Spatial & Temporal Mapping of Socio-economic Development Using Satellite Data for India

ABSTRACT

Satellite data is being actively used to build indicators that can act as a proxy for socio-economic data, in the absence of such data being available from ground-level surveys. We do a comprehensive evaluation in this paper to assess the efficacy for prediction of socio-economic indicators in India, using different sources of satellite data, at spatial scales or villages and districts, with different kinds of classifiers, and to also predict indicators over time. We make several useful observations and are able to build a method for assessing the aggregate longitudinal trend in the development of districts. This can be useful to build applications that can red-flag areas that might be deteriorating in important indicators, and commission targeted studies for these areas.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → *Computing in government*.

KEYWORDS

poverty mapping, socio-economic development, landsat, VIIRS, convolutional neural networks, census

ACM Reference Format:

. 2019. Spatial & Temporal Mapping of Socio-economic Development Using Satellite Data for India. In *COMPASS'19: ACM SIGCAS Conference on Computing and Sustainable Societies, July 03–05, 2019,Accra, Ghana.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The use of satellite imagery to serve as a proxy for socio-economic indicators has seen a lot of attention in recent years [6, 8, 15, 16, 19, 23]. Satellite data is collected at a high spatial frequency, several times a year for many satellites, and is available over many years. This can be useful to address problems in the collection of socio-economic data at high-frequency and fine-spatial scales. In India, the national population census which covers a wide range of indicators at the village level is done only every ten years, and the subsequent release of compiled census data takes several more years. The economic census which is again at the village level, is also conducted at gaps of more than ten years. Further, several indicators are changed between successive rounds of the census, which makes it hard to do longitudinal comparison. Other surveys done by the National Sample Survey Office (NSSO) are done on a sample basis at the state level. There is certainly a push towards more frequent data collection in areas like health, with annual National Family Health Survey (NFHS) surveys being conducted, or maintenance of individual case-data of mothers and children to track progress during pregnancy and early childhood, but by and large it remains a problem to study socio-economic phenomenon at fine spatial and temporal scales.

Much of the analysis using satellite data to predict socio-economic indicators has been facilitated with rapid advances in machine learning systems which allow for the processing of very large datasets and to mine predictive patterns from the data. However, not much work has been done to evaluate whether models learned on data from one time period are able to work on other time periods as well, and similarly to understand which socio-economic indicators are likely to be predicted well through satellite data. In this paper, we conduct a comprehensive evaluation of several different models. We first identify six socio-economic household indicators from the Indian population census of 2011, that we aim to predict: asset ownership, fuel used for cooking, type of employment, bathroom and sanitation facilities, main source of drinking water, and main source of light. We use three types of models for this prediction: Using nightlights data which captures the brightness of lights at night, using spectral data of different bands with hand-crafted features created on the spectral bands, and using RGB imagery with a Convolutional Neural Network (CNN) model architecture. For each of these models and indicators, we compare how well they work at the district level and the village level. Finally, we also evaluate how well the models are able to predict the indicators at different times.

Our anticipation was that the prediction of indicators like the main source of light could benefit from nightlights data, the main source of water could be related with features for surface-level water-bodies that could appear in both day-time spectral and RGB data, indicators for asset ownership could be related with urbanization features of dense residential construction seen in RGB data, etc. We found that deep-learning models are generally able to perform better than other models, but all the models perform differently for different socio-economic indicators. We were not able to identify any discernible patterns behind why some models would work better on some indicators. The performance also dropped significantly when trying to predict the indicators for 2001, which was the round of census data collection previous to 2011. In the process, we found that nightlights data works better when chosen from winter months, and that several regions which have a frequent cloud cover are the ones for which the models typically fail. It seems therefore that it may not be straightforward to generalize the use of satellite data for the prediction of socio-economic indicators. On the upside, we were able to predict with reasonable robustness the overall temporal trend of how socio-economic development of an area is improving or deteriorating over time, ie. trend prediction seems to

COMPASS'19, July 03–05, 2019, Accra, Ghana

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn

work when aggregated across multiple indicators, and smoothed to remove sporadic noise in the predictions. This we found to be useful to identify locations that were significantly deteriorating or improving, and would warrant further investigation.

We next described related work, followed by a description of the dataset, and then the evaluation of different models to solve the various classification problems.

2 RELATED WORK

The use of satellite imagery is now actively used in economics to build proxy indicators for socio-economic development. Studies have shown a strong correlation between nightlights and GDP at the country level [10]. However, while national level predictions are reasonably accurate, the performance at sub-national levels is low [2]. This is believed to be because nighttime lights have a tendency of spillover to neighboring regions, called the blooming effect, and hence do not work well at small spatial scales [11]. We use a new satellite system for nightlights, called VIIRS, and treat is as a baseline model upon which to improve the accuracy of our models.

Other than nightlights, day-time imagery has also been used by researchers. A village-level analysis to monitor development using high resolution satellite images was done through a deep CNN based regression model for a hand crafted asset vector for six Indian states [19]. Transfer-learning approaches have also shown good performance, where day-time images were trained on nighttime lights and then applied with model tuning for the prediction of socio-economic indicators [13, 19]. Commune level predictions (communes are an administrative subdivision in Vietnam, similar to villages in India) have been shown to work as well [6]. We follow a similar approach with using CNN based models.

Related research has shown that CNN based models learned for some socio-economic indicators in a country, are not able to predict the indicators accurately for countries in other continents [8]. This is because of extensive parameter tuning of the models. Other researchers have tried to predict land-use and land-cover classes, such as industrial areas, residential areas, cropland, forest, etc, or even just built-up and non-built-up areas, instead of predicting socio-economic indicators [6, 9], but these models have not been tested for reliability across time.

3 DATASET

3.1 Satellite data

We used data from three different satellites for our analysis, all obtained from the Google Earth Engine (GEE).

3.1.1 VIIRS nighttime data. This consists of monthly average radiance composites of nighttime lights obtained from the Visible Infrared Imaging Radiometer Suite (VIIRS) of sensors, which provides data at an approximately 500m (15 arc seconds) resolution. We use a stray-light corrected composite of this data, available via GEE at a monthly basis [12]. Note that this data has not yet been filtered to screen lights from aurora, fires, boats, and other temporal lights.

3.1.2 Landsat 7. The Landsat-7 satellites provide daytime spectral data at a 30m resolution, from 1999 onward. We downloaded the Tier 1 top-of-atmosphere reflectance [3, 22] data (highest available quality) from GEE. It contains ten primary bands shown in Table 1, and several additional bands can be derived from these primary bands. Most of our analysis is done using the Landsat-7 data since it goes far back in time, but a studies have reported deteriorated data in recent times in the RGB bands due to a correction error in Landsat-7 data [20]. Therefore, in one instance we use RGB data from the new Landsat-8 system instead of Landsat-7. Landsat-8 data is available from 2013 onward [21].

Band	Type / Expression	Resolution
B1	Blue	30m
B2	Green	30m
B3	Red	30m
B4	Near Infrared	30m
B5	Shortwave Infrared 1	30m
B6_VCID_1	Low-gain Thermal Infrared	30m
B6_VCID_2	High-gain Thermal Infrared	30m
B7	Shortwave infrared 2	30m
B8	Panchromatic	15m
BQA	Quality Assessment Bitmask	
NDVI (derived)	(B4-B3)/(B4+B3)	30m
MNDWI (derived)	(B2-B5)/(B2+B5)	30m
NDBI (derived)	(B5-B4)/(B5+B4)	30m

Table 1: Landsat 7 bands

3.2 Census of India: 2001 and 2011

The Government of India conducts a population census every ten years. We used data from the 2001 and 2011 censuses, available from the official website [4]. The census reports the number of households in each spatial unit (village, district, state), belonging to 90 different categories spanning from the type of construction of the house, the cooking fuel which is used, assets owned by the household, type of employment, sector of employment, etc. The 2001 data was available only at the district level, while for 2011 we had village-level data as well. Between 2001 and 2011, 47 districts were split into smaller districts due to administrative changes, and for our analysis we grouped them back into the original 593 districts that existed as of 2001. We have district level shapefiles for all the 593 districts, which we use to demarcate the corresponding satellite data. We also have village-level shapefiles for approximately 150,000 villages from across 166 districts; we therefore conduct our village level analysis on this smaller set.

3.2.1 Discretization of socio-economic variables. Each socio-economic category variable provided by the census may have multiple parameters, with the number of households reported for each parameter. For example, for the variable regarding the primary type of fuel for cooking used by households, the census reports the number of households in each spatial unit using firewood, kerosene, LPG (Liquid Petroleum Gas), PNG (Piped Natural Gas), biogas, etc. For our analysis, we wanted to reduce these to a single value for each variable and we developed the following procedure. We first group the mutually exclusive parameters within a variable into three broad types of rudimentary, intermediate, and advanced. For example, firewood is considered as a rudimentary type of fuel for cooking, kerosene and cowdung are grouped together as an intermediate type, and PNG and LPG are grouped together as advanced types of fuel for cooking. This grouping is showing in Table 2 at the district level, along with the range of values for different parameters.

Variable	% of population employed in	Non Agricultural (in %)	Agricultural (in%)	High Unemployment (in %)
Employment	Unemployed	50-60	40-50	55-65
	Agricultural labour	5-10	25-35	15-20
	Non Agricultural	22-35	10-15	8-15
Variable	**Using/Access to**	**Level-1 (in %)**	**Level- 2 (in %)**	**Level -3 (in %)**
Asset Ownership	TV	15-30	30-50	60-85
	Telephone	35-55	40-60	50-60
	Cycle	50-70	10-30	35-50
	2 Wheeler	5-12	5-18	20-40
	4 Wheeler	0-2	0-5	2-12
Bathroom Facility	No Latrine facility	65-82	20-40	18-40
	Pit Latrine	0-5	30-45	0-10
	Piped Sewer/Septic Tank	15-28	25-40	50-70
Fuel for Cooking	Firewood	60-80	0-12	10-25
	Cow Dung/Kersosene	30-50	40-60	5-20
	LPG/PNG/Bio gas	15-40	5-20	45-65
Condition of Household	Dilapidated House	5-10	0-5	0-5
	Livable House	55-65	40-50	25-35
	Good House	30-40	45-55	65-75
Main Source of Light	No source of light	0-5	0-5	0-5
	Kerosene oil/Other oil	70-80	30-50	5-15
	Electricity/Solar Light	20-30	50-70	85-95
Main Source of Water	Well/Spring/River	40-70	2-20	5-15
	Hand Pump/Tube Well	2-25	55-80	10-28
	Tap Water/Treated water	20-40	10-28	60-85

Table 2: Census variables: % of households mapped to coarse types for each variable

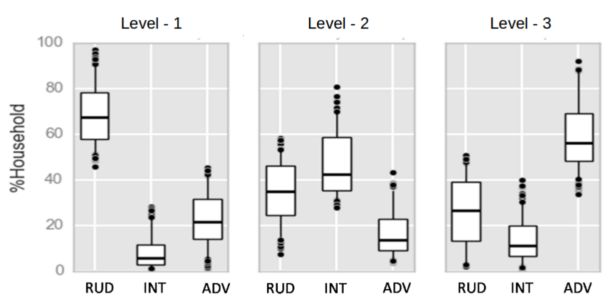

Figure 1: Fuel for cooking: Box-plot for districts at three levels

We then do a k-means clustering on the districts based on the percentage of households in each district that are of type rudimentary, intermediate, and advanced. Figure 1 shows a box-plot for the distribution of districts across three levels (k = 3) in terms of their use of different types of fuel for cooking. This allows us to label each district as a level-1/2/3 district: Level-1 districts predominantly use rudimentary types of fuel for cooking, level-2 districts primarily use intermediate types of fuel, and level-3 districts predominantly use advanced types of fuel for cooking. This method therefore allows us to map each district to a single coarse value for each variable.

We experimented with different values of k for different variables in terms of the quality of clusters obtained, and eventually settled on k = 3 as a reasonable and uniform mapping of districts for all the variables. This method of discretization is useful for several reasons. First, as shown in the book *Factfulness* by Hans Rosling [17], who used a similar 4-level mapping for different stages of development of countries and regions, such a coarse mapping is easy for people to interpret and to compare different districts with one another. Second, it reduces the variables to a single quantity without assigning arbitrary weights to club together different parameters for each variable. Each district may be marked as belonging to a different level for different variables, for example, a district could be at level-1 in terms of its dominant use of fuel for cooking, but at level-3 in terms of asset ownership, and so on. We apply the same method to also classify districts in terms of the dominant type of employment: agricultural, non-agricultural, or high unemployment.

The same method was used to generate 3 levels for the villages, for each socio-economic indicator.

4 ANALYSIS FRAMEWORK

We present our results organized in the following manner. We first distinguish between cross-sectional and temporal classification performance. Cross-sectional classification is done for the year 2011 to see which socio-economic indicators can be predicted accurately from different satellite data sources. Temporal classification is to see if the cross-sectional classifiers learned for 2011 can accurately predict the conditions that prevailed in 2001. Second, for both cross-sectional and temporal classification, we build classifiers to operate at two levels: District and village. Third, for each of the combinations of district/village-levels and cross-sectional/temporal, we build three types of classifiers: Using VIIRS data, using Landsat data with hand-crafted features, and using Landsat data in a CNN (Convolutional Neural Network) model. We present these results in the forthcoming sections.

5 CROSS-SECTIONAL CLASSIFICATION

5.1 District-level cross-sectional classification for 2011

5.1.1 Problem formulation. We formulate this as a multi-class classification problem. The input is a satellite image of a district, and the output is the socio-economic census-label for an indicator. Different models are built for every socio-economic indicator.

5.1.2 VIIRS-based classification. Monthly VIIRS data for the entire country is obtained at a 500m resolution for 2014. This is the year closest to 2011 for which the census-labels are available. We assume that the slow-changing nature of these indicators will not cause much bias due to this temporal offset between the two data sources. The VIIRS data provides average radiance (avg_rad) values for each 500m x 500m pixel, between (-1.5, 340573). Upon studying the distribution of these values, we found that 90% of the pixels had values between (-1.5, 1), hence to obtain more diverse binning we use $log(2 + avg\_rad)$ values. Shown in Figure 2 is the distribution of these values for two sample districts.

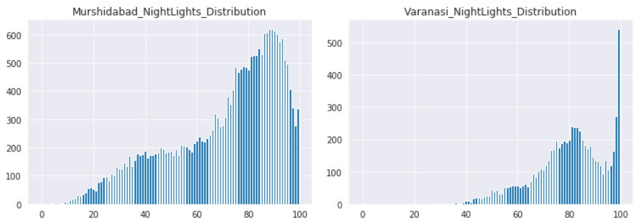

Figure 2: VIIRS Distribution of two Districts

We make a feature vector for each district based on such histograms, using a binning strategy that creates bins having an equal frequency of pixels in each bin. We then run a random forests classifiers for the district feature vectors to predict class labels for each socio-economic indicator separately. We ran such a model for each month of the year, doing a 80:20 split for training & testing, and shown in Table 4 is the performance of the classifiers for different socio-economic indicators, using a 5-fold cross validation.

Month	Jan	Feb	Mar	Apr	May	Jun
F1 Score	0.576	0.583	0.582	0.560	0.523	0.534
Month	Jul	Aug	Sep	Oct	Nov	Dec
F1 Score	0.567	0.557	0.599	0.584	0.541	0.582

Table 3: Monthly F1-scores(macro) using VIIRS data at the district level, averaged over all six indicators

We use data only from the month of December for computing VIIRS features because the classifiers seem to operate better on winter data, possibly due to the days being shorter and hence more use of lighting in the households making them more visible at night. Table 3 shows the average macro F1 scores over models for the different indicators, and we can see in general that using data from the winter months seem to give a better performance.

Indicator	**VIIRS-based**		**Landsat-coarse**	
	Macro	**Weighted**	**Macro**	**Weighted**
Fuel for Cooking	0.59	0.65	0.69	0.74
Employment	0.59	0.60	0.68	0.70
Bathroom Facility	0.59	0.58	0.65	0.69
Main Source of Water	0.64	0.69	0.73	0.78
Main Source of Light	0.46	0.54	0.68	0.73
Assets	0.59	0.60	0.68	0.67

Table 4: District level F1-scores for 2011

5.1.3 Landsat-based coarse classification. Landsat-7 data is obtained at a 500m resolution for the year 2011. The cloud cover is removed through a standard process suggested in GEE, by filtering out images with a small cloudcover value and then taking a median of the spectral bands in the pixel-level data [7]. Histogram-based features are then created using all the primary and derived bands. A similar equal-frequency binning strategy is used as done for VIIRS-based features, with 10 bins for each band. We then ran multiple classifiers (logistic regression, random forests, SVM, k-neighbours) on four different combinations of feature vectors: RGB bands only (B1,B2 and B3), derived bands only (NDVI,NDBI and MNDWI), RGB and derived bands (6 bands in total), and all the bands. The best performance was obtained when all bands was used, with a random forests classifier. The F1 scores are shown in Table 4.

5.1.4 Observations. We have reported in Table 4 both the F1 weighted scores (more weight is given to classes having more elements) and the F1 macro scores (average of class-wise F1 scores). Macro scores are useful to check whether minority classes are being predicted correctly, whereas the weighted scores give a preference to each class based on the size of the class. We can see that Landsat-based classifiers are able to outperform VIIRS-based classifiers for all the socio-economic indicators. This is probably not surprising because many nightlight maps of India (for example, [1, 12, 18]) have shown that vast areas of the country hardly show much nightlights activity due to low electrification or low provisioning of lighting especially in rural homes, hence nightlights based features are not able to capture sufficiently diverse information. The absence of cloud-cover removal methods and other spurious lighting activities like fires, also probably makes nightlights less useful for prediction of socio-economic indicators. Among the various indicators, we see that main source of water, fuel for cooking, and main source of light, are among the better predicted indicators using Landsat data, when considering weighted F1 scores. We are unable to guess why this might be so, but it is important to point out that the performance is quite varied across different indicators. These observations are consistent with macro F1 scores as well.

5.2 Village-level cross-sectional classification for 2011

5.2.1 Problem formulation. As before, we formulate this as a multi-class classification problem with the input being the satellite image of a village and the output being the census-label for different socio-economic indicators for the village. We have shapefiles for approximately 150,000 villages across 166 districts, for the states of Bihar, Orissa, Karnataka, Gujarat, Kerala and Maharashtra.

5.2.2 VIIRS-based classification. As before, we use VIIRS data for the month of December, and build histograms using the same binning strategy. 5-fold cross-validation scores are shown in Table 5.

5.2.3 Landsat-based coarse classification. This time we use Landsat-7 images at a 30m resolution, and build histogram based features as before. The performance is reported in Table 5.

Indicator	**VIIRS-based**		**Landsat-coarse**		**Landsat-InceptionV3**		**Landsat-ResNet50**	
	Macro	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**
BF	0.36	0.73	0.48	0.79	0.48	0.73	0.50	0.75
EMP	0.51	0.60	0.52	0.66	0.54	0.63	0.56	0.65
FC	0.49	0.67	0.58	0.75	0.57	0.71	0.58	0.72
MSL	0.43	0.49	0.62	0.66	0.64	0.67	0.65	0.69
MSW	0.47	0.54	0.59	0.61	0.60	0.64	0.61	0.65
ASSET	0.46	0.51	0.55	0.60	0.55	0.56	0.57	0.58

Table 5: Village level F1-scores for 2011

5.2.4 Landsat-based CNN classification. Recent advancements in the performance of CNN-based deep-learning methods motivated us to try them for our analysis. The Resnet50 [**?**] and InceptionV3

[**?**] architectures trained on Imagenet, have been actively used in a transfer learning approach to apply them to different domains. We follow the same strategy, using the Landsat RGB bands to obtain village-level images, and predict the different socio-economic classes. Separate models are learned for each socio-economic indicator. Since we only use RGB bands here, we used the Landsat-8 data given the problem of degraded quality reported for Landsat-7.

Since villages vary in shape but we need to provide a constant-size input to the deep-learning models, we checked that 224x224 sized shapes were able to accommodate 98% of the villages. We therefore decided to allow for some loss of information by cropping the remaining larger villages to this size. Areas outside the village boundaries were padded with zero values.

The dataset was then divided into 80% training and 20% test images. Since there was class imbalance, the class ratio was kept the same in both the training and test data, and class_weights (to penalize the classes based on their size each time they are misclassified) were used to offset the imbalance.

Figure 3 shows the model architecture. The final layers in the Resnet50 and InceptionV3 architectures were removed, and a global average pooling layer [**?**] followed by 2 fully connected layers with dropout parameters [**?**] were added. The last layer is a fully connected layer with 3 nodes, and softmax activation for the predictions. These is a standard architecture used for transfer-learning in image classification [**?**].

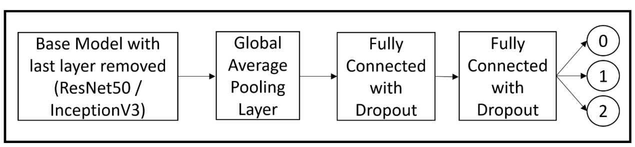

Figure 3: CNN model architecture

The model training was done in a 2-step process. First, we initialized our network weights with those of ResNet50 and InceptionV3 pre-trained on the Imagenet dataset. The layers from these base models were then frozen, and the newly added layers were trained with a batch-size of 64. After this transfer-learning, we further fine-tuned the entire model. Tables 5 shows the performance for different socio-economic indicators. We see that the ResNet50 model performs better than InceptionV3, and use it for the temporal analysis described in the next section.

5.2.5 Observations. We can see that deep learning models outperform the VIIRS and Landsat-coarse classifiers. However, it is interesting that the socio-economic indicators which were being predicted better at the district level (main source of water, fuel for cooking, and main source of light), are not the same ones that are being predicted better at the village level. Consistently across all the classifiers at the village level, the indicators for bathroom facilities and fuel for cooking, are predicted much better when considering the weighted F1 scores. We are unable to guess why this might be so, and it raises a concern on the seeming arbitrariness of why classification of some indicators works better than others. Further, although the macro F1 scores at the district level seemed to be more or less consistent with the weighted F1 scores, this does not hold

Indicators	F1-scores (2011)		F1-scores (2001)	
	Macro	Weighted	Macro	Weighted
Bathroom Facility	0.65	0.69	0.57	0.71
Employment	0.68	0.70	0.60	0.62
Fuel for Cooking	0.69	0.74	0.60	0.69
Main Source of Light	0.68	0.73	0.45	0.46
Main Source of Water	0.73	0.78	0.54	0.56
Assets	0.68	0.67	0.30	0.28

Table 7: District-level F1-scores for 2001 & 2011, using Landsat-coarse classification

true when classifying at the village level. A possible reason could be that since we use k-means clustering, we may have obtained poorer quality clusters at the village level where villages close to the cluster boundaries may be getting misclassified. This does seem to be the case for the indicator on bathroom facilities for which there is heavy spillover for level-2 and level-3 districts, that seems to explain the difference in performance for this indicator when considering macro and weighted scores. Yet even for the other indicators there does not seem to be a similarity in the performance of classifiers for the indicators at different spatial scales.

Indicator	District Direct Landsat Coarse 2011		District Aggregated Landsat CNN 2011	
	Macro	Weighted	Macro	Weighted
Bathroom Facility	0.54	0.69	0.76	0.82
Employment	0.72	0.75	0.76	0.77
Fuel for Cooking	0.61	0.75	0.77	0.84
Main Source of Light	0.75	0.77	0.82	0.85
Main Source of Water	0.69	0.78	0.81	0.86
Assets	0.76	0.76	0.53	0.52

Table 6: Village level F1-scores for 2011

6 TEMPORAL CLASSIFICATION

We next check how well the models trained on 2011 data can predict the values for 2001. This will help understand if indeed the value of satellite data being available over time, can be realized. Since we have district level census data for 2001, we can evaluate the district-level models directly. For village-level models, we first predict the village-level census-labels for 2001, and then build a village-to-district aggregation classifier which yields the district-level census-labels from the predicted village-level labels. Further, we check for every two years between 2001 and 2011, if there is a consistency in the longitudinal trend for village and district level predictions: Since all the indicators we study are likely to be slow-moving indicators, it would be surprising to find both upward and downward movements in levels for the same district or village.

6.1 District-level classification for 2001

We start with using the Landsat-coarse district-level model learned for 2011, to predict the 2001 census-labels for various socio-economic indicators. Table 7 shows the F1-scores for 2001, and for comparison also shows the corresponding F1-scores for 2011. The scores consistently dip sharply for 2001, showing that the district-level models do not translate in a straightforward manner across time.

Indicators	Increasing	Constant	Doubtful		
			Non-Negative Slope	Negative Slope	High Error
Bathroom Facility	12.03%	48.43%	11.40%	16.71%	11.40%
Fuel for Cooking	10.46%	46.09%	9.06%	22.65%	11.71%
Main Source of Light	1.71%	49.21%	0.0%	8.43%	40.62%
Main Source of Water	4.68%	45.93%	14.06%	25.46%	9.84%
Employment	8.75%	51.87%	8.90%	16.87%	13.59%
Assets	3.75%	39.06%	10.78%	17.96%	28.43%

Table 8: District level trend analysis between 2001 to 2011

To understand this in more detail, we use the 2011 models to predict the classes for every two years, until 2001. For each socio-economic indicator, this gives us six datapoints for every district, of the level for the district in a year. Since we expect the census-label changes from 2001 to 2011 to mostly be either positive (movement to a higher level) or stay the same, we show in Table 8 in the first column, the percentage of districts that showed a movement from a lower level in the initial years to a higher level in the later years. Similarly, the second column shows the percentage of districts that stayed at the same level over the years.

We can see that for most indicators almost 40-50% of the districts show some negative movement at some point in time, and we mark these districts as 'doubtful'. We then attempt to remove any uncorrelated noise from these districts in the following way. We fit a regression line between the six datapoints from 2001 to 2011 on the x-axis, and the district level on the y-axis. We assume an equal spacing between the levels, and similarly an equal spacing between the years. We then find the *stderr* for each district and plot the distribution of these residuals to find a knee-point[1] (Figure 4): We consider that districts with a small residual may just have had some sporadic noise. Note that the socio-economic levels for the districts are actually ordinal variables and hence our method is not strictly correct, but it can be argued to be a way to distinguish cases with sporadic uncorrelated noise from cases which may be truly unexpected. Based on the method, the third and fourth columns in Table 8 show the additional percentage of districts which have a non-negative or negative slope but with a small standard error.

[1]We identify the kneepoint by joining the first and last points on the sorted *stderr* values, and find the furthest point from this line.

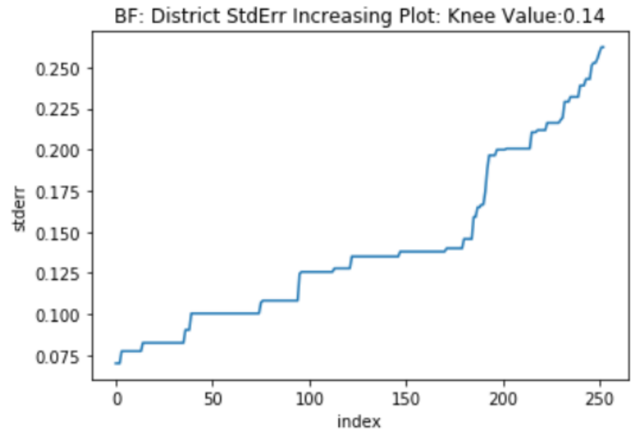

Figure 4: Kneepoint identification to identify districts showing an error in longitudinal trend

We see that other than indicators for the main source of light and for assets, the remaining indicators have roughly only 10% of the districts appearing to have errors in their longitudinal trends. Figure 5 shows in darker shades of red those districts that have more indicators showing large errors in the trend. We can see that districts along the Himalayan ranges, the North East regions, and coastal regions, have low confidence scores of trend prediction. We hypothesize that this could be because of frequent cloud cover in these regions, and figure 6 shows in darker shades of blue those districts with a larger percentage of pixels having their cloud bit set to 1 (using the Quality Assessment (BQA) Band) [22]. We can see visually that districts with high errors show a strong co-incidence with districts having higher cloud cover: The Pearson correlation coefficient between the error scores and cloud scores is 0.18 but is highly significant with a p-value less than 0.00001. This shows that it may not be very reliable to use satellite data for often cloudy districts.

Next, we leave aside the districts that have at least one high error trend prediction for any indicator. For the remaining districts, we show in Figure 7 in a darker shade of green those districts that have a larger number of indicators showing a positive trend over time. We can see clusters around large metropolitan cities (Kolkata, Mumbai, Bangalore, Chennai), which is not unexpected since concentrated economic growth in these areas is known to have led to growth in socio-economic indicators, and also pulled migrants from neighboring regions who continue to repatriate savings to their homes or engage in seasonal rural-urban migration cycles [5, 14].

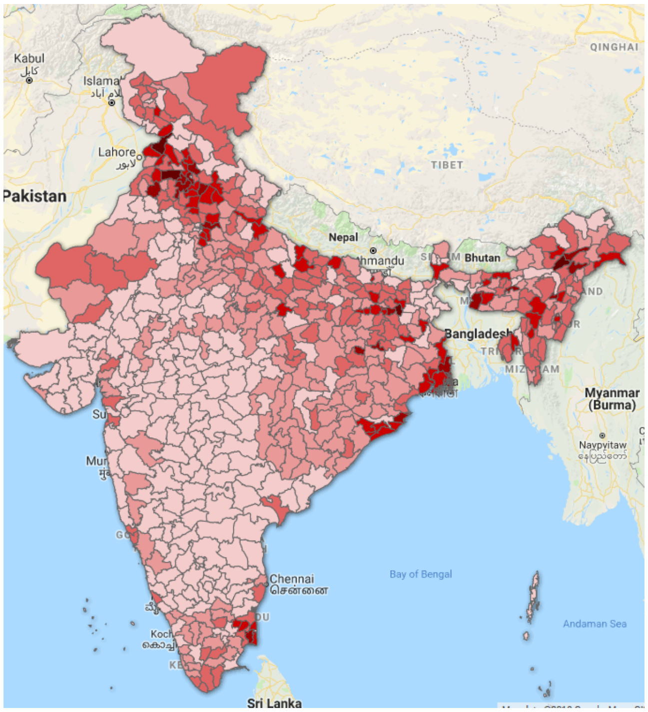

Figure 5: Districts with more number of unsure indicators are shown in a darker shade of red

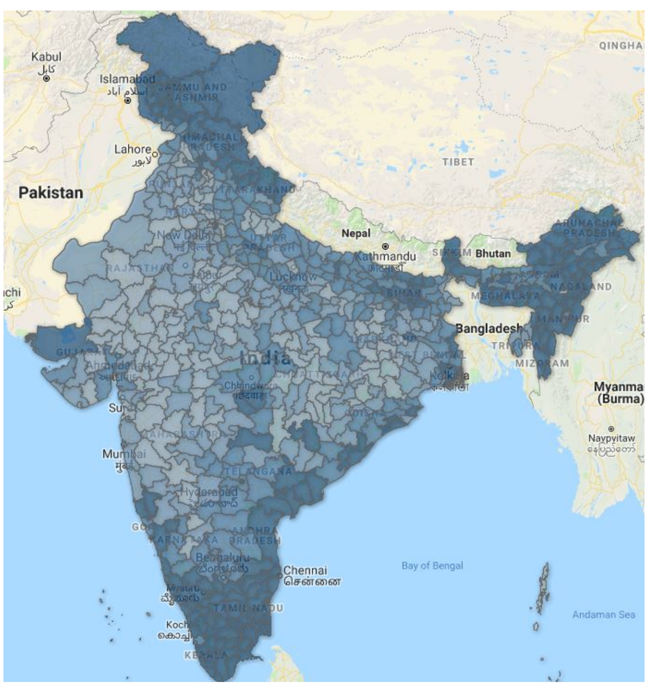

Figure 6: Districts with more average cloud cover across the years are shown in a darker shade of blue

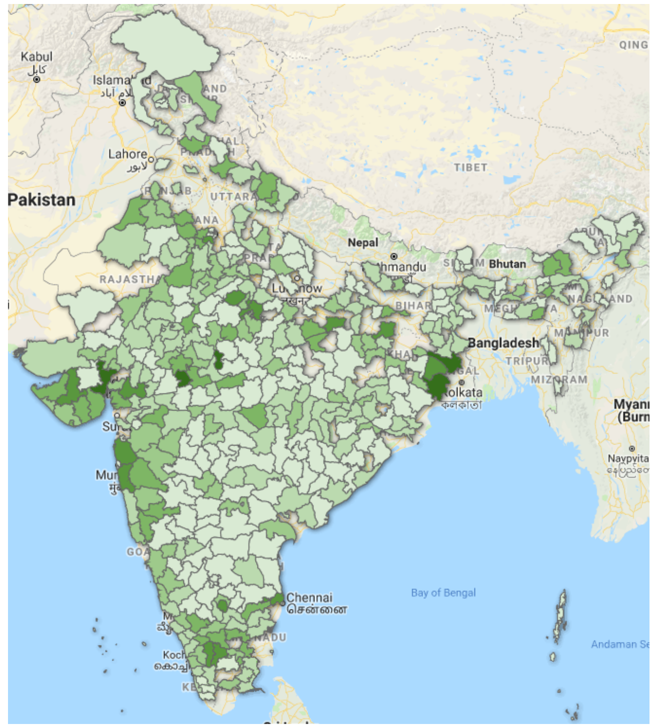

Figure 7: Districts with more number of indicators moving in a positive direction are shown in a darker shade of green

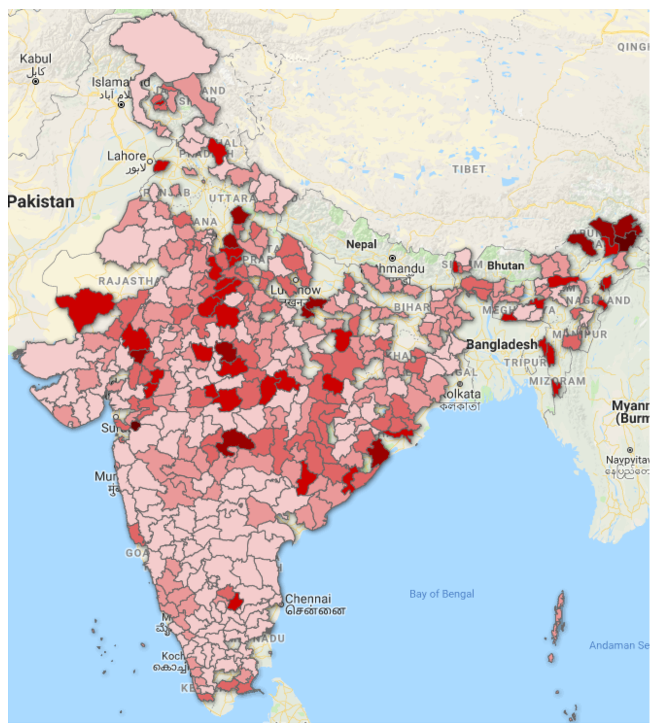

Figure 8: Districts with more number of indicators moving in a negative direction are shown in a darker shade of red

We can also see some tier-2 cities having grown strongly, such as Nagpur and Indore. Other clusters include the garments and textile belt in Tamil Nadu in South India, a large part of Gujarat which is known to have had high industrialization over the years, and the Mewar region in Rajasthan and Bundelkhand region in Uttar Pradesh which are drought prone areas and have seen significant proactive efforts by the government. This seems to indicate that while it may not be very reliable to use classifiers trained on a particular year to predict socio-economic indicators in any other given year, but a trend prediction may be more reliable. We evaluate this by checking the correlation between trend prediction using the Landsat-coarse classifier, and changes that have happened between 2001 and 2011 according to the census ground-truth data. A Pearson correlation between the count of indicators that show a positive movement, and the corresponding count based on actual census data, show a strongly significant correlation of 0.20. For individual indicators, bathroom facilities and fuel for cooking show a much higher correlation of 0.50 and 0.58, both significant.

Figure 8 similarly helps identify a few districts that have shown a negative movement for several socio-economic indicators. The correlation for negative movement with actual data from the census is also strongly significant with a value of 0.12. We are understanding more about these districts of what could have been the key factors for such a fall in socio-economic development between 2001 and 2011. This brings out the potential for using satellite data to raise alerts about areas that could be of concern.

6.2 Village level classification for 2001

Following the same approach as in the previous section, we use the Landsat-CNN village-level classifiers to predict results for 2001. Since village-level census-labels are not available for 2001, we aggregated the village-level predictions to a district-level prediction. This was done by weighting the predicted village-levels with the population of the villages, to find the amount of populations living at different levels. The district was then assigned the majority population level. This was done for the 166 districts for which we had village-level shapefiles. The F1-scores for different socio-economic indicators are shown in Table 9.

Indicator	District Level Direct		District Level Aggregated CNN	
	Macro	Weighted	Macro	Weighted
Bathroom Facility	0.49	0.78	0.50	0.79
Employment	0.56	0.60	0.64	0.67
Fuel for Cooking	0.45	0.68	0.58	0.81
Main Source of Light	0.40	0.43	0.57	0.59
Main Source of Water	0.62	0.64	0.57	0.58
Assets	0.27	0.49	0.29	0.55

Table 9: District-level F1-scores for 2001: Landsat-coarse direct classification and Landsat-CNN village classification aggregated to the district level

We can see that these district-level F1-scores aggregated based on the predicted village-levels, are better for all the socio-economic indicators (except main source of water) than the Landsat-coarse district-level temporal prediction. Although these are still lower than the cross-sectional predicted F1-scores for 2011, and raise concerns about the ability to use satellite data to predict socio-economic development in any given year, but as we show next, the Landsat-CNN village-level models seem to give more consistent trend prediction across the years.

Indicator	Increasing	Constant	Doubtful		
			Non-Negative Slope	Negative Slope	High error
BF	5.96%	69.20%	6.11%	14.00%	4.70%
FC	11.97%	59.75%	5.82%	9.47%	12.98%
MSL	12.48%	33.42%	15.03%	23.74%	15.32%
MSW	13.08%	40.86%	8.87%	20.16%	17.01%
EMP	5.33%	56.49%	6.86%	19.10%	12.21%
Asset	12.59%	43.99%	10.78%	17.32%	13.10%

Table 10: Village level trend analysis between 2001 to 2011

Following a similar method as at the district level, we run the Landsat-CNN village-level models on alternate year, between 2001 to 2011. Table 10 shows the % of villages categorized under different degrees of plausible error. Comparing Table 10 with Table 8, we can see that not only more villages fall in the first two columns of non-doubtful cases, but even the last column of high error villages has lower percentages than the percentages of high error districts. This is somewhat comforting that trend prediction using Landsat-CNN classifiers at the village-level may be feasible and can be used for district-level trend prediction as well, instead of using a direct district-level prediction using Landsat-coarse classifiers. This encourages us to attempt predicting the future: The trend from 2011 to 2018.

6.3 Trend prediction between 2011 to 2018

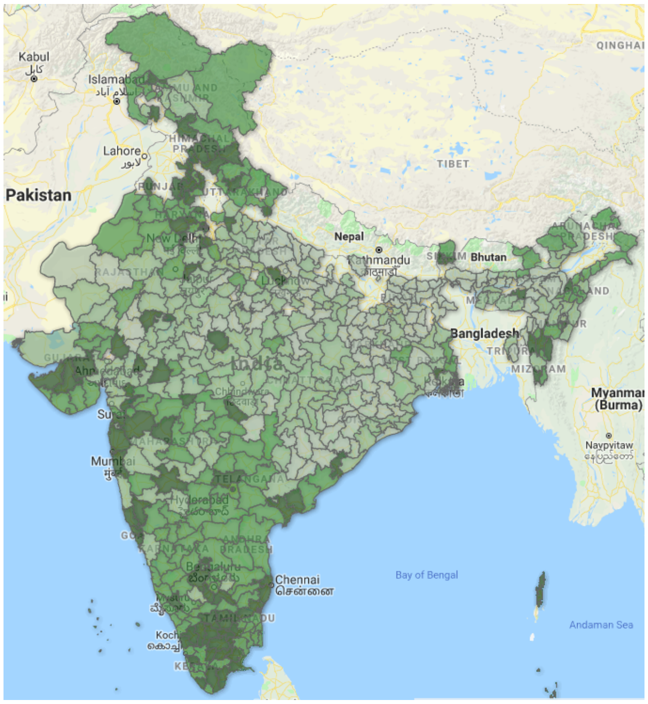

Figure 9: Status as of 2011: Districts with more positive indicators are shown with a darker shade of green

We show in Figure 9 the aggregate socio-economic development using the census data as of 2011. Most development is around the metropolitan areas and those with industries.

We then use the Landsat-CNN classifier to predict socio-economic indicators in 2013, 2015, and 2017, using the satellite data. We use the same method as earlier to remove districts for which the stderr in the trendline is large, and we also remove districts which were already at level-3 in 2011 for most indicators. Shown in Figure 10 is the degree of change based on the trend: Districts improving in most indicators are shown in green, those deteriorating are shown in red, and the remaining in yellow. Most positive movement is seen in districts of Madhya Pradesh and Gujarat, and a few in Bihar and Jharkhand, all of which were at lower levels in 2011. Many districts however of Orissa, Uttar Pradesh, Bihar, and Maharashtra, which were at lower levels in 2011 have not changed over the years. We plan to validate these predictions using articles from mass media, to see if there are reports on any possible reasons for a lack of growth in these areas.

7 CONCLUSIONS

We presented a comprehensive analysis of the potential to use nightlights and daytime satellite data for the prediction of socio-economic indicators at different spatial scales, and whether the models can predict indicators over time. We found that nightlights-based models are sensitive to the months for which data is used, and even models for daytime-imagery are sensitive to whether

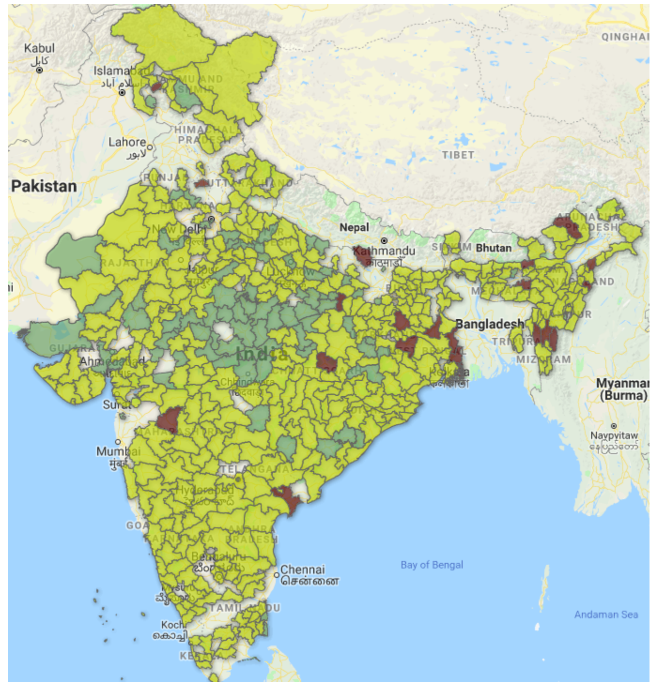

Figure 10: Change between 2011 to 2017: Improving districts are shown in green, deteriorating in red, and no change in yellow

the locations have a tendency for cloud cover. In general, deep learning models using day-time imagery work better than other models, but we were not able to discern any patterns about why some models work better for some socio-economic indicators. The models also did not translate well to predict indicators for some other specific time. However, the models are able to predict temporal trends reasonably accurately, especially in aggregate to convey whether districts have by and large improved or deteriorated on most indicators. This is a useful application which can help identify hotspot locations that should be investigated further, and as part of future work we plan to analyze mass media and other data about these locations to evaluate the hotspot predictions.

REFERENCES

[1] Thushyanthan Baskaran, Brian Min, and Yogesh Uppal. 2015. Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections. *Journal of Public Economics* 126 (2015), 64–73.
[2] Frank Bickenbach, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder. 2016. Night lights and regional GDP. *Review of World Economics* 152, 2 (2016), 425–447.
[3] Gyanesh Chander, Brian L Markham, and Dennis L Helder. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote sensing of environment* 113, 5 (2009), 893–903.
[4] C Chandramouli and Registrar General. 2011. Census of India 2011. *Provisional Population Totals. New Delhi: Government of India* (2011).
[5] Priya Deshingkar, Rajiv Khandelwal, and John Farrington. 2008. *Support for migrant workers: The missing link in India's development.* ODI London.
[6] Ran Goldblatt, Alexis Rivera Ballesteros, and Jennifer Burney. 2017. High Spatial Resolution Visual Band Imagery Outperforms Medium Resolution Spectral Imagery for Ecosystem Assessment in the Semi-Arid Brazilian Sertão. *Remote Sensing* 9, 12 (2017), 1336.
[7] Google. [n. d.]. SLC-off Products: Background. https://bit.ly/2F3vlfO.
[8] Andrew Head, Mélanie Manguin, Nhat Tran, and Joshua E Blumenstock. 2017. Can Human Development be Measured with Satellite Imagery?. In *ICTD*. 8–1.
[9] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. 2017. Eurosat: A novel dataset and deep learning benchmark for land use and land

cover classification. *arXiv preprint arXiv:1709.00029* (2017).

[10] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028.

[11] Tengyun Hu, Jun Yang, Xuecao Li, and Peng Gong. 2016. Mapping urban land use by using landsat images and open social data. *Remote Sensing* 8, 2 (2016), 151.

[12] IDFC Institute. [n. d.]. Spatial distribution of IndiaâĂŹs GDP. https://bit.ly/2CmbyYC.

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

[14] Kunal Keshri and Ram B Bhagat. 2013. Socioeconomic determinants of temporary labour migration in India: A regional analysis. *Asian Population Studies* 9, 2 (2013), 175–195.

[15] Brian Min and Kwawu Gaba. 2014. Tracking electrification in Vietnam using nighttime lights. *Remote Sensing* 6, 10 (2014), 9511–9529.

[16] Brian Min, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. 2013. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. *International journal of remote sensing* 34, 22 (2013), 8118–8141.

[17] Hans Rosling. 2019. *Factfulness*. Flammarion.

[18] Ajai Sreevatsan. 2018. What night lights reveal about the Indian economy. https://bit.ly/2GYsqDY.

[19] Potnuru Kishen Suraj, Ankesh Gupta, Makkunda Sharma, Sourabh Bikash Paul, and Subhashis Banerjee. 2017. On monitoring development using high resolution satellite images. *arXiv preprint arXiv:1712.02282* (2017).

[20] USGS. [n. d.]. SLC-off Products: Background. https://on.doi.gov/2CmzKtw.

[21] USGS. [n. d.]. USGS Landsat 8 Collection 1 Tier 1 and Real-Time data Raw Scenes. https://bit.ly/2FggkJ2.

[22] USGS/Google. [n. d.]. USGS Landsat 7 Collection 1 Tier 1 and Real-Time data TOA Reflectance. https://bit.ly/2u9EGxt.

[23] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. 2016. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*.