Socioecologically informed use of remote sensing data to predict rural household poverty

Gary R. Watmough[a,b,1], Charlotte L. J. Marcinko[c], Clare Sullivan[d,2], Kevin Tschirhart[e], Patrick K. Mutuo[f,g], Cheryl A. Palm[g], and Jens-Christian Svenning[a]

[a]Section for Ecoinformatics and Biodiversity, Center for Biodiversity Dynamics in a Changing World, Department of Bioscience, Aarhus University, 8000 Aarhus, Denmark; [b]School of Geosciences, University of Edinburgh, EH8 9XP Edinburgh, United Kingdom; [c]GeoData, University of Southampton, SO17 1BJ Southampton, United Kingdom; [d]Agriculture and Food Security Center, Earth Institute, Columbia University, Palisades, NY 10964; [e]Center for International Earth Science Information Network, Columbia University, New York, NY 10964; [f]International Institute of Tropical Agriculture, Nairobi, Kenya; and [g]Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32603

Edited by Assaf Anyamba, National Aeronautics and Space Administration Goddard Space Flight Center, Greenbelt, MD, and accepted by Editorial Board Member Susan Hanson December 3, 2018 (received for review July 27, 2018)

Tracking the progress of the Sustainable Development Goals (SDGs) and targeting interventions requires frequent, up-to-date data on social, economic, and ecosystem conditions. Monitoring socioeconomic targets using household survey data would require census enumeration combined with annual sample surveys on consumption and socioeconomic trends. Such surveys could cost up to $253 billion globally during the lifetime of the SDGs, almost double the global development assistance budget for 2013. We examine the role that satellite data could have in monitoring progress toward reducing poverty in rural areas by asking two questions: (*i*) Can household wealth be predicted from satellite data? (*ii*) Can a socioecologically informed multilevel treatment of the satellite data increase the ability to explain variance in household wealth? We found that satellite data explained up to 62% of the variation in household level wealth in a rural area of western Kenya when using a multilevel approach. This was a 10% increase compared with previously used single-level methods, which do not consider details of spatial landscape use. The size of buildings within a family compound (homestead), amount of bare agricultural land surrounding a homestead, amount of bare ground inside the homestead, and the length of growing season were important predictor variables. Our results show that a multilevel approach linking satellite and household data allows improved mapping of homestead characteristics, local land uses, and agricultural productivity, illustrating that satellite data can support the data revolution required for monitoring SDGs, especially those related to poverty and leaving no one behind.

SDGs | remote sensing | poverty | socioecological systems | population environment

The Sustainable Development Goals (SDGs) focus on reducing poverty as well as reducing global inequalities and protecting the Earth's life support systems (1). The range of issues covered by the 17 goals and 169 targets will require more data and higher frequency of collection than is currently available (2). Household surveys are the standard approach to collecting detailed socioeconomic data but are expensive and time consuming. Most countries conduct a household census every 10 y to support government planning. Given the rapid nature of socioeconomic change, additional information is required between census enumeration periods to monitor socioeconomic indicators and targets. It has been suggested that monitoring the SDGs would require census enumeration every 10 y combined with annual sample surveys on consumption behavior and socioeconomic trends (3). Following these guidelines could cost close to $253 billion globally during the lifetime of the SDGs, almost double the official global development assistance budget for 2013 (3). This has recently led to discussions on Data for Sustainable Development at the United Nations' High-Level Political Forum (4).

The frequency of survey and census data collection varies between countries, preventing standardized approaches to monitoring progress and planning resource allocation (5). Thus, additional approaches are needed for high-frequency data collection to monitor progress toward the SDGs (6) and to provide more locally relevant recommendations and targeted SDG interventions. Recent studies have examined the role that remotely sensed (RS) satellite data could play in monitoring development in low and middle-income countries (LMICs) by producing spatial estimates of human well-being (7–9). Satellite sensors provide synoptic data on a range of biophysical parameters and land use/land cover information, which can be used for environmental monitoring and mapping. Satellite-derived data also have the potential for monitoring aspects of socioeconomic development at fine spatial and temporal resolutions (*SI Appendix*, Table S1 identifies RS features that could be used as proxies for socioeconomic conditions). This is especially clear for rural communities in LMICs that rely on natural resources and environmental products for food, fuel, building materials, and medicines (10, 11). Relationships exist between different aspects of human well-being and local environmental characteristics (12, 13), notably natural and physical capital stocks that are utilized as part of rural livelihood strategies (14). These stocks include agrobiodiversity (15), woodlands (16), and access to market infrastructure (17).

Significance

Understanding relationships between poverty and environment is crucial for sustainable development and ecological conservation. Annual monitoring of socioeconomic changes using household surveys is prohibitively expensive. Here, we demonstrate that satellite data predicted the poorest households in a landscape in Kenya with 62% accuracy. A multilevel socioecological treatment of satellite data accounting for the complex ways in which households interact with the environment provided better prediction than the standard single-buffer approach. The increasing availability of high-resolution satellite data and volunteered geographic data means this method could be modified and upscaled in the future to help monitor the sustainable development goals.

Author contributions: G.R.W., C.A.P., and J.-C.S. designed research; G.R.W., C.L.J.M., and J.-C.S. performed research; C.S. contributed new reagents/analytic tools; P.K.M. was involved in field data collection; G.R.W., C.L.J.M., and K.T. analyzed data; and G.R.W., C.L.J.M., P.K.M., C.A.P., and J.-C.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.A. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. Email: gary.watmough@ed.ac.uk.

[2]Present address: Department of Geography, University of Wisconsin–Madison, Madison, WI 53706.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812969116/-/DCSupplemental.

Published online January 7, 2019.

Data for monitoring the SDGs need to be at fine spatial and temporal scales to enable decision makers and researchers to track and understand the trajectories in development progress (1). Mismatches in scale could be a problem for understanding socio-ecological systems (18) because human uses of, and dependencies on, natural resources may differ depending on the scale at which analysis is performed (19). Past studies have highlighted the potential for RS data to be used for poverty mapping at aggregated community levels such as the village (9), groups of villages (8), or census enumeration districts (7). Aggregating household and landscape information can result in the modifiable areal unit problem (20), due to the need to construct artificial boundaries. This effectively means that the same set of data can produce different results depending on how data are aggregated and lead to erroneous conclusions. In general, the average values from single polygons used to link RS and socioeconomic data in the past mask the multilevel interactions that occur between households and environmental resources. Aggregating environmental resources into a single polygon covering multiple households assumes that all households have the same opportunity to use the landscape to pursue livelihood strategies. This could have substantial consequences for policy recommendations based on understandings of the relationship between wealth and environment resulting from these analyses (21). Wealth can vary between neighboring households. Therefore, it is reasonable to expect that the relationships between wealth and RS features will differ at the community and household level. To examine these complex relationships requires analysis of wealth and RS features at finer spatial scales than done previously.

Fine spatial resolution satellite data could be helpful for monitoring SDG1 "Ending Poverty"; in particular, it could contribute to identifying extreme poverty and those areas likely affected by poverty, targeting resource allocation, and building rural resilience to climatic and environmental impacts. In this study, we hypothesized that, fine-grained socio-economic and environmental data allow a more mechanistic understanding of human–environment interactions. We tested this hypothesis using a case study in rural Kenya by predicting household level wealth using environmental characteristics extracted from RS data. We examine two study questions crucial to understand whether RS data can be used to bridge the data gaps in monitoring aspects of household wealth: (*i*) Can the variance in household wealth be explained with RS data? (*ii*) Does a socioecologically informed approach to treating RS data increase the ability to explain the variance in household level wealth?

Results

We used a classification tree to examine if RS data could be used to predict household level wealth in the rural village of Sauri, Kenya. Within the study area, households typically live in homesteads, small areas with several structures, gardens or woodlots, and a surrounding hedge. Agricultural fields are interspersed between homesteads. Agriculture is the primary livelihood, with maize the main crop and bananas, beans, cassava, kale, and sorghum also grown. Rainfall is bimodal, allowing two cropping seasons: the long rains (March–June) during which the majority of maize crops are grown and the short rains (September–December), which are highly variable. This area is typical of many small-holder farming landscapes in East Africa; it is highly fragmented, densely populated, and topographically varied, with a complex mosaic of land cover classes. In 2005, 79% of the Sauri population was living below $1 per day (1993 PPP) and 89.5% below $2 per day (22).

We developed a multilevel approach to examine the relationships between household wealth and RS features at four spatial levels: level 1 homestead, level 2 agricultural land, level 3 village cluster, and level 4 wider village periphery (Fig. 1 and described in *SI Appendix*, Fig. S1). This method was compared with the single-level approach previously used for predicting wealth with aggregated socioeconomic data. Overall model accuracy for the multilevel approach was 60% using the training data and 45% using the testing data, between 6 and 12% higher than that using the single-level approach (Table 1). The predictive accuracy for explaining the variance in the poorest households increased from 52% in the single-level approach to 62% using the multilevel approach. *t* tests indicated that the overall test accuracy and accuracy of wealth group 1 were significantly different between multilevel and single-level approaches (*SI Appendix*, Table S3).

The statistical relationships between household level wealth and multilevel RS features are shown in Fig. 2. The most important predictor variable appears at the top of the tree, meaning that building size was the most important RS variable for explaining the variance in household wealth. Other important variables in decreasing order of importance were amount of bare agricultural land and planted agricultural land adjacent to the homestead (level 2), amount of bare land in the homestead (level 1), the count of years that the number of agricultural growing days was lower than the 14-y average for that pixel, the growing period for year 2005 of the HH survey (level 4) and the amount of land classed as homestead within the common pool resource buffer (level 3).

The poorest households were characterized by a small building size (level 1), a relatively large proportion (almost half) of bare agricultural land in level 2 and bare ground in level 1 (Fig. 2). If a household had less than 43% bare ground within the homestead area, but with less than 163 growing days in the year, it was classified in the poorest household category. Poor households that had a large building size (37/92) had less than 21% of the agricultural land planted in September, but experienced over 6 y of below-average growing periods during the 14-y time series of Normalised Difference Vegetation Index (NDVI) and had over 16% of the common pool resource buffer (level 3) covered in homestead areas. Overall, 60% (55 households of a total of 92) of group 1 households, 31% (29 households of 92) of group 2 households and only 9% of group 3 households had a building size under 140 m^2.

Fig. 1. The multilevel approach to linking households and landscape characteristics. Households have individual access to homestead areas (A, B, and C: level 1) and agricultural fields (A1–A3; B1–B3, C1–C3: level 2) surrounding the homestead. These levels should be linked to a single household. Households will also make use of common pool resources (level 3) around the village, which can be linked to multiple households. The wider regional level (level 4) considers infrastructure access. X, Y, and Z indicate fields that are adjacent to multiple households or no households, which would be split using our current method.

Table 1. Accuracies from multilevel and single-level approaches to predicting wealth using satellite features

Approach	Tree size	Test accuracy, %	Training accuracy, %	Group 1, %	Group 2, %	Group 3, %
Multilevel	7.7	45	59	62	51	55
Single-level	10.4	38	59	50	49	52

Results are averaged from 1,000 iterations of the model trained on 80% of the household sample and tested using the remaining 20%. Group 1 is the poorest 40% of households, group 2 the middle 40%, and group 3 the wealthiest 20% of households.

The majority of wealthy households were characterized as having a large building size (>140 m^2), less than 21% of the agricultural area planted by September 2004—the beginning of the short rainy season, more than 6 y of below average growing period, and less than 16% of the level 3 common pool resource area classed as homestead. Wealthy households with a small building size only had a small amount of unplanted agricultural land within the agricultural fields (level 3).

Discussion

The multilevel approach included more complex types of land use and resource access based on the spatial arrangement of homesteads and agricultural fields, compared with a traditional single-level analysis. Our results show that considering socioecological conditions at multiple levels increases the accuracy of predicting wealth from RS data.

Can Household Wealth Be Predicted from RS Data? This study considers if wealth can be predicted from RS data at the household level. Predicting wealth in this area from RS data using a multilevel approach had an overall accuracy of 45% averaged over 1,000 model iterations. This is similar to past studies that predicted socioeconomic outcomes from RS data at coarser spatial resolutions (7–9). However, the multilevel approach developed here explained 62% of the variation in household wealth for the poorest group. A relatively high accuracy considering the complexities of household wealth and predictor variables that were derived from a single satellite image.

Does a Multilevel Treatment Increase the Ability To Explain Variances in Household Poverty? The multilevel approach maps homestead characteristics, local land uses, and agricultural productivity and relates them to a single household. Results indicate that splitting the RS features into different levels can have a positive impact on model accuracy as the optimal classification trees used features derived from all four levels (Fig. 2). There was a 10% increase in predictive capacity between the multilevel and single-level approaches for group 1, but little or no difference when predicting groups 2 and 3. Wealthier households may be less reliant on agriculture for food and income with nonfarm incomes such as salaries, business enterprises, and remittances contributing more to income in wealthier Kenyan households.

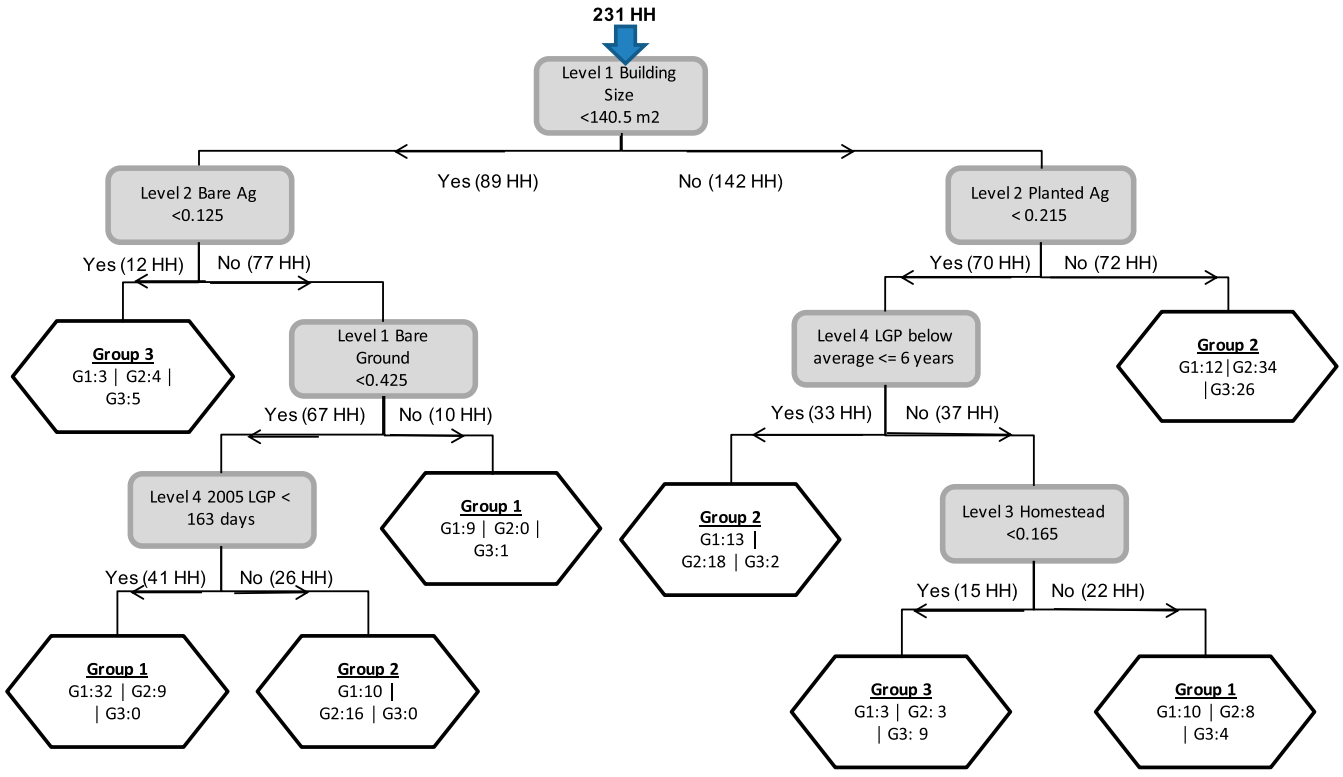

Fig. 2. Tree derived from cross-validation with an overall classification accuracy of 52%. Brackets after Yes/No indicate the number of households (HH) that met the split criteria. Group 1 = poorest, group 2 = middle, and group 3 = wealthiest households correspond to the predicted wealth group using the preceding data splits. G1/G2/G3 indicate the number of households observed in each wealth group at that terminal node. LGP, length of growing season. Level 1, homestead; level 2, agriculture; level 3, common-pool resource area; level 4, wider region for accessibility and length of growing period; bare ag, proportion of bare agricultural land within level 2.

The single-level approach assumes that all land within the buffer zone can be accessed and utilized by a given household. If an RS feature appears in multiple buffer zones, it will be linked to multiple households (Fig. 3), while in reality access to resources may be restricted to a single household. For example, homestead areas will most likely only be used by the household embedded within it. Of the 1,150 homesteads in the study area, 1,149 had more than one overlapping buffer zone with an average of 17 overlaps and maximum of 38. Thus, RS features within a homestead, which should only be linked to a single household, could be associated with up to 37 different households when using the single-level approach. This risks misestimating many households' resource access and introduces error into predictive models. The multiscale method can account for common pool resources such as hedges that are accessed by multiple households and separate them from agricultural fields and homesteads, which are likely used by single households. This result indicates that work using open data with displaced GPS coordinates such as that available from the Demographic and Health Surveys (DHS) may not be as useful for monitoring socioecological systems at fine spatial resolutions.

Relationships between RS variables and household wealth. The most important variables for explaining variance in household wealth were size of the household's buildings (level 1) and proportion of agriculture and bare land in level 2 (Fig. 2). The majority of households with small building sizes were from the poorest wealth categories (*SI Appendix*, Table S2). Small buildings likely indicate that a household has limited financial capital stock or has a small family size (human capital) with reduced labor pool and a lower diversity of livelihood strategies. Building size is not a seasonally dependent variable and could therefore provide a consistent RS variable for predicting rural wealth. The small number of households that had a small building size and were from the wealthiest group were differentiated from the poorer households by having a relatively small amount of bare agricultural land surrounding the homestead (level 2 nonvegetated <12.5%).

Tree regression allows for complexities to be identified in the relationships between wealth and RS variables. Households characterized by large building sizes had a lower proportion of bare agricultural land and a lower proportion of planted agricultural land at the start of the second planting season. Wealthier households derived 71% of their incomes from nonagricultural sources (23). Therefore, the results may indicate that these households do not need to plant second crops during the short rainy season. Poorer households were characterized as having more bare land within the agricultural fields (level 2) in September, which means that the land has likely been prepared for planting for the short rains. This is an important finding because planting during the short rains is a high-risk strategy as around 50% of harvests fail due to drought (23). This result is consistent with poorer households planting second crops through necessity due to a lack of options for growing food or generating incomes (21% of the poorest households income was derived from nonfarm activities) (23).

The main growing period in the study area is around 155 d long (between March and July) and Moderate Resolution Imaging Spectro-radiometer (MODIS) data indicate a double cropping pattern. Therefore, the model prediction that poorer households had a total growing period of <163 d is indicative of two short agricultural seasons. This could be because poorer households delay planting while hiring themselves out to plant other farmers' fields for cash payments that are used to fund their own planting. This would result in late planting and a shorter main growing period compared with wealthier households. However, it could also be due to poorer households planting different crops with different maturing periods.

A large proportion of bare ground within a homestead (level 1) was associated with the poorest households. While it cannot be determined from the imagery, bare ground in the homestead would have different uses in different homesteads. Households use this space for socializing, and drying crops among other uses. Field observations indicated that wealthier households were more likely to invest in "greening" the homesteads to provide fencing poles, wind breaks, and pasture.

The role of remote sensing in the data revolution for SDG monitoring programs. The increasing availability of high-resolution satellite data means that methods, such as those developed in this study, could support the SDG "data revolution" (4) and provide a more cost-effective way of monitoring development than annual household surveys. The World Bank estimates the costs for a household survey at $322.99 (USD 2014 prices) per household in Sub-Saharan Africa (24). This is the gold standard for surveys as it includes multiple modules and household visits. If the World Bank cost estimates were used to collect the socioeconomic information of the 330 households originally surveyed in our study site in Sauri, the total cost would be in the region of $106,500 per year. In comparison, acquisition of high-resolution satellite imagery for the 100-km^2 site ranged from $1,750 to $5,000 per year (*SI Appendix*, Table S4). The World Bank proposes to survey countries every 3 y using sample surveys of between 3,000 and 10,000 households depending on the country (24). However, to monitor socioeconomic conditions sufficiently, some form of annual survey is recommended (3). Therefore, the World Bank approach leaves up to 10 y during the 15-y SDG period with no household surveys during the SDG timeframe, which could risk our understanding of the dynamics of change. If the sampled households are a panel, satellite data covering these households could be acquired every year to provide continual monitoring of some socioecological conditions and potentially provide $100,000s worth of savings compared with household survey costs.

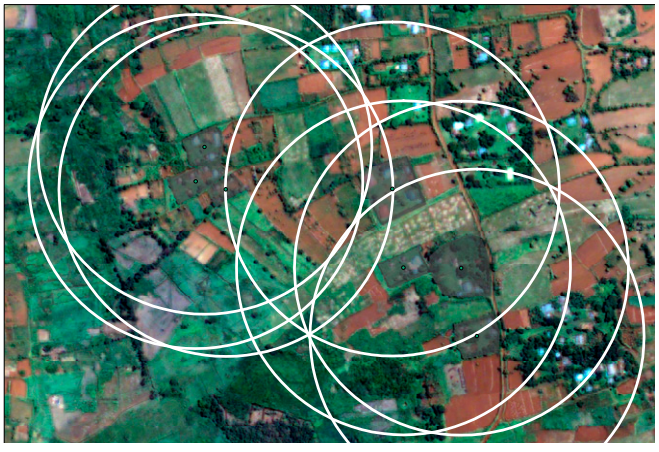

Fig. 3. The single-level approach to linking satellite and household data often uses a single radial buffer zone. This can be problematic as it results in overlapping regions and multiple pixels being assigned to multiple households when households would not have access to some land parcels such as multiple homesteads.

Future Work. The methodology developed here would need to be tested in multiple places, with different spatial arrangements of homes and agricultural fields, configurations of common resource areas, road networks, and market access. The approach still lacks detailed land tenure information but could vary the size of level 3 based on land ownership. Not all households have the same access to land and common pool resources across the landscape (25). Local and regional institutions can also impact the ways different actors access and utilize natural resources (26). Therefore, future work should examine how protected land areas, tenure rights, and institutional arrangements could be integrated into the multilevel approach. This could result in more accurate links between individual households and the parcels of land which they use. Developments in data and technology availability since the household survey was collected in 2005 provide significant future opportunities that should be explored for mapping wealth, health, and life on land. We

highlight three areas that could lead to improvements in the way that wealth is mapped and monitored through time.

i) Agricultural productivity from space: Field-level agricultural yields have recently been predicted from fine-spatial resolution RS data (27). We were unable to estimate growing period or yield at the individual field level due to a lack of data availability. Despite this, the coarse resolution (500-m MODIS pixels) growing period was an important predictor for household wealth. Since large numbers of households across the developing world rely on agriculture for food security and livelihoods, time-series information at the field level could add valuable information and increase the predictive accuracies of estimating wealth from space. The RS data required to achieve this is increasingly available from new high-resolution satellites such as the 3-m resolution Planet constellation (available from 2014, 9 y after the household survey we used) and the 10-m resolution Sentinel-2 dataset (launched 2015). Yield information at the field level could help to further examine the impacts of crop failures on other natural capital stocks. Often households use forest resources as a safety net to plug the gaps in food and income (28). A multilevel treatment of RS data may show changes in the local and regional common pool resources consistent with them being used more heavily (grassland browning over time or woodland areas reducing in size or thinning) as well as time-series data showing a drop in agricultural growing period or yield in a particular year.

ii) Document RS variables relevant for socioeconomic outcomes: Some RS variables will be seasonally sensitive such as agricultural cycles, meaning they may not be significant predictors of wealth at all times of the year. Therefore, documenting the RS variables that are significant predictors of wealth at different times of the year would seem a worthwhile activity. Knowing this information before analyzing a region would allow users to target a subset of variables from RS data, limiting time-consuming land use classifications and instead focusing on methods that target these particular variables. For example, if building size is an important predictor, a filtering algorithm could be used to identify buildings.

iii) Explore if volunteered geographic information could be used to identify agricultural field ownership: Over time, new technologies such as volunteered geographic information (29) and mobile phone location data (30) may allow for the development of models to characterize how individual households utilize landscapes. If we know the regular routes that individuals take to get to fields, roads, markets, or other resources, we can begin to think about which additional resources are being collected along these routes. For example, hedges along paths and field boundaries may be providing fuelwood, fodder, or fruit. This would provide vital information on ecosystem service availability, and any changes in particular parts of the landscape could be identified and the potential impact on livelihoods and wealth estimated. These assessments could be supported by species-level vegetation maps using hyperspectral and LiDAR data.

Conclusion

Frequent monitoring of socioeconomic changes using household surveys is prohibitively expensive. Here, we demonstrate that satellite data can predict the poorest households in a site in rural Kenya with 62% accuracy. We developed an approach to consider how spaces within the landscape are utilized by human populations, when examining the relationships between household wealth and RS features. We investigated these relationships at four spatial levels (homestead, agricultural land, village cluster, and wider village periphery) and compared this with the single-level approach previously used for predicting wealth with aggregated socioeconomic data. Our results show that considering how rural populations derive livelihoods from different spaces from within the landscape and isolating household characteristic in fine-grained RS data increases the accuracy in predicting household poverty using satellite imagery. The method can be adapted to other rural regions by examining the societal community structure and ways in which the landscape is utilized. High-resolution satellite data could provide a faster and cheaper way to track several SDGs than classic survey methods, especially those related to poverty, food security, and leaving no one behind.

Materials and Methods

The study used household survey data covering 231 households collected in 2005 and a high-resolution satellite image acquired a few months earlier in Sauri village, Yala County, western Kenya (22). GPS data were taken at each household, which permits the establishment of relationships between land parcels and households (see *SI Appendix*, section S4 for details of the household survey dataset). At the time of survey, consent to use data for related research was obtained from each household respondent. The research was compliant with European Union and Danish data protection and the Institutional Review Board of Columbia University approved all experiments involving human subjects (New York). To protect confidentiality, figures do not depict households for which we have survey data. Socioeconomic data from the household surveys were used to create a weighted relative wealth index using the approach outlined in ref. 31. The wealth index was comprised of 52 household assets such as furniture, appliances, electrical items, transport availability, and farm equipment. The index for households was grouped into three categories: poorest 40% (group 1), middle 40% (group 2), and wealthiest 20% (group 3). Splitting the wealth scores into more than three groups was not possible due to the small number of households in the survey (231) resulting in small sample sizes (*SI Appendix*, section S4 for details of assets and method for categorising the index).

Satellite Data. Features were extracted from a fine spatial resolution land use/land cover (LULC) map derived from a QuickBird image from September 2004. The image acquisition date in September coincided with the end of the main "long-rains" season and preparation for the "short-rains" season. It is likely that any bare agricultural land at this time has been prepared for the second season, and any vegetated agricultural land has been left to a natural fallow and will not be planted or is covered in perennials such as bananas. The data were pan-sharpened to 0.6-m spatial resolution covering a spatial extent of 10 × 16 km (32). Eight land use classes were identified using a combination of object-based image analysis, fuzzy-classification, and a random forest classifier. Classes included the following: Agriculture, Building, Grassland, Nonvegetated (bare ground), Road, Shrub, Water, and Woodland (details of class definitions in the *SI Appendix*, section S2). Overall classification accuracy was 90.5% (kappa coefficient 0.878), with class accuracies ranging from 79% for shrub vegetation to 98% for nonvegetated. The LULC classification method and data description can be found in ref. 32. In addition, a NDVI time series from 500-m resolution MODIS data (*SI Appendix*, Fig. S1) were used to examine the agricultural growing period between 2001 and 2006.

Linking Household and Environmental RS Data. RS features were used as proxies for livelihood capital stocks in a similar way to that described in refs. 33 and 34. The list of features derived from QuickBird and MODIS data in this study and the livelihood capitals for which they may serve as proxies are listed in *SI Appendix*, Table S1.

Single-Level Approach. The single-level approach extracts land use and environmental variables within a given distance of the socioeconomic entity under study, for example, a 1-km radial buffer zone around each village centroid (9). The size of buffer zones can be socioecologically informed using data such as the distance traveled to a particular amenity (market, hospital) or resource (firewood, agricultural fields) or by the extent of a village. We used a 200-m radial buffer zone around each household GPS location, as this was the median distance traveled by households for firewood collection in Sauri. RS features were extracted within each buffer zone (*SI Appendix*, Table S1) using the "isectpolyrst" function in the Geospatial Modeling Environment (GME; www.spatialecology.com), by calculating the proportion of the buffer zone covered in the different land use classes.

Agricultural Growing Periods. The number of agricultural growing days per year were estimated within the 200-m radial buffer zone around each household using a time series of MODIS NDVI. The MODIS 16-d MCD43A4 surface reflectance composite data were extracted from Google Earth Engine

from January 2001 to December 2006. Each 200-m buffer zone was linked to the 500-m MODIS pixel in which it was contained; if a buffer zone was on the boundary of two or more pixels, it was given the average value. A Savitzky–Golay filter with a window size of six was used to smooth the data to estimate the length of growing period per year for each pixel (*SI Appendix*, section S3). The growing period was defined as the sum of the length of both growing periods in each year. Season start and end points were identified as the point where NDVI increased/decreased by 10% of the distance between the minimum and maximum and was computed in the TIMESAT software (35).

Multilevel Approach. We developed a mechanistic approach to represent the complexity of land and resource availability by considering capital endowments used exclusively by single households, those used by multiple households, common pool resources, and community infrastructure (*SI Appendix*, section S1 for more details). For this case, we identified four levels highlighted in a stylized landscape model in Fig. 1 and *SI Appendix*, Fig. S2. At each level, particular RS features are extracted, e.g., land use within the area, vegetation productivity, and access measures such as distance to roads and market. Unless otherwise stated, land use proportions were extracted at each level using the isectpolyrst tool in the GME.

Predicting Household Wealth Using Remotely Sensed Features. Household wealth was predicted with RS predictor variables using classification trees in R 3.3.2 (R Development Core Team 2016) and the "tree" package (36). (Some analytical steps could only be performed in a single software package at the time of analysis and so multiple software packages were used. eCognition, the only software allowing for multilevel object-based image classification and region growing of the homesteads; ESRI ArcMap, industry standard GIS software; GME tool, was able to deal with overlapping radial buffers which is not possible in ArcMap Buffer tool.) Classification trees have several benefits for this type of analysis. They are simple to implement and interpret and do not assume a normal error distribution. Classification trees are also hierarchical, allowing each variable to be used for splits multiple times (37), effectively meaning that nonlinear relationships can be handled, important for modeling population–environment relationships (9). To reduce the problem of overfitting, we split the data into training/calibration (80% of the total data) and testing/validation (20% of the total data) samples (37). Each of the three wealth groups were sampled independently to ensure that the testing dataset contained 40% of households from the poorest wealth group, 40% from the middle, and 20% from the wealthiest group. The optimal tree was identified using a cross-validation approach, which prevents the model algorithm overfitting and predicting random noise in the data. The full tree was pruned to the size of the optimal tree; pruning is an essential step for generating useful predictions and ensures the most parsimonious tree with the highest predictive accuracy is obtained. The *y* variable was the wealth group of the household (1–3), and the *x* variables were the various RS features (*SI Appendix*, Table S1). The model was applied to the testing sample and a confusion matrix created using the "caret" package (38) to identify the overall model prediction accuracy as well as the accuracy of each wealth group. We repeated this process 1,000 times with the seed changed in each iteration to ensure a different set of households were included in the training and testing samples. This number of iterations ensured convergence in the calculated model prediction accuracies. This process was repeated for models using RS features extracted from the single-level approach and the multilevel approach for comparison.

ACKNOWLEDGMENTS. We thank Dr. Mark Musumba and Prof. Mat Williams for comments on an earlier version of the manuscript and Sombras Blancas Art and Design for converting Fig. 1 to digital graphics. The project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant agreement 656811. J.-C.S. considers this work a contribution to the Danish National Research Foundation Niels Bohr professorship project Aarhus University Research on the Anthropocene and a contribution to his VILLUM Investigator project "Biodiversity Dynamics in a Changing World" (Grant 16549). C.A.P. considers this work a contribution to the Millennium Villages Project, from which all the household data and high-resolution images were obtained.

1. Griggs D, et al. (2013) Policy: Sustainable development goals for people and planet. *Nature* 495:305–307.
2. Jacob A (2017) Mind the gap: Analyzing the impact of data gap in Millennium Development Goals' (MDGs) indicators on the progress toward MDGs. *World Dev* 93: 260–278.
3. Jerven M (2014a) Benefits and costs of the data for development targets for the Post-2015 Development Agenda, Data for Development Assessment Working Paper, September 16, 2014.
4. IEAG (2014) A world that counts: Mobilising the data revolution for sustainable development. *Independent Expert Advisory Group on a Data Revolution for Sustainable Development* (United Nations, New York).
5. Devarajan S (2013) Africa's statistical tragedy. *Rev Income Wealth* 59:9–15.
6. Jerven M (2014b) Poor numbers and what to do about them. *Lancet* 383:594–595.
7. Engstrom R, Hersh J, Newhouse D (2016) Poverty in HD: What does high resolution satellite imagery reveal about economic wealth? *Annual Bank Conference on Development Economics 2016: Data and Development Economics World Bank, September 2016*. Available at documents.worldbank.org/curated/en/610771513691888412/pdf/WPS8284.pdf. Accessed December 20, 2018.
8. Jean N, et al. (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353:790–794.
9. Watmough GR, Atkinson PM, Saikia A, Hutton CW (2016) Understanding the evidence base for poverty-environment relationships using remotely sensed satellite data: An example from Assam, India. *World Dev* 78:188–203.
10. Okwi PO, et al. (2007) Spatial determinants of poverty in rural Kenya. *Proc Natl Acad Sci USA* 104:16769–16774.
11. Tallis H, Kareiva P, Marvier M, Chang A (2008) An ecosystem services framework to support both practical conservation and economic development. *Proc Natl Acad Sci USA* 105:9457–9464.
12. Angelsen A, et al. (2014) Environmental income and rural livelihoods: A global-comparative analysis. *World Dev* 64:S12–S28.
13. Kristjanson P, et al. (2010) Understanding poverty dynamics in Kenya. *J Int Dev* 22: 978–996.
14. Scoones I (2009) Livelihoods perspectives and rural development. *J Peasant Stud* 36: 171–196.
15. Zimmerer K, Vanek S (2016) Toward the integrated framework analysis of linkages between agrobiodiversity, livelihood diversification, ecological services and sustainability amid global change. *Land (Basel)* 5:10–38.
16. Mamo G, Sjaastad E, Vedeld P (2007) Economic dependence on forest resources: A case from Dendi district, Ethiopia. *For Policy Econ* 9:916–927.
17. Stifel D, Minten B (2017) Market access, well-being and nutrition: Evidence from Ethiopia. *World Dev* 90:229–241.
18. Cumming GS, Cumming DHM, Redman CL (2006) Scale mismatches in socio-ecological systems: Causes, consequences and solutions. *Ecol Soc* 11:14.
19. Mcsweeney K (2002) Who is 'Forest-Dependant'? Capturing local variation in forest product sale, Eastern Honduras. *Prof Geogr* 54:158–174.
20. Jelinski DE, Wu J (1996) The modifiable areal unit problem and implications for landscape ecology. *Landsc Ecol* 11:129–140.
21. Seguin A-M, Apparicio P, Riva M (2012) The impact of geographical scale in identifying areas as possible sites for area-based interventions to tackle poverty: The case of Montreal. *Appl Spat Anal Policy* 5:231–251.
22. Sanchez P, et al. (2007) The African Millennium villages. *Proc Natl Acad Sci USA* 104: 16775–16780.
23. Mutuo P, et al. (2007) Baseline Report: Millennium Research Village Sauri, Kenya, *The Earth Institute at Columbia University*, p92. Available at mp.convio.net/site/DocServer/Sauri_Baseline_Report_final_3-7-07.pdf. Accessed date May 9, 2018.
24. Kilic T, Serajuddin U, Uematsu H, Yoshida N (2017) Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity, Policy Research Working Paper, WPS 7951, The World Bank Group, Washington DC.
25. Sen A (1999) *Development as Freedom* (Oxford Univ Press, Oxford), 366 p.
26. Leach M, Mearns R, Scoones I (1999) Environmental entitlements: Dynamics and institutions in community-based natural resource management. *World Dev* 27:225–247.
27. Burke M, Lobell DB (2017) Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc Natl Acad Sci USA* 114:2189–2194.
28. Wunder S, Börner J, Shively G, Wyman M (2014) Safety nets, gap filling and forests: A global-comparative perspective. *World Dev* 64(Suppl 1):S29–S42.
29. Norman P, Pickering C (2017) Using volunteered geographic information to assess park visitation: Comparing three online platforms. *Appl Geogr* 89:163–172.
30. Steele JE, et al. (2017) Mapping poverty using mobile phone and satellite data. *J R Soc Interface* 14:20160690.
31. Michelson H, Muniz M, DeRose K (2013) Measuring socio-economic status in the Millennium villages: The role of asset index choice. *J Dev Stud* 49:917–935.
32. Watmough GR, Sullivan C, Palm CA (2017) An operational framework for object-based land use classification of heterogeneous rural landscapes. *Int J Appl Earth Obs Geoinf* 54:134–144.
33. Watmough GR, Atkinson PM, Hutton CW (2013a) Predicting socioeconomic conditions from satellite sensor data in rural developing countries: A case study using female literacy in Assam, India. *Appl Geogr* 44:192–200.
34. Watmough GR, Atkinson PM, Hutton CW (2013b) Exploring the links between census and environment using remotely sensed satellite sensor imagery. *J Land Use Sci* 8: 284–30.
35. Jönsson P, Eklundh L (2004) TIMESAT–A program for analysing time-series of satellite sensor data. *Comput Geosci* 39:833–845.
36. Ripley B (2014) tree: Classification and Regression Trees, R Package, version 1.0-39. Available at https://cran.r-project.org/web/packages/tree/index.html. Accessed December 20, 2018.
37. James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning: With Applications in R* (Springer, New York), 426 p.
38. Kuhn M, Johnson K (2008) Building predictive models in R using the caret package. *J Stat Softw* 28:1–26.