Supplementary Document for Aggregate Development Index (ADI)

Adhya Dagar, Chahat Bansal

Indian Institute of Technology- Delhi, India

September 2020

1 Introduction

Indicators act as a metric to measure the progress and performance of a process and its implementation and inform the development of policies and plans. They could be used to measure change in a system such as tracking the government and public efforts towards achieving Sustainable Development Goals (SDG) over time or could be limited to be used as sample parameters to provide insights about the bigger picture.

There exists a spectrum of indicators to track the economic, political, ecological, cultural and social development of any region. These indicators might be specific to a particular sector for example an indicator that measures the percentage of households with access to sanitary latrines might help us evaluate the health of a community or might be more universal like the Human Development Index(HDI) which aggregates information and spans across multiple dimensions like quality of life, education and economic well being.

Hence, it is important to select these indicators meticulously as only well chosen indicators can effectively implement, monitor, evaluate and offer feedback to highlight important work areas, help make strategic decisions and allow targeted interventions to address problem areas.

This article aims to re-compute one such index, the Aggregate Development Index, introduced in Bansal et al. [2], where open-source satellite images that are available at frequent intervals and offer a lesser resource-intensive alternative to ground surveys and the census are used to predict socioeconomic development in Indian districts.

Our target index is called Aggregate Development Index (ADI), recently coined by Bansal et al. [2]. We propose a revised formula for the computation of this index and add another dimension to its existing form, hence making it more homogeneous to the universally accepted indicator- the Human Development Index. The scripts for calculating the revised index along with all the data files are available for public use on Github.

2 Background

Previously, many different approaches have been used to design indices that measure socioeconomic progress such as the Human Development Index by the United Nations(UN) and the Wealth Index by Demographic and Health Surveys(DHS).

Human Development Index (HDI), is a statistical tool used to measure a country's overall achievement in its social and economic dimensions. These dimensions try to capture an overall sense of well being based on the health of people, their level of education, and their standard of living. The health dimension is assessed by life expectancy at birth, the education dimension is measured by mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age. The standard of living dimension is measured by gross national income per capita. The HDI uses the logarithm of income, to reflect the diminishing importance of income with increasing GNI. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean.

Similarly, wealth indices are composite measures of a household's cumulative standard of living. They are computed using data on ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities. Each household asset for which information is collected is assigned a weight or factor score generated through principal components analysis. The resulting asset scores are standardized in relation to a standard normal distribution with a mean of zero and a standard deviation of one. These standardized scores are then used to create the break points that define wealth quintiles as: Lowest, Second, Middle, Fourth, and Highest.

While HDI and the Wealth Index are computed at the spatial granularity of a country, Bansal et al. proposed an Aggregate Development Index (ADI) to capture the development status of districts and villages in India [2]. It was simply computed as a sum of levels of six indicators from the Indian census- Assets, Bathroom Facilities (BF), Condition of Household (CHH), Fuel for Cooking (FC), Main Source of Light (MSL), and Main Source of Water (MSW). The value of this index ranged from 6 (all six indicators at the lowest level 1) to 18 (all six indicators at the highest level 3) for every district. On comparing the values of ADI computed directly from the census with those predicted using satellite images, they report a normalized root mean square error (RMSE) values of 0.1738 for 2011 using their forward classification and 0.1160 for 2001 using backward classifier over Landsat images of districts. We propose to revise the formula for computing ADI and show comparison of our normalized RMSE values against those previously reported.

3 Revised Formula for ADI

There is a need for aggregate indicators which capture spatial inequalities and comment upon the holistic well-being within a region as standard economic indicators like the Gross Domestic Product(GDP) do not provide information about these disparities and are sector specific.

In order to make ADI more comprehensive and representative of various socioeconomic dimensions, we choose three spheres of growth: education, health, and standard of living. We have included Literacy as the seventh indicator to the six indicators considered in the calculation of ADI in [2]. We map these seven indicators to the following dimensions for the purpose of computation-

- *Health:* Bathroom Facility (BF) and Fuel for cooking (FC)
- *Education:* Literacy (LIT)
- *Standard of Living:* Assets (ASS), Condition of Household (CHH), Main Source of Water (MSW) and Main Source of Light (MSL)

These indicators have been mapped to the respective dimensions only after considering evidence from published literature that shows correlation and sometimes causation between the indicators and the dimensions.

According to [6] low and middle income countries witness many deaths as a result of inadequate water, sanitation, and hygiene each year, representing 60% of total diarrhoeal deaths. Open defecation perpetuates a vicious cycle of disease and poverty. Poor Bathroom Facilities in a country are often interlinked with highest levels of malnutrition and poverty, and big disparities of wealth. Similarly, the type of fuel used for cooking can have a deleterious effect on health. In [5], the study showed that the women using biomass fuel for cooking suffered from respiratory diseases and other morbidities. Furthermore, such morbidities were found to be increased with increase in duration of cooking. Thus we can establish a direct impact of the condition of bathroom facilities and fuel for cooking on the health of a person.

As per [1] an individual's quality of life is fundamentally bounded by their access to material resources, some of which are transformed into well-being. Having an adequate accommodation for example, is at the top of the hierarchy of human material needs followed by additional assets owned by person which might be commensurate with the spending capacity or the 'ability to afford'. Low quality housing may be associated with reduced well-being and increased levels of psychological stress. Housing conditions and assets owned by a household have obviously an important impact on the quality of life.

While [3] concludes that a sufficiently large supply of electricity can ensure that a higher level of economic growth, [4] identifies the impact of availability of water resources on securing economic investments. Since electricity and water are two fundamental human needs to live a quality life we assume these indicators should be counted when determining a person's standard of living.

Now, having established our dimensions and selecting the associated indicators for each district, we sum up the levels of all indicators as done in [2]. The revised formula thus becomes-

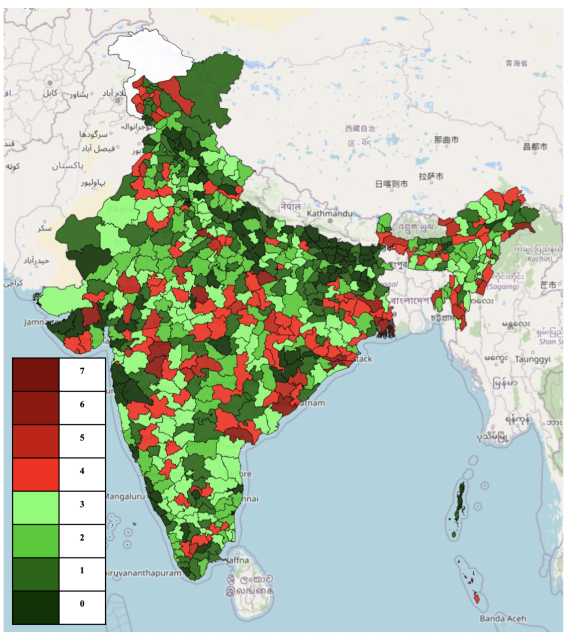
$$ADI = ASS + BF + CHH + FC + LIT + MSL + MSW \tag{1}$$

4 Results and Analysis

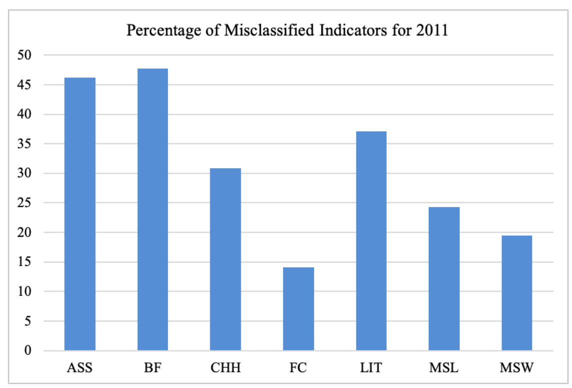
With the above mentioned formula, the values of ADI will range from 7 (all seven indicators at the lowest level 1) to 21 (all seven indicators at the highest level 3) for a district.

On comparing the values of ADI for 2011 predicted by the forward classifier [2] over Landsat images with the actual values computed from the census data, we get a normalized RMSE (root mean square error) value of **0.1548** across the districts. Similarly, a normalized RMSE value of **0.1102** is achieved using the predictions from the backward classifier for 2001. While the old formula gave us normalized RMSE value of **0.1738** for 2011 and **0.1160** for 2001, the revised formula gives us lower error rate, hence reaffirming that it is a better composite metric to quantify development.

Figure 1a shows the districts in darker shades of red on which the two-step classification works incorrectly when going from the base year of 2001 to the target year of 2011. We observe that only 25.17% of the districts have four or more indicators predicted incorrectly and this count for the remaining 74.65% districts is atmost three. On further analysis we find that indicators like FC and MSW have less incorrectly predicted values as compared to BF and Assets. At state level, we find that majority of districts from Bihar, Delhi, Kerela, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and Uttrakhand have 3 or few misidentified indicators while states like Jammu and Kashmir, Madhya Pradesh and West Bengal have relatively 4 or more incorrectly predicted indicators. This calls for a deeper understanding of why the classifier works in certain regions and fails to predict in others.



(a) Count of mis-classified indicators in 2011: Districts with fewer indicators predicted correctly are shown in darker shades of red.

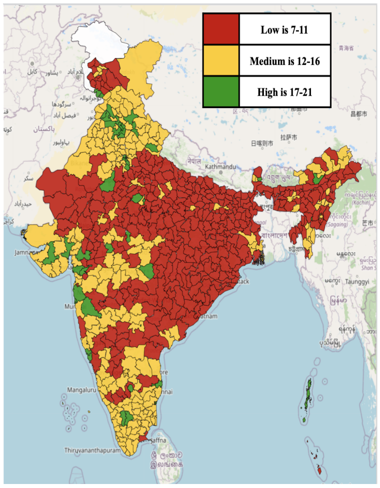


(b) Percentage of Misclassified Indicators in 2011

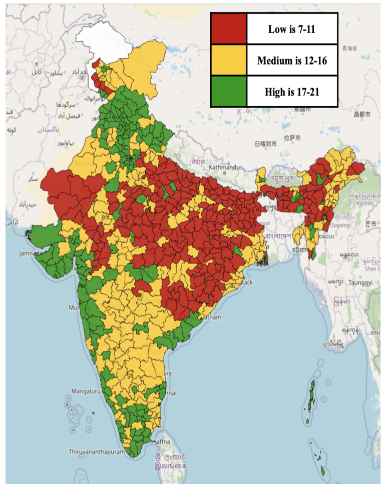
Figure 1: Visualizing mis-classifications in indicator prediction using satellite imagery

To provide a high level understanding and for the purpose of visualising development across spatial and temporal scales as in Figure 2, we classify the districts into low, medium, or high level of development by dividing the range of obtained ADI values into three equal intervals. A fixed set of intervals will allow us to track and compare the changes in development across multiple years.

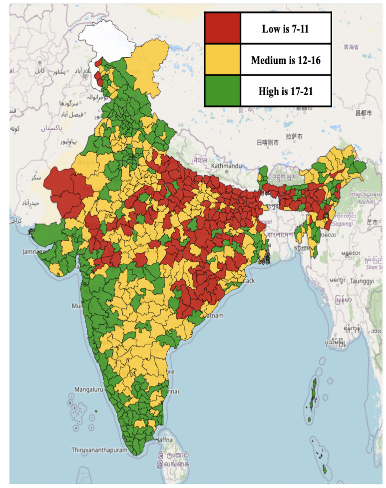
We classify a district with ADI value in the range [7-11] as a low development district, range [12-16] as a medium development district and [17-21] as a highly developed district. To validate our three intervals we perform unsupervised K-Means clustering on 2001 and 2011 ADI data(as per census) and observe three clusters segmented by similar range of values as seen in Figure 3.



(a) ADI in 2001(as per census)



(b) ADI in 2011(as per census)



(c) ADI in 2019 (predicted)

Figure 2: Visualizing district development in India through ADI

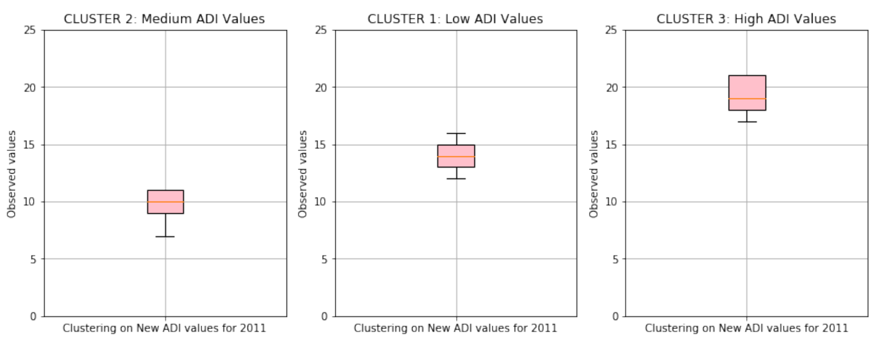


Figure 3: Clustering Performed on ADI values(as per census) of 2011

Some best performing districts in terms of positive change in development(2001 to 2011) are Visakhapatnam, Dadra & Nagar Haveli, Guntur, Solan while in districts such as Kaushambi, Balrampur, Lower Subansiri and Rajgarh the ADI values have depreciated over time.

5 Conclusion

We see how by adding the Literacy dimension to the index introduced in [2] we can offer a better approach to understanding development. The updated ADI formula helps us understand the interplay of health, literacy and standard of living on the over all well being and socioeconomic development in a country.

References

[1] Saamah ABDALLAH and Laura STOLL. "Review of Individual-level Drivers of Subjective Wellbeing". In: *Analysis, Implementation and Dissemination of Wellbeing Indicators. Eurostat* (2012).

[2] Chahat Bansal et al. "Temporal Prediction of Socio-economic Indicators Using Satellite Imagery". In: *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 2020, pp. 73–81.

[3] Sheng-Tung Chen, Hsiao-I Kuo, and Chi-Chung Chen. "The relationship between GDP and electricity consumption in 10 Asian countries". In: *Energy policy* 35.4 (2007), pp. 2611–2621.

[4] Charles W Howe. "THE EFFECTS OF WATER RESOURCE DEVELOPMENT ON ECONOMIC GROWTH THE CONDITIONS FOR SUCCESS". In: *Natural Resources Journal* 16.4 (1976), pp. 939–955.

[5] Ipsa Mohapatra, Sai Chandan Das, and Sonia Samantaray. "Health impact on women using solid cooking fuels in rural area of Cuttack district, Odisha". In: *Journal of family medicine and primary care* 7.1 (2018), p. 11.

[6] World Health Organization. *Sanitation*. URL: `https://www.who.int/news-room/fact-sheets/detail/sanitation`. (accessed: 27.09.2020).