Predicting poverty through time with publicly available data

Aamir Shakir, Martin Jaggi, Mary-Anne Hartley
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne
aamir.shakir@epfl.ch, martin.jaggi@epfl.ch
mary-anne.hartley@epfl.ch

Matthieu Gani
The Essential Medical Technology Foundation
mga@essentialmed.org

Abstract

Fighting poverty remains challenging due to laborious and expensive tracking and targeting methods, especially over time. Our work presents an accurate, scalable, inexpensive method to estimate consumption expenditure from publicly available data using surveys, satellite images and OpenStreetMap features from four African countries, Nigeria, Tanzania, Ethiopia and Malawi. Our approach is capable of predicting consumption through time. The features explain up to 75% of the variation in local-level economic outcomes, and for the temporal prediction, up to 60%. Our method presents a novel way to predict poverty over time. It could transform efforts to understand the development of poverty in developing countries and the tracking and targeting of poverty.

Poverty is the first Sustainable Development Goals by the UNO. To fight poverty it is important to monitor it. Non-Profit Organizations, governments and many others require accurate, quantitative data on the local distribution of wealth and poverty to make critical decisions or to build new plans and strategies [6]. It is costly to get accurate economic data, and barely half of all nations have access to sufficient data on poverty [12]. Even if data is collected it is mostly desegregated, geographically and temporal scarce.

Several years pass between nationally representative wealth surveys. These surveys are limited in the repeated observation of individual household or locations, which makes it difficult to measure the wealth development over time. To fill the geographical gaps researchers developed approaches to use nontraditional data to construct poverty maps. The methods include small area statistics that combine household sample surveys with comprehensive census data. More recent is the use of satellite nighttime images, records of mobile phone data, high-resolution satellite imagery, or all combined.

1 Related Work

Traditionally, the situation of poverty is recorded through national and representative surveys. These surveys collect information such as the income or consumption of a household. However, these surveys have some disadvantages; they are labour-intensive, expensive, time-consuming and complex. For this reason, states conduct surveys irregularly and on a small scale. As a result, the data collected is geographically limited, and for some regions of a country, no data is available. In their work, Jean et al. [9] present a method with which it is possible to predict poverty at a geographically fine-grained level with the help of transfer learning. To predict poverty, Jean et al. use three data sources. First: images of luminosity at night time which can signal economic level; second: high-resolution images of the daytime from which valuable information on landscape features can be extracted; finally, survey data on household consumption expenditure and asset wealth as validation for the findings extrapolated from the analysis of satellite images. To train their model, they use a convolutional neural network (CNN) model trained on ImageNet. This CNN is further trained to predict the nighttime luminosity level based on the daytime images, with the idea that the model can detect specific features in the daylight images, such as land use segmentation, signs of human assets and activities, and other useful geographic information that indicates the variation in night light brightness. This variation can be a proxy for economic activity. The trained weights are extracted and used as a feature for a Ridge Regression, which predicts the per capita consumption (pcc) in one cluster. The pcc is extracted from surveys which also defines the cluster. Their work focused on five African countries: Malawi, Nigeria, Rwanda, Tanzania and Uganda. Based on the transfer learning approach included more features such as proprietary connectivity data provided by Meta. Chi et al. presented a model which predicts the wealth index for 135 low- and middle-income countries [3]. They used the transfer learning approach and other data to train a gradient boosting tree.

2 Aim and Objectives

The work presented is primarily based on data from private providers. For example, the work by Jean et al. uses satellite imagery provided by Google Maps. These services are costly, and funding is a significant challenge for NGOs and poorer countries. In order to create applications based on the proposed methods, the data must be as inexpensive as possible. This challenge is the first concern to be addressed in this paper: To what extent can we predict poverty using only publicly available data? We will use a similar proposed approach by Jean et al. by using publicly available data to answer this. Furthermore, the work presented solves one drawback of the surveys - the geographic scope of the data. Another disadvantage is that the temporal dimension has not yet been addressed and is highly relevant. If they are conducted, surveys have mostly several years between them. The temporal observation of poverty would allow poverty development to be monitored at fine-grained intervals and possible simulations to measure the extent of specific projectnnnhgfdsa hjs, for example. The work should address the second concern of the temporal scope: To what extent can we predict poverty through time?

3 Methodology

3.1 Data Sources

To address the main questions, the data sources must fulfill two constraints. First, they have to be publicly available; second, they should be available over a long period, and the data should also be suitable for comparison over time. Our model relies on multiple inputs. Selected inputs and possible alternatives will be discussed in the following. A graphical overview and usage of the data sources are shown in Figure 1.

3.1.1 Living Standards Measurement Study

The Living Standards Measurement Study (LSMS) collects data from households across a specific country [2]. The country is divided into clusters, and the household in those clusters have the cluster's longitude and latitude to protect the household's privacy. The clusters and households are selected to have a robust statistical representation of the country. Like the Demographic and Health Surveys (DHS) program, the survey questionnaires contain multiple questions about health, education and consumption [4]. Other than the DHS, which is mainly used in literature as ground truth to predict poverty, calculates a unit-less wealth index, the LSMS collects the total consumption of each household. The DHS "is constructed as a relative index within each country at the time of the survey. Each wealth index has a mean value of zero and a standard deviation of one. Thus, specific scores cannot be directly compared across countries or over time" [11]. Since we want to address the temporal scope, we cannot use the DHS. The LSMS provides the nominal household consumption, allowing us to compare the consumption over time. LSMS can address the temporal scope and is publicly available and used as the ground truth. An exemplary distribution of the clusters is shown in

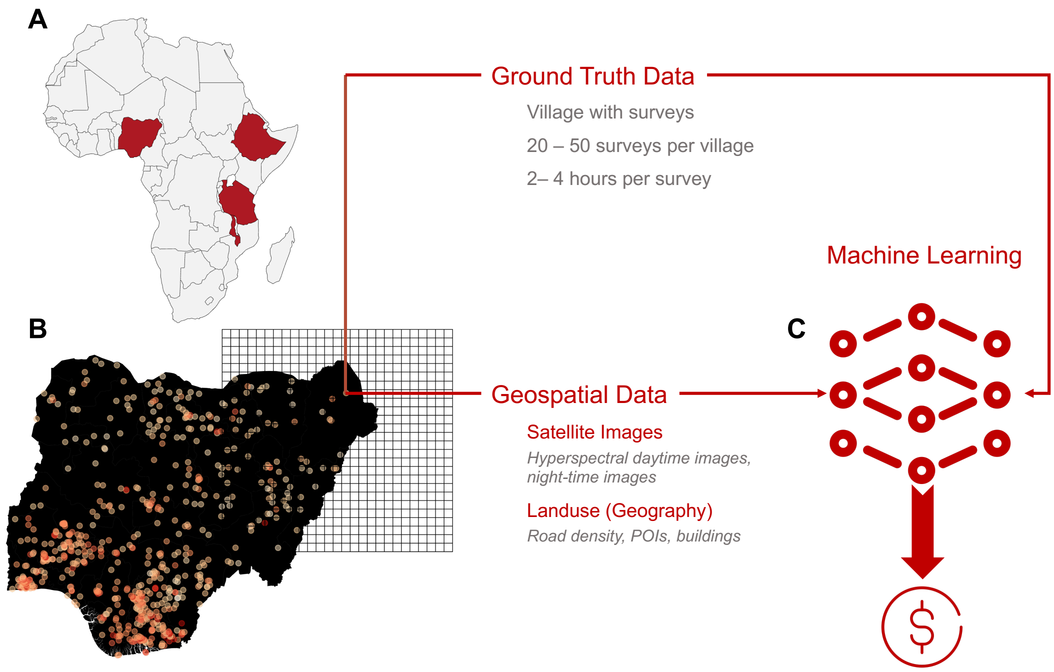

Figure 1: Overview of the approach. (**A**) Data from Living Standards Measurement Study surveys are taken from four countries (Nigeria, Ethiopia, Tanzania and Malawi). (**B**) Clusters of a country are shown for Nigeria as an example. (**C**) For each cluster, geospatial data is extracted, which is used as features to predict consumption in the clusters. Predictions are compared against the ground truth (the surveys) to estimate the machine learning model's performance.

Figure 2, on which the geographical scarcity is also clearly shown. Predicting the consumption helps to understand which clusters live under the poverty line and need urgent humanitarian help.

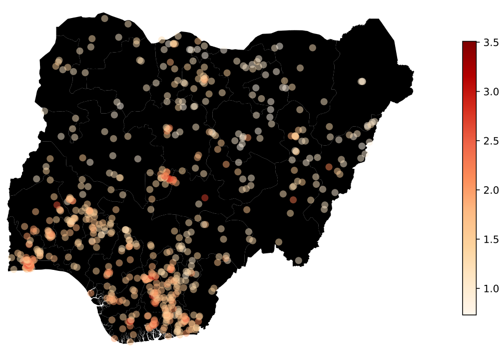

Figure 2: Distribution of clusters in Nigeria in 2015. Points indicate clusters where surveys where taken. Color indicates capita consumption. Scale (volume) is showing the per capita consumption per day per person.

3.1.2 Daytime Satellite Images

Provider of satellite images such as Google Maps violets, both points of the requirements for the data sources. The public access to the data and the temporal scope since it is a paid service in which we cannot specify the time of the image collection. The National Aeronautics and Space Administration (NASA) and European Space Agency (ESA) each have earth observation programs and images of the earth. Sentinel-2 provided high-resolution optical imaging for land services and was launched in 2014 by the ESA [5]. The data is available from June 2015, which is not covering a long time. For example there are surveys from early 2000s. Surveys are usually conducted in intervals of several years, limiting the application to a few countries that have surveys from June 2016 to today. Observations from one year are aggregated to obtain reliable data on a cluster. An alternative is NASA's Landsat program. The first generation was launched in 1992, and reliable imagery data is available from 1999 on [13]. With these properties, Landsat is ideally suited. On the one hand, it is publicly accessible, and on the other, it provides data from a protracted-time period.

3.1.3 Nighttime Satellite Images

Low-light images of the Earth from space have been possible since the mid-1960s with sensors on board the Defense Meteorological Satellite Program (DMSP) satellite platforms [1]. The data is openly available, covers an extended period, and suits our requirements.

3.1.4 Landuse

Landuse features such as buildings and streets can be extracted using satellite images. This presents us with multiple challenges. First, we need high-resolution data which is openly accessible. Then we need to annotate the data and train our model, which also introduces errors. The other alternative is using community-contributed data such as OpenStreetMap (OSMP), which provides open, accessible data. It is also possible to use historical versions, which meet the temporal scope. However, community-driven data depends on the activity of individuals. However the missing information can also be utilized as a proxy for poverty.

3.2 Data Extraction

The presented data sources provide data in domain-specific formats. This section covers the extraction of the sources, which will be used to predict poverty.

3.2.1 Day- Nighttime Satellite Images

Satellite data is mainly in the GeoTIFF format and is difficult to process. The files can be large and split into tiles. The splitting causes further complications, e.g. if the cluster is located in different tiles. We mainly used Google Earth Engine (gee) to extract the data and focused on the prediction part. Gee provides a free and easy-to-use environment to process geospatial data by hiding the complexity of the data processing behind a powerful API [7]. The computation is done on gee server, which also solves resource limits. gee has prepared datasets of Landsat and DMSP data. For Landsat, we use a composite of Landsat 7 and Landsat 8 Surface Reflectance by taking the mean of every capture in a specified year. The capture aggregate ensures better data quality with fewer clouds and artefacts. The Surface Reflectance collection is atmospherically corrected, which allows us to compare different collections with different capture dates. The Nighttime data only contains one band, which is extracted. The size of the extracted cluster is 7.65km x 7.65km. The size is determined by the required input dimension of the feature generator (CNN). It requires an input shape of 255 pixel x 255 pixel. Landsat provides a resolution of 30m/pixel. The required input shape and the Landsat resolution determine the size of the extracted cluster (255 pixel · 30m/pixel). Later processing such as cropping reduces the cluster size to 6.75km x 6.75km. A selection of extracted data samples are shown in Figure 3.

3.3 OpenStreetMap

OpenStreetMap data formats changed over time. Since we also want to explore historical entries, we must deal with the different formats. An alternative to processing the files locally is to use the Open-Source API ohsome, which is "a generic web API for in-depth analysis of OpenStreetMap (OSM) data with a focus on its history. It allows aggregated statistics about the evolution of OSM data and the contributors behind the data. Furthermore, data extraction methods are provided to access the historical development of individual OSM features." [10, 8].

3.4 Feature Generation from Satellite Images

To generate features from the Satellite Images for the models, the transfer learning approach presented by Jean et. al is used [9]. We use a pre-trained ResNet18, which takes the daytime satellite images

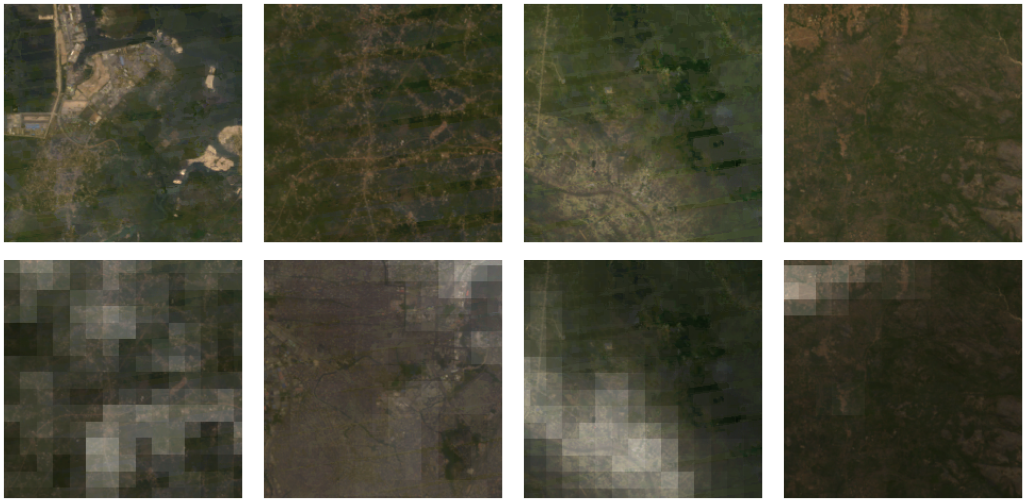

Figure 3: Four random selected daytime (top row) and nighttime (bottom row) satellite images of Nigeria in 2015 using RGB and radiance bands. Nighttime images are overlapped with daytime images. White colour stands for light activity.

as input and tries to predict the nightlight intensity. For this the mean on each nighttime entry is calculated and by using a Gaussian Mixture Model divided into five classes - very low, low, medium high and very high luminosity. The input layer is modified to 7 channels to use all available data, as shown in Figure 4. The preprocessing parameters of the network are set to the mean and standard derivation of the training data. After the training we perform a forward pass for each cluster and extract the weights from the penultimate layer. These extracted weights are used as part of the features for the models to predict poverty. Please refer to Appendix A.3 for more information regarding the training. Features generated from the CNN are in the following referred to as CNN or CNN features.

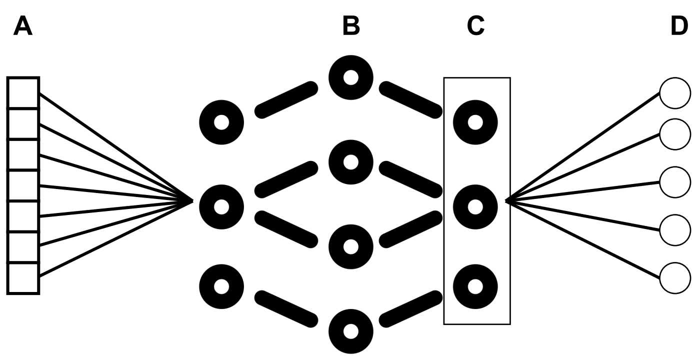

Figure 4: Figure shows high-level structure of model. (**A**) Input layer is modified to process images with 7 channels. (B) Pre-trained ResNet18 model, used to train on daytime satellite images to predict luminosity based on nighttime satellite images. (D) Output-layer is modified to predict one of the 5 bins of the input. (C) After the training a forward pass is performed on images of cluster and weights from penultimate layer are extracted and used as features for further models.

3.4.1 OpenStreetMap Features

The features for OpenStreetMap are divided into three categories i. Roads: here, we extract the count and length of different roads. ii. Point Of Interest (POI) are objects and buildings of daily life such as schools, waste bins etc. We collected the occurrences of a predefined list of POIs in a cluster iii. Buildings, for which we collected amount and area. All data points are collected in the cluster size 6.75km x 6.75km to align the cluster of the satellite images. The detailed features per category and their types are provided in Appendix A.1.

3.5 Preprocessing

3.5.1 Groundtruth

We use the per capita consumption (pcc) per person to predict poverty. Since we are addressing a temporal scope, it is crucial to consider macroeconomic effects such as inflation. The LSMS provides us with the yearly nominal consumption of one household. To compare different years with each other, we will predict the nominal pcc. A base year is selected, and the other years' data is scaled with the consumer price index. An example of the conversion is in Appendix A.2. Wealth is unequally distributed, which also affects our data distribution. Most people have a low income and lower consumption than wealthy people. To balance the distribution, we perform a log transform. As shown in the example in Figure 5 the transform leads to a better-balanced distribution. The log-transform keeps the relative differences between the entries, allowing further analyses.

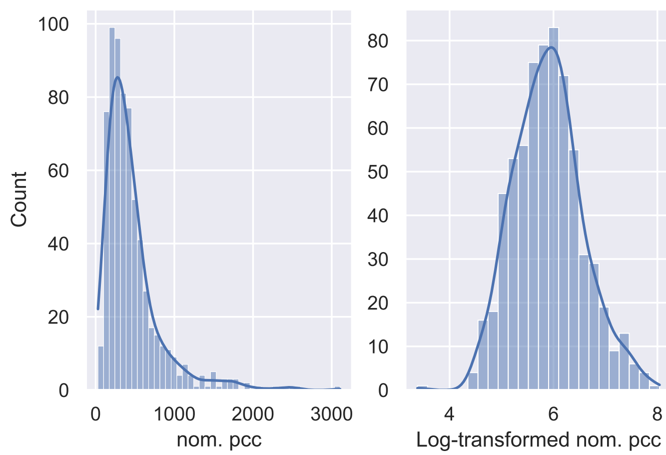

Figure 5: The left figure shows the distribution of nominal consumption. The right figure shows log-transformed nominal consumption. X axis presents per capita consumption, and y axis the occurrences in the dataset.

3.5.2 Features

PCA showed that 194 components of the features extracted from the CNN explain 95% of the variation. However, we found that our models perform better without dimension reduction. The CNN features are standardized. If combined features (CNN and OSM feature merged), the standardization is performed over the whole dataset.

4 Evaluation

We developed an approach to estimate temporal microregional wealth using publicly available data. We use data of 63260 households, aggregated to 6854 6.75km x 6.75km clusters collected in 4 countries in the period of 2012 to today. For each of the cluster we estimate the per capita consumption (pcc) per day in dollars (see A.2 for additional comparisons).

As presented in Figure 1, our approach relies on traditional face-to-face surveys as a ground-truth measurement for each household. The groundtruth is provided by Living Standards Measurement Study (LSMS) of the World Bank. One of the wealth indicators collected in the surveys is the consumption of the household. We use consumption as a wealth measurement since we want to address the temporal scope. Households are aggregated into clusters to protect privacy. The clusters contain geographical markers linked to the vast majority of nontraditional public available data, such as satellite images and community-driven map data. These data are processed using deep learning and other algorithms, which convert the raw data to a set of quantitative features of each cluster. We use the features to predict the consumption per cluster.

4.1 Predictive performance using available public data

We found that the best-performing models use a combination of features based on satellite images and land use features provided by the community-driven OpenStreetMap (OSM) service. A ridge regression trained on the four countries individually shows that the models can explain 23% to 55% of the substantial variation in cluster-level wealth (Figure 6).

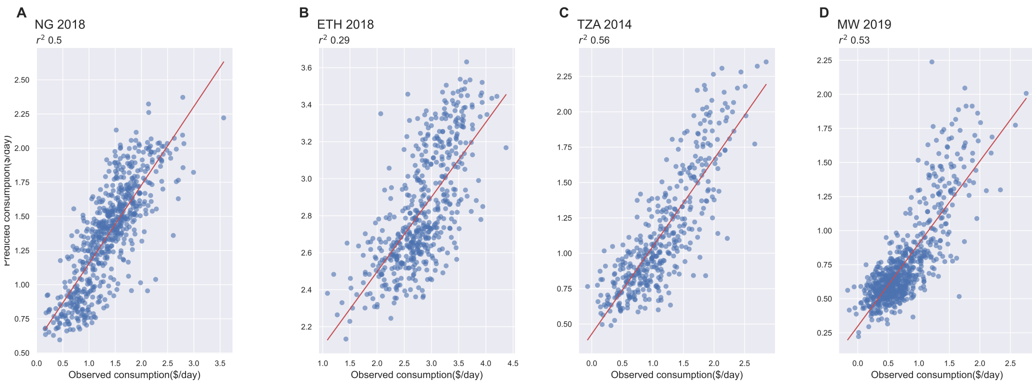

Figure 6: Predicted cluster-level consumption from transfer learning approach (y axis) and OpenStreetMap features compared to survey-measured consumption (x axis). Results are shown for Nigeria (**A**), Ethiopia (**B**), Tanzania (**C**), and Malawi (**D**). Predictions and reported r^2 values in each panel from 10-fold cross-validation. Best fit line shown in blue. Values for both axis are log-transformed. Countries are ordered by population size.

For those predictions, the recent surveys of the countries are selected. Ethiopia's performance is lower than for other countries. Table 1 shows that Ethiopia's performance is lower than the other countries, independent of the selected features. For the OSM features, a possible explanation is the lack of an active community, while for the CNN features, the captures lack enough sound quality cloud-free images. OSM features have a strong prediction power since they directly allow us to access traditional proxies such as roads, buildings, etc. Currently, a great effort is made by the OSM community to have better support for African countries. The OSM features can be an easy, cheap and accessible method to predict consumption and poverty in large parts of the world. We believe that the features can also be used to train complex models. Another benefit of the OSM features is that they are more explainable than the CNN features; here, we can find a direct correlation between real-world features and poverty. However, since the features are not available everywhere and the temporal availability is limited, we still need the CNN features. As Table 1 shows, combining both features (integrating both features) gives the best results.

Table 1: Predictions using a Ridge Regression. Scores are r^2 values from 10-fold cross-validation.

Type	CNN	OSM	CNN + OSM
Nigeria 2018	0.47	0.34	**0.50**
Ethiopia 2018	0.28	0.25	**0.29**
Tanzania 2014	0.51	0.44	**0.56**
Malawi 2019	0.44	0.47	**0.53**

Furthermore, we created a model using the data of all four countries. The model trained on the pooled dataset explains 54% of the variation (Figure 7 B). Our model can divide rural and urban areas to a certain degree. However, for the combined CNN and OSM features, our model explains 45% of the variation for rural regions and 43% for urban areas. This is an unexpected result since we would assume that urban areas should have more valuable features in satellite images and OSM. We found that the low resolution of the satellite images makes it hard for the model to learn features specific to urban areas. Also, due to the scarcity of OSM Features, the difference between rural and urban can be marginalized. Overall a high predictive power is achieved despite the imperfect knowledge of the location of the clusters, as up to 10 km of random noise was added to cluster coordinates by the data collection agencies to protect the privacy of survey respondents. Our model's performance improves with more per cent of the data used; by using more than 40% of the poorest cluster in the dataset, our model gains almost linearly explanatory power. An other geographical scope, we address is, how well we can predict consumption in other countries than the ones on which we trained the model. This would allow us to expand the prediction to nations in which we don't have any ground truth.

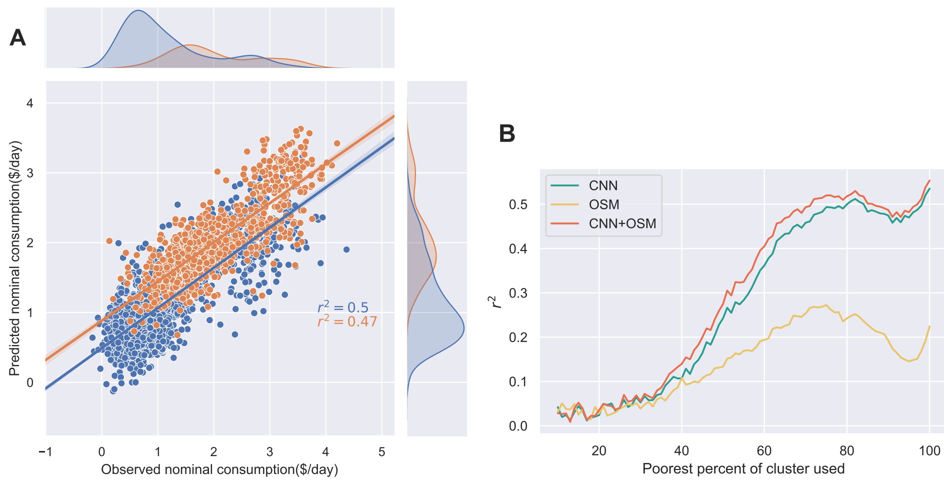

Figure 7: (**A**) Performance of the combined features for recent survey data for rural versus urban area. Each dot is a rural (blue) or urban (orange) cluster, with density showing the distribution of observed (x axis) and predicted (y axis) consumption. Predictions and reported r^2 values in each panel are from tenfold cross-validation. (**B**) Performance of CNN, OSM and combined features for estimating consumption. Trials ran separately for increasing percentages of available clusters (e.g., x-axis value of 50 indicates that all clusters below 50th percentile in consumption were included).

To verify the approach, we trained our model on the combined features of one country and cross-validated it with the data of other countries. (Figure 8). We found that the model does generalize well over country borders. This is unexpected since trying to predict consumption which has a unit like dollars, is dependent on the economics of one country; this should make it difficult for our model to generalize over countries. However, the results here are similar to the ones from Jean et al. [9].

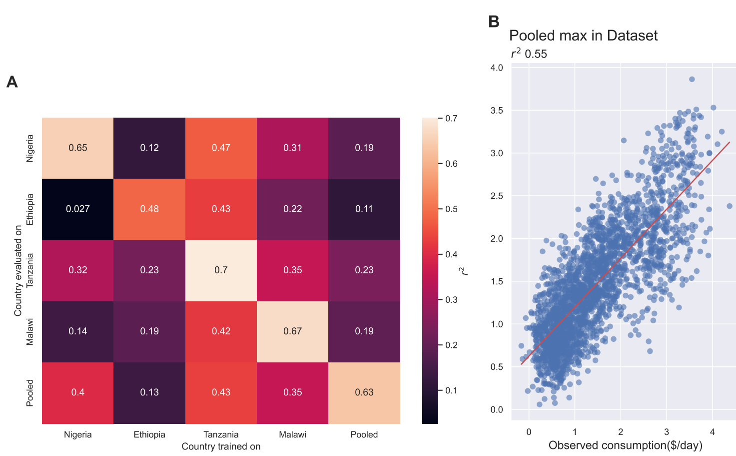

Figure 8: **(A)** Cross-validated r^2 values for consumption predictions for models trained in one country and applied in other countries. Countries on x axis indicate where model was trained, countries on y axis where model was evaluated. Reported r^2 values are averaged over 10-fold cross validation. (B) Pooled model trained on all countries, using the most recent year. Reported r^2 values are averaged over 10-fold cross validation.

4.2 Temporal predictive performance

Predicting through time is possible by using the nominal consumption of a household and scaling it to the inflation rate of a base year (see A.2 for further details). The combined features are used based on the geographical scope finding. To predict through time, we conducted three settings: A) direct time travelling, a model is trained on data of time t and is evaluated on another time t'. B) In A Leave One Out setting, when there are more than two time spans available, we left one dataset from time t out, trained the model on the data from another time, and evaluated the model on time t. C) Combined model, we merged all features from different time points (adjusted to inflation to 2010) and evaluated the performance. For direct time travelling (A) we found that the model's performance varies from country to country (see Table 2). The explanation of the model is better when we predict the future. The model relies on a combination of CNN features and OSM features. The OSM features change over time. In this case, they get more precise, and this leads to a better performance in the future prediction. The model can predict poverty through time and get results that are close to the non-time travel predictions of the combined features (Table 2) for some years. Starting from an early base year, our model can help to understand the development of consumption and poverty.

Table 2: Prediction result of predicting through time on 10-fold. Past Prediction means that model is trained on larger year and evaluated on the lower year. Vice versa for Future Predictions. Reported values are r^2.

Base Year	Year	Past Prediction	Future Prediction
		Nigeria	
2012	2015	0.11	0.26
2012	2018	0.09	0.31
2015	2018	0.37	0.49
		Ethiopia	
2013	2015	0.14	0.20
2013	2018	0.10	0.21
2015	2018	0.20	0.27
		Tanzania	
2012	2014	0.22	0.43
		Malawi	
2016	2019	0.48	0.48

Leave One Out setting (B) can only be performed on Nigeria and Ethiopia because those are the only countries with more than two datasets from different times available. For Nigeria and Ethiopia, our model outperforms the models from previous experiments (compare Table 3). The additional temporal data makes the model more robust and helps to generalize better since features change over time. Also, the prediction of the past gets better with more data from different times. Again we get worse results, if we try to predict data from another satellite for example for Nigeria trained on 2015 and 2018 predicting 2012. For 2015 and 2018 the data is from the same nighttime satellite and for 2012 from a different one.

Based on the finding of setting B, that the models seem to generalize well over time, we conduct setting C, in which we merge the different data from different times together and predict the consumption. All consumptions are scaled to the year 2010 concerning the specific country. However, we find that even with more temporal data, our model can generalize over borders (see Figure 9). However the performance could be better to use a unit less wealth index score similar to DHS surveys for this task by performing a PCA over the survey data. The results indicate, at least for the presented countries, that we can perform the consumption prediction using available public data, which is limited in its predictive power across borders in time. The unit-less prediction value could be a possible solution for predicting through time. If the same PCA is performed for every dataset, the values should be comparable over time too.

Table 3: Predictions using a Ridge Regression using Leave One Out cross-validation. Model trained on Training Years and evaluated on Evaluation Year. Scores are r^2 values from five-fold cross-validation.

Training Years	Evaluation Year	r^2	Base Performance
	Nigeria		
2015, 2018	2012	0.10	0.15
2013, 2018	2015	0.52	0.41
2013, 2015	2018	0.60	0.50
	Ethiopia		
2015, 2018	2013	0.14	0.07
2013, 2018	2015	0.35	0.21
2013, 2015	2018	0.41	0.29

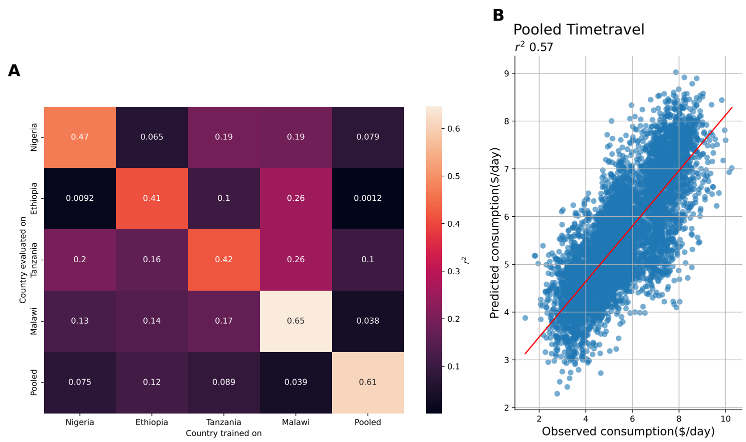

Figure 9: (**A**) Cross-validated r^2 values for consumption predictions for models trained in one country and applied in other countries. Countries on x axis indicate where model was trained, countries on y axis where model was evaluated. Data of all available time-points for one country was used. Reported r^2 was performed on 10-fold cross validation. (**B**) Performance of 10-fold cross validated pooled model over all countries. Observed consumption on x axis and predicted consumption on y axis.

5 Conclusion and Feature Work

In our work, we have shown that publicly available data can be used to predict consumption, which indicates wealth and poverty. Our approach does generalize across countries, which allows us also to predict consumption in countries, for which we do not have the ground truth. Our model only uses a simple Ridge Regression, first experiments with Gradient Boosting have shown better results. A next step would be to test the approach with different, more complex models. For example, the method could merge the CNN model with a Regression layer to have direct feedback on the predictions instead of using transfer learning. Our work showed that consumption can be predicted through time, and the models perform better if data from different time points is used. Our models lack performance when predicting the past. Future work should investigate the exact causes of the poorer performance. One possibility would be to explore the features' importance, existence and accuracy and find a way to harmonize data from different satellites. Nevertheless, to our knowledge, our proposed model is the first model that can predict consumption through time and lays the foundation for many future applications. For example, it could be used to understand the distribution and development of poverty better. Also, it can be used as impact measurement to certain for projects which takes actions against poverty. The temporal component should allow practical simulations for policymakers to find the most effective strategies. The ground truth is interchangeable. The approach could be used to estimate entities related to healthcare.

References

[1] World Bank. Introduction to nighttime light data (20 min) - open nighttime lights. https://web.archive.org/web/20220628104002/https://worldbank.github.io/OpenNightLights/tutorials/mod1_2_introduction_to_nighttime_light_data.html.

[2] World Bank. Living standards measurement study. https://web.archive.org/web/20220628104051/https://www.worldbank.org/en/programs/lsms/overview.

[3] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3):e2113658119, January 2022.

[4] MEASURE DHS et al. Demographic and health surveys. *Calverton: Measure DHS*, 2013.

[5] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[6] Chris Elbers, Tomoki Fujii, Peter Lanjouw, Berk Özler, and Wesley Yin. Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics*, 83(1):198–213, 2007.

[7] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

[8] GIScience Research Group and HeiGIT. ohsome-py: A python client for the ohsome api. https://web.archive.org/web/20220628103905/https://github.com/GIScience/ohsome-py/blob/master/README.md.

[9] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[10] Martin Raifer, Rafael Troilo, Fabian Kowatsch, Michael Auer, Lukas Loos, Sabrina Marx, Katharina Przybill, Sascha Fendrich, Franz-Benjamin Mocnik, and Alexander Zipf. Oshdb: a framework for spatio-temporal analysis of openstreetmap history data. *Open Geospatial Data, Software and Standards*, 4(1):1–12, 2019.

[11] Shea Oscar Rutstein and Sarah Staveteig. *Making the demographic and health surveys wealth index comparable*, volume 9. ICF International Rockville, MD, 2014.

[12] Umar Serajuddin, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. Data deprivation: another deprivation to end. *World Bank policy research working paper*, (7252), 2015.

[13] Michael A Wulder, Thomas R Loveland, David P Roy, Christopher J Crawford, Jeffrey G Masek, Curtis E Woodcock, Richard G Allen, Martha C Anderson, Alan S Belward, Warren B Cohen, et al. Current status of landsat program, science, and applications. *Remote sensing of environment*, 225:127–147, 2019.

A Appendix

A.1 OSM Categories

Buildings: building (any), residential, commercial, education, health, buildings

Types: count, area, density

Point Of Interests: monument, kindergarten, town_hall, stadium, optician, post_box, laundry, playground, computer_shop, outdoor_shop, florist, prison, atm, mall, camp_site, gift_shop, community_centre, veterinary, greengrocer, bar, sports_centre, university, jeweller, bank, mobile_phone_shop, camera_surveillance, drinking_water, pitch, track, toilet, water_tower, chalet, car_rental, dentist, furniture_shop, artwork, beauty_shop, library, tourist_info, park, viewpoint, motel, graveyard, hospital, comms_tower, shelter, hostel, beverages, public_building, museum, swimming_pool, kiosk, college, hairdresser, attraction, water_well, bookshop, recycling, pharmacy, sports_shop, cafe, theatre, guesthouse, stationery, picnic_site, clothes, pub, hotel, nightclub, fire_station, cinema, restaurant, waste_basket, shoe_shop, bicycle_shop, police, school, butcher, doityourself, chemist, car_wash, telephone, car_dealership, toy_shop, fast_food, food_court, tower, bakery, memorial, others, supermarket, post_office, courthouse, doctors, convenience, embassy, bench, department_store, travel_agent, fountain, water_works

Types: count

Roads: residential, track, living_street, trunk, primary, secondary, tertiary, service, pedestrian, intersection

Types: count, length, density

More about the categories: https://wiki.openstreetmap.org/wiki/Map_features

More about the types: https://docs.ohsome.org/ohsome-api/v1/endpoints.html#users-aggregation

A.2 Scaling of nominal the pcc

In economics, we can distinguish between two values, the real value and the nominal value. In comparison, the nominal value is taken without any normalization. For the real consumption, we take the purchasing power parity (ppp) to scale it. In short: It's a basket which contains the price of different goods. We can get the real consumption by dividing the nominal consumption by the ppp. To reduce the prediction over time, we use the nominal consumption, which is comparable by scaling with the consumer price index. Inflation is measured by the relative difference of the underlying Price Index for a given time frame. The consumer price index allows us to compare the difference for any given year. We select a base year and a target year and calculate the difference between the consumer price indexes. This value is used to divide through the nominal value to scale it.

For example, we want to predict poverty in Nigeria. As the base year, we select 2012 and as target 2015. The consumer price index has grown from 124.4 to 158.9 (27.8%). One household has a consumption of 1000 in the year 2015. To get the value of this 1000 in 2012, we divide by 1.278. The result is 782.57. The household would have to spend 782.57 to have the same consumption in 2015. This fact allows us to compare consumption over time.

A.3 Training Configuration of CNN

The modified ResNet18 is trained on all images of the clusters across all time points. Which gives us more trainingsdata. In total we used 6854 images. Model is trained with Stochastic Gradient Descent with momentum. Data is split into test, train and validation sets. Weights of model with best performance on validation set is chosen. Our best model had an accuracy of 66%. According to CUMULATOR each training generates an overall carbon footprint of 13.4 gCO2eq.