area to central Kolkata and the nearest town were generated using the Google Maps API (application programming interfaces) for the time period August 16th, 2019 at 8:00AM (Fig. 2 a & b). Travel time information (duration in minutes) encoded into the routes were joined to village polygons. As the analysis in this study was focused upon 2011 due to the availability of socioeconomic data (Section 3.1), 2019 travel times to Kolkata were adjusted based on the presence of bridges that post-dated 2011. The adjustment of 2019 travel time was achieved through a visual inspection of the routes against the ESRI base map imagery to identify river crossings. These crossings were cross referenced with the imagery archive available in Google Earth Pro (version 7.3.2) and the earliest date extracted so a year could be associated with each bridge crossing. The earliest available date that the bridge was observed in Google Earth Pro (which has an imagery archive going back to 2003) was assigned to each bridge. Twelve bridges were identified as being built after 2011. River crossings that showed no bridges in 2019 were tagged as ferry

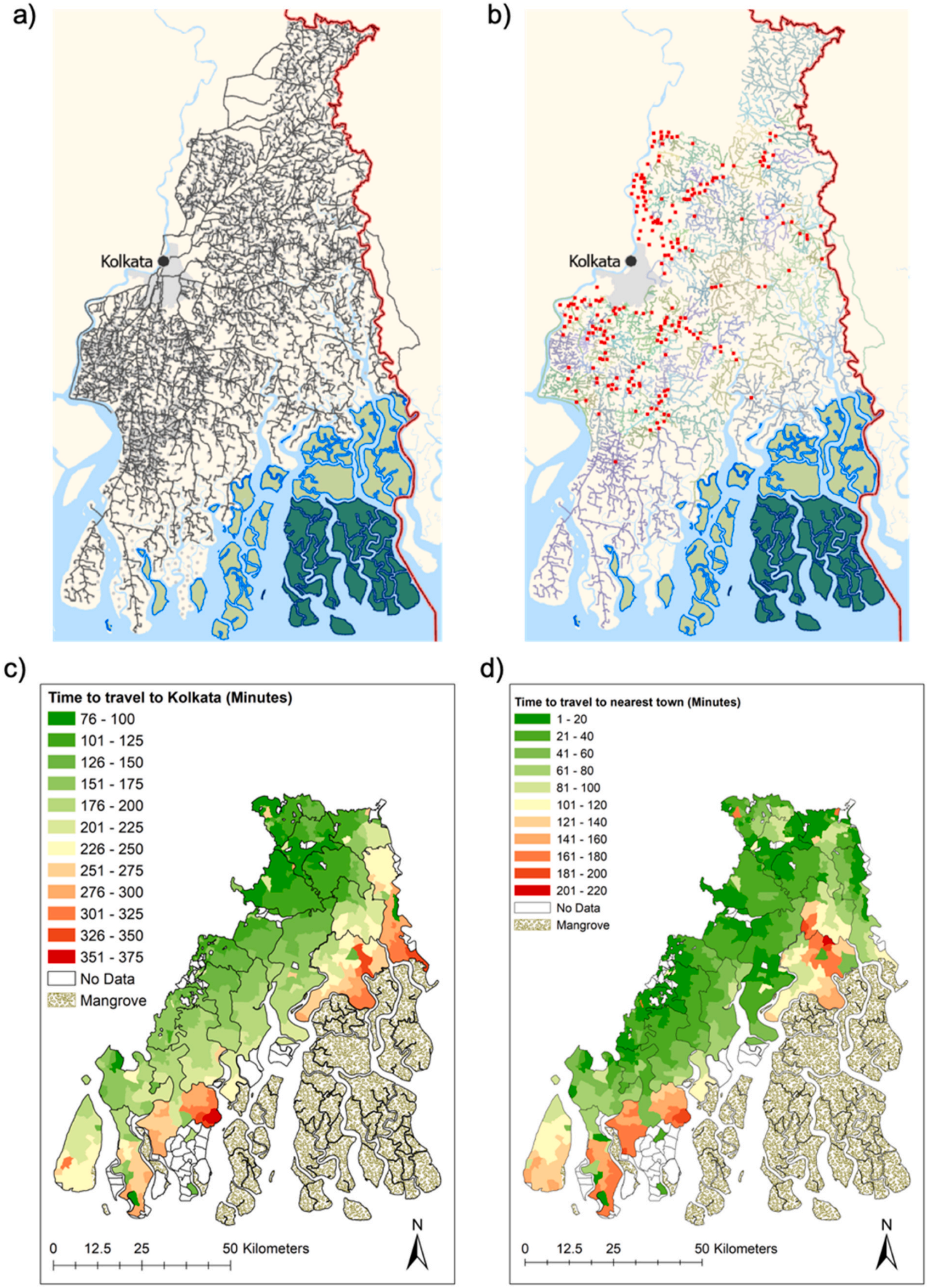

Fig. 2. a) Google Maps API output on the August 16, 2019 at 08:00 for single route from each village to Kolkata, b) Google Maps API output for the August 16, 2019 at 08:00, routes are colour coded to emphasize villages routing to their nearest town (red points), c) Time to travel from each village to Kolkata based on access to bridges in 2011 and d) Time to travel to nearest town. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

crossings as it was possible to observe in the imagery associated pier structures and roads terminating at the river banks on both sides. Thus, bridges that appeared in Google Earth Pro after 2011 would have been ferry crossings before 2011. For these crossings, the 2019 travel times for the routes were adjusted based on the assumption that a ferry crossing would take 1 h (loading, transport across the channel and disembarking). No adjustment was made for improved road conditions as it was not possible to source if/when road surfaces were improved since 2011. The addition of bridges post 2011 affected less than 5% of the villages in the study region and visually there is little impact on travel times. Travel durations from each village to Kolkata and to the nearest town used in the analysis are shown in Fig. 2 c & d.

3.2.2. Land use land cover

Multitemporal cloud-free Landsat-5 Thematic Mapper (TM) data at 30 m resolution were used to classify the land use/land cover (LULC) of the SBR. Satellite images were downloaded for the years 2000–2001 and 2010–2011 from the USGS website (https://earthexplorer.usgs.gov/). Radiometric calibration and atmospheric correction were performed for reducing atmospheric and scattering effects using the FLAASH tool in ENVI (Module, 2009; Sibanda and Ahmed, 2021). Maximum likelihood supervised classification techniques were executed on the Landsat images (Giri et al. 2007, 2021; Datta and Deb, 2012). Training samples (380 in total) were manually selected on satellite images based on field knowledge and high-resolution Google Earth imagery to run the classification. LULC classes identified included rural settlement; monocrop, double-crop and triple crop agricultural land; aquaculture; waterbody; mangrove; river/stream, urban settlement, barren land and mud flat. Classification of the three types of cropland (monocrop, double crop, and triple crop) was achieved by classifying images obtained in Kharif (July–Nov), Rabi (Dec–April), and Rabi Summer (March–June) seasons into crop (given value = 1) and non-crop (given value = 0). As almost all cropland is used for the Kharif rice crop within the SBR, the Kharif cropland raster was used to clip the Rabi and Rabi summer rasters. The Kharif, Rabi and Rabi summer rasters were then added together to identify areas of multiple cropping. Post-editing was performed for better accuracy based on field knowledge and historic high-resolution Google Earth imagery following methods set out in Thakkar et al. (2017). The proportion of LULC classes within each village were obtained using zonal statistics.

3.2.3. Village\Block level NDVI values

Normalised Difference Vegetation Index (NDVI) 16-day composite data associated with Julian day 289 in 2011 were extracted from the MODIS satellite product MOD13Q1 sourced from the LP DAAC website (https://lpdaac.usgs.gov/; Didan, 2015). Data had a spatial resolution on 250 m and captured the NDVI over the two weeks prior to the start of harvest of the main season rice crop across the region. Analysis was carried out in ESRI ArcPro. Data were clipped to the study area and quality controlled using the pixel reliability layers; unreliable data were excluded. The land use land cover map, described in section 3.2.2, was simplified into two classes: cultivated and non-cultivated. The polygons representing cultivated land were converted into a binary raster at the same resolution and alignment as the NDVI data to create an agricultural mask. This mask was used with the zonal statistics tool to extract the NDVI of cultivated land within each village boundary. Finally, a scaling factor of 0.0001 was applied.

3.2.4. Hazard indices

Landsat-5 TM data at 30 m spatial resolution was acquired from https://earthexplorer.usgs.gov for the date of the February 19, 2009 and the May 26, 2009 directly after the severe cyclone Aila. The imagery was used to map the consequent extensive inundation by the storm surge of 2–3 m high over and above the high tide within the SBR. Imagery was atmospherically corrected using ENVI FLAASH tool and clipped to the SBR boundary. The water area was extracted using the Normalised Difference Water Index (NDWI) as described in McFeeters (1996). Positive values of NDWI were classified as water and negative values were classified as non-water (Guo et al., 2017). The binary raster of the water area was converted to a polygon vector layer which was then clipped by the previous month's river polygon to exclude perennial water bodies and obtain the surge inundation layer. Post-editing was performed whereby, small clusters of pixels were removed in cases where inundation could not occur due to the distance from a river. Overall accuracy was estimated at 94.4% with a kappa value of 0.88. The kappa coefficient values indicate the classification results are in strong agreement between classified and original class. Village specific inundation values were obtained using zonal statistics. A flood index was generated by scaling the percentage inundation per village using min-max normalisation (Fig. 3a).

Cyclonic wind hazard was calculated using the Tropical cyclone risk model (TCRM) (Arthur et al., 2008) and following the methodology described in Ghosh et al. (2019). The model generates synthetic cyclone tracks across the study area based on the properties of historic cyclone events. Maximum wind speed is derived over the life of each synthetic cyclone event using a parametric wind field and boundary layer model (Kepert, 2001; Powell et al., 2005). The generalised extreme value (GEV) distribution was used to specify the maximum wind speed for each grid location across the Indian Bengal Delta (IBD). Tropical cyclones of multiple return periods from 5 to 2000 years were modelled to estimate the speed of cyclones. Wind speeds were aggregated to the village level using zonal statistics and values were scaled using a min-max normalisation to create cyclone hazard index for the IBD (Fig. 3b).

3.3. Machine learning to identify associations between poverty and environmental variables

The vMPI was grouped into three equally sized categories: poorest 33% (group 1), middle 33% (group 2), and wealthiest 33% (group 3). Analysis of vMPI distribution and preliminary results indicated that this choice of categorisation offered the greatest predictive capacity, whilst balancing the needs of appropriate distinction between categories and sufficient category sample size. Poverty groups were predicted from the explanatory variables using a Random Forest model in R 3.6.0 and the randomForest (Liaw and Wiener, 2002) and caret packages (Kuhn, 2008). Random Forest is a machine learning algorithm (Breiman, 2001) that combines Classification and Regression Tree (CART) methods with bagging (bootstrap aggregating). CART methods have several advantages for this type of analysis, they do not assume normal error distribution and are suitable for examining highly non-linear and complex relationships (Auret and Aldrich, 2012), important when considering human-natural systems (Liu et al., 2007). Random forest models produce predictions by averaging the results of many classification trees which have been trained independently on a random subset of data (e.g. bootstrapped samples). Furthermore, when growing each classification tree, only a random selection of explanatory predictor variables is considered at each split in the tree. This prevents correlations in predictions and reduces variance in the final mode and prevents overfitting to the data. To determine model performance data were split into training/calibration (80% of total data) and testing/validation (20% of the total data) samples. Each of the three poverty groups were sampled independently to ensure that the testing dataset consisted of an equal proportion of each poverty category. A k-fold cross validation approach (set at k = 10) was used to estimate model accuracy from the training data (Svetnik et al., 2004). The final model was applied to the testing sample and a confusion matrix created using the "*caret*" package was used to identify the overall model prediction accuracy as well as the accuracy of each poverty group.

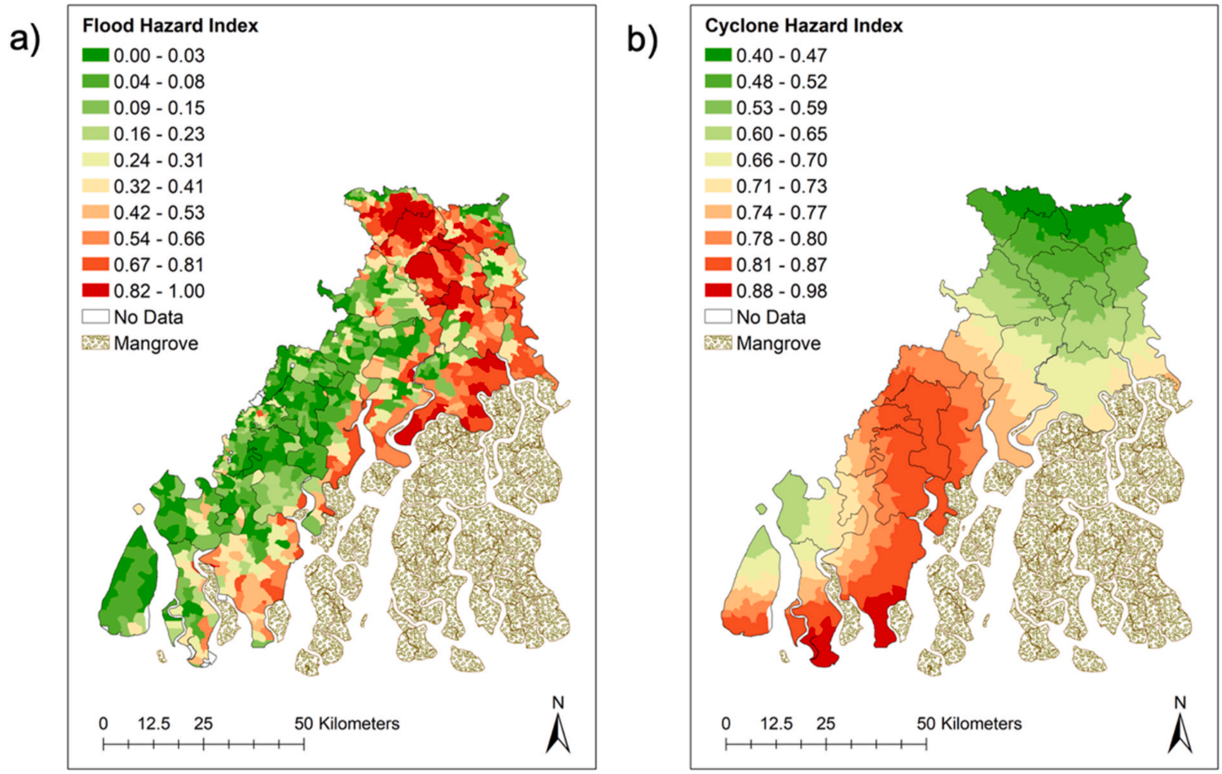

Fig. 3. Village level hazards indices of a) flood based on storm surge after cyclone Alia in 2009 and b) modelled cyclone wind speed.

4. Results

4.1. Multidimensional poverty index

The distribution of poverty across the SBR varied considerably. Villages with higher multidimensional poverty index scores (red) were clustered in the central regions of the SBR. Whilst the majority of villages with relatively low levels of poverty (green) were on the western side of the SBR (Fig. 4). Areas of extreme poverty include the northern region of Basanti and the southwest region of Canning II, along with northern areas of Kultali and the eastern areas of Jaynagar I and Jaynagar II blocks, all administrative areas within the South 24 Parganas District. Other blocks indicating relatively high levels of multidimensional poverty in the South 24 Parganas District included Canning I and Gosaba. In contrast, relatively low poverty scores occurred in the southwest of the SBR in Sagar, Namkhana, Patharpratima and Kakdwip blocks in the South 24 Paganas district. Within the North 24 Parganas district, Sandeshkahali II block and the southern regions of Hingalganj block were regions of high poverty. In contrast, relatively low poverty scores were seen in the northwest of the SBR in Haroa, Hasnabad, Minakhan and Sandeshkhali I blocks of the North 24 Parganas district.

Results indicated that across the SBR 15.8% of villages were highly deprived in education, 52.5% were highly heath deprived and 40.5% of villages were highly deprived in terms of living standards. Within the poorest villages 46.5% were highly deprived in education, 76.1% were highly heath threatened and 75.8% were highly deprived in living standards.

4.2. Predicting village level poverty from environmental variables

The Random Forest model was built on 778 observations (villages) and 14 explanatory variables. The model was run using 1000 trees as this was sufficient for convergence in results. The out-of-bag (OOB) training accuracy assessed through k-fold cross validation was 57.1%. The optimum *mtry* parameter was 4, this represents the number of randomly selected parameters considered at each split when growing trees. The overall accuracy of the model when applied to the unseen test dataset (n = 193) was 61.5% with a kappa coefficient of 0.42 (Table 3). Both test and cross-validation training accuracies are much greater than the no information rate (33%). The breakdown of model accuracy per group indicated the Poorest villages (Group 1) were predicted well with an accuracy of 78%. The predictive accuracy of the Middle and Wealthiest poverty groups (Group 2 and 3) was lower at 41 and 66%, respectively. Results indicate no evidence of spatial clustering in correct/incorrect predictions (Fig. 5a). There is little confusion between the poorest and wealthiest groups (identified by values of −2 or +2 in Fig. 5b). The confusion matrix (not shown) indicated that the relatively lower accuracy for the middle group was due to confusion with the poorest group and wealthiest group in near equal measure. The accuracy of the wealthiest group was affected by confusion with the middle group.

4.3. Associations between village level poverty and environmental variables

The importance of each environmental or population variable to the model prediction was calculated. Overall, cyclone hazard, proportion of rural settlement, time to travel to Kolkata, NDVI and percentage of agricultural land were among the most important variables for model predictions (Fig. 6). However, if importance is disaggregated for each poverty group (Fig. 6), the proportion of rural settlement within a village was the most important variable for the Poorest group followed by percentage agricultural land, time to nearest town and NDVI. Additionally, cyclone hazard and proportion of aquaculture land were also important for prediction of the Poorest group. Cyclone hazard was the most important variable for predicting the Wealthiest group followed by the time to travel to Kolkata and the NDVI of agricultural land. Other important variables for the prediction of the Wealthiest group included the proportion of rural settlement and the time to nearest town. For the Middle group, time to Kolkata was most important followed by population density, proportion of river or stream within a village, proportion

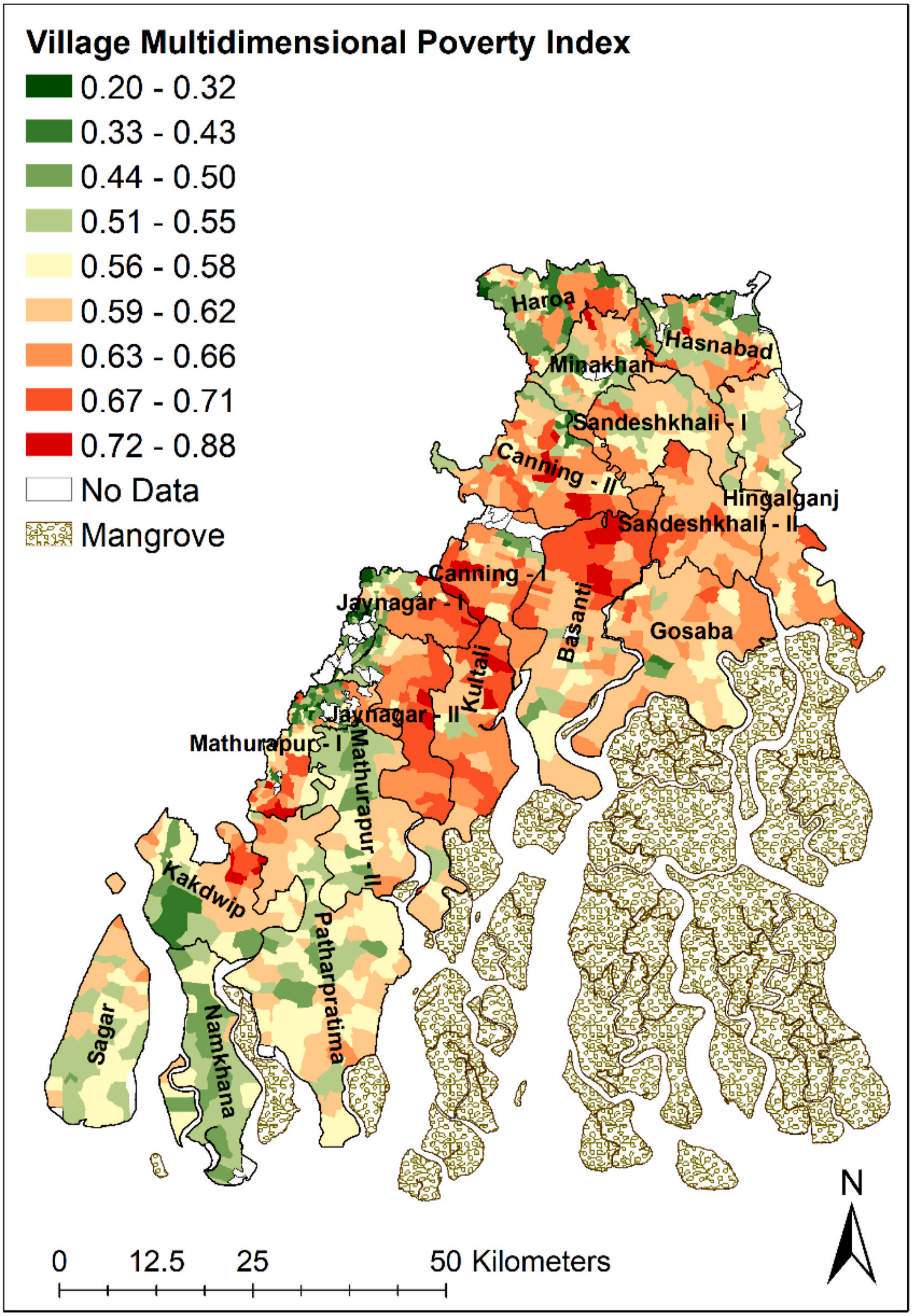

Fig. 4. Village level Multi-dimensional Poverty Index (vMPI) across the Sundarban Biosphere Reserve.

of multi cropped land, mangrove and rural settlement.

Partial dependence plots (PDP- Fig. 7) allow insights from the Random Forest in terms of the effect of each explanatory variable on the predicted outcome. They show the type of relationship (e.g. whether it is linear or more complex) and the direction of influence between the vMPI group and each explanatory variable. For a classification result, the partial dependence plots display the probability for a certain class (y-axis) given different values of the explanatory variable (x-axis) (Molnar, 2019). Partial dependence plots for the five most important variables in Poorest and Wealthiest groups are shown in Fig. 7. Partial dependence plots are not shown for the Middle group because the predictive accuracies were considered to be too low and therefore relationships were subject to a higher level of uncertainty.

Results indicated that a village was more likely to be classified as poorest if it had relatively low amounts of rural settlement (<30%) (Fig. 7a), high amounts of agricultural land (>50%) (Fig. 7b) and

Table 3
Predictive accuracy of the Random forest model calculated from unseen test dataset and internal cross validation estimate.

	Overall Accuracy (%)	Poorest (Group 1) Accuracy (%)	Middle (Group 2) Accuracy (%)	Wealthiest (Group 3) Accuracy (%)
Test Data (n = 193 Withheld)	61.5	78.1	40.6	65.6
K-fold Cross validation	57.1	65.8	39.4	60.4

experienced moderate to high cyclone hazard (index value between ~ 0.55–0.75) (Fig. 7c). Furthermore, a village was more likely to be classified as poorest if it was between 25 and 125 min away from the nearest town (Fig. 7e) and was within 200 min of Kolkata (Fig. 7f). In contrast, the PDP plots for the Wealthiest group of villages indicated that prediction had a bimodal relationship with cyclone hazard, with a village more likely to be classified as wealthiest if it was in a region where cyclone hazard was low (<0.6) or very high (>0.75) (Fig. 7c). A village was also more likely to be classified in this group when the proportion of land within the village boundary classified as rural settlement was relatively high (>25%) (Fig. 7a) and there were relatively lower amounts of agricultural land (<30%) (Fig. 7b). The relationship between the Wealthiest group and time to travel to the nearest town was multimodal, perhaps indicating the complex nature between village level poverty and access to urban centres. Fig. 7f indicates a linear relationship with the probability of being classified as wealthiest decreasing as time to travel to Kolkata increases. Villages <100 min from Kolkata were more likely to be predicted in the Wealthiest group. Analysis of the PDP for NDVI suggests that agricultural land for villages classified as poorest was less productive when compared to the wealthiest villages. Fig. 7d indicates that the Poorest group was more likely to be classified if NDVI was between 0.4 and 0.65 compared to the Wealthiest group which was more likely to be classified at higher NDVI values (between 0.6 and 0.8).

5. Discussion

5.1. Poverty prediction

This study has identified statistical associations between village level poverty and environmental variables, determined from EO and non-traditional geospatial data, within a complex socio-ecological deltaic system. Although it is widely recognised that physical environment and climatic factors can play a significant role in the distribution of poverty in LMIC countries (Vista and Murayama, 2011; Okwi et al., 2007), quantification of these relationships has, to-date, been limited. Here we

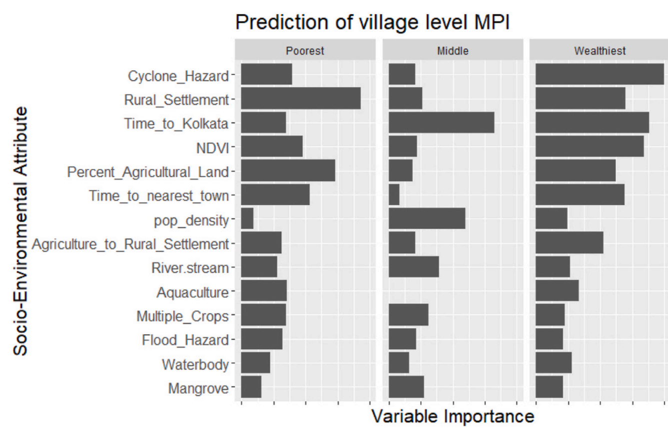

Fig. 6. Variable importance for the prediction of each poverty group ranked in terms of overall importance to model prediction.

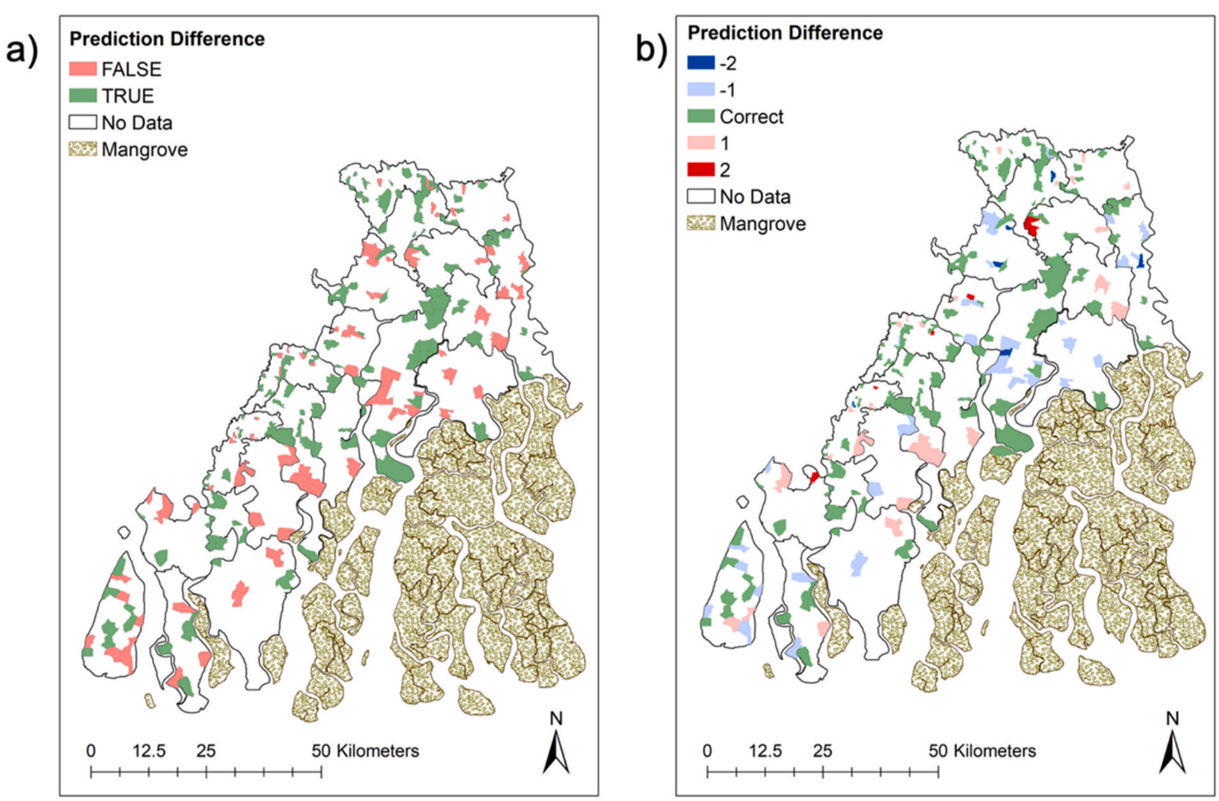

Fig. 5. a) Spatial distribution of correct (n = 118) and incorrect predictions (n = 74) of poverty categories in the withheld test data (n = 193). b) Predicted poverty category – actual poverty category where poverty categories were numbered as follows: Poorest = 1, Middle = 2, Wealthiest = 3.

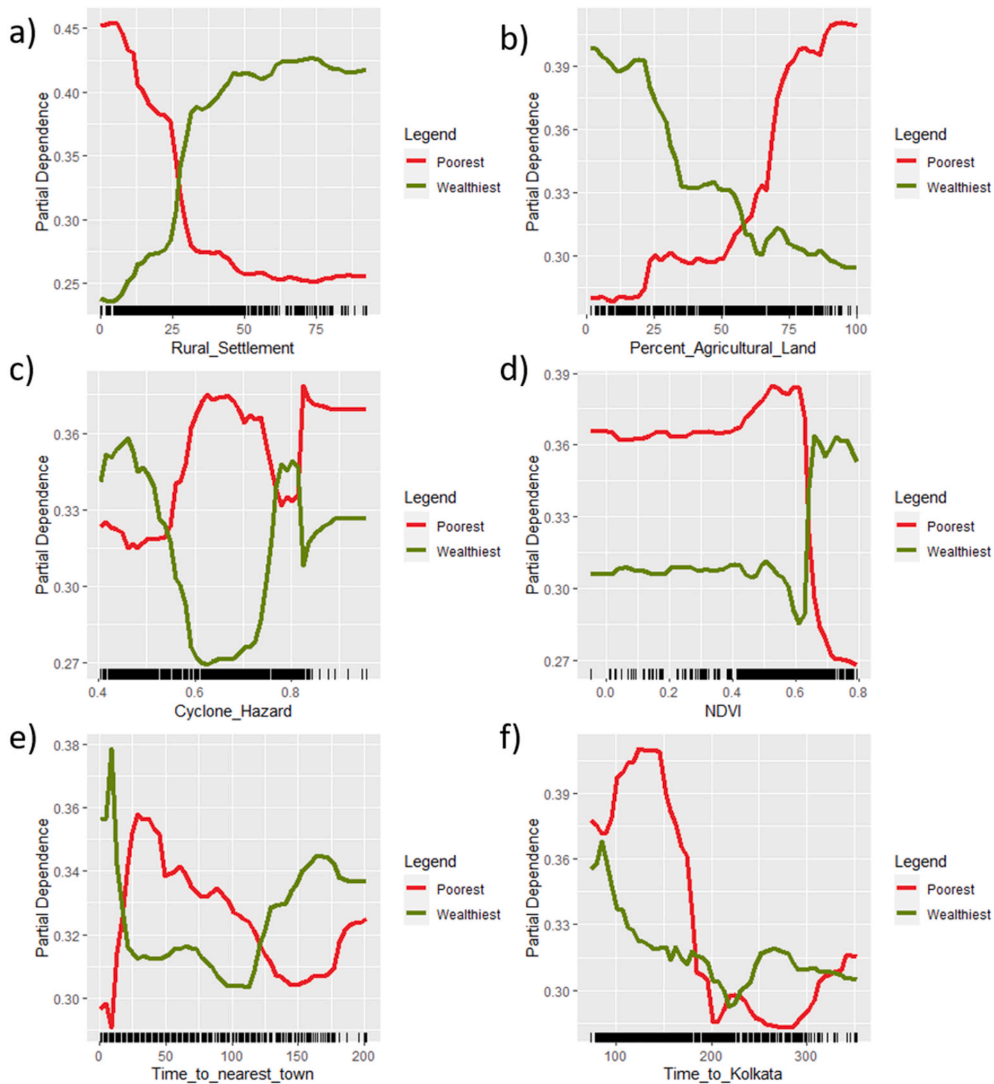

Fig. 7. Partial dependence plots indicating the probability for Poorest and Wealthiest villages prediction given different values of the top 5 most important variables in each of these groups. a) Rural settlement, b) % agricultural land c) cyclone hazard, d) NDVI of agricultural land e) time to travel to nearest town and f) time to travel to Kolkata. Note estimates are less reliable in regions where there are relatively few data points as indicated at the bottom of each plot.

have demonstrated that environmental variables that characterise the physical and climatic conditions are able to predict village level poverty with an accuracy of up to 61.5 percent within the SBR, India. Results showed a very high level of predictive accuracy for the poorest villages, up to 78.1% were predicted correctly in the test dataset withheld from model development. Relatively high accuracy was also achieved for the wealthiest group of villages (up to 65.6%). The middle poverty group had lower predictive accuracy (up to 40.6%) and results indicated this was due to high rates of confusion with both the poorest and wealthiest group. This is perhaps unsurprising given the rather arbitrary boundaries set by the categorisation of the vMPI. There is a good distinction between the vMPI values of the poorest and the wealthiest groups but less so between the poorest and the middle group and the middle and the wealthiest group. Similar issues have been found in previous studies using categorisation approaches (Watmough et al., 2019). The pattern in prediction capacity across the three poverty categories was consistent in the k-fold cross validation accuracy estimates. However, the overall and individual accuracies were slightly lower compared to those calculated from the independent test dataset. This may be because out of bag and cross validation methods can potentially overestimate the true prediction error in classification problems especially when there is an equal number of observations from all response classes, a relatively small sample size, a large number of predictor variables, small correlation between predictors and weak effects (Janitza and Hornung, 2018). The magnitude of model accuracy compares well with previous studies (Engstrom et al., 2017; Steele et al., 2017; Watmough et al., 2016) and are notable considering the complex nature of poverty and the limited set of predictor variables used.

5.2. *Examining the associations between environment and poverty*

In addition to assessing predictive capacity, the outputs from the model allowed variable importance and the nature of associations with poverty to be examined. Our results indicated that the poorest villages within the SBR were associated with relatively small proportions of rural settlement (i.e. built up areas) and were dominated by agricultural land (>50% of village area). Conversely, the wealthiest villages were associated with higher proportions of rural settlement and lower proportions of agricultural land. This potentially suggests that poorer villages are associated with a higher dependency on agricultural livelihoods, as supported by Hajra and Ghosh (2018). Furthermore, results indicated that agricultural land within the poorest villages was associated with a lower NDVI prior to harvest, suggesting that, although the proportion of agricultural land was high, agricultural yield per hectare may be less than land in the wealthiest villages. However, the differentiation in NDVI signal taken from a single snapshot in time could indicate other factors at play including differing farming practices.

Several of the blocks, including Kakdwip, Mathurapur and Namkhana, which have lower poverty scores in the vMPI, are connected to the mainland via bridges and roads. The infrastructure and facilities in these blocks are more developed than in blocks without direct connection to the mainland and the populations have more diverse livelihood options (Nishat, 2019). Namkhana, in particular, is a destination for many tourists from Kolkata that come to visit both the beaches and use locations within the block as an entry point to the mangrove forest reserve and this activity provides related livelihood opportunities to the local communities. Similarly, Sagar Island sees large amounts of religious pilgrimage tourism every year which significantly contributes to the local economy (Nishat, 2019). A lack of diversity in livelihood opportunities is known to perpetuate poverty and limit the ability of people to cope with external shocks (de Sherbinin et al., 2008; Berchoux et al., 2019). Therefore, communities within these more connected blocks may be better able to cope with shocks events such as cyclones, compared to communities with less diverse opportunities, that are more dependent on precarious agricultural livelihoods.

Exposure to cyclone hazard had a significant impact on the poverty distribution within the SBR and results indicated that the relationships with poverty were nonlinear (Fig. 7c). The SBR is highly susceptible to medium to severe tropical cyclones originating within the Bay of Bengal and they can have devastating effects on infrastructure and livelihoods, particularly those relating to agriculture. High wind speeds can lead to storm surges and the failure of embankments which allows ingression of saline water onto agricultural land (Ghosh and Mistri, 2021), whilst strong winds can lead to extensive crop damage. Classification of the poorest group was associated with moderate to high cyclone hazard with a unimodal relationship centred around a hazard score of 0.65. This perhaps reflects the high cyclone hazard scores seen for communities within the central region of the SBR (Fig. 4b), including within Kultali and northern Basanti blocks, which had relatively high poverty in the vMPI (Fig. 5). These blocks are known to be almost completely detached from the mainland and therefore have limited livelihood opportunities. Conversely, a bimodal relationship was seen between the wealthiest villages and cyclone hazard, which indicated that classification in to the wealthiest group was more likely at low and very high cyclone hazard scores (Fig. 7c). It may be that although cyclone hazard is very high in the southeast region of the SBR, including southern Namkhana, that these communities have more diverse livelihoods and are relatively wealthier than communities in high to moderate cyclone hazard areas with limited livelihood options and high dependency on agriculture.

Associations have been found in several studies between poverty and remoteness or access to markets (Barrett, 2005; Okwi et al., 2007; Steele et al., 2017; Watmough et al., 2016). However, the nonlinear associations found with travel time to nearest town may indicate a more complex situation within the SBR. The relationship between travel time to the nearest town and prediction of the wealthiest group was bimodal in nature and associated with travel times less than 25 min or over 110 min. The first peak in the bimodal pattern may be capturing the villages along the western edge of the SBR that are closer to towns (Fig. 3b) and have relatively lower levels of poverty. Improved access to towns can provide increased access to markets which have been associated with increased wealth (Muktar et al., 2013) providing more opportunities for incomes and for people to develop different or multiple livelihood strategies (Berchoux et al., 2020; Weiss et al., 2018). Increased access to urban centres can also mean better access to educational and health facilities (Blanford et al., 2012; Gabrysch and Campbell, 2009). The second peak may be capturing those villages in the far south of the SBR which, as previously discussed, although remote have lower poverty scores.

5.3. Limitations and future work

This study has demonstrated that environmental and population variables can achieve good predictive capacity of the poorest and wealthiest villages across the SBR and lends support to the assertion EO and non-traditional geospatial data could play an important role in improving the monitoring of socio-economic conditions and the progress being made towards achieving the SDGs (Anderson et al., 2017; Kavvada et al., 2020). However, it must be remembered that environmental information alone cannot predict all aspects of poverty and they do not capture important contributions to poverty reduction such as government safety nets or assistance. The poverty-environment dynamics identified within this study can have value for local policy and decision makers in that they increase understanding of the factors that are contributing to poverty across the region and better understanding will allow for more effective targeting of policy and resources. Furthermore, results could have application as an exploratory tool to examine impacts of potential future LULC change or policy upon regional poverty. Although this type of analysis would assume the strength and nature of the association described remain constant going forward. The complex nature of poverty and the fact that it varies overtime based on people's livelihood strategies and ability to adapt to changes in climate, markets and policies across multiple scales (local, regional, national) means that there is uncertainty around the timeframes over which associations between poverty and individual environmental variables hold. Therefore, future work should address these uncertainties by expanding analysis to identify if and how poverty-environment dynamics change over time and what implications this may have for monitoring socioeconomic conditions between survey periods.

6. Conclusion

Effective poverty monitoring is a well established problem. The gold standard method is the census which are collected once every 10 years in most countries. The decadal gap in enumeration means that rapidly changing socioeconomic conditions can be missed. Census data are also not always readily available to researchers or consultants alike. As such the development of an approach which allows the use of standardised spatial data, whilst no substitute for census, has the potential to allow for inter-survey monitoring or augmentation with the census. Clearly the outcomes are associative, there being no particular way that we might directly link some spatial environmental variables with poverty in a process driven manner. For example, the road access variables form part of the association with poverty through a series of plausible assumptions around access to market, education, health and institutions (linking social capital). However, the linkages between reliance on agricultural productivity and increased poverty may have as much to do with remoteness as income derived from selling agricultural products. Indeed, a key conclusion of the work, is that reliance on agriculture, remoteness and a lack of livelihood alternatives may substantially contribute to poverty outcomes. Whilst avoiding environmental determinism and the issue of causality, the methodological approach has shown, with a particular efficacy in the extremes of the vMPI (poor/less poor), that a spatial statistical approach can contribute in a constructive way to the estimation of poverty in a rural region of an LMIC country. The associations between poverty and environmental factors found in this study have the potential to inform policy linked to socio-environmental scenarios and to support monitoring of work towards SDG1 across the region.

Funding

This research was funded by NERC Grant NE/S012478/1, Formas Grant 2019–00045 and the UKIERI-DBT (Grant No BT/IN/TaSE/70/SH/2018–19) Under UK- India Education Research Initiative.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to acknowledge Prof. Sugata Hazra for insight into the human-environment dynamics within the SBR to aid interpretation of results, Dr. Tim Daw for insightful discussions that aided presentation and interpretation of results, Prof. Robert Nicholls for advice on structure and manuscript edits, Ian Waldock for contribution to the preliminary analysis to assess an appropriate statistical methodology and Partho Protim Mondal for providing technical assistance in calculating the village level multidimensional poverty index. This research is funded under the "Towards a Sustainable Earth: Environment-human systems and the UN Global Goals" (TaSE) programme in the project "Opportunities and trade-offs between the SDGs for food, welfare and the environment in deltas". Funding has been provided by NERC Grant NE/S012478/1, Formas Grant 2019–00045 and the UKIERI-DBT (Grant No BT/IN/TaSE/70/SH/2018–19) Under UK- India Education Research Initiative.

References

Alkire, S., Santos, M.E., Seth, S., Gaston, Y., 2010. Is the multidimensional poverty index robust to different weights. In: Oxford Poverty & Human Development Initiative (OPHI) Research in Progress Series. www.ophi.org.uk/publications/ophi-research-in-progress.

Alkire, S., Roche, J.M., Santos, M.E., Seth, S., 2011. Multidimensional poverty index 2011: brief methodological note. https://ophi.org.uk/multidimensional-poverty-index-2011-brief-methodological-note/.

Amoako Johnson, F., Hutton, C.W., Hornby, D., Lázár, A.N., Mukhopadhyay, A., 2016. Is shrimp farming a successful adaptation to salinity intrusion? A geospatial associative analysis of poverty in the populous Ganges–Brahmaputra–Meghna Delta of Bangladesh. Sustain. Sci. https://doi.org/10.1007/s11625-016-0356-6.

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., Friedl, L., 2017. Earth observation in service of the 2030 agenda for sustainable development. Geo Spatial Inf. Sci. 20, 77–96. https://doi.org/10.1080/10095020.2017.1333230.

Arthur, W., Schofield, A., Cechet, R., Sanabria, L., 2008. Return Period Cyclonic Wind Hazard in the Australian Region 28th AMS Conference on Hurricanes and Tropical Meteorology.

Auret, L., Aldrich, C., 2012. Interpretation of nonlinear relationships between process variables by use of random forests. Miner. Eng. 35, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008.

Banerjee, A., 1998. Environment, Population, and Human Settlements of Sundarban Delta. Concept Pub. Co., New Delhi.

Barrett, C.B., 2005. Rural poverty dynamics: development policy implications. Agric. Econ. 32, 45–60. https://doi.org/10.1111/j.0169-5150.2004.00013.x.

Bedi, J.K., Ghuman, R.S., Bhullar, A.S., 2015. Health and economic impact of unsafe drinking water: a study of ludhiana. Econ. Polit. Wkly. 50, 23–26.

Berchoux, T., Hutton, C.W., 2019. Spatial associations between household and community livelihood capitals in rural territories: an example from the Mahanadi Delta, India. Appl. Geogr. 103, 98–111. https://doi.org/10.1016/j.apgeog.2019.01.002.

Berchoux, T., Watmough, G.R., Hutton, C.W., Atkinson, P.M., 2019. Agricultural shocks and drivers of livelihood precariousness across Indian rural communities. Landsc. Urban Plann. 189, 307–319. https://doi.org/10.1016/j.landurbplan.2019.04.014.

Berchoux, T., Watmough, G.R., Amoako Johnson, F., Hutton, C.W., Atkinson, P.M., 2020. Collective influence of household and community capitals on agricultural employment as a measure of rural poverty in the Mahanadi Delta, India. Ambio 49, 281–298. https://doi.org/10.1007/s13280-019-01150-9.

Blanford, J.I., Kumar, S., Luo, W., MacEachren, A.M., 2012. It's a long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger. Int. J. Health Geogr. 11, 24. https://doi.org/10.1186/1476-072X-11-24.

Blumenstock, J.E., 2016. ECONOMICS. Fighting poverty with data. Science 353, 753–754. https://doi.org/10.1126/science.aah5217.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chakraborty, S., 2015. Investigating the Impact of Severe Cyclone Aila and the Role of Disaster Management Department - A Study of Kultali Block of Sundarban.

DasGupta, R., Shaw, R., 2013. Changing perspectives of mangrove management in India – an analytical overview. Ocean Coast Manag. 80, 107–118. https://doi.org/10.1016/j.ocecoaman.2013.04.010.

Datta, D., Deb, S., 2012. Analysis of coastal land use/land cover changes in the Indian Sunderbans using remotely sensed data. Geo Spatial Inf. Sci. 15, 241–250. https://doi.org/10.1080/10095020.2012.714104.

de Sherbinin, A., VanWey, L.K., McSweeney, K., Aggarwal, R., Barbieri, A., Henry, S., Hunter, L.M., Twine, W., Walker, R., 2008. Rural household demographics, livelihoods and the environment. Global Environ. Change 18, 38–53. https://doi.org/10.1016/j.gloenvcha.2007.05.005.

Demombynes, G., Sandefur, J., 2014. Costing a Data Revolution. Center for Global Development Working Paper No. 383, Available at: SSRN: https://ssrn.com/abstract=2622756 or https://doi.org/10.2139/ssrn.2622756.

Devarajan, S., 2013. Africa's Statistical Tragedy. Rev. Income Wealth, vol. 59, pp. S9–S15. https://doi.org/10.1111/roiw.12013.

Didan, K., 2015. MOD13Q1 MODIS/terra vegetation indices 16-day L3 global 250m SIN grid V006 [data set]. NASA EOSDIS Land Process. DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006.

Duflo, E., Greenstone, M., Hanna, R., 2008. Indoor air pollution, health and economic well-being. S.A.P.I.EN.S, 1.1. http://journals.openedition.org/sapiens/130.

Emmanuel, R., 1997. Urban vegetational change as an indicator of demographic trends in cities: the case of detroit. Engstrom et al 2017 Environ. Plann. Plann. Des. 24, 415–426. https://doi.org/10.1068/b240415.

Engstrom, Ryan, Hersh, Jonathan, Newhouse, David, 2017. Poverty from Space : Using High-Resolution Satellite Imagery for Estimating Economic Well-Being. Policy Research Working Paper. World Bank, Washington, DC. No. 8284. https://openknowledge.worldbank.org/handle/10986/29075.

FAO, 2005. Fertilizer Use by Crop in India. First Version, Published by Food and Agriculture Organization of the United Nations. Rome. http://www.fao.org/docrep/009/a0257e/A0257E05.htm=.

Gabrysch, S., Campbell, O.M.R., 2009. Still too far to walk: literature review of the determinants of delivery service use. BMC Pregnancy Childbirth 9, 34. https://doi.org/10.1186/1471-2393-9-34.

Ghosh, S., Mistri, B., 2021. Assessing coastal vulnerability to environmental hazards of Indian Sundarban delta using multi-criteria decision-making approaches. Ocean Coast Manag. 209, 105641. https://doi.org/10.1016/j.ocecoaman.2021.105641.

Ghosh, A., Das, S., Ghosh, T., Hazra, S., 2019. Risk of extreme events in delta environment: a case study of the Mahanadi delta. Sci. Total Environ. 664, 713–723. https://doi.org/10.1016/j.scitotenv.2019.01.390.

Giri, C., Pengra, B., Zhu, Z., Singh, A., Tieszen, L.L., 2007. Monitoring mangrove forest dynamics of the Sundarbans in Bangladesh and India using multi-temporal satellite data from 1973 to 2000. Estuar. Coast Shelf Sci. 73, 91–100. https://doi.org/10.1016/j.ecss.2006.12.019.

Giri, S., Samanta, S., Mondal, P.P., Basu, O., Khorat, S., Chanda, A., Hazra, S., 2021. A geospatial assessment of growth pattern of aquaculture in the Indian Sundarbans Biosphere Reserve. Environ. Dev. Sustain. https://doi.org/10.1007/s10668-021-01612-9.

Guha, I., Ghosh, S., 2007. Does Tourism Contribute to Local Livelihoods? A Case Study of Tourism, Poverty and Conservation in the Indian Sundarbans, pp. 973–993. SANDEE Working Paper No. 26-07.

Guo, Q., Pu, R., Li, J., Cheng, J., 2017. A weighted normalized difference water index for water extraction using Landsat imagery. Int. J. Rem. Sens. 38, 5430–5445. https://doi.org/10.1080/01431161.2017.1341667.

Hajra, R., Ghosh, T., 2018. Agricultural productivity, household poverty and migration in the Indian Sundarban Delta. Elem. Sci. Anthr. 6 https://doi.org/10.1525/elementa.196.

Hargreaves, P.K., Watmough, G.R., 2021. Satellite Earth observation to support sustainable rural development. Int. J. Appl. Earth Obs. Geoinf. 103, 102466. https://doi.org/10.1016/j.jag.2021.102466.

HDR, 2009. District Human Development Report. Development and Planning Department, Government of India.

Heinrich, G.A., 1998. The Prince and the Pauper Revisited: A Bootstrap Approach to Poverty and Income Distribution Analysis Using the PACO Data Base. (Research Papers on Comparative Analysis of Longitudinal Data; No. 21 (CEPS/INSTEAD).

Janitza, S., Hornung, R., 2018. On the overestimation of random forest's out-of-bag error. e0201904 PLoS One 13.

Jerven, M., 2017. How much will a data revolution in development cost? Forum Dev. Stud. 44, 31–50. https://doi.org/10.1080/08039410.2016.1260050.

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., Chuvieco, E., 2020. Towards delivering on the sustainable development goals using earth observations. Remote Sens. Environ. 247, 111930. https://doi.org/10.1016/j.rse.2020.111930.

Kepert, J., 2001. The dynamics of boundary layer jets within the tropical cyclone core. Part I: linear theory. J. Atmos. Sci. 58 (17), 2469–2484. https://doi.org/10.1175/1520-0469(2001)058<2469:tdoblj>2.0.co;2.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software 1 (5).

Kumar, A., Das, K., 2014. Drinking water and sanitation facility in India and its linkages with diarrhoea among children under five: evidences from recent data. Int. J. Humanit. Soc. Sci. 3, 50–60.

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 Int. 23, 53–66. https://doi.org/10.1080/10106040701494575.

Jianguo, L., Thomas, D., R, C.S., Marina, A., Carl, F., Emilio, M., N, P.A., Peter, D., Timothy, K., Jane, L., Elinor, O., Zhiyun, O., William, P., L, R.C., H, S.S., W, T.W., 2007. Complexity of coupled human and natural systems. Science 317 (80), 1513–1516. https://doi.org/10.1126/science.114400407. Liu et al 20.

Liaw, A., Wiener, M., 2002. Forest. Classification and Regression by RandomForest, 2/3. R News, pp. 1–5.

Marcinko, C.L.J., Nicholls, R.J., Daw, T.M., Hazra, S., Hutton, C.W., Hill, C.T., Clarke, D., Harfoot, A., Basu, O., Das, I., Giri, S., Pal, S., Mondal, P.P., 2021. The development of

a framework for the integrated assessment of SDG trade-offs in the sundarban Biosphere reserve. Water. https://doi.org/10.3390/w13040528.
McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Rem. Sens. 17, 1425–1432. https://doi.org/10.1080/01431169608948714.
Module, F.L., 2009. Atmospheric Correction Module: QUAC and FLAASH User's Guide, vol. 4, p. 44. Version.
Mohanty, S.K., 2011. Multidimensional poverty and child survival in India. PLoS One 6, e26857.
Molnar, C., 2019. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book -. (Accessed 15 September 2021).
Muktar, S.T., Mohamad, J.H., Woldemichael, Y.M., 2013. Dimensions and determinants of agro-pastoral household poverty in dembel district of Somali regional state, Ethiopia. J. Econ. Sustain. Dev. 4, 13–19.
Nishat, B., 2019. Landscape Narrative of the Sundarban : towards Collaborative Management by Bangladesh and India.
Njuguna, C., McSharry, P., 2017. Constructing spatiotemporal poverty indices from big data. J. Bus. Res. 70, 318–327. https://doi.org/10.1016/j.jbusres.2016.08.005.
Okwi, P.O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A., Henninger, N., Benson, T., Kariuki, P., Owuor, J., 2007. Spatial determinants of poverty in rural Kenya. LP – 16774 Proc. Natl. Acad. Sci. Unit. States Am. 104, 16769. https://doi.org/10.1073/pnas.0611107104.
OPHI-UNDP Handbook, 2019. How to Build a National Multidimensional Poverty Index (MPI): Using the MPI to Inform the SDGs. United Nations Development Programme (UNDP and Oxford Poverty and Human Development Initiative (OPHI), University of Oxford. https://ophi.org.uk/wp-content/uploads/How_to_Build_Handbook_2019_PDF.pdf -. (Accessed 15 September 2021).
PCA, 2011. Primary Census Abstract. Census of India. Census of West Bengal. Census of Odisha. Directorate of Census Operations, Government of India, New Delhi. http://censusindia.gov.in/pca/.
Powell, M., Soukup, G., Cocke, S., Gulati, S., Morisseau-Leroy, N., Hamid, S., Dorst, N., Axe, L., 2005. State of Florida hurricane loss projection model: atmospheric science component. J. Wind Eng. Ind. Aerod. 93 (8), 651–674. https://doi.org/10.1016/j.jweia.2005.05.008.
Sánchez-Triana, E., Paul, T., Leonard, O., 2014. Building Resilience for Sustainable Development of the Sundarbans. The International Bank for Reconstruction and Development, The World Bank, Washington, DC, USA.
Scoones, I., 2015. Sustainable Livelihoods and Rural Development. Practical Action Publishing, UK.
Santos, M.E., Alkire, S., 2011. Training material for producing national human development reports. Multidimens. Poverty Index (MPI). https://www.ophi.org.uk/wp-content/uploads/MPI-Primer1.pdf.
Sibanda, S., Ahmed, F., 2021. Modelling historic and future land use/land cover changes and their impact on wetland area in Shashe sub-catchment, Zimbabwe. Model. Earth Syst. Environ. 7, 57–70. https://doi.org/10.1007/s40808-020-00963-y.
Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A.M., Hadiuzzaman, K.N., Lu, X., Wetter, E., Tatem, A.J., Bengtsson, L., 2017. Mapping poverty using mobile phone and satellite data. J. R. Soc. Interface 14, 20160690. https://doi.org/10.1098/rsif.2016.0690.
Svetnik, V., Liaw, A., Tong, C., 2004. Variable selection in random forest with application to quantitative structure–activity relationship. In: Proc. 7th Course Ensemble Methods Learn. Mach.
Thakkar, A.K., Desai, V.R., Patel, A., Potdar, M.B., 2017. Post-classification corrections in improving the classification of Land Use/Land Cover of arid region using RS and GIS: the case of Arjuni watershed, Gujarat, India. Egypt. J. Rem. Sens. Sp. Sci. 20, 79–89. https://doi.org/10.1016/j.ejrs.2016.11.006.
UNDP, 2015. United Nations development programme, human development report, 2015 Work Hum. Dev. http://hdr.undp.org/sites/default/files/2015_human_development_report.pdf.
Vista, B., Murayama, y., 2011. Spatial determinants of poverty using GIS-Based Mapping. In: Murayama, Y., Thapa, R. (Eds.), Spatial Analysis and Modeling in Geographical Transformation Process, vol. 100. GeoJournal Library, Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0671-2_16.
Watmough, G.R., Marcinko, C.L.J., 2021. Towards achieving the UNs data revolution: combining earth observation and socioeconomic data for geographic targeting of resources for the sustainable development goals. In: Dunn, J., Balaprakash, P. (Eds.), Data Science Applied to Sustainability Analysis. Elsevier, pp. 229–254. https://doi.org/10.1016/B978-0-12-817976-5.00012-7 (Chapter 11).
Watmough, G.R., Marcinko, C.L.J., Sullivan, C., Tschirhart, K., Mutuo, P.K., Palm, C.A., Svenning, J.-C., 2019. Socioecologically informed use of remote sensing data to predict rural household poverty. Proc. Natl. Acad. Sci. Unit. States Am. 116, 1213–1218. https://doi.org/10.1073/pnas.1812969116.
Watmough, G.R., Atkinson, P.M., Saikia, A., Hutton, C.W., 2016. Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. World Dev. 78, 188–203.
Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Gething, P.W., 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553 (7688), 333–336. https://doi.org/10.1038/nature25181.
World health statistics, 2018. Monitoring Health for the SDGs, Sustainable Development Goals. World Health Organization, Geneva, p. 2018. Licence: CC BY-NC-SA 3.0 IGO.