GEO-AMAZON EARTH OBSERVATION CLOUD CREDITS PROGRAMME

Integrating Earth Observation Data with Censuses and Sample Surveys to Estimate Development Indicators for India

Submitted by

iihs™
INDIAN INSTITUTE FOR HUMAN SETTLEMENTS

Indian Institute for Human Settlements

Lead Author: Krishnachandran Balakrishnan, Ph.D.

Institution: Indian Institute for Human Settlements (IIHS)

Mailing address: 197/36, 2nd Main, Sadashivanagar, Bangalore, India 560080

Email: kbalakrishnan@iihs.co.in

Phone: +91 – 9632653377

Collaborator: Satyukt Analytics

Contact person: Sat Kumar Tomer, Ph.D.

Mailing address: #117, Sanjay Nagar Main Road, Aswathnagar, Bangalore, India 560094

Email: sat@satyukt.com

Phone: +91 **-** 9986568525

PROPOSAL SUMMARY

In most developing countries, tracking development indicators is a challenging task due to lack of data at sufficient temporal and spatial resolution. While India has regular censuses and sample surveys, the censuses are limited in temporal resolution and the sample surveys are limited in spatial resolution. Advances in the collection, management and analysis of large scale earth observation datasets presents a unique opportunity to integrate earth observation data with these censuses and sample surveys to overcome their spatial and temporal limitations. We propose to develop such a system for India which will make use of statistical and machine learning techniques to integrate data from earth observation, census and sample surveys to generate land-cover maps, population distribution maps and development indicator datasets at sufficient spatial and temporal resolution. The GEO-AWS cloud credit will be used to kickstart this project. We propose to solve three main problems as described below. Each problem and its solution constitute one work package (WP) of the proposed project.

WP 1. Lack of Land Cover Data at Sufficient Resolution

Problem: The land cover datasets currently available for India are lacking either in spatial or temporal resolution. For instance, while annual land cover data is available from global land cover maps prepared using MODIS data, this data is of 500m resolution and therefore of insufficient spatial resolution for analysis of human settlements (Friedl et.al., 2010). Globeland30 and the Indian Space Research Organization (ISRO) have generated 30m resolution land cover data, but both these datasets are available only for two time periods (Chen et.al, 2015; NRSC, 2014). Even these 30m datasets are unable to capture the granularity and heterogeneity of land cover within the larger villages and urban areas of India.

Solution: We will develop a biennial time series land cover data starting from 1985, using freely available earth observation data from the Landsat, Sentinel and Resourcesat programs. An Open Data Cube will be set up to compile and process the earth observation data. A three step land cover classification approach will be implemented. In the first step we will use 30m land cover data which we have generated for the most populous 100 cities—as part of the India Urban Atlas project at IIHS (Malladi et.al., 2017)—to train and validate a machine learning algorithm which can identify built-up cells. In the second step, the ISRO land cover dataset will be used to train and validate a machine learning algorithm to identify other land cover classes. In the third step we will use a sub-pixel prediction algorithm to generate degree of built-up and degree of vegetation data for all the built-up cells identified in the first step. The output from all three steps will be combined to generate a 30m resolution land cover map with sub-pixel information for the built-up cells.

WP 2. Lack of Population Distribution Data at Sufficient Resolution

Problem: The population data from the Census of India can currently be mapped only at coarse resolutions—at the district and sub-district levels. This is primarily due to the lack of availability of spatial boundaries below this level. As a result, at present it is very difficult to systematically map intra-urban population distribution, or population distribution at the village level.

Solution: We will build on the recent release of spatial boundaries for the more than 600,000 villages and cities of India by Meiyappan et.al. (2018) for 1991 and 2001 Censuses. Using the 2001 dataset as a starting point, we will build a similar spatial boundary dataset for the 2011 Census also. After this we will use land cover data from WP 1 to implement a dasymetric mapping (Eicher & Brewer, 2001; Mennis, 2003; Wright, 1936) approach to redistribute population to the built-up cells within village and city boundaries. In cities, we will use degree of built-up and degree of vegetation data from WP1 along with street density for each built-up cell to redistribute population.

Initial results show that our method is a substantial improvement on the 100m population maps available from Worldpop which is the best available dataset at present (Worldpop, 2017).

WP 3. Lack of Development Indicators at Sufficient Resolution

Problem: The Censuses, National Sample Surveys and National Family Health Surveys are the primary source of data on development indicators in India. But the decennial census is of insufficient temporal resolution while the sample surveys are limited in spatial resolution.

Solution: Using the outputs from WP1 and WP2, we will develop a system for integrated analysis of data from censuses, sample surveys and earth observation data. We will use a Bayesian Hierarchical Modeling approach to build on the strengths of each dataset to generate maps of development indicators and associated uncertainties at the settlement scale and at a temporal resolution which matches the sample surveys.

APPLICATIONS AND EXTENSIONS OF PROPOSED PROJECT

The outputs from this project will form a set of base datasets which will be used by cities and states to co-develop and co-implement more detailed data portals for tracking and communicating development indicators and exposure of populations to various risks. We have already initiated work with various city and state governments towards this.

The proposed project will contribute to monitoring indicators related to Sustainable Development Goals 1, 2, 3, 6, 9, 11 and 15. The list of goals and full list of relevant indicators our project addresses is provided in the main text of the proposal. The project will also enable India to address the following Target, Guiding Principle and Priorities for Action under the Sendai Framework for Disaster Risk Reduction which relate to understanding and disseminating disaggregated information about disaster risk. This includes Target (g) under Article 18, Guiding Principle (g) under Article 19 and Articles 23 and 24 under Priorities for Action #1.

Using the outputs of the proposed project we will be able to generate development trajectories at a settlement scale starting from the 1991 Census. This can then be used to produce forecasts of developmental trends based on various policy scenarios. Other high frequency earth observation data like soil moisture, cultivation patterns and reservoir storage level data can also be integrated into this in the future to improve estimates of development indicators and economic activity.

THE TEAM

This project is envisaged as a transdisciplinary collaboration between researchers from the natural sciences, social sciences and technology domains. The team working on this project consists of researchers with expertise in Computer Vision, Machine Learning, Statistics, Economics, Public Policy, Geographical Information Science, Disaster Risk Reduction, Hydrology, Agricultural Remote Sensing and Environmental Planning. In addition, we will be working with representatives from city and state governments to co-develop and co-implement specific projects related to monitoring of development indicators and disaster risk reduction which builds on the outputs of this project.

ABOUT IIHS

The Indian Institute for Human Settlements (IIHS) is a national education institution committed to the equitable, sustainable and efficient transformation of Indian settlements. IIHS is actively engaged in research, consulting, teaching and professional training. IIHS faculty and researchers come from a wide range of disciplinary backgrounds and work collaboratively to address complex developmental challenges. IIHS is also actively engaged with city, state and national governments in consulting, professional training and advisory roles.

FULL PROPOSAL

INTRODUCTION

Since 2015, several landmark international framework agreements have been arrived at which together are shaping global development priorities. This includes the Sustainable Development Goals (United Nations, 2015a), the Sendai Framework on Disaster Risk Reduction (UNISDR, 2015), the Paris climate accord (United Nations, 2015b) and Habitat III (United Nations, 2017). But the effective implementation and monitoring of these depend on availability of reliable data at sufficient spatial and temporal resolution (IEAG, 2014; United Nations, 2015a). Even when data is available it may often exist in a siloed manner, making integrated analysis difficult.

The data related inadequacies are especially serious in least developed and developing countries (Serajuddin, 2015). Within developing countries, the problem is especially acute in cities since they are often highly heterogeneous and rapidly growing. At the same time, given the concentrations of population and potential for economic and social transformation, cities are key to achieving many of these development goals. They are also often the sites of extreme deprivation and expose large numbers of people to risks of various kinds (United Nations, 2017).

Data on population and development indicators in India also has similar inadequacies. While India does conduct a well regarded and robust national census every ten years, the spatial and temporal resolution at which data is made available presents significant challenges for research and policy. The Census of India releases data in tabular form at the village level for rural areas but does not provide georeferenced spatial boundaries for these villages, thereby making spatial analysis at this resolution impossible. In urban areas, data from the Census is released only at the level of the lowest administrative units called 'wards'. But data aggregated to the ward level often conceals substantial variation within them (Balakrishnan, 2016). The Census does provide basic population data in tabular form at the substantially finer resolution of 'Enumeration Blocks' for several cities (Census of India, 2011a)**,** but it does not make spatial boundaries available for these either.

On the temporal front, the decennial census is often inadequate for policy making given the rapid rate of change, especially in cities. For instance, Bangalore almost doubled in population from 4.3 million to 8.4 million during the period between the 2001 and 2011 Censuses (Census of India, 2001; Census of India, 2011b). India does have a National Sample Survey Organization (NSSO) which conducts surveys with greater frequency, but the samples are representative only at the state or national levels. India also conducts a more infrequent National Family Health Survey (NFHS) as part of the international program of Demographic and Health Surveys (DHS) (ICF, 2017) which is also representative only at the state and national levels. In addition to these, India conducts a quinquennial economic census and agricultural census, both of which are rich sources of socio-economic data.

Together the censuses and sample surveys mentioned above provide the most reliable national scale data on development indicators for India. But most of these datasets exist in their own silos due to their differing protocols, mandates and data formats used, making it difficult to conduct integrated analyses which draws on the strength of each type of data. In addition, combining these datasets with remote sensed imagery to derive more information about development indicators is an area that holds tremendous potential (Jean et.al., 2016; Perez et.al, 2017; Utazi et.al., 2018). But India is yet to take advantage of the large amounts of free remote sensed data which is currently available from international space agencies—as well as data from its own satellites—for this purpose. For instance, even for land cover mapping, currently there is no easily accessible and analyzable time series data for the country.

Our proposal addresses the challenges discussed above by building on recent advances in geospatial technology, machine learning and statistical data analysis methods. The project is structured as three work packages (WP). In WP1 we will develop a freely accessible time series land cover database for India at 30m resolution for the period from 1985 till date. For every built-up cell, we will further generate information on degree of built-up and degree of vegetation based on sub-pixel analysis of the earth observation data. In WP2, this land cover data will be used in combination with recently released spatial boundaries of villages and towns of India (Meiyappan et.al., 2018) and data from the national census to generate high resolution population maps. In WP3, the land cover time series data and high resolution population maps developed in previous work packages will be combined with data from sample surveys to develop high resolution maps of development indicators like poverty, health and literacy.

In the sections below we provide details about the identified problems and how we propose to address them, project deliverables and timelines related to the project. Each problem and its solution constitute one work package of the proposed project.

PROBLEMS IDENTIFIED AND SOLUTIONS

WP 1. Lack of Land Cover Data at Sufficient Resolution

Problem: Land cover time series data is crucial for monitoring the nature and extent of land surface transformations occurring across the country. This is especially important in light of the significant socio-economic, demographic and rural-to-urban transitions which are occurring in India today. But such a time series data at adequate resolution is lacking for India. Although the MODIS data is a freely available time series of land cover, its 500m resolution limits its applicability, especially in urban settings (Fiedl et.al., 2010).

The national scale 30m land cover dataset available from the Government of India's official geospatial data portal is only for the years 2005 and 2011 (NRSC, 2014). Besides, this is only available through a Web Map Service system and is not freely available for download. Globeland30 provides 30m global land cover data, but only for two time periods—2000 and 2010 (Chen et.al, 2015). Using the 2005 land cover data from ISRO, Roy et.al. (2015) developed 30m land cover maps for India for 1985 and 1995. The data for all these three years has been made freely available, but at 100m resolution.

Urban areas present a challenge even in the case of the 30m datasets available from ISRO and Globeland30. As seen in Fig. 1a and Fig. 1b, these datasets are unable to capture the heterogeneity and fine grain mix of land cover types in urban areas. In comparison to the heterogeneity visible in the Google Earth image in Fig. 1d, they tend to show urban areas as homogeneously built-up patches. The 30m land cover dataset we have prepared for the 100 most populous cities of India, as part of the India Urban Atlas (IUA) project (Malladi et.al., 2017), captures the heterogeneity in urban areas better (Fig. 1c and 1d),. The land cover classification approach described below uses data from ISRO and from the IUA project as reference datasets.

Solution: We propose to build a 30m resolution land cover time series dataset for India for every alternate year starting from 1985, using freely available earth observation data from the Landsat, Sentinel and Resourcesat programs. We will set up an Open Data Cube to compile and analyze the raw time series data from these programs. The land cover classification will be done in three steps.

<u>Step1</u>: In the first step, we will train and validate a Convolutional Neural Network (CNN) to identify built-up cells using data from the India Urban Atlas (IUA). The IUA contains 30m resolution land cover maps for the 100 most populous cities in India for 2001, 2011 and 2017. For each city, a 60km X 60km classified scene is available, centered on the centroid of the administrative boundary. Each scene is classified into four land cover categories: Built-up, Water, Vegetation, and Others. In this first stage, we will use only the Built-up category from the 2011 data from IUA to build a CNN which will take earth observation data as input and produce a binary map of Built-up and Unbuilt cells as output.

*<u>Step 2</u>**:*** In the second step, we will train a CNN for all areas which the CNN from the first step has classified as Unbuilt. The 2011 ISRO land cover dataset which has seven first level classes will be used to train this CNN. The seven classes available are Built-up, Agriculture, Forest, Grassland, Barren, Wetland/Water, and Snow/Glaciers. To start with, since the first stage CNN trained on the IUA data will be superior to the 2011 ISRO data in terms of ability to identify built-up cells, all cells which are classified as Built-up in Step 1 will be removed from the ISRO dataset. Along with this all cells which are classified as Unbuilt in Step 1, but are classified as Built-up by the ISRO dataset will also be removed from the ISRO dataset. The remaining cells from the ISRO dataset will be deemed to consist only of Unbuilt cells and will be used for training and cross validation of the second stage CNN. Therefore, the second stage CNN will take earth observation data as input and provide a six class map as output—which will have all the above mentioned classes except Built-up. This output from the second stage will be combined with the Built-up layer produced by the first stage CNN to generate a composite seven class map.

The CNNs trained in the above two stages will then be used to conduct land cover classification for the time series data. Validation of temporal extension of the first stage CNN which generates the Built-up class will be conducted for 2001 using land cover data from IUA. For all other six classes validation of temporal extension will be conducted for 2005 using the reference dataset for this year from ISRO.

*<u>Step 3</u>**:*** Once the 30m land cover time series is completed, we will generate degree of built-up and degree of vegetation data for the built-up cells in the 30m land cover maps. In the literature, information unmixing models have been widely used to estimate degree of built-up and vegetation of 30m cells (see Pesaresi et.al. 2016). In comparison, our method for this relies on the temporal overlap between the lower resolution imagery from Landsat, Sentinel and Resourcesat programs with the 5.8m resolution Linear Imaging Self Scanning sensor IV (LISS IV) data from the Resourcesat program by ISRO (ISRO, 2003). Below we describe how the temporal overlap between the Landsat 5 TM sensor and the LISS IV sensor will be used for generating degree of built-up and degree of vegetation data. A similar procedure will be followed for other medium resolution sensors also.

Landsat 5 TM sensor was operational from 1984 to 2011 while the LISS IV data from Resourcesat 1, 2, and 2A is available from 2003 onwards (USGS, 2012; ISRO, 2003). Since the LISS IV data is provided downsampled to 5m, we will first generate 5m land cover maps for 30 selected urban scenes using LISS IV data for 2011. Ground truthing will be done with freely available Google Earth imagery. Thus for each 30m cell of the corresponding Landsat 5 TM data, we will have 36 cells (6 x 6 grid) of 5m resolution land cover information. This land cover information will be used to calculate degree of built-up and degree of vegetation for each 30m cell. This data will then be used

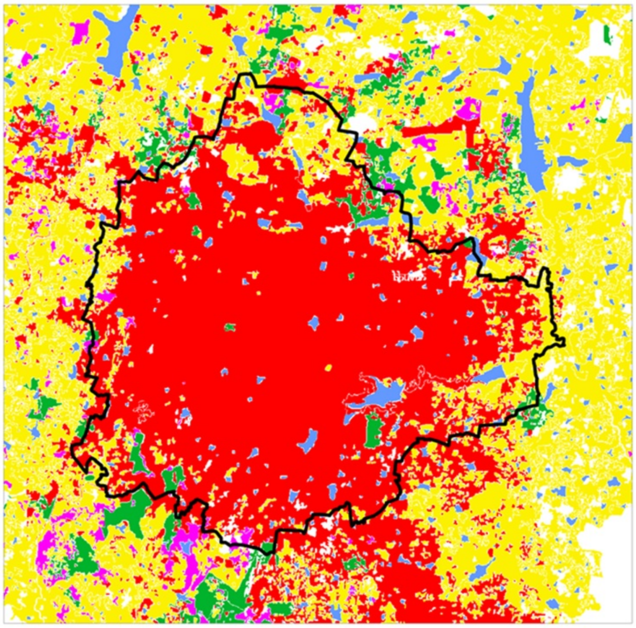


Fig 1a. *2011 land cover map for Bangalore area from 30m resolution ISRO dataset*

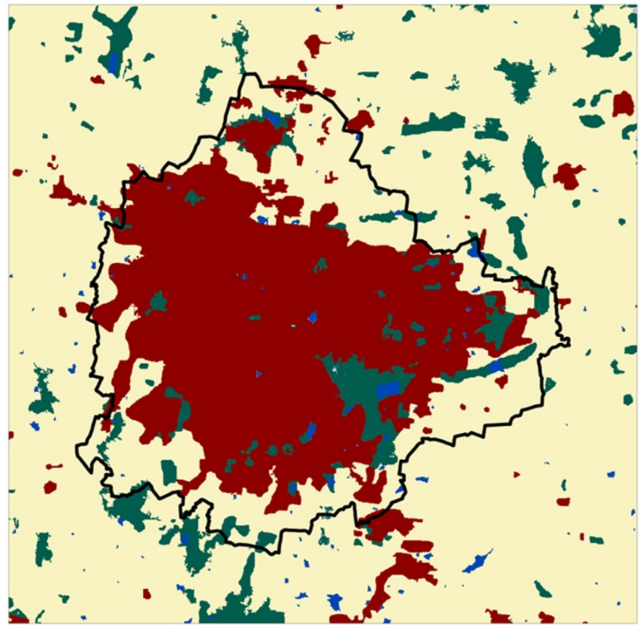


Fig 1b. *2011 land cover map for Bangalore area from Globeland 30*

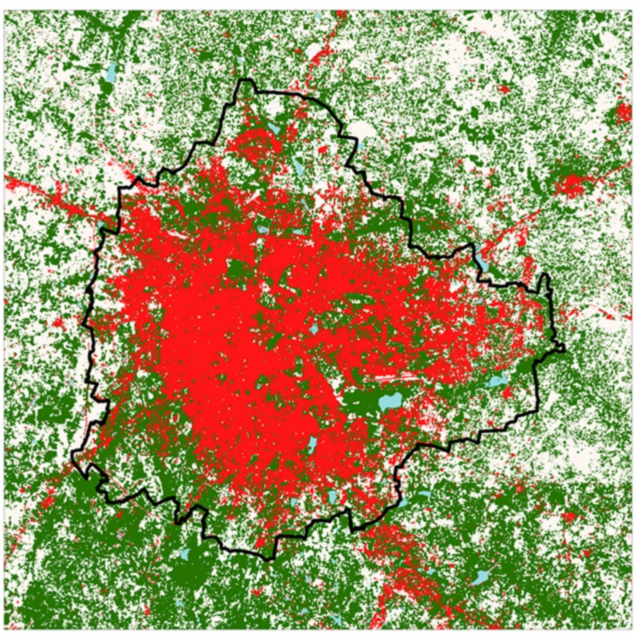


Fig 1c. *2011 land cover map for Bangalore area from India Urban Atlas*

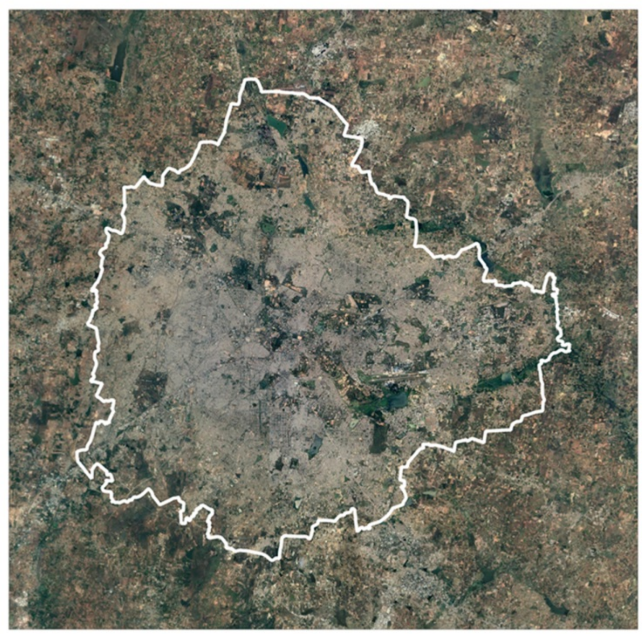


Fig 1d. *Google Earth imagery for Bangalore area*

NOTE: *The boundary shown in all images indicates the extent of the Corporation of Bangalore* to train and validate a CNN which will use the raw 30m Landsat 5 TM data as input and predict the degree of built-up and degree of vegetation for each cell.

After this we will use this CNN to predict degree of built-up and degree of vegetation for the built-up cells in the classified 30m land cover maps from other years. This temporal extension will be validated using classified LISS IV data for the years from 2011 to 2003. Figure 2b shows the initial results we have obtained using Landsat 5 TM data from 1988. Degree of built-up predicted by our CNN model can also be partially validated with the release of the 10m resolution World Settlement Footprint (Esch et.al, 2018). Pouliot et.al. (2018) demonstrates how temporal extension of such a CNN for super resolution mapping of Landsat data is feasible.

We have already implemented our approach using data for the city of Bangalore for 2011. Figure 1c shows classified Landsat 5 TM data from 2011 March while Figure 2a shows predicted degree of built-up for each of the 30m Landsat cells. The 4-fold cross validation approach we implemented provided results with an R-square value of 0.88 with 78% of the predicted degree of built-up values being within 15% percent of actual value.

In summary, at the end of WP1, we will have an Open Data Cube based 30m resolution time series of classified land cover maps for India for every alternate year from 1985 onwards. In addition we will also have degree of built-up and degree of vegetation information for all the 30m built-up cells in the land cover maps. This will enable us to classify built-up cells into various types based on their built-up and vegetation density and then track what types of built-up cells have grown and where. By generating a time series of such maps, we can begin to analyze how areas with varying built-up densities within a city have grown over time thereby enabling a neighborhood level analysis of the nature and extent of urban expansion. In comparison much of the existing efforts at studying urban expansion have focused on city level densities and aggregate spatial characteristics of urban expansion (see Angel et.al., 2016).

WP 2. Lack of Population Distribution Data at Sufficient Resolution

Problem: As discussed in the introduction, population data from the Census of India is made available only in tabular form at the village and Enumeration Block levels for rural and urban areas respectively. Since georeferenced boundaries for these spatial units are unavailable, population can currently be mapped only at the much coarser scale of district or sub-district level in rural areas and ward level in urban areas. In both cases, aggregation of population to such coarser spatial units renders many densely populated areas practically invisible in maps (IEAG, 2014). In the case of urban areas, since even ward boundaries are very difficult to obtain, it is nearly impossible to systematically map urban population across India to understand intra-urban heterogeneity. The lack of population data at sufficient resolution makes it difficult to evaluate access to basic infrastructure related to healthcare and education and exposure of populations to various disaster risks.

Currently, the finest resolution population maps available for India today are from global population mapping projects like Landscan and Worldpop (Dobson et.al., 2000; Stevens et.al., 2015). Landscan uses data on roads, slope of land, night lights and land cover to generate global population maps at approximately 1km resolution (Dobson et.al., 2000). Worldpop uses a wide range of datasets including land cover, slope, elevation, water bodies, mean temperature, mean precipitation, night lights to generate population maps at approximately 100m resolution. Below we describe how we will develop a population map at 30m resolution. While Landscan and Worldpop generate 'ambient' population maps where some population is distributed to entities like roads and highways, our population maps will rely on the definitions used by the Census of India and focus on mapping night time residential population distribution (Census of India, 2011c).

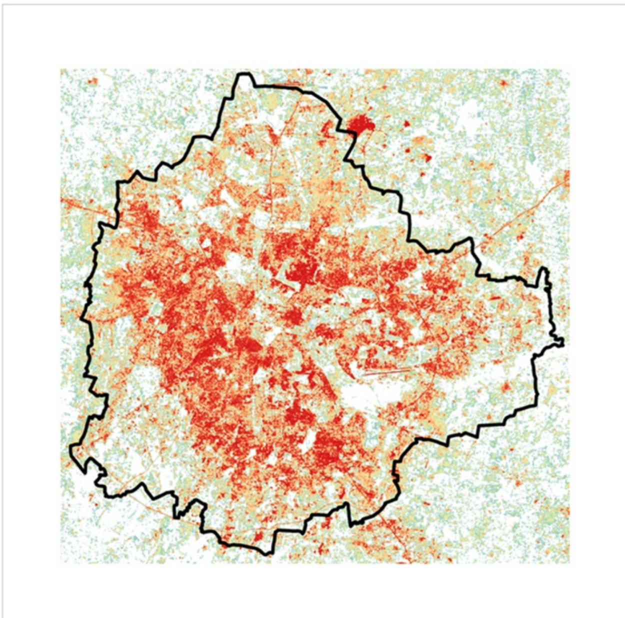
Solution: In 2018, the Socio-economic Data and Applications Center of NASA (SEDAC) in collaboration with Center for International Earth Science Network (CIESIN) at Columbia University, released spatial boundaries and socio-economic data for all of the more than 600,000 villages and towns of India (Meiyappan et.al., 2018). The datasets correspond to the 1991 and 2001 censuses. Building on this release we are in the process of developing 2011 village boundaries also. We will use this dataset in combination with the land cover maps generated in WP 1 and street network data from OSM to generate population density maps corresponding to the census years 1991, 2001 and 2011. Below we describe how this will be implemented in urban and rural areas.

Urban areas: For urban areas we will implement a simple population distribution model which proposes that the population within any 30m cell is proportional to the degree of built-up value for the cell and the street density of the cell (based on OSM streets) and inversely proportional to the degree of vegetation of the cell. The degree of built-up and degree of vegetation within a cell will be available from the land cover mapping method described as part of WP1. Street density will be calculated using OSM data on streets.

The method proposed here is based on the principle of dasymetric mapping (Eicher & Brewer, 2001; Mennis, 2003; Wright, 1936) and redistributes known population within the urban boundary to 30m cells. The population to be assigned to each cell is determined using ancillary data--which in this case includes degree of built-up, degree of vegetation and street density. Prior work demonstrates that type of residential development and street density are good predictors of population density (Mennis & Hultgren, 2006; Reibel & Agrawal, 2007; Reibel & Bufalino, 2005; Xie, 1995). Balakrishnan (2016) shows that population at the scale of 30m cells also depends on wealth since in wealthier neighborhoods the same amount of built-up space may be occupied by fewer people. In our method, degree of built-up is used as an indicator of type of residential development while degree of vegetation is used as a proxy for wealth.

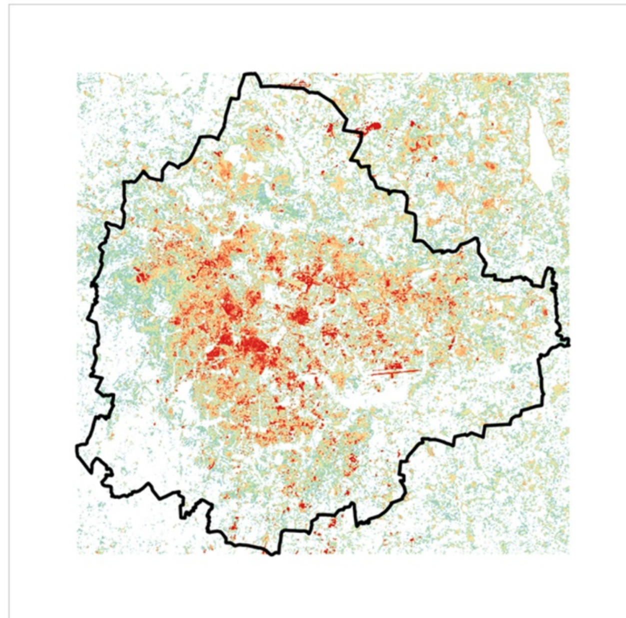
As discussed earlier, the finest resolution population data currently available for India is from the Worldpop program (Worldpop, 2017). Worldpop implements a Random Forests based approach to disaggregate sub-district level population to approximately 100m resolution cells. In the case of India it uses 17 ancillary datasets for this purpose (Stevens, 2015; Worldpop, 2017). As Figures 2c and 2d shows, initial results for Bangalore indicate that although our model is substantially simpler it captures much of the heterogeneity visible in the Google Earth Imagery in Figure 1d. We use known ward level population as a validation check. Root Mean Square Error and Mean Absolute Error for Worldpop are 24463 and 19159 while for our method these are 18571 and13388 respectively. We will use ward level data from 50 other cities of various population sizes across the country for further calibration and accuracy assessment before applying it for all urban areas.

Rural areas: In rural areas settlements are considerably smaller in size in comparison to urban areas and we do not have sufficient heterogeneity in street density or degree of built-up and degree of vegetation to use them for population redistribution. Therefore, we will apply a binary dasymetric approach and uniformly distribute population of each village to the 30m built-up cells within the spatial boundary of each village. In addition, since villages are relative homogeneous in comparison to cities, we can potentially map other census variables also onto the built-up cells. Therefore at the village level we can have population and other census variables dasymetrically mapped to built-up cells, while in urban areas only population will be disaggregated to 30m resolution cells.



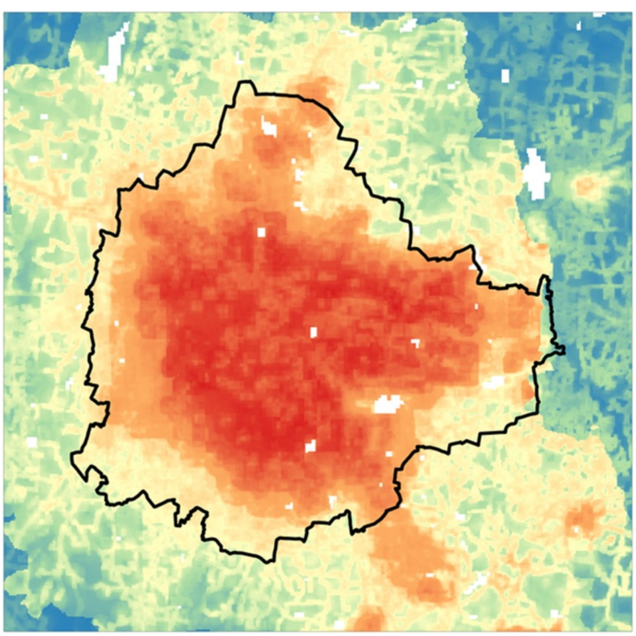
0 – 6.5 | 6.5 – 29.2 | 29.2 – 59.9 | 59.9 – 85.7 | 85.7 - 100

Fig 2a. *Degree of built-up map for 2011. Legend indicates degree of built-up as percentage of total area of each 30m cell*



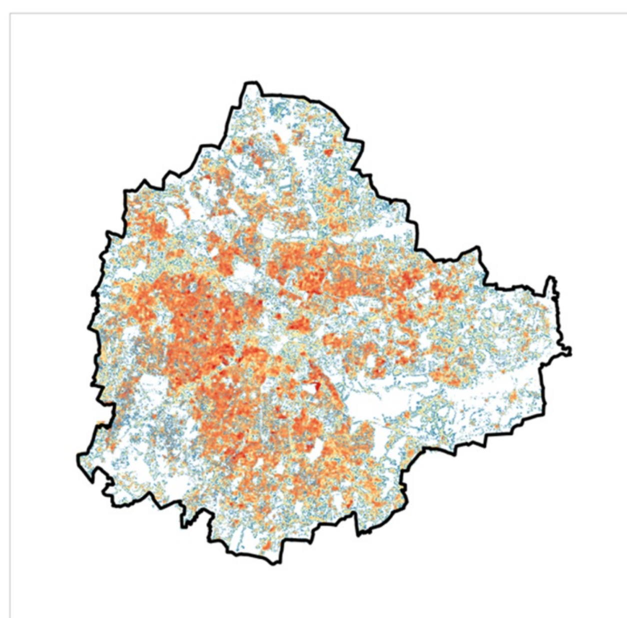
0 – 6.5 | 6.5 – 29.2 | 29.2 – 59.9 | 59.9 – 85.7 | 85.7 - 100

Fig 2b. *Degree of built-up map for 1988. Legend indicates degree of built-up as percentage of total area of each 30m cell*



0 – 1.2 | 1.2 – 7.1 | 7.1 – 14.6 | 14.6 – 34.2 | 34.2 – 217

Fig 2c. *Population density map from Worldpop. Legend indicates persons per 100m cell*



0 – 1 | 1 – 11 | 11 – 26 | 26 – 59

Fig 2d. *Population density map prepared using our proposed method. Legend indicates persons per 30m cell*

NOTE: *The boundary shown in all images indicates the spatial extent of the Corporation of Bangalore*

Within village boundaries considerable number of built-up cells can belong to roads which connect villages to other settlements. There can also be built-up cells which indicate other unpopulated areas like industrial zones. These two types of built-up cells will be removed prior to distributing population to 30m cells within each village boundary. The built-up cells which are on roads will be removed using OSM data on streets and highways, and using computer vision methods which can detect linear strips of built-up cells between dense settlements which may represent streets and highways. To avoid removing buildings which may be located along them, we will define a buffer width from the center line of streets and such linear built-up strips. The buffer width will be decided based on an analysis of typical widths of different categories of streets and highways.

We will remove built-up cells associated with industrial areas based on the fact that industrial areas tend to have buildings with larger footprints, have a well structured layout, and have buildings which mostly have roofs made of lightweight materials like metal sheet roofing. Therefore using a dataset of industrial areas developed using information from state industrial development boards, we will train a CNN to identify patches of industrial areas based on the spatial and spectral characteristics of such areas.

At the end of WP2, we will have 30m resolution population distribution map for India for 1991, 2001 and 2011. In addition, for rural areas, we will also map other census variables to the built-up cells within each village. Our proposed method for population distribution will be a considerable improvement over Worldpop population maps since their method relied on distributing population aggregated to approximately 600 districts to 100m cells and then using 5967 sub-district polygons to conduct accuracy assessment. We propose to directly distribute population from more than 600,000 villages and towns to 30m cells. As discussed earlier, in the case of cities, we will use ward level population data from 50 cities to assess the accuracy of our method. For rural areas accuracy assessment at finer scales is unfortunately not possible, but may also not be necessary since villages are much more homogeneous and settlement and population sizes are considerably smaller, in comparison to cities.

The availability of spatial boundaries for all villages will also help us solve the problem of lack of development indicators for parliamentary and assembly constituencies. This problem arises because the Census data is made available for administrative units of the country (district, sub-district and villages), while parliamentary and assembly constituencies are delineated with the aim of ensuring evenness in population distribution. Over the past decade researchers working on India have been grappling this problem (Alam, 2010; Swaminathan et.al., 2019). Since in rural areas electoral constituencies are aggregations of villages, Census data mapped to village spatial boundaries can be aggregated to electoral constituencies by performing a simple overlay analysis.

WP 3. Lack of Development Indicators at Sufficient Resolution

Problem: Although India conducts a decennial census, the temporal resolution is not sufficient to track development indicators, as the country is undergoing rapid socio-economic and urban transitions. Moreover, during the time period from 2015 to 2030 when Sustainable Development Goals and the Sendai Framework for Disaster Risk Reduction are supposed to guide global and national development agenda, India will have only one Census—in the year 2021. While India does have more frequent sample surveys, their spatial resolution is limited since most of the surveys are representative only for larger geographical areas and cannot be used at the settlement scale. In addition to these, India also conducts quinquennial Economic Censuses and Agricultural Censuses but these datasets have so far not been used in combination with the sample surveys and census datasets in a systematic way to estimate development indicators.

Solution: We will develop a system which will integrate data from all the above sources with earth observation data to generate estimates of development indicators at sufficient spatial and temporal resolution for India. Below we first provide an overview of the sample surveys available in India before describing our approach for developing such a system. Data from the Economic and Agricultural Censuses can be integrated with this system since they are collected using the same spatial units as the population census.

India conducts two types of national scale sample surveys. The first is the National Sample Surveys (NSS) which have been conducted since 1950 and focus on socio-economic aspects. The second is India's version of Demographic and Health Surveys which is called the National Family Health Survey (NFHS). These sample surveys collect much more detailed information on socio-economic indicators, education and healthcare compared to the Census. Since they only cover a representative sample, while the Census of India covers the entire population, they are able to canvass a more detailed questionnaire and are able to carry out surveys at greater frequency than the Census. To illustrate, the NSS conducts quinquennial Household Consumption Expenditure Surveys (CES) as well as Employment and Unemployment Surveys (EUS). These surveys are used for poverty measurement as well as for assessing the state of employment in the country, and form the basis for policy design as well as the targeting of beneficiaries. In addition to these, the NSS also conducts occasional surveys on topics such as migration, condition of urban slums and situation assessment of agricultural households.

While much of the data collected by the NSS and NFHS is directly relevant to monitoring progress on the SDGs, the sampling is not representative at the settlement scale. For instance, the data from the NSS is representative only at the scale of NSS regions—which are clusters of eight to ten districts—or at state and national levels. The NFHS which has a larger sample size, is more infrequent, and is representative only at the district level. But both NSS and NFHS data is available at the unit level while the Census is only available aggregated to the settlement level in rural areas and at the ward level in urban areas.

We will use a Bayesian Hierarchical Modeling (BHM) approach to draw on the strengths of the censuses, sample surveys and earth observation data to generate development indicators at the near-settlement spatial resolution and at a temporal resolution which matches the frequency of sample surveys. We will draw on methods from the field of Small Area Estimation (Rao and Molina, 2015) to achieve this. The basic premise of such methods is that the socioeconomic indicator being estimated for a small area would be correlated with a several other variables like those available from the census or from earth observation data. Besides, the spatial and temporal variation of such socioeconomic indicators is usually smooth, without large jittery spikes or discontinuities. This makes it possible to build statistical models to integrate multiple types of data to bring in more information into the estimation of the indicators for small areas.

BHMs have been widely used in spatial analysis due to their flexibility in capturing the spatial correlation structure and the uncertainty estimates they give. In recent years, software tools like Stan and INLA have made it possible to fit complex models within a reasonable amount of time and computing resources (Carpenter et. al., 2017; Rue et. al. 2009). Anjoy et. al. (2018) demonstrate that a BHM based small area estimation method for poverty in the Indian state of Odisha which combines data from the census and from NSS outperforms estimates generated only using the sample survey data. Graetz et al. (2018) use a BHM framework for estimating educational attainment of women at a five by five kilometer grid across a large part of Africa. They compiled an extensive database consisting of 173 survey and census data sets, and systematically combined them using an ensemble model specified as a BHM. Molina et. al. (2014) developed a BHM for estimating poverty in Spanish provinces using the data from Spanish Survey on Income and Living Conditions.

With the availability of free earth observation data and recent advances in the management and analysis of large scale datasets, it is now possible to use information derived from earth observation datasets as additional auxiliary datasets for small area estimation and spatial disaggregation applications at a national scale. Over the past few years several researchers have attempted to estimate socioeconomic conditions using earth observation data at a national and continental scale. While Jean et. al. (2016) used high resolution, three band Google Earth imagery to predict poverty across Africa, Perez et.al. (2017) extend this approach to show that medium resolution multispectral imagery from the Landsat program can also be used to estimate socioeconomic indicators. Watmough et al. (2019) demonstrates how high resolution satellite imagery can help in estimating poverty, using the case of Kenya.

We will build a system that does small area estimation for various indicators from the NSS and NFHS data for the cities and villages of India. The proposed system will use a BHM framework that combines information from earth observation data, censuses and sample surveys to produce the required small area estimates. The datasets developed in WP1 and WP2 will feed into this BHM framework. The BHM will also account for the spatial correlation structure that is often present in such data.

To validate the model, we will use data from 2011—a year for which both Census and NSS datasets are available. For instance, if female literacy is the indicator of interest, we will set aside data for this variable from the 2011 Census, and then use the remaining census variables and NSS data to develop small area estimates of female literacy using a BHM framework. Then we will use the Census data on female literacy to validate the model output. This way of direct validation will be superior to other indirect ways of validation adopted by many papers cited above.

APPLICATIONS AND EXTENSIONS OF THE PROPOSED WORK

IIHS is actively engaged in working with several city and state governments in advisory and consulting capacities. The outputs from this project will form a set of base datasets which will be used by cities and states to co-develop and co-implement more detailed data portals for tracking and communicating development indicators and exposure of populations to various risks. Below we describe some of our ongoing work which will use outputs form this proposed project.

IIHS is currently working with Guwahati Municipal Corporation in the state of Assam, on developing a City level Geospatial portal for Guwahati that will act as a decision support system for City and Regional Planning and Resilience building. As part of this project we are working with city administration on bringing together datasets related to urban infrastructure, hazard maps and vulnerability for the city. IIHS has also been developing City Portals as part of the Ministry of Home Affairs (Government of India) and UNDP project on 'Enhancing Community Resilience to Disaster Risk Reduction and Climate Change Adaptation' with a focus on mainstreaming disaster risk reduction and climate change adaptation. IIHS is supporting UNDP to develop portals for seven cities: Cuttack, Dharamshala, Navi Mumbai, Shillong, Shimla, Vishakapatnam, and Vijayawada. Like in the case of Guwahati, this involves bringing together a wide range of datasets to assess exposure and vulnerability to various disasters. The outputs from WP1 and 2 in particular will feed into this work.

At the state level, IIHS has been working with more than 10 states to conduct training programs for administrators at various levels. These include programs related to sustainable development, climate adaptation, public health and urban planning. The outputs from this project will be used to develop new training modules for state government administrators such that they can use these datasets and analytical outputs to inform policy decisions.

The proposed project can also be extended in several ways. Using the outputs of the proposed project we will be able to generate development trajectories at a settlement scale starting from the 1991 Census. This can then be used to produce forecasts of developmental trends based on various policy scenarios. Such forecasts will be valuable in policy formulation and in tailoring policies to specific state and regional contexts.

The proposed BHM framework described in WP3 has the potential to integrate many other types of datasets to improve estimation of development indicators and to monitor a wider range of indicators. For instance, as mentioned earlier, data from Economic and Agricultural Censuses can be integrated into this framework. Similarly other high frequency earth observation data like soil moisture, cultivation patterns and reservoir storage level data can also be integrated into this in the future to improve estimates of development indicators and economic activity.

CONTRIBUTION TO NATIONAL EFFORTS TO ADDRESS GLOBAL POLICY FRAMEWORKS

The proposed project will contribute to monitoring indicators related to Sustainable Development Goals and to addressing the targets, guidelines and priorities for action under the Sendai Framework for Disaster Risk Reduction.

Sustainable Development Goals

The proposed project will help monitor progress towards SDGs 1, 2, 3, 6, 9, 11 and 15. The outputs generated in WP 1, 2 and 3 related to land cover, population distribution and other development indicators can fully or partially support in monitoring the indicators related to these goals. More information on the relevant goals and indicators is given below.

Goal 1. End poverty in all its forms everywhere

1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)
1.2.1 Proportion of population living below the national poverty line, by sex and age
1.2.2 Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
1.4.1 Proportion of population living in households with access to basic services

Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture

2.1.1 Prevalence of undernourishment
2.2.1 Prevalence of stunting
2.2.2 Prevalence of malnutrition

Goal 3. Ensure healthy lives and promote well-being for all at all ages
3.1.2 Proportion of births attended by skilled health personnel
3.7.1 Proportion of women of reproductive age (aged 15–49 years) who have their need for family planning satisfied with modern methods
3.7.2 Adolescent birth rate (aged 10–14 years; aged 15–19 years) per 1,000 women in that age group
3.b.1 Proportion of the target population covered by all vaccines included in their national programme

Goal 6. Ensure availability and sustainable management of water and sanitation for all
6.1.1 Proportion of population using safely managed drinking water services

Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
9.1.1 Proportion of the rural population who live within 2 km of an all-season road

Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable
11.1.1 Proportion of urban population living in slums, informal settlements or inadequate housing
11.2.1 Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities
11.3.1 Ratio of land consumption rate to population growth rate
11.7.1 Average share of the built-up area of cities that is open space for public use for all, by sex, age and persons with disabilities

Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
15.1.1 Forest area as a proportion of total land area

Sendai Framework for Disaster Risk Reduction
The outputs from this project, especially in WP 2 and 3, will be very useful in understanding exposure to various hazards and adaptive capacity of populations. As part of our ongoing work with state and city governments we will use these outputs to co-develop data portals for analyzing and communicating disaster risk related information. This work will then feed into policy formulation and implementation related to disaster risk reduction at the regional and local scales.

Through this the project will enable India to address the following Target, Guiding Principle and Priorities for Action under the Sendai Framework for Disaster Risk Reduction which relate to understanding and disseminating disaggregated information about disaster risk. This includes Target (g) under Article 18, Guiding Principle (g) under Article 19 and Articles 23 and 24 under Priorities for Action #1.

PROJECT TIMELINE

		2019							2020													2021													2022				
	Activity	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	April	May
WP 1	Set up Open Data Cube with all relevant earth observation data	X	X	X	X	X	X	X	X	X	X	X																									
WP 1	Train and validate a CNN to identify built-up cells	X	X	X																																	
WP 1	Train and validate a CNN to identify other types of land cover				X	X	X	X																													
WP 1	Train and validate a CNN to estimate degree of built-up and degree of vegetation for the built-up cells				X	X	X	X	X	X																											
WP 1	Validate temporal extension of CNNs							X	X	X	X	X																									
WP 1	Use the CNN to generate time series of land cover maps												X	X																							
WP 2	Develop village and town boundaries for 2011	X	X	X	X	X	X	X	X	X	X	X																									
WP 2	Assemble OpenStreetMap data for all urban areas										X	X																									
WP 2	Generate 30m population maps for urban areas												X	X	X																						
WP 2	Train and validate a CNN to identify industrial areas													X	X	X																					
WP 2	Remove industrial areas and streets from land cover dataset																X	X	X																		
WP 2	Generate 30m population maps for rural areas																			X	X	X															
WP 3	Procure unit level NSS and NFHS data								X	X	X	X	X	X																							
WP 3	Develop BHM based approach for estimating development indicators														X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
WP 3	Validate BHM based approach																													X	X	X	X	X			
WP 3	Generate maps of development indicators																																		X	X	X

CLOUD CREDIT BUDGET

Given the scale of this project and its computational requirements, we will most likely exceed the $60,000 limit for single nation projects. But this is a project which we are anyway undertaking, hence whatever we use over and above $60,000 will be paid for using other funds.

TEAM

Krishnachandran Balakrishnan (Project Lead), Consultant, IIHS
Focus areas: Spatial Analysis, Environmental Planning, Water Resources Management

Krishna works on developing new methods in the field of spatial analysis by bringing together approaches from Computer Vision, Machine Learning and Geographic Information Science. His recent work focuses on methods for sub-pixel analysis of Landsat imagery using Convolutional Neural Networks and building height extraction from satellite stereo imagery. His doctoral work focused on fine-grained analysis of heterogeneity within cities and urban spatial inequality. As part of his doctoral work, he developed a new method for high-resolution population density estimation in Indian cities. The method combines data on land cover, land use, building height, street network and wealth estimates to develop 30m resolution population density maps.

As part of the Climate Development Knowledge Network project on Future Proofing Cities, he led the spatial analysis work for Bangalore city. Currently, as part of a Global Challenges Research Fund (GCRF) project at IIHS, Krishna leads the thematic area of spatial modeling and water. In this project he is currently working on developing a Bayesian Hierarchical Modeling framework for downscaling census data from the ward level to 30m resolution cells. Krishna is also part of a team consisting of researchers from the Indian Institute of Science, Indian Institute of Management and the Stockholm Environment Institute which is working on developing a detailed understanding of the groundwater dynamics within Bangalore city. Recently he also led the development of a framework for assessing suitability of on-site sanitation systems across for the Government of Tamil Nadu.

As part of the Urban Practitioners Program at IIHS, Krishna has been actively involved in conducting training programs for urban professionals from various states across the country. His training sessions have focused on the use of fine-grained data in urban planning and on urban water resources management.

Krishna received his Ph.D. in Environmental Planning from UC Berkeley. He holds previous degrees in Architecture and Landscape Architecture from School of Planning and Architecture, New Delhi and UC Berkeley.

Aromar Revi, Director, IIHS
Focus areas: Sustainable Development, Climate Change, Disaster Risk Reduction

Aromar Revi is the founding Director of the Indian Institute for Human Settlements (IIHS). Over a decade, he has built IIHS into one of the world's leading education, research, training, advisory and implementation-support institutions located in the global south, focusing on the multi-sectoral and multi-dimensional challenges and opportunities of sustainable urbanisation.

Aromar is a global expert on implementing Sustainable Development; Co-Chair of the UN Sustainable Development Solutions Network (SDSN), where he led a successful global campaign for an urban Sustainable Development Goal (SDG 11) as part of the UN's 2030 development agenda. This brought major global urban institutions (UN Habitat, UCLG, C-40, ICLEI, Metropolis, Cities Alliance, SDI and WIEGO) and over 300 cities and organisations together.

He is a member of the UCLG-Ubuntu and policy Advisor to the UCLG Presidency on localization of the SDGs. UCLG is the global voice and advocate of local and regional governments, representing 0.24 million towns, cities, metropolises and regions across the world. Aromar is also a member of the Managing Board of Cities Alliance the global partnership for sustainable cities and urban poverty reduction.

Aromar is one of the world's leading experts on global environmental change, especially climate change. He is a Coordinating Lead Author (CLA) of the seminal 2018 IPCC Special Report on Global Warming of 1.5 °C (SR15) that was released in 2018 to global scientific and media acclaim. He has worked extensively to bring the global urban and climate agendas together, including as a coordinating lead author of the SR1.5 Summary for Urban Policymakers, released at CoP24 in Poland in 2018.

Aromar is a CLA of the synthesis chapter on Climate Resilient Development Pathways of the IPCC Working Group Assessment Report 6 on Adaptation, to be released in 2021. He was earlier a CLA of the IPCC Assessment Report 5 on Urban Areas that established the role of cities and regions in addressing climate risks in 2014. He was a co-PI of a large five-year international Climate Adaptation research programme that spanned India and eight countries in Africa, that explored mechanisms to take climate adaptation to scale in semi-arid regions.

Aromar' policy, practice and research work lie at the interface of sustainability and climate science; and the emerging discipline of 'urban science', that he is helping define internationally. He is a member of the UCL-Nature Sustainability Expert Panel on science and the future of Cities that issued its seminal 2018 report on the global state of the urban science-policy interface. He is Co-PI on two significant international urban research programs that will help define the future of urban science (PEAK) and responses to urban inequality (KNOW) bringing together leading universities and researchers from four continents.

Aromar was a long-standing Board member of the Balaton Group, that pioneered the development of sustainability. He was a Fellow of the India China Institute at the New School University, where he worked on long-range development pathways for India and China.

Aromar has been a senior advisor to multiple ministries of the Government of India, since the late 1980s; and consulted with a wide range of international development institutions, national and transnational firms on economic, environmental and social change at global, regional and urban scales. This includes extensive in-country and regional experience across key UN, multilateral and bilateral agencies including: UNDP, UNICEF, UNEP, UN Habitat, UNISDR, UNESCO, UNU; the World Bank and ADB, DFID, CIDA, GiZ, IDRC, NORAD, SDC, SIDA, USAID, and AusAID.

He is one of South Asia's most experienced risk and disaster management professionals having led teams to plan & execute rehabilitation programs for ten major earthquake, cyclone and flood events affecting over 5 million people and was a member of the Advisory Board of UNISDR' Scientific & Technical Advisory Group (STAG) and its bi-annual Global Assessment of Risk, from 2008.

Aromar is an alumnus of IIT-Delhi and the Law and Management schools of the University of Delhi.

Amir Bashir Bazaz, Lead – Practice, IIHS
Focus areas: Economics, Climate change, Energy

Amir works on issues at the intersection of economics, climate change mitigation & adaptation and sustainable development. He has substantial experience of working with various top-down & bottom-up economy-energy-environment modelling frameworks/architectures. His current research interests are low carbon societies/infrastructure, climate change adaptation and mitigation (across scales) with specific focus on urban-climate change linkages. He has previously been the National Expert Consultant to the Ministry of Environment, Forests and Climate Change, Government of India for the Second National Communication to the UNFCCC and taught courses in Development & Environmental Economics during his academic engagements at Symbiosis International University, Pune.

At IIHS, Amir is the Regional Research Lead for a multi-partner, multi-year climate adaptation research project – Adaptation at Scale in Semi-Arid Regions (ASSAR). This project is a part of an IDRC/DfID funded global climate adaptation research program – Collaborative Adaptation Research Initiative in Africa and Asia (CARIAA), operational across the regions of West, South and East Africa as well as South Asia. In addition, Amir is a part of many practice-based engagements at IIHS, notably; on 'Energy Innovation' (on a project led by Cambridge University), 'Sustainability of Ecosystem Services' (in collaboration with the Nature Conservancy India and Keystone Foundation) and ‘Migration-Climate Resilience dynamics for Indian cities’ (supported by the Swiss Agency of Development & Cooperation). He has been a regular team member for many ‘Disaster and Climate Resilience’ projects at IIHS.

Amir Bashir Bazaz holds a PhD in Management from Indian Institute of Management Ahmedabad, with a specialization in Public Systems.

Sat Kumar Tomer, CEO – Satyukt Analytics
Focus areas: Hydrology, Agricultural Remote Sensing, Microwave Remote Sensing

Sat Kumar is a hydrologist with more than a decade of experience in the application of multi-satellite remote sensing in the areas of water resources and agriculture. In particular he has considerable experience in microwave remote sensing and soil moisture monitoring and modeling.

Sat Kumar has worked extensively with various state governments in India for the preparation of State Specific Action Plan for Water Resources. He has also worked on projects related to adaptation options for irrigated agriculture to climate change.

He received his Ph.D. from Indian Institute of Science, Bangalore.

Teja Malladi, Head – Geospatial Lab, IIHS
Focus areas: Geospatial analysis, Disaster Risk Reduction

Teja is part of the Practice team and heads the Geospatial Lab at IIHS. He holds a Master’s in Geo-Information Science and Earth Observation with a specialization in Natural Hazards and Disaster Risk Management from the University of Twente, Netherlands. He works in the areas of natural hazard, risk and vulnerability assessment using remote sensing and geographic information systems, and also has experience in post-disaster reconstruction and rehabilitation.

Some of his recent work includes an 18-month long research project titled Disaster-related Resettlement and Relocation in Urban Areas, and geospatial analysis for the state of Odisha for identifying locations for high-intensity economic agglomerations. His recent publications include Urban India Evidence – 2015 and 2016 and Urban Atlas 2017. He has developed cases and films on urban risk. He teaches as part of the Urban Practitioners’ Programme and is the course director for ‘Working with Maps’, a course on GIS.

Shriya Anand, Senior Consultant – Academics & Research, IIHS
Focus areas: Economics, Public Policy

Shriya Anand is a faculty member at the Indian Institute for Human Settlements, teaching topics related to urban economic development and quantitative research methods. She anchors the Urban Informatics Lab (UIL) at IIHS, which analyses, communicates and disseminates data and information related to India’s urbanization. She has worked extensively with the Census of India, the National Sample Survey Organisation (NSSO), and the Economic Census datasets on various urban problems.

Her research at IIHS is primarily centered on the Indian urban economy and economic geography, with a particular focus on the role of employment in urban development and poverty reduction. Her current research focuses on mapping Bangalore’s industrial transition and its implications for employment trends and land use change. She has recently been studying large industrial infrastructure projects such as the Delhi-Mumbai Industrial Corridor, their relationship with urbanization, and associated choices about development pathways.

Shriya holds a Master in Public Affairs with a concentration in Economics from Princeton University, and a Master in Mathematics from Cambridge University, UK.

Amruth Kiran, Consultant, IIHS
Focus areas: Computer Science, Remote Sensing, Open Data Cube

Amruth has an undergraduate degree in Computer Science and a Masters degree in Geoinformatics. He works with the Geospatial Lab at IIHS on developing online, scalable and user-friendly applications.

For his Masters dissertation at the Indian Institute for Remote Sensing, he developed a web-based geospatial modelling platform using Open Data Cube as part of a project on Indian Bioresource Information Network. The platform utilised the scalable and robust concept of Data Cubes to model species distribution in the Indian state of Uttarakhand. This work was presented and published at the International Society for Photogrammetry and Remote Sensing annals conference, Dehradun 2018.

Earlier he worked at the Transdisciplinary University, Bengaluru for developing mobile applications in the realm of GIS and primary healthcare which was consequently published in IJREAT 2016.

Krishna Kumar P., Consultant, IIHS
Focus areas: Computer Vision, Machine Learning

Krishna Kumar is part of the Geospatial Lab at IIHS. His current work focuses on applying computer vision and machine learning methods to geospatial analysis. In particular he has been working on sub-pixel analysis of Landsat imagery and land cover classification using Convolutional Neural Network based approaches.

Kirshna Kumar received his Ph.D. in Computer Science and Engineering from National Institute for Technology Surathkal. His doctoral dissertation focused on the segmentation and classification of medical images. He has secured an M.Tech. degree in Computer Vision and Image Processing from Amrita Vishwavidyapeetham, Coimbatore.

Pratyush Tripathy, Research Assistant – Geospatial Lab, IIHS
Focus area: Geoinformatics, Image processing

Pratyush holds a five year integrated masters in Geoinformatics. He works with the geospatial lab on image processing for spatial analysis, microwave remote sensing applications, street network analysis, understanding urban spatial inequalities, and establishing relationship between socio-economic factors and remotely sensed data. His key skills are algorithm development, automation, geostatistical analysis, and handling large scale remote sensing datasets.

His previous work involves multi-temporal urban growth analysis and modelling using cellular automata, markov chain, and logistic regression; understanding the diversity of the physical structures across the built environment, and SAR (synthetic aperture radar) interferometry. As a part of his master`s thesis, he developed method to delineate neighborhoods based on physical characteristics like street density, building height and built-up vegetation mix.

Sooraj M. Raveendran, Consultant, IIHS
Focus areas: Machine Learning, Statistics

Sooraj Raveendran is part of the Urban Informatics Lab at IIHS. At the lab, he employs state-of-the-art computational tools on urban data to develop a comprehensive understanding of the complex urban transformation in India and then to make useful predictions that could assist decision making at various levels.

Earlier in his career, at DISH Network, Sooraj applied statistical modelling and machine learning on business analytics problems like personalized product recommendations and offer optimization for multiple satellite and internet TV platforms. This required dealing with huge volumes of data and occasionally with extremely sparse or poor quality data. He is experienced in using Amazon cloud services like EMR and Redshift. Sooraj holds a Bachelor's degree in computer science and a graduate certificate in data science. Currently, he is pursuing a master's in statistics with a focus on biostatistics and epidemiology.

ABOUT IIHS
The Indian Institute for Human Settlements (IIHS) is a national education institution committed to the equitable, sustainable and efficient transformation of Indian settlements. IIHS is actively engaged in research, consulting, teaching and professional training. IIHS faculty and researchers come from a wide range of disciplinary backgrounds and work collaboratively to address complex developmental challenges. IIHS is also actively engaged with city, state and national governments in consulting, professional training and advisory roles.

The Research program at IIHS has developed an international reputation for inter-disciplinary, cross-scale and cross-institutional urban and regional research. Our Practice program works on complex urban challenges with a wide range of partner institutions ranging from national and state governments, UN agencies, international development financing institutions, foundations, INGOs, NGOs, and trade unions. We are therefore, uniquely positioned to bring research to bear on practical processes that are part of India's ongoing transitions.

Over the years IIHS has been involved in global projects in the area of Sustainable Development. IIHS lead one of the largest climate adaptation research programs in South-Asia. The 5-year (2014–2018) Adaptation at Scale in Semi-Arid Regions (ASSAR) research program was part of the Collaborative Adaptation Research Initiative in Africa and Asia (CARIAA) international climate adaptation research program (funded by IDRC and DfID).

In addition, we are currently part of two UKRI-GCRF funded grants. The KNOW (Knowledge in Action for Urban Equality) project is a collaboration with University College London (UCL) and other academic and community organizations in countries as diverse as Peru, Sierra Leone, Sri Lanka and Uganda. This collaboration aims to address the challenge of different kinds of disparities, and the development of locally-led strategies, by recognizing context-specific problems and identifying context-specific solutions and inform various levels of stakeholders ranging from urban planners to community groups.

IIHS is also part of a global consortium of partners on the PEAK (Prediction and projection, Emergent urbanisms, Adopted knowledge, Knowledge exchange) Urban project. The project brings together leading scholars from Oxford University, University of Cape Town, Peking University, EAFIT University in Colombia and IIHS in an effort to address the challenges inherent to rapid urban transitions, including limited resources, complex governance structures and inadequate capacities particularly in the cities of Africa, Asia and Latin America. Over the period of the project, our work focuses on opportunities for urban transformation and implementing the urban SDGs: health, energy, water, infrastructure, land planning, and informality.

IIHS is committed to capacity building across different career stages. IIHS' Urban Practitioners Program (UPP) is one of the largest training and capacity development program in the world. Last year, it trained nearly 600 members of India's Administrative Service apart from over 3,500 public officials and practitioners, on urban-related themes. Further, the Digital Blended Learning Program at IIHS is a key initiative that helps in achieving our core objectives of access and inclusion to critical knowledge that informs sustainable practice. This program has curated a specialized global Massive Open Online Course (MOOC) on Sustainable Cities, in partnership with the SDG Academy, which has been accessed by 19,000 learners from across the world in the first three iterations.

REFERENCES

Alam, M. S. (2010). On Matching Census Tracts and Electoral Boundaries: The Bottom-up Aggregation Approach. Economic and Political Weekly, xlv(34), 64–72.

Angel, Shlomo, Alejandro M. Blei, Jason Parent, Patrick Lamson-Hall, and Nicolás Galarza Sánchez. 2016. Atlas of Urban Expansion, Vol. 1, Areas and Densities. New York: New York University Urban Expansion Program; United Nations Programme for Human Settlements; Lincoln Institute of Land Policy.

Anjoy, P., Chandra, H., & Basak, P. (2018). Estimation of Disaggregate-Level Poverty Incidence in Odisha Under Area-Level Hierarchical Bayes Small Area Model. Social Indicators Research, (0123456789). https://doi.org/10.1007/s11205-018-2050-9

Balakrishnan, K. (2016). Heterogeneity within indian cities: Methods for empirical analysis University of California, Berkeley.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan : A Probabilistic Programming Language. Journal of Statistical Software, 76(1). https://doi.org/10.18637/jss.v076.i01

Census of India. (2011a). District Census Handbook for Bangalore Urban District, Directorate of Census Operations, Karnataka, Ministry of Home Affairs, Government of India.

Census of India. (2011b). Primary Census Abstracts, Registrar General of India, Ministry of Home Affairs, Government of India.

Census. (2011c). Instruction Manual for Population Enumeration, Registrar General of India, Ministry of Home Affairs, Government of India.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., … Mills, J. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7–27. https://doi.org/10.1016/j.isprsjprs.2014.09.002

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. Cartography and Geographic Information Science, 28(2), 125-138.

Esch, T., Bachofer, F., Heldens, W., Hirner, A., Marconcini, M., Palacios-Lopez, D., … Gorelick, N. (2018). Where we live-A summary of the achievements and planned evolution of the global urban footprint. Remote Sensing, 10(6). https://doi.org/10.3390/rs10060895

Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X. Y., Muchoney, D., Strahler, A. H., ... & Baccini, A. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*(1-2), 287-302.

Graetz, N., Friedman, J., Osgood-Zimmerman, A., Burstein, R., Biehl, M. H., Shields, C., … Hay, S. I. (2018). Mapping local variation in educational attainment across Africa. Nature, 555(7694), 48–53. https://doi.org/10.1038/nature25761

ICF. (2017). India National Family Health Survey NFHS-4 2015-16 . Mumbai, India : IIPS and ICF . Retrieved from http://dhsprogram.com/pubs/pdf/FR339/FR339.pdf

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Machine Learning To Predict Poverty. Science, 353(6301), 790–794. Retrieved from http://science.sciencemag.org.www2.lib.ku.edu/content/sci/353/6301/790.full.pdf

IEAG (2014). A World that Counts. Data Revolution Group: The UN Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. Retrieved from http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

ISRO (2003), Indian Remote Sensing Satellite – P6, Resourcesat -1 https://www.isro.gov.in/Spacecraft/irs-p6-resourcesat-1. Accessed on 10 February 2019.

Malladi, T., Chatterji, N., & Jana, A. (2017). India Urban Atlas: Mapping the Growth and Expansion of India's 100 Largest Cities. Indian Institute for Human Settlements, Bengaluru.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping*. The Professional Geographer, 55(1), 31-42.

Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. Cartography and Geographic Information Science, 33(3), 179-194.

Molina, I., Nandram, B., & Rao, J. N. K. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical bayes approach. Annals of Applied Statistics, 8(2), 852–885. https://doi.org/10.1214/13-AOAS702

NRSC (2014), Land Use / Land Cover database on 1:50,000 scale, Natural Resources Census Project, LUCMD, LRUMG, RSAA, National Remote Sensing Centre, ISRO, Hyderabad

Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., … Ermon, S. (2018). Infrastructure Quality Assessment in Africa using Satellite Imagery and Deep Learning. 616–625. https://doi.org/10.1145/3219819.3219924

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D., & Ermon, S. (2017). Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning. (Nips). Retrieved from http://arxiv.org/abs/1711.03654

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., … Syrris, V. (2016). Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. https://doi.org/10.2788/253582

Rao, J.N.K, Molina, Isabel (2015) Small Area Estimation (2nd Ed), Wiley.

Reibel, M., & Agrawal, A. (2007). Areal interpolation of population counts using pre-classified land cover data. Population Research and Policy Review, 26(5-6), 619-633.

Reibel, M., & Bufalino, M. E. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. Environment and Planning A, 37(1), 127-139.

Roy, P. S., Roy, A., Joshi, P. K., Kale, M. P., Srivastava, V. K., Srivastava, S. K., … Kushwaha, D. (2015). Development of decadal (1985-1995-2005) land use and land cover database for India. Remote Sensing, 7(3), 2401–2430. https://doi.org/10.3390/rs70302401

Rue, H. and Martino, S (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations J. R. Statistical Soc. 71 part 2, 319-392

UNISDR (2015). Sendai Framework for Disaster Risk Reduction 2015-2030., Third UN World Conference on Disaster Risk Reduction, Sendai, Japan, March 18, 2015. Retrieved from http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). Data Deprivation: Another Deprivation to End. (April). https://doi.org/10.1596/1813-9450-7252

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. PLoS

ONE, 10(2), 1–22. https://doi.org/10.1371/journal.pone.0107042
Swaminathan, A., Subramanian, S. V, Blossom, J. C., Joe, W., Venkataramanan, R., & Kumar, A. (2019). Burden of Child Malnutrition in India. Economic and Political Weekly, 54(29), 7–8. Retrieved from https://www.epw.in/journal/2019/2/special-articles/burden-child-malnutrition-india.html?0=ip_login_no_cache%3D8a60daca19072f37a489f8542bec1c58
United Nations. (2015a). Transforming our world : the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/1. Retrieved from https://www.refworld.org/docid/57b6e3e44.html
United Nations (2015b). The Paris Agreement, United Nations Framework Convention on Climate Change. Retrieved from https://unfccc.int/sites/default/files/english_paris_agreement.pdf
United Nations (2017). The New Urban Agenda, Habitat III, The United Nations Conference on Housing and Sustainable Urban Development, 17-20 October. Retrieved from http://habitat3.org/wp-content/uploads/Habitat-III-the-conference.pdf
U.S. Geological Survey. (2012). Landsat-A Global Land-Imaging Mission: U.S. Geological Survey Fact Sheet 2012–3072, 4 p. (Revised July 19, 2012). Retrieved from https://pubs.usgs.gov/fs/2012/3072/fs2012-3072.pdf
Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Takahashi, S., Metcalf, C. J. E., … Tatem, A. J. (2018). High resolution age-structured mapping of childhood vaccination coverage in low and middle
Watmough, G. R., Marcinko, C. L. J., Sullivan, C., Tschirhart, K., Mutuo, P. K., Palm, C. A., & Svenning, J.-C. (2019). Socioecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences*, *116*(4), 1213–1218. https://doi.org/10.1073/pnas.1812969116
Wright, J. K. (1936). A method of mapping densities of population: With cape cod as an example. Geographical Review, 26(1), 103-110.
WorldPop. 2017. India 100m Population, Version 2. University of Southampton. DOI: 10.5258/SOTON/WP00532.
Xie, Y. (1995). The overlaid network algorithms for areal interpolation problem. Computers, Environment and Urban Systems, 19(4), 287-306.