Analyzing Power infrastructure shortage in African & Asian Countries using night time Satellite Imagery

Bhavya Ghai, Sabir Ismail and **Venkata Kedarnath Pakala**
Stony Brook University
Stony Brook, NY 11790, USA
{ bghai, saismail, vpakala } @cs.stonybrook.edu

Abstract

This project tries to identify & rank different countries in Africa & Asia suffering from serious power infrastructure shortage. Apart from World Bank data, this study specifically focuses on Satellite imagery data collected by NOAA from 1992-2013. Intensity of light in night time satellite images for a specific country is considered as a proxy for power infrastructure in that country. It is combined with demographic, economic and development indicators from world bank data to cluster countries into different categories depending on power infrastructure. This study aims to bolster transperancy & awareness about the state of power infrastructure in data poor countries

1 Introduction

Every year billions of dollars are spent by UN & other charities for the development & upliftment of African & other countries in Asia. Ideally, funds should be allocated to those countries who most need such funds.But this requires knowledge and analysis of the infrastructure shortages per region so that right amount of investments can be made based on region and problems can be addressed based on their acuteness. Unlike the developed countries, many african & asian countries lack a robust & transperant data portal. Without official numbers, it's difficult to analyze the state of a country & make optimal decisions.

One of the basic issues people face in developing world is the shortage or lack of electricity. In the entire world, nearly 2 out of 10 people do not have access to electricity. But this problem is more acute in Sub-Saharan Africa where 7 out of 10 people do not have access to electricity. In actual figures, 622 million out of 1 billion people live in darkness. Such a problem can be addressed if we have information per square km of the region. Acquiring data at that level for entire continent like Africa would involve considerable effort, expenditure and time. There is cheaper alternative, satellite images. With rapid advances in technology, organizations like NASA, NOAA are able to provide high resolution night light satellite images of earth per day. These images are well processed and to counter . Analysis of night light images can help us gather information on electricity infrastructure of very large regions like Africa at a granular level and this processing can be done within stipulated time and expenditure. Night time satellite images can serve as a proxy to the power generation & transmission infrastructure of a country.

Moreover, some countries might project false numbers for political gains or to misled international organizations to gain more funds, etc. In such situations, satellite images can be used an alternate to verify original numbers or can be used as a replacement altogether. With recent advances in data science, it is now easier than ever to store & process huge amount of spatial data. In this project, we will be using spark to store & process satellite images. Furthemore, we will using ML techniques like linear modeling, clustering & dimensionality reduction for our analyses.

2 Sustainable Development Goal and Background

United Nations has adopted 17 development goals to address different pressing social problems as part of their agenda. (Wikipedia, 2017) These goals have specific targets to be achieved for next 15 years. These goals are quite diverse & touch problems like poverty, health, infrastrucutre, life in water, life of land, etc.

The focus of our project is to address some of the targets of SDG Goal 9 (UN, 2017) Industry, Innovation and Infratructure. Investments in infrastructure like transportation, irrigation , energy, technology - are important for sustainable development of the countries. Facts show that 2.6 billion people across world lack access to electricity and basic sanitation. Inadequate infrastructure limits access to health care and education. It also limits access to markets , jobs ,training and creates a major hurdle in business. Some of the targets that this goal includes are developing quality,reliable and resilient infrastructure, promote sustainable industralization by 2030,enhance scientific research , upgrade technological capabilities of industrial sectors and adopt clean and environment friendly technologies. In this project, we are addressing the problem of ensuring reliable and sustainable electricity infrastructure for all using night light satellite images.

As we are using the dataset form the (DMSP, 2017) there are some problems. Night Light satellite images are being suing quite long for the research. They try to find out light pollution to economic activity, greenhouse gas emissions and using night-time lights to help with disaster management. (Elvidge et al., 2009a) use nighttime lights to modeling the spatial distribution of population density, carbon emissions, and economic activity. But there are some problems with satellite image as mentioned in the paper (Samson,). But, a well-documented deficiency of this dataset is the lack of intra- and inter-annual calibration between satellites, which makes the imagery unsuitable for temporal analysis in their raw format. There are several proposed methods form the doing thin intra and inter-annual calibration between satellites. (Doll, 2008) use night light images and show some potential use of them. There are several techniques for calibration of these night light satellite images. In the paper (Savory et al., 2017), they used quadratic regression model: $X_{i,0} = C_0 + C_1\ X_{i,j} + C_2\ X_{i,j}$ Where $X_{i,0}$ is the DN values of the i-th grid cell in the reference image, $X_{i,j}$ is the DN values of the of the i-th grid cell in satellite-year image j, and the parameters C_0, C_1 and C_2 are the intercept, linear, and quadratic coefficients. (Elvidge et al., 2009b) also use same quadratic regression model. In the (Samson,) use some additional calibration techniques. We also use this quadratic regression model in our first part. For calibration, all those paper use Sicily, as reference image. The reason is, there is not much change of population and night light intensity in the year 2000-2010, so they take this as a reference image. So we also use Sicily image of year 1999, as our reference image.

3 Data

We are using the two types of data 1) Image data - night light satellite images from NOAA website. 2) Shape files 3) Socio economic data from world bank.

Satellite Images: Satellite data is fetched via DMSP-OLS program hosted by NOAA. DMSP or defense meteorological satellite program started as a classified operation and became public in 1973. There are 2 sets of image data a) images for the years 1992-2012 which contain a single image for the entire world with resolution of 43201 * 16801 pixels. These images include cloud free, averaged over all days in the year and constant light intensity b) images for years 2013- 2017 . These images include 6 images for the entire world. Resolution horizontally and vertically is 100 dpi.

Shape files Shape files are used to extract the images of specific countries from the main image. So, we downloaded the shape files of all 49 African countries and 12 Asian countries.

We automated the download of images and shape files using linux script files.

Once we extract the image per country we compute the DN values. DN(Digital Number) is used as a measure of light intensity of a region or country. These DN values are calibrated from 0 to 63. High DN value indicates the country has more night lights.

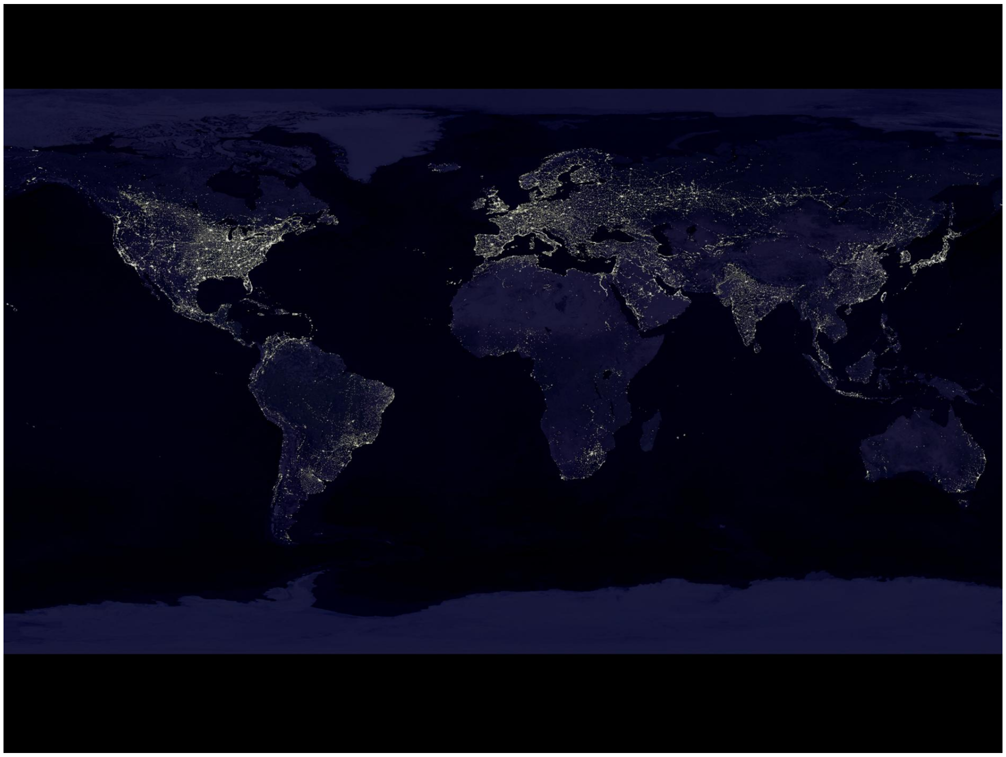

Figure 1: World view using night time satellite images

World Bank: We have gathered data for access to electricity,population, socio-economic data like GDP per capita . We used the data to compute the dnvalue per capita which requires population density in a region. This gives a feature to quantify the supply of electricity . While GDP per capita, electricity access, population density in a region are used for computing demand for electricity. Higher is the difference in supply and demand higher is the problem of infrastructure in the region. As the data is time series , we could use that to compute the change in the problem over a period of 15 years.

4 Methodology

For analyzing the power infrastructure shortage in Asian & African countries, we have used night time satellite images from NOAA. Firstly, we analyzed satellite images to capture the intensity of light for a particular country. Then we'll combine this value with other world band development indicators to arrive to a final ranking. NOAA provide night time satellite images from year 1992 to 2017. The files are cloud-free composites made using all the available archived DMSP-OLS smooth resolution data for calendar years. In cases where two satellites were collecting data - two composites were produced. The products are 30 arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude. There are two version of the images, we used the first version which include images from 1992 to 2013, (DMSP, 2017). A well-documented deficiency of this dataset is the lack of intra- and inter-annual calibration between satellites, which makes the imagery unsuitable for temporal analysis in their raw format. There are several proposed methods form the doing thin intra and inter-annual calibration between satellites. In the paper (Savory et al., 2017), they used quadratic regression model: $X_{i,0}$ = C_0 + C_1 $X_{i,j}$ + C_2 $X_{i,j}$ Where $X_{i,0}$ is the DN values of the i-th grid cell in the reference image, $X_{i,j}$ is the DN values of the of the i-th grid cell in satellite-year image j, and the parameters C_0, C_1 and C_2 are the intercept, linear, and quadratic coefficients. (Elvidge et al., 2009b) also use same quadratic regression model. In the (Samson,) use some additional calibration techniques. We also use this quadratic regression model in our first part. For calibration, all those paper use Sicily, as reference image. The reason is, there is not much change of population and night light intensity in the year 2000-2010, so they take this as a reference image. So we also use Sicily image of year 1999, as our reference image.

Next we do another calibration, that is shifting as introduced in the paper, (Samson,). Because shifting the image to some pixel make more calibrated, So again we use Sicily as reference image, and calculate the shifting value for the all years.

First, we use shape file of the Sicily, to mask out portion of the Sicily from the original image. In our satellite images, we have only pixel values, but it does not contain any longitude and latitude information. So we use some mathematical equation to get longitude and latitude from the pixel value. We mask out Sicily from the all our satellite images, year 1992 to 2013. In some years, we have several satellite images, we took both and later part we took average of them. So we run a quadratic regression model on our reference image with all the images and find the coefficient values for the each (year, satellite).

Table 1: Correlation coefficient for Calibration

Year	c0	c1	c2	R square
1992	-2.057	1.5903	-0.009	0.9075
1993	-1.0582	1.5983	-0.0093	0.936
1994	-0.3458	1.4864	-0.0079	0.9243
1994	-0.689	1.177	-0.0025	0.9071
1995	-0.0515	1.2293	-0.0038	0.9178
1996	-0.0959	1.2727	-0.004	0.9319
1997	-0.3321	1.1782	-0.0026	0.9245
1998	-0.0608	1.0648	-0.0013	0.9536

Then we mask out countries from the original images using shape file of the targeted countries. We focus on the Africa continent, which is economically more backward compared to other continent. In this continent we have all 49 countries for our analysis. We also target some parts of the Asian region as there are more divers countries are available. We have selected Afghanistan, Bangladesh, Bhutan, Nepal, India, China, Singapore, Malaysia, Myanmar, Maldives, Sri Lanka, Pakistan. Some of these countries are very well developed and some are not. So we also choose them to analyze their night time satellite images.

For, all the countries, next we do that calibration as stated previously. Our all pixel have DN values on range 0-63 as described in the (DMSP, 2017) . We just sum up them to find Sum of Light (SOL) for every countries for each year. Now, our dataset is ready for comparing with World Bank dataset.

Power infrastructure of a country can be viewed in different aspects. To rank a country for power infrastructure we need to consider different aspects, quantify them & come to a final rating for each country. In this study we have considered, three different aspects to power infrastructure of a country. One aspect can be the availability of electricity to its citizens. To quantify this, we use the ration of DN values and the population of the country. this gives an estimate of the power consumed by each citizen. Second aspect is efficiency. We have used the ration of power generated from renewable sources to quantify this aspect. Third aspect We merged the datasets and computed scores based on amount of lighting per capita and the demand for lighting. We used it to perform clustering. We determined number of clusters using elbow method. The labels generated are used for grouping the countries into categories like 'worse','bad','good','better' based on acuteness of the problem of electricity infrastructure.

5 Results

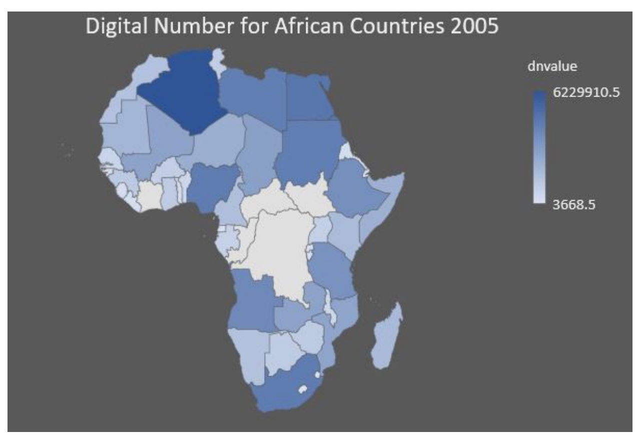

Figure 2: Processed Digital number values for African countries in 2005

We have generated DN(digital number) values for various Asian and Sub-Saharan countries. For this we extracted country wise image from the main satellite image using shape files. Then we calculated DN value for each country that shows how much bright the country is during night. This data provides a way to quantify the supply of electricity in a country. For demand, we used multiple datasets from world bank for information on access to electricity in percentage of population , population , GDP per capita, percentage of transmission loss and percentage of renewable sources.

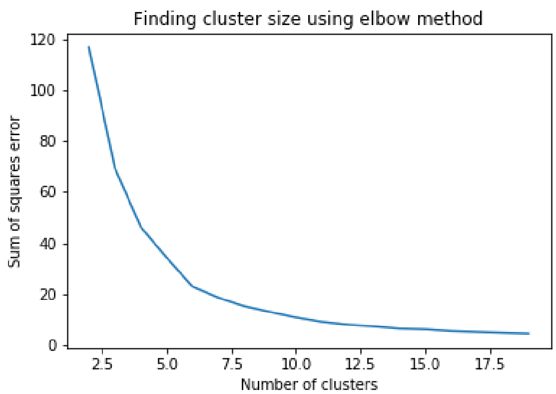

Figure 3: Clustering countries into different sections using K-means algorithm based on their feature parameters

Table 2: Best rated countries for Power Infrastructure

Country	Rating Score
Gabon	1.1631
Tunisia	1.3816
South Africa	1.3918
Singapore	1.6329
Mauritius	1.8311

Table 3: Worst rated countries for Power Infrastructure

Country	Rating Score
Benin	0.03718
Togo	0.06932
Mozambique	0.07181
Nigeria	0.08264
Namibia	0.38274

6 Discussion

Satellite images are a very cheaper and convenient way to perform analysis on a large scale. Through this project we show and learn that satellite images can be used for various purposes other than just infrastructure. Using satellite images has saved cost, effort and time. With rapid advancement in technologies and provision of high quality images , we can use image processing to make various predictions pertaining to different sectors.

7 Conclusion

Our team has thrived for novelty right from the beginning. Reading different papers in this area, we found literature describing current state of power infrastrucutre in africa, predicting cost to improve infrastructure, etc. Africa has still got some attention in literature but many asian countries who are facing similar situation have not been covered that much. Hence, we came up with our own problem statement where we included african & asian countries together. We ana tried to quantify the state of power infrastructure & tried to identify countries that are in immediate need in Africa & Asia over a period of 20 years. Based on our background study, night time satellite images have been used to predict GDP, poverty, etc. The use of night time satellite images to capture power supply in africa is another unique apect of our project.
Using geo-spatial data along with other statisitical data can greatly echance decision making. Our approach can help UN or other international organizations to build a more transparent & accurate perspective about the power state infrastrucutre in developing countries. Hence, it will indirectly help to develop the power infrastrucutre in those countries by promoting better allocation of funds.

8 Future Work

In the future, this work can be extended by adding more parameters for measuring power infrastructure. Satellite image data can be used to analyze if the distribution of power in a country is equitable or not. Furthermore, we can also try to quantify quality of power by measuring things like hours of powercut/day, hours of low voltage power, etc. This might give even better picture of the state of a country.

9 Team members Contribution

Initially, we thought about dividing work among different team members and aggregating the work together in the end. In reality, we sat together & worked closely with each other all the time. There's so much give and take that we can't really segregate which task was done by whom. In a nutshell, this entire project is a team effort with everyone playing some role in every aspect of it.

References

DMSP. 2017. Version 4 dmsp-ols nighttime lights time series, December.

Christopher NH Doll. 2008. Ciesin thematic guide to night-time light remote sensing and its applications. *Center for International Earth Science Information Network of Columbia University, Palisades, NY*.

Christopher D Elvidge, Edward H Erwin, Kimberly E Baugh, Daniel Ziskin, Benjamin T Tuttle, Tilottama Ghosh, and Paul C Sutton. 2009a. Overview of dmsp nightime lights and future possibilities. In *Urban remote sensing event, 2009 Joint*, pages 1–5. IEEE.

Christopher D. Elvidge, Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. 2009b. A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3):595–622.

Naizhuo Zhao ; Yuyu Zhou ; Eric L. Samson. Correcting incompatible dn values and geometric errors in nighttime lights time-series images. *IEEE*.

David J. Savory, Ricardo Andrade-Pacheco, Peter W. Gething, Alemayehu Midekisa, Adam Bennett, and Hugh J. W. Sturrock. 2017. Intercalibration and gaussian process modeling of nighttime lights imagery for measuring urbanization trends in africa 20002013. *Remote Sensing*, 9(7).

UN. 2017. Goal 9: Build resilient infrastructure, promote sustainable industrialization and foster innovation, December.

Wikipedia. 2017. Sustainable development goals, December.