Augmenting data for Urban Metabolism of cities Tool using Machine learning and Satellite Image Analysis of city

by

Emmanuel Havugimana

B.S. Electrical Science and Engineering
Massachusetts Institute of Technology, 2020

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author . . . .
Department of Electrical Engineering and Computer Science
August 6, 2021

Certified by . . . .
John E. Fernandez
Director of MIT Environmental Solutions Initiative
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Augmenting data for Urban Metabolism of cities Tool using Machine learning and Satellite Image Analysis of city

by

Emmanuel Havugimana

Submitted to the Department of Electrical Engineering and Computer Science
on August 6, 2021, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

We use image analysis to augment data about a city's material flow or material stock.We take existing data about cities such as energy consumption,biomass,water consumption,energy production and construction material either at the city level or national level and add data from satellite based remote sensing. From remote sensing we can get data like built area,population distribution across the region,and night light intensities.

We do this by coupling the insights from images which indicate a proxy for where resources are concentrated.We increase data available for the Urban metabolism tool database in resources correlated to satellite data. We show how data can be collected and may be integrated.

Thesis Supervisor: John E. Fernandez
Title: Director of MIT Environmental Solutions Initiative

Acknowledgments

Here, I would like to express my deepest gratitude to my thesis supervisor Prof John E. Fernandez for his support and guidance and also for his research group team,and autonomy he made possible.

I would also like to thank Anthony Hernandez for the work done on umtool, Mainly, UI and other parts of the tool, and for setting up the Github repo. I would also like to thank Matias Williams for his technical support on geoprocessing. I would also to thank Norhan Bayomi and Alejandro Jimenez Jaramillo for managing the team.

My family for insulating me from the details of pandemic and lockdown. God for Keeping me safe. For last but not least, Mostafa Negm and Alma Dogalova for keeping me sane during MEng.

Contents

List of Figures

List of Tables

Chapter 1

Background and Introduction

More and more people are living in cities and some cities across the world have similar challenges and situations. But given a huge number of cities across the world, it is hard to get data across all cities for comparative and grouping reasons. The study of grouping of cities,,city typology is interested in putting together similar cities so that policies can be shared or case studies can be used on similar cities. This makes it easy for cities to learn from each other in same typology [9].

Here we describe a methodology and a tool to allow classifying cities across the globe using more increasingly ubiquitous satellite data. Surveys take a lot time and resources to compile and different statistical bodies collecting data, units/conventions,dates and frequency of data collection, and integration of data across cities is slow. Satellite data collection is expected to grow more and we provide a possible data pipeline that can be used to make estimates of data about cities that can then be used to do typological analysis of cities then policy.

1.1 Need for satellite data

Satellites offer synchronicity that surveys can not match even on national scale. Satellites samples may be taken across the globe almost on a same time period which makes easy to compare them, and a local or private satellite imagery by corporation, city or country can also be used to enrich globally available data for in-house processing.

The uniformity of sampling method across a region makes comparison easier to do.

1.2 Spatial up scaling using satellite data

Most data published by organizations like World Bank, IMF or UN are at national level. Data like Human Development Index,GDP, population, Trade. But inside a country,the population can be distributed in an interesting way. For example,in the US the ratio of people living in coastal states or close to the coast is higher than one in center.

In countries like Russia, some regions might almost be underpopulated due to the huge size of the country and different conditions across country. Figures like population density on a national scale might not be interesting as people do not feel or have less agglomeration in less dense countries, but regional or city level population density would describe the population density in places that people are living in.

For example,even though there exists classifications like developed countries and underdeveloped countries,a city in an underdeveloped country like Democratic Republic of Congo might have better Internet access,stabler electricity, more closer hospitals, etc than some rural region in the United States.

Up-scaling national data can help in having a more detailed view of data;For example population distribution, resource usage across space and allow more comparison and data as there are more cities than countries and **cities are smaller and more homogeneous than countries**.Hence,city data is better summarized than a country can be summarized by simple statistics.

Population rasters have been created by [10] using national population surveys and up-scaling that using built-up area that can be estimated using remote sensing [11]. We are expanding on that work in a special angle by focusing on a policy oriented region of interested where by instead of regular grid output,we can take grided data from satellites or derivatives like work by [10] and get values for smaller administrative units for this case city[by an arbitrarily region of interest can be chosen to be aggregation region from gridded data] and putting together different data sources.

With up-scaling ,we use available data for cities and countries to make good estimates of cities for which we don't know the values.

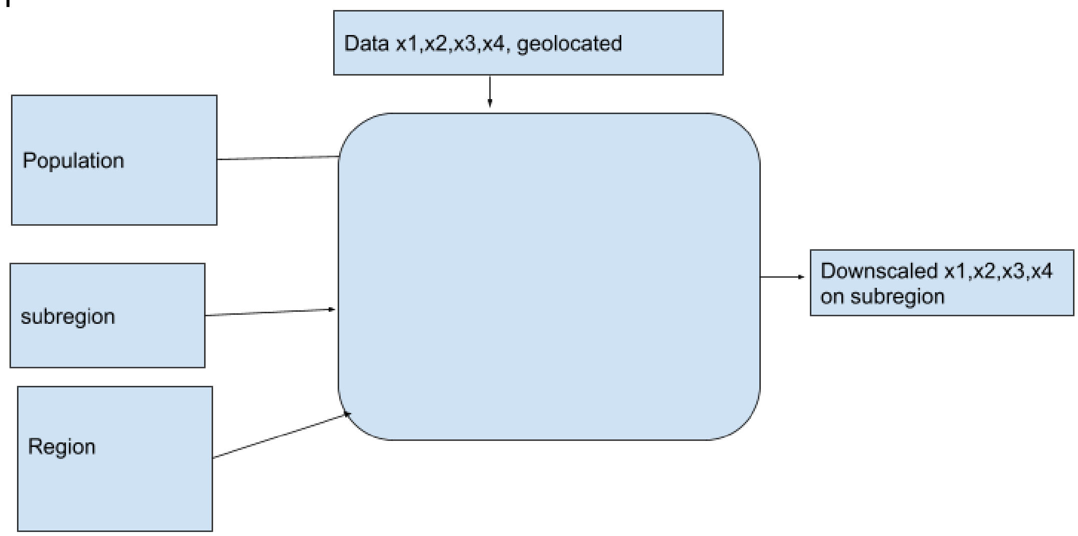

Figure 1-1: Up-scaling National Data to city data using data derived from satellites

Chapter 2

Description of the tool

2.1 Overview

We are in process of developing a tool for policy makers to get good information for reasoning about solutions in their communities. The tool uses the web as platform for access. Web application is better suited for this tool as it does not require installation which lowers learning curve of the tool. Being on the web mostly served from our server on an AWS EC2 instance means the tool can evolve or get new features faster and to all users at the same time.

We currently are not tracking usage statistics as it is in development stage but in future being on web also means we can learn who is accessing our tool either through asking them to fill forms or using minimal data like device location/type from cookies which will be used to make the tool more optimal.

We are using the D3.js library for the frontend [2] mainly for graphs, charts and maps and Flask server for tool logic and data access API on the backend.[6]

2.2 Database and API

Some of structured data is hosted on MYSQL database is served by a Flask app endpoint runing on Apache2 server on EC2 Instance MYSQL for database.

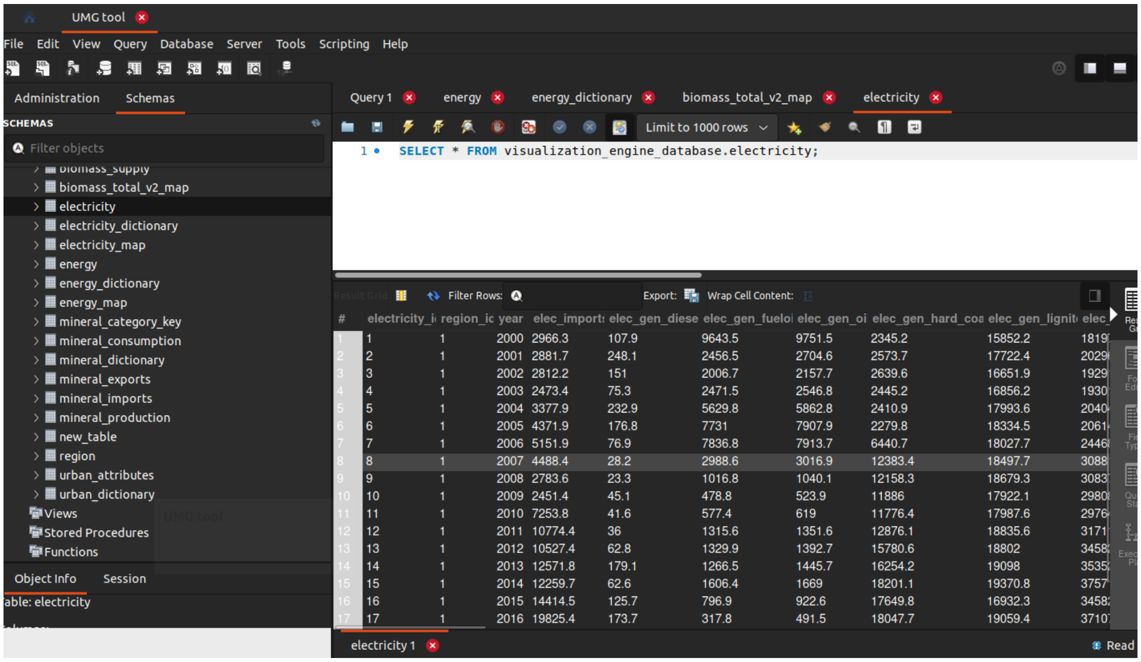

Figure 2-1: MYSQL database :hosted on Amazon Web Service(AWS

2.3 Sankey diagrams

This part of the tool one can select a city and type of resources to look at in a flow of resources model from production/extraction of resource to its usage and wastage.

Figure 2-2: Resources usage of Thailand . From sources which comprise production and import to different consumption or usage of resource

2.4 Maps

Visualizing location of interest and then displaying information related to the place searched or selected is envisioned in the final tool. We have developed a prototype with database of 9000+ cities and selection of countries. Tool will continue from there to the full operational mode envisioned.

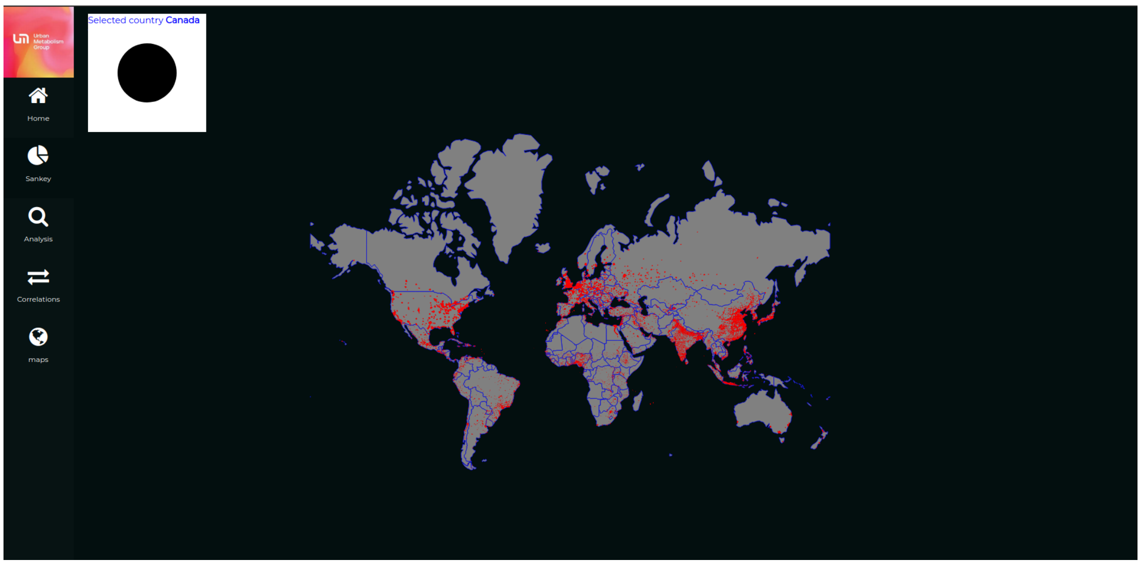

Figure 2-3: Mapping tool prototype. Selecting a region or search display results about the region

2.5 How this fit into the tool

Since the tool is fundamentally data dependent,This thesis is a way to find data at large scale so that the tool can offer rough guidance or have a potential for global reach. Working on tool, we realized that city data was not readily available and reading each city report from different websites,data formats and sources might not be the primary feeder mechanism to the tool as it is prohibitively expensive to do so. This is because it takes a lot of time time both to find the initial dataset, but also to find updates to the data sets. Stale data is not something a tool developer might want as being up to date with current data and models is what make a tool usable over time. Satellite data offers that in that most satellites data are published over time and once published it cover a wide area. Potential for spatial upscaling and downscaling also offer a way to include reports data which can be done by incorporating different levels of accuracy or [data preferences] by preferring reported data and then showing on tool the source of data[either derived by tool or the report from which it can be read]

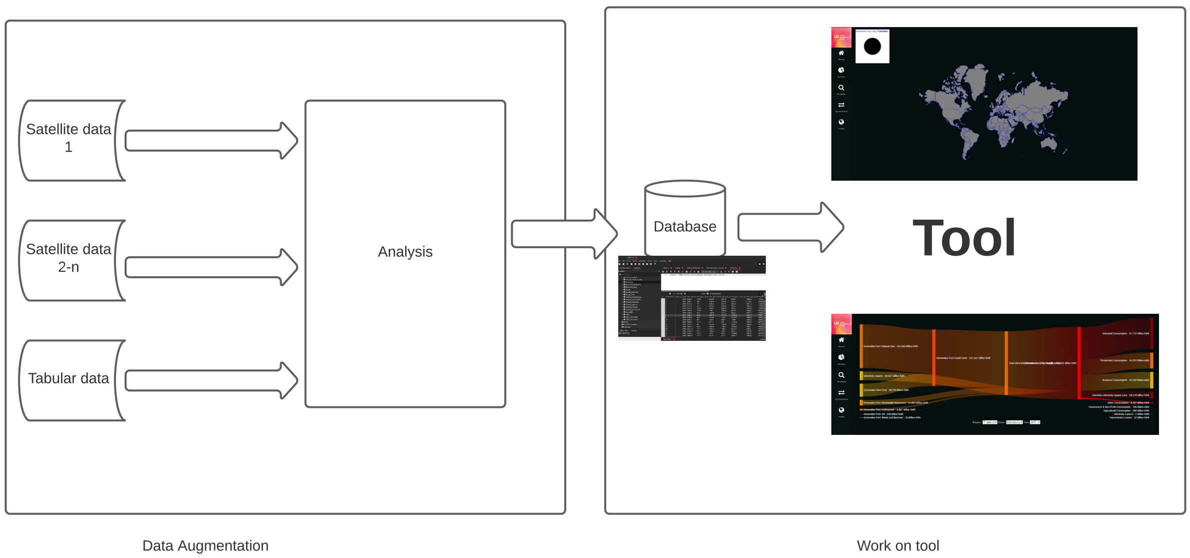

Figure 2-4: Overall system

Chapter 3

Data Collection

Satellite data, though abundant, it is hard to put together for processing as there are multiple formats and each satellite has its own encoding. So we using already processed data that abstracts low level details of satellite imagery; details like cloud removal, piecing together different images taken at different location or time together and mapping original pixels values to physical interpretation like temperature or light intensity. Hence, we mostly starts with data at processing level 2 [1] and above .

3.1 Tabular data

For tabular data, we are leaning on previous research group work and UN data. Tabula [1] was used to extract tables from pdf and PDFSam Basic [13] was used to split pdf's into data section to be processed by tabula.

[1]Data is made of derived geophysical variables at satellite resolution. Details about levels of processing can be found at `https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels`

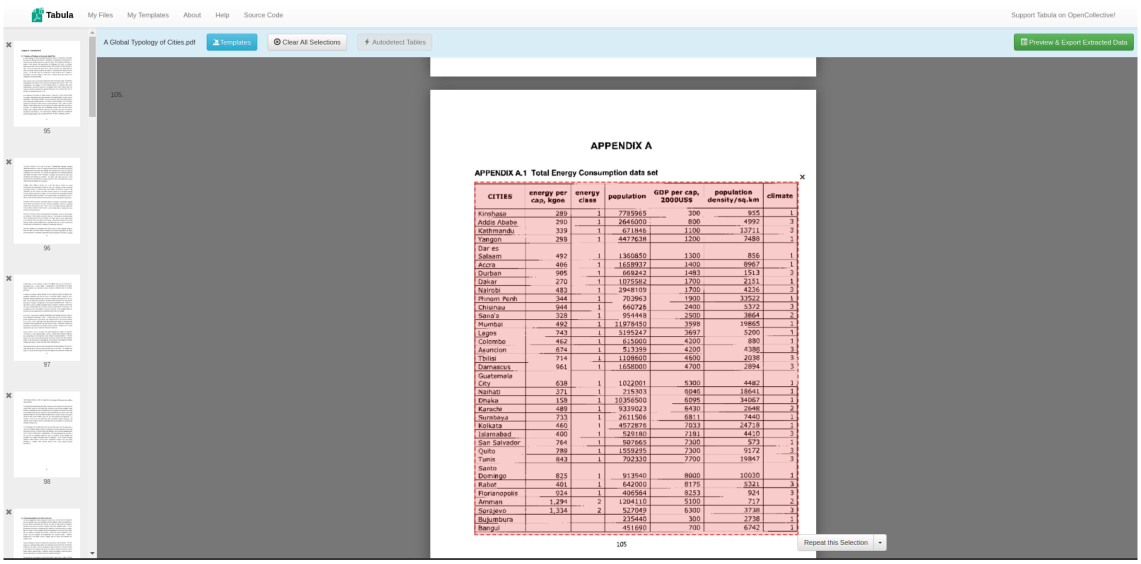

Figure 3-1: selecting table to extract the data from pdf using tabula

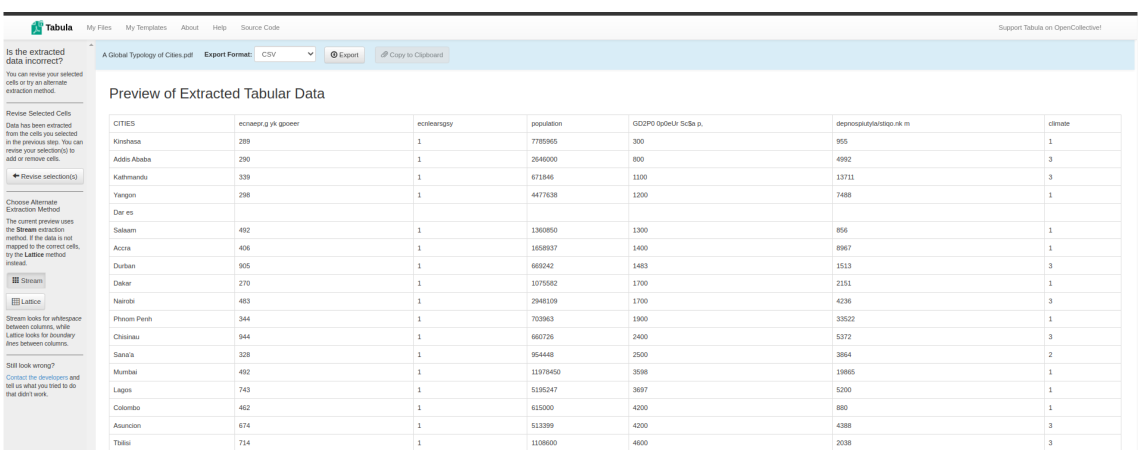

Figure 3-2: Extract data from pdf document using tabula. This is exported and cleaned

Previous student in group working on global typology of cities [3] in table 3.1 and UN Cities Database [7] described in table 3.2. The following tables describes the availability of data as extracted using tabula[2] from the documents, followed by cleaning(like making sure cities with more than one word name are well interpreted and table joining) and checked by crossing checking samples from extracted

Data	Number of cities
CO2	150
Tot DMC	150
Water	150
Biomass	150
Fossil Fuel	150
Construction	150
Industrial Minerals	150
electricity	150
Energy	150

Table 3.1: Urban typology thesis

Data	Number of cities
population (millions)	100
GDP(billions)	97
Land Area (sq. km)	63
Human Development Index (country)	88
GDP per capita	99
Real GDP growth rate (% per annum)	96
Income Distribution (Gini Index)	56
City unemployment rate (%)	58
GHG Emissions per capita (tCO2e/cap)	89
GHG Emissions per GDP (ktCO2e/$bn)	89
PM2.5 Concentration (mcg/cu.m)	35
PM10 Concentration (mcg/cu.m)	76
Total energy consumption per capita (GJ)	23
Total energy consumption per GDP (MJ/$)	23
Total electrical use per capita (kWh)	67
Population with authorized electrical service (%)	53

Table 3.2: UN cities Data

[2] Tabula is a software that extract tables from PDF document. It can be downloaded at https://tabula.technology/,

3.2 Satellite derived data

First, the Global data is downloaded and then clipped to match the city geometry and location. After that statistics are extracted. This is generalize-able to any selected area. The pipeline can be you select an area on map (by area selector) and get statistics, we used cities as areas of interest. used rasterio for raster data processing [5], geopandas [8] and shapely [4] for tabular and shapefiles processing.city boundaries used were from GHS global settlement layer [12]

Data	Number of cities
City light	9000
Population	9000
Built Area	9000
Temperature	9000
NDVI	9000
Geometry(shape)	9000
Area	9000

Table 3.3: Satellite derived

pseudocode for processing Satellite Image data into a different values for city

```
data=dict()
for each_data_file in [cityLight,pop,landuse]:
    # read and process each satellite image file
    dataMatrix=read(each_data_file)
    cityshapes=read(cityshapefile)  # read city geometries
    aggregates=dict()
    for city in city_shapes:
        # extra a city from global data
        cityData=clip(dataMatrix,with=city.geommetry)
        save(cityData,file_name_<city>_data) # save the the extracted data for each city
        aggregates[city]=compute_aggregates(cityData)
    # store the put together values for each city
    data[each_data_file]=aggregates
```

Code Listing 3.1: clip city from global satellite image

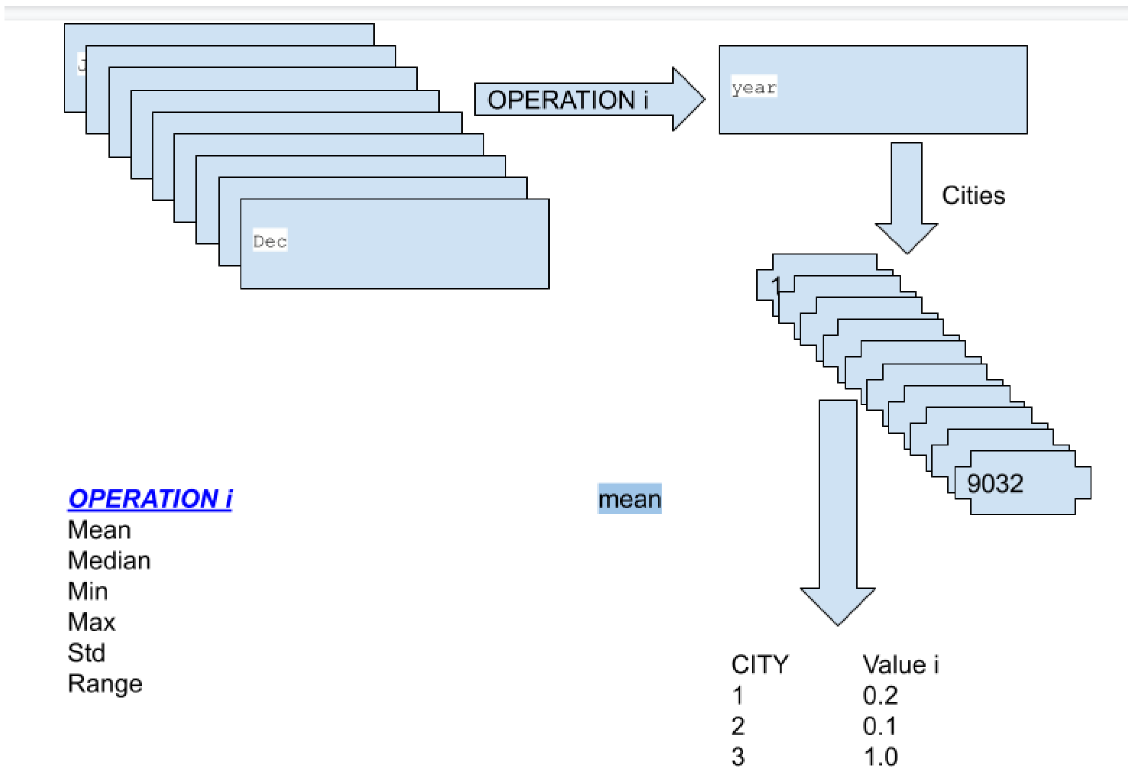

Figure 3-3: Example of Satellite image data processing . To top left, different rasters for different months are aggregated into a single raster for each year aggregated with either mean, median, max, min, std values of the months for each point on raster. Then year raster is clipped into 9000+ different city rasters which are then further processed to give one or more values for each city

Chapter 4

Analysis

Region geometries can change from that of cities to that of countries and the processing would not change. The model on cities can also work on smaller entities like counties in a city or large scales like countries. That means training on any of set with rich data sets can guide or help in inferring relationships between other geometries. For example train on country level data and test or validate on cities.

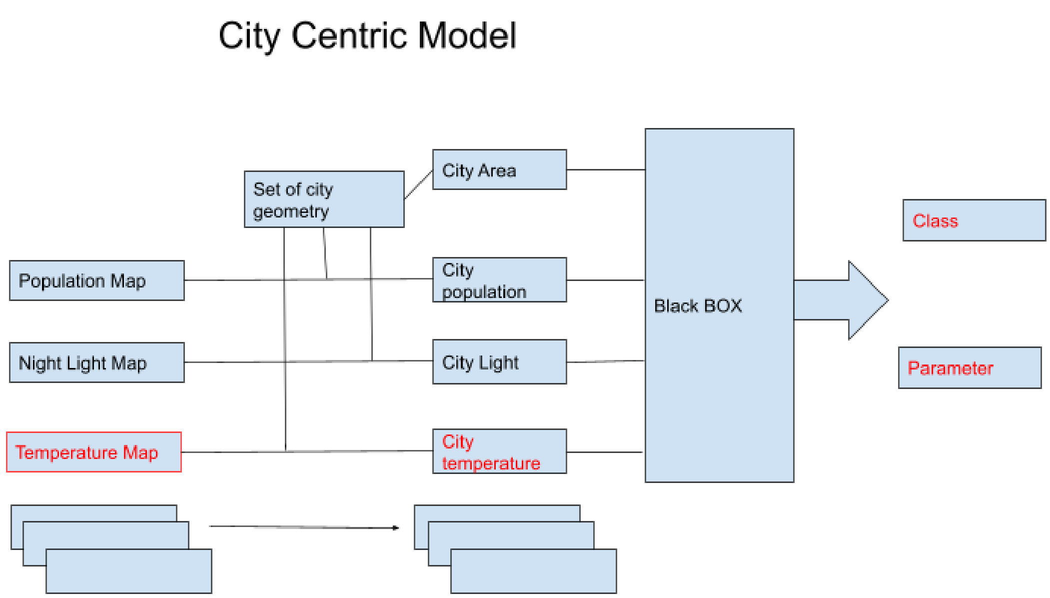

Figure 4-1: pipeline for city data approximation

4.1 Country satellite data and country statistics

Find relationship between variables derived from satellite imagery. temperature, built area, night light, population and the data from UN, world bank on other variables for which a satellite can not measure directly like GDP, Emissions, HDI index, and resources usage. use that data to calibrate a model on cities or countries with missing data and then downscale country data partially using insights from country level analysis

Crude GDP estimation from the flowing figure of correlation matrix can include population size, which people agglomerate more there they are opportunities or economic activities hence high population in a region might indicate total output or total GDP. Whereas from correlation matrix, the population density [on national scale] does not affect GDP per capita or total GDP. For GDP per capita it is correlated with night light per person, total area(interesting)

Given that total GDP of country can be estimated to some extent from nightlight.

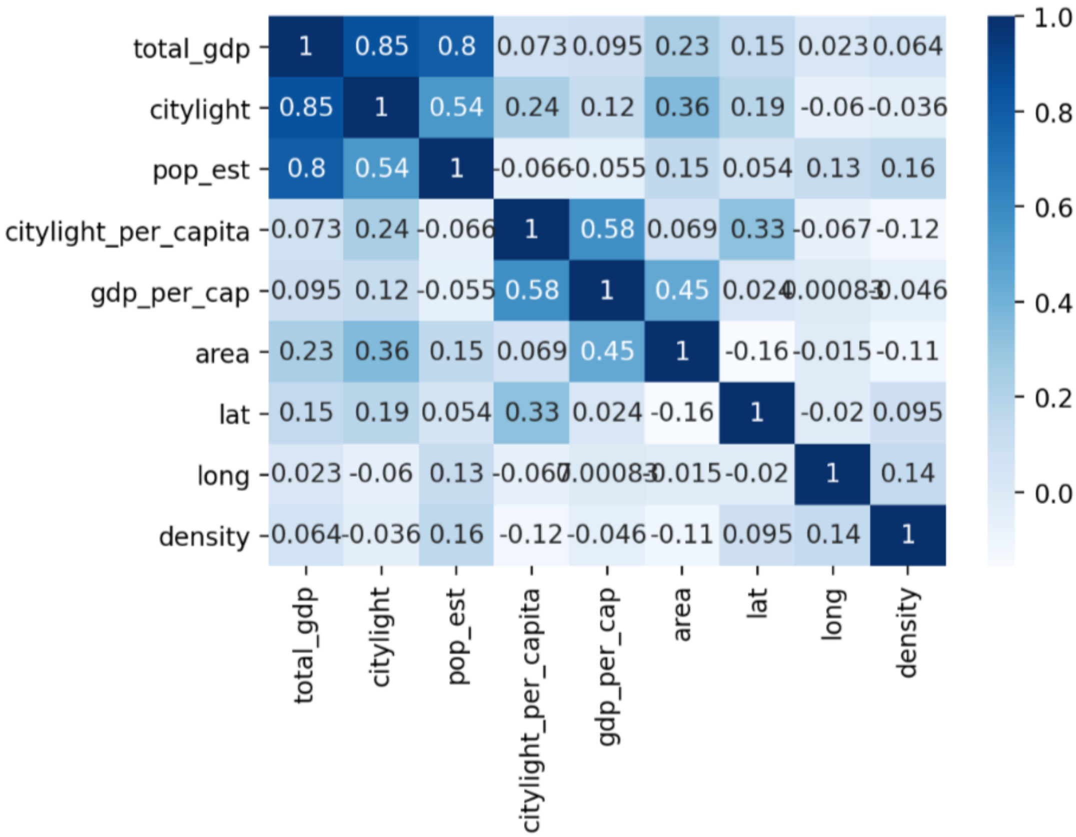

Figure 4-2: Country Correlations Focusing on GDP, population, city_night_light, area, latitude, longitude

4.2 City satellite data and stats

City values are then usable to model learnt on country level. From country level, 4-3, GDP is highly correlated with population, and city light. That is learnt in country and used to predict in city. for city City light

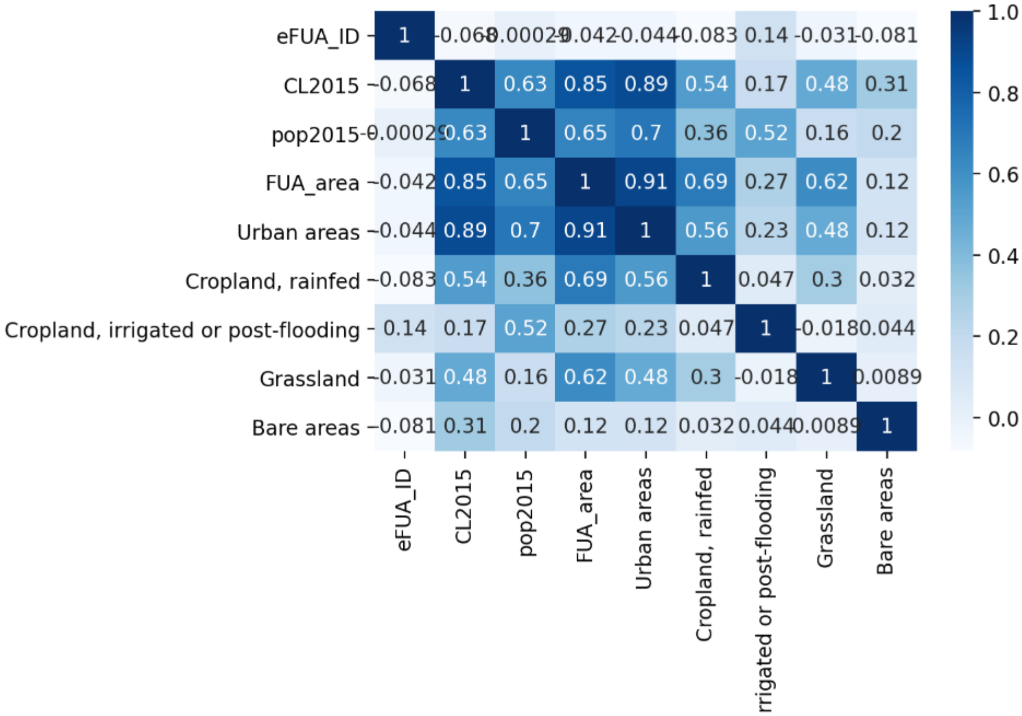

Figure 4-3: city Correlations Matrix with Land Usage, population, city_night_light

4.3 UN city database and night light correlations

This correlations show that tabular data can help from merging UN cities database with data from satellites like citylight. Data is merged using city name and 97 cities were found in both database by merging with same name 97 of 100 cities

Figure 4-4: city Correlations Matrix with Emissions,Land Area,population,GDP,Total Energy usage

Chapter 5

Conclusion

We motivated the need for global scale data and have shown how satellites can help in making data at city level available.

We have shown how to get data of cities from global satellites using image processing and how that can help in getting global scale data for other non-satellite data. For example, data that might be correlated with economic data and environmental data like emissions as in 4-4. We demonstrated that the data from satellites can be found for more cities and that its is correlated or can be used to get derive approximates of other variables that can not be measured directly by satellites like GDP, but can it can be found from signals that satellites can pick from the location related to it. We showed that data from satellites can be merged with data from other sources given they share identifiers.

Appendix A

Tables

Following table show samples of satellite related data we have.

Table A.1: Sample of population(in **millions**) of cities 1990,2000,2015 for the 60 most populous in 2015

	eFUA_ID	eFUA_name	Cntry_name	FUA_area	pop1990	pop2000	pop2015
1	5129.0	Tokyo	Japan	11971.0	30.4	33.2	36.5
2	7466.0	Delhi [New Delhi]	India	5569.0	16.8	21.6	30.1
3	4897.0	Jakarta	Indonesia	5292.0	15.9	20.8	29.8
4	10097.0	Shanghai	China	5878.0	11.5	16.7	26.9
5	2845.0	Quezon City [Manila]	Philippines	4873.0	13.8	18.9	25.0
6	94.0	Seoul	SouthKorea	7053.0	18.2	20.6	24.3
7	3981.0	Cairo	Egypt	4348.0	14.9	17.9	23.5
8	10452.0	Kolkata	India	4586.0	18.3	20.7	23.1
9	7163.0	Mumbai	India	2367.0	16.2	18.7	22.3
10	6567.0	São Paulo	Brazil	6077.0	16.3	18.9	21.7
11	5006.0	Mexico City	Mexico	4831.0	17.8	19.6	21.4
12	7704.0	Beijing	China	5536.0	8.5	12.7	21.3
13	3705.0	Dhaka	Bangladesh	3268.0	8.6	12.8	20.4
14	7244.0	New York	UnitedStates	17489.0	18.2	18.9	19.5
15	3956.0	Osaka [Kyoto]	Japan	10090.0	17.1	17.6	17.6
16	9711.0	Guangzhou	China	4413.0	8.6	11.7	16.7
17	2262.0	Moscow	Russia	8459.0	11.9	13.6	16.4
18	1894.0	Bangkok	Thailand	5161.0	6.6	9.9	16.3
19	1642.0	Los Angeles	UnitedStates	10407.0	12.4	14.2	15.7
20	4215.0	Buenos Aires	Argentina	5376.0	10.8	12.4	15.0
21	1637.0	Istanbul	Turkey	2540.0	8.3	10.6	14.8
22	2034.0	Karachi	Pakistan	1348.0	7.9	10.0	13.4
23	4686.0	Tehran	Iran	3332.0	20.7	12.9	13.4
24	5038.0	Ho Chi Minh City	Vietnam	3036.0	5.0	7.6	12.8
25	10153.0	Jieyang	China	5922.0	9.1	10.8	12.7
26	5288.0	London	UnitedKingdom	6605.0	10.0	10.8	12.6
27	865.0	Lagos	Nigeria	2284.0	6.3	8.3	12.3
28	9099.0	Bengaluru	India	2054.0	4.7	7.1	11.9
29	6458.0	Lahore	Pakistan	3154.0	6.5	8.6	11.8
30	5723.0	Chengdu	China	4034.0	6.6	8.6	11.8
31	9963.0	Suzhou	China	5056.0	6.7	8.6	11.4
32	2702.0	Paris	France	6778.0	9.8	10.3	11.2
33	6961.0	Rio de Janeiro	Brazil	3598.0	8.6	9.8	10.8
34	6401.0	Surabaya	Indonesia	4278.0	8.6	9.6	10.7
35	9850.0	Chennai	India	2070.0	6.7	8.2	10.6
36	2327.0	Lima	Peru	2121.0	5.8	7.3	9.7
37	4532.0	Nagoya	Japan	7350.0	8.8	9.3	9.6
38	10078.0	Hangzhou	China	4634.0	5.6	7.2	9.6
39	1141.0	New Taipei [Taipei]	Taiwan	2463.0	7.3	8.3	9.5
40	5107.0	Bandung	Indonesia	2312.0	6.1	7.4	9.3
41	4048.0	Bogota	Colombia	1863.0	5.5	7.0	9.1
42	9045.0	Hyderabad	India	1904.0	5.6	6.8	8.9
43	4316.0	Comilla	Bangladesh	4493.0	6.0	7.4	8.8
44	6654.0	Chicago	UnitedStates	13185.0	8.1	8.5	8.8
45	9725.0	Guangzhou	China	1573.0	3.7	5.3	8.7
46	8989.0	Wuhan	China	2682.0	5.6	6.8	8.5
47	2207.0	Hanoi	Vietnam	3293.0	5.7	6.8	8.3
48	8055.0	Tianjin	China	2413.0	4.1	5.7	8.3
49	3337.0	Johannesburg	SouthAfrica	3795.0	3.8	5.5	8.2
50	1322.0	Kuala Lumpur	Malaysia	3314.0	3.8	5.3	7.5
51	5918.0	Yogyakarta	Indonesia	4298.0	6.5	7.1	7.5
52	506.0	Luanda	Angola	1941.0	0.1	0.5	7.3
53	7617.0	Pune	India	1558.0	3.8	5.1	7.3
54	5991.0	Ahmedabad	India	1133.0	4.4	5.6	7.3
55	2626.0	Toronto	Canada	8223.0	5.0	5.8	7.3
56	2072.0	Santiago	Chile	2923.0	5.3	5.9	7.1
57	5543.0	Dallas	UnitedStates	19826.0	4.5	5.2	7.1
58	9701.0	Nanjing	China	2419.0	3.5	4.8	6.9
59	9718.0	Guangzhou	China	2458.0	4.7	5.6	6.7
60	7094.0	Xi'an	China	2083.0	4.5	5.4	6.6

Table A.2: Sample of population(in **thousands**) of cities 1990,2000,2015 for the 60 least populous in 2015 among 9000 cities in the database

	eFUA_ID	eFUA_name	Cntry_name	FUA_area	pop1990	pop2000	pop2015
1	3008.0	Nawnghkio	Myanmar	7.0	32.2	45.8	49.1
2	8732.0	Kurara	India	2.0	69.2	53.3	49.8
3	6277.0	Mehrestan	Iran	6.0	25.5	38.5	50.1
4	4665.0	Chongdan	NorthKorea	5.0	57.5	53.9	50.2
5	2736.0	Amagoro B Central	Uganda	3.0	25.2	33.6	50.7
6	7316.0	Anguwan Ari	Nigeria	4.0	33.8	40.3	50.8
7	8042.0	UNNAMED	India	2.0	68.7	60.0	50.8
8	7587.0	UNNAMED	India	5.0	56.8	47.2	50.8
9	7902.0	Mingu	China	9.0	62.9	59.3	50.8
10	549.0	Bole	Ghana	5.0	31.9	40.3	50.8
11	597.0	Tulsipur	Nepal	7.0	22.1	43.1	50.9
12	1008.0	Huye	Rwanda	7.0	26.9	33.4	51.0
13	5956.0	Dargaz	Iran	8.0	33.6	49.7	51.0
14	8232.0	UNNAMED	India	7.0	10.8	34.2	51.1
15	371.0	Caranavi	Bolivia	7.0	35.5	42.8	51.2
16	5451.0	Kanyabayonga	DemocraticRepublicoftheCongo	3.0	47.7	39.9	51.2
17	1619.0	Cutervo	Peru	6.0	51.8	54.3	51.3
18	7982.0	Yawal	India	5.0	34.6	40.9	51.4
19	2127.0	Kyela	Tanzania	6.0	25.1	33.9	51.4
20	2823.0	UNNAMED	Ethiopia	4.0	9.3	35.0	51.5
21	2498.0	Sangmélima	Cameroon	12.0	29.9	40.0	51.5
22	3017.0	Onchon	NorthKorea	5.0	46.7	50.1	51.6
23	8145.0	Fuquan	China	9.0	72.5	66.4	51.6
24	3083.0	Chitembo	Angola	6.0	65.1	72.3	51.6
25	1717.0	Darou Mousti	Senegal	12.0	69.8	77.2	51.6
26	7629.0	Uttarkashi	India	9.0	43.5	49.2	51.8
27	10176.0	Nirmali	India	5.0	26.2	34.7	51.8
28	4934.0	Sinijhoro	Pakistan	2.0	48.6	46.1	51.9
29	729.0	Kaédi	Mauritania	7.0	23.7	34.2	52.0
30	4803.0	Rahŭng	NorthKorea	14.0	46.4	50.8	52.1
31	4959.0	Totalai	Pakistan	8.0	47.4	45.9	52.1
32	7990.0	UNNAMED	India	4.0	32.5	40.2	52.1
33	8537.0	Huvina Hadagali	India	7.0	36.2	43.2	52.1
34	8006.0	UNNAMED	India	7.0	30.9	39.2	52.1
35	3193.0	Yakoma	DemocraticRepublicoftheCongo	9.0	27.9	40.4	52.2
36	10184.0	Chipurupalle	India	9.0	52.8	57.3	52.3
37	1833.0	Ouèssè	Benin	8.0	25.3	34.2	52.3
38	4787.0	Pamplona	Colombia	9.0	47.5	52.0	52.4
39	1585.0	El Empalme	Ecuador	14.0	33.6	41.5	52.4
40	1788.0	UNNAMED	Myanmar	11.0	36.0	48.8	52.4
41	414.0	Dwarka	India	14.0	46.9	47.0	52.5
42	8912.0	Makhtal	India	10.0	40.6	46.5	52.5
43	3587.0	Wulensi	Ghana	2.0	26.1	31.6	52.5
44	2920.0	UNNAMED	Zambia	6.0	138.7	98.6	52.6
45	1386.0	Owando	RepublicofCongo	10.0	9.2	28.8	52.6
46	6877.0	Anju Town	China	14.0	56.2	56.1	52.6
47	9363.0	Jiahe	China	12.0	43.5	51.8	52.7
48	891.0	Akon	SouthSudan	3.0	102.7	85.5	52.7
49	4920.0	Saravena	Colombia	11.0	24.6	35.4	52.7
50	9534.0	Huzurnagar	India	11.0	46.0	47.4	52.8
51	7825.0	UNNAMED	India	2.0	30.3	54.1	52.8
52	3389.0	Breves	Brazil	13.0	35.3	43.0	52.9
53	1237.0	Daru	SierraLeone	9.0	31.0	33.7	52.9
54	8178.0	Yellapur	India	15.0	39.8	53.0	52.9
55	7907.0	Sahaspur	India	4.0	37.9	51.5	53.0
56	12.0	Cobli	Benin	3.0	33.9	44.2	53.0
57	2808.0	Cinkassé	BurkinaFaso	5.0	23.7	33.2	53.0
58	4824.0	Thanatpin	Myanmar	10.0	35.9	57.2	53.0
59	1770.0	Ejura	Ghana	14.0	47.3	50.8	53.1
60	1097.0	Andoany	Madagascar	5.0	7.0	16.6	53.1

Table A.3: Sample of night light intensities of cities 2012,2015,2020 for the 60 most bright at night as view from sky in 2015

	eFUA_ID	eFUA_name	Cntry_name	FUA_area	CL2012	CL2015	CL2020
1	6654.0	Chicago	UnitedStates	13185.0	691846.1	1454388.9	1288821.8
2	7244.0	New York	UnitedStates	17489.0	560460.4	1326884.0	1201834.4
3	1642.0	Los Angeles	UnitedStates	10407.0	1193105.6	1308393.0	1336983.4
4	2262.0	Moscow	Russia	8459.0	524714.8	1231479.6	1395706.0
5	5511.0	Houston	UnitedStates	15114.0	1062434.9	1114128.2	1197741.0
6	5543.0	Dallas	UnitedStates	19826.0	1059535.1	1055885.9	1195434.0
7	296.0	Rotterdam [The Hague]	Netherlands	2943.0	381281.2	1034745.2	903204.8
8	4215.0	Buenos Aires	Argentina	5376.0	799542.4	951705.2	857490.9
9	3327.0	Riyadh	SaudiArabia	3239.0	701998.4	894906.5	925972.3
10	2626.0	Toronto	Canada	8223.0	209245.7	728933.7	787215.8
11	6451.0	Atlanta	UnitedStates	12348.0	710351.5	728675.1	692690.1
12	5129.0	Tokyo	Japan	11971.0	548772.8	716287.2	695306.3
13	6442.0	Miami	UnitedStates	6494.0	597970.0	695216.1	740240.6
14	750.0	Saint Petersburg	Russia	2764.0	339931.8	671115.6	563930.1
15	7165.0	Philadelphia	UnitedStates	11525.0	276083.7	654815.6	593529.4
16	94.0	Seoul	SouthKorea	7053.0	557014.7	644933.2	761669.1
17	2702.0	Paris	France	6778.0	128127.0	640633.1	566155.4
18	3981.0	Cairo	Egypt	4348.0	656296.9	636930.2	616244.8
19	7068.0	Washington D.C.	UnitedStates	7446.0	222203.3	615047.9	545722.2
20	6567.0	São Paulo	Brazil	6077.0	590541.1	606855.8	586862.9
21	6948.0	Detroit	UnitedStates	8184.0	282293.6	599506.1	629174.4
22	3485.0	Montreal	Canada	5027.0	191952.4	593822.6	573552.6
23	4211.0	Phoenix	UnitedStates	9707.0	555052.5	573115.6	604271.9
24	5288.0	London	UnitedKingdom	6605.0	144858.3	553904.4	486457.9
25	10097.0	Shanghai	China	5878.0	489561.9	552448.8	614694.8
26	5006.0	Mexico City	Mexico	4831.0	468765.8	490080.4	547618.8
27	6433.0	Minneapolis [Saint Paul]	UnitedStates	9833.0	299354.0	445687.2	549800.8
28	6192.0	St. Louis	UnitedStates	6517.0	314592.6	444949.0	393387.9
29	7466.0	Delhi [New Delhi]	India	5569.0	313071.2	429496.5	407924.1
30	357.0	Dubai	UnitedArabEmirates	2177.0	351147.1	425849.6	506955.5
31	3147.0	Madrid	Spain	4270.0	253868.4	419890.5	378417.4
32	5332.0	Denver	UnitedStates	6917.0	183106.1	410082.7	426068.8
33	1637.0	Istanbul	Turkey	2540.0	187417.1	399612.0	482269.0
34	429.0	Kuwait City	Kuwait	1147.0	316379.2	399120.1	489511.3
35	4143.0	Las Vegas	UnitedStates	1623.0	323069.2	392557.1	415767.5
36	5961.0	Kansas City	UnitedStates	7540.0	216353.1	391642.5	387626.6
37	146.0	Doha	Qatar	1100.0	252264.4	382648.5	556671.1
38	4686.0	Tehran	Iran	3332.0	247099.7	371912.9	423922.3
39	7704.0	Beijing	China	5536.0	161208.2	369183.8	462417.8
40	6961.0	Rio de Janeiro	Brazil	3598.0	395698.6	368383.9	292080.8
41	9963.0	Suzhou	China	5056.0	337132.2	368188.5	448078.3
42	3956.0	Osaka [Kyoto]	Japan	10090.0	294001.8	367546.4	401982.9
43	6990.0	Cleveland	UnitedStates	4770.0	161736.1	362631.6	310835.2
44	1600.0	Milan	Italy	3393.0	112445.9	360886.2	368219.6
45	4510.0	Seattle	UnitedStates	8328.0	78183.5	358347.2	336200.2
46	3691.0	Dammam	SaudiArabia	1198.0	276826.7	350109.5	350703.5
47	1572.0	Edmonton	Canada	2816.0	151860.0	346196.2	487101.3
48	1234.0	Jeddah	SaudiArabia	1517.0	287569.2	338384.0	361123.9
49	1894.0	Bangkok	Thailand	5161.0	1991.6	337089.2	354809.5
50	6487.0	Orlando	UnitedStates	5531.0	312162.1	333751.9	369005.4
51	6887.0	Columbus	UnitedStates	8211.0	172180.5	329345.0	290472.6
52	3337.0	Johannesburg	SouthAfrica	3795.0	313999.1	328048.4	319083.9
53	6720.0	Indianapolis	UnitedStates	7597.0	186236.6	325124.6	326888.7
54	6237.0	Tampa	UnitedStates	4168.0	300770.5	320653.5	324904.5
55	1255.0	San Jose	UnitedStates	4038.0	211809.5	307738.5	288297.3
56	1322.0	Kuala Lumpur	Malaysia	3314.0	106178.8	301771.9	342951.0
57	6851.0	Charlotte	UnitedStates	5134.0	227597.1	297881.0	276850.6
58	4829.0	Kazan	Russia	1113.0	75501.5	296110.4	328100.5
59	7123.0	Baltimore	UnitedStates	4070.0	117130.8	289617.0	272870.3
60	4532.0	Nagoya	Japan	7350.0	229783.7	278887.1	281610.9

Table A.4: Sample of 60 cities with the highest mean m2 temperature(in **Celsius**) in 2015 [in 9000 cities database]. m2 temperatures for 1990,200,2015 are shown

	eFUA_ID	eFUA_name	Cntry_name	m2_1990_average	m2_2000_average	m2_2015_average
1	4507.0	El Hawata	Sudan	30.6	29.9	31.7
2	4342.0	Al Fao	Sudan	30.8	30.2	31.6
3	4375.0	Al Quwaysi	Sudan	30.4	30.0	31.6
4	6702.0	Asaita	Ethiopia	31.0	30.5	31.6
5	4275.0	El Suki	Sudan	30.3	29.9	31.5
6	4208.0	Sennar	Sudan	30.4	29.9	31.5
7	4242.0	UNNAMED	Sudan	30.4	29.9	31.5
8	4308.0	Singa	Sudan	30.2	29.8	31.5
9	4070.0	Wad Madani	Sudan	30.8	30.1	31.5
10	4104.0	UNNAMED	Sudan	30.5	29.9	31.5
11	3965.0	Al Hasahisa	Sudan	30.8	30.1	31.4
12	4573.0	New Halfa	Sudan	30.8	30.0	31.4
13	3891.0	Al Uk	Sudan	30.8	30.1	31.4
14	3853.0	Al Hilaliyah	Sudan	30.8	30.1	31.3
15	3692.0	UNNAMED	Sudan	30.2	29.6	31.3
16	3648.0	Al Managil	Sudan	30.7	30.0	31.3
17	4606.0	Khashm El Girba	Sudan	30.7	29.8	31.3
18	3777.0	Al Kamilin	Sudan	30.7	30.1	31.2
19	3738.0	Wad Rawah	Sudan	30.6	30.0	31.1
20	3605.0	Rabak	Sudan	30.1	29.5	31.0
21	3517.0	Giad Industrial Complex	Sudan	30.5	29.9	31.0
22	729.0	Kaédi	Mauritania	30.9	30.0	30.8
23	3562.0	Kosti	Sudan	29.9	29.3	30.8
24	4540.0	Al-Qadarif	Sudan	29.9	29.0	30.7
25	3422.0	Jebel Aulia	Sudan	30.1	29.5	30.6
26	4667.0	Kassala	Sudan	30.0	29.2	30.6
27	4637.0	UNNAMED	Sudan	29.7	28.9	30.6
28	4174.0	Atbara	Sudan	29.8	29.8	30.6
29	2120.0	Renk	SouthSudan	29.7	28.9	30.6
30	4140.0	ad-Damer	Sudan	29.8	29.7	30.6
31	4408.0	Ad-Damazin	Sudan	29.3	28.5	30.5
32	350.0	Hodeidah	Yemen	30.0	30.0	30.5
33	861.0	Kiffa	Mauritania	30.2	29.8	30.4
34	3815.0	Shendi	Sudan	29.7	29.5	30.4
35	4000.0	UNNAMED	Sudan	29.7	29.5	30.4
36	113.0	Kayes	Mali	30.7	29.9	30.4
37	51.0	Teseney	Eritrea	29.8	28.9	30.4
38	4441.0	Er Roseires	Sudan	29.2	28.4	30.4
39	3226.0	Khartoum	Sudan	29.8	29.3	30.4
40	3019.0	Abu Arish	SaudiArabia	29.5	29.4	30.2
41	2851.0	Sabya	SaudiArabia	29.4	29.6	30.2
42	3375.0	Ed Dueim	Sudan	29.8	29.1	30.1
43	2963.0	Damad	SaudiArabia	29.3	29.2	30.1
44	3071.0	Samitah	SaudiArabia	29.3	29.4	30.0
45	2041.0	Al Lith	SaudiArabia	29.2	29.4	30.0
46	1871.0	Gao	Mali	30.4	30.0	30.0
47	2116.0	Qunfudhah	SaudiArabia	29.0	29.5	29.9
48	6744.0	Dolo	Ethiopia	29.4	29.7	29.8
49	3277.0	Ghaliya	Sudan	29.2	28.8	29.8
50	589.0	Podor	Mauritania	29.7	29.0	29.8
51	891.0	Akon	SouthSudan	29.0	28.0	29.7
52	501.0	Abs	Yemen	29.1	28.8	29.7
53	188.0	Harad	Yemen	29.1	29.1	29.7
54	2537.0	Podor	Senegal	29.6	28.9	29.7
55	2853.0	Tambacounda	Senegal	29.4	29.3	29.7
56	3175.0	Tandalti	Sudan	29.0	28.3	29.7
57	1720.0	Malakal	SouthSudan	28.5	28.0	29.7
58	293.0	Tillabéri	Niger	29.9	29.4	29.7
59	759.0	Malualkon	SouthSudan	28.9	27.9	29.6
60	2605.0	Umm Hajar	Chad	29.7	28.6	29.6

Table A.5: Sample of 60 cities with the lowest mean m2 temperature(in **Celsius**) in 2015 [in 9000 cities database]. m2 temperatures for 1990,200,2015 are shown

	eFUA_ID	eFUA_name	Cntry_name	m2_1990_average	m2_2000_average	m2_2015_average
1	6118.0	Norilsk	Russia	-10.1	-10.9	-7.7
2	6441.0	Yakutsk	Russia	-6.8	-8.9	-6.6
3	6468.0	Magadan	Russia	-5.0	-5.7	-4.8
4	6058.0	Novy Urengoy	Russia	-7.1	-7.7	-4.6
5	1812.0	Khorugh	Tajikistan	-2.0	-3.9	-2.4
6	2983.0	Nagqu	China	-5.2	-5.2	-2.2
7	6154.0	Nizhnevartovsk	Russia	-1.7	-2.3	-0.3
8	5876.0	Yugorsk	Russia	-0.4	0.2	0.4
9	6450.0	Chita	Russia	-0.6	-1.5	0.4
10	117.0	Ulaanbaatar	Mongolia	-0.7	-1.3	0.5
11	4785.0	Yakeshi/Yaysi	China	0.1	-0.9	0.8
12	6432.0	Ulan-Ude	Russia	0.3	-0.7	0.9
13	5267.0	Jiagedaqi	China	0.7	-0.4	1.1
14	6486.0	Komsomolsk-on-Amur	Russia	1.4	-0.5	1.2
15	6395.0	Bratsk	Russia	0.3	-2.0	1.3
16	4583.0	Hulunbuir	China	0.6	-0.5	1.4
17	6549.0	Petropavlovsk-Kamchatsky	Russia	1.7	0.3	1.4
18	316.0	Murmansk	Russia	0.4	1.1	1.5
19	5840.0	Serov	Russia	1.1	1.4	1.6
20	4580.0	Ushuaia	Argentina	1.8	1.7	1.6
21	4046.0	Hezuo	China	1.1	0.6	2.0
22	5996.0	Zlatoust	Russia	1.9	1.7	2.0
23	5790.0	Lysva	Russia	2.2	2.2	2.2
24	3486.0	Manzhouli	China	1.0	0.0	2.2
25	5900.0	Nizhny Tagil	Russia	2.1	1.9	2.2
26	5573.0	Berezniki	Russia	2.4	2.3	2.2
27	6385.0	Kyzyl	Russia	1.5	0.8	2.2
28	5960.0	Novouralsk	Russia	2.3	2.1	2.3
29	4806.0	Anchorage	UnitedStates	-0.5	1.5	2.3
30	6423.0	Irkutsk	Russia	1.9	0.5	2.4
31	6024.0	Dayangshu	China	1.9	0.5	2.4
32	6405.0	Angarsk	Russia	2.1	0.7	2.4
33	6477.0	Birobidzhan	Russia	2.7	1.3	2.6
34	6247.0	Anzhero-Sudzhensk	Russia	1.5	0.1	2.6
35	6301.0	Achinsk	Russia	1.6	-0.1	2.7
36	5972.0	Yekaterinburg	Russia	2.7	2.4	2.7
37	6376.0	Kansk	Russia	1.6	-0.7	2.7
38	6495.0	Khabarovsk	Russia	3.0	1.6	2.7
39	6331.0	Mezhdurechensk	Russia	1.9	1.4	2.7
40	5702.0	Gilgit	Pakistan	3.5	3.1	2.7
41	6142.0	Ishim	Russia	3.0	2.0	2.7
42	6757.0	Nenjiang	China	2.4	0.9	2.7
43	8430.0	Nancha	China	2.8	1.1	2.7
44	4241.0	Syktyvkar	Russia	1.9	2.3	2.8
45	6179.0	Tomsk	Russia	1.5	0.2	2.8
46	3466.0	UNNAMED	NorthKorea	3.0	1.6	2.8
47	6340.0	Krasnoyarsk	Russia	1.8	-0.1	2.8
48	1575.0	Altay	China	2.1	1.3	2.9
49	6008.0	Miass	Russia	2.7	2.4	2.9
50	5718.0	Sylva	Russia	2.9	2.9	2.9
51	6106.0	Tyumen	Russia	3.1	2.3	3.0
52	6980.0	Blagoveshchensk	China	2.3	0.8	3.0
53	6459.0	Blagoveshchensk	Russia	2.4	0.8	3.1
54	3344.0	Huangjiazhai	China	2.3	2.8	3.1
55	6167.0	Omsk	Russia	3.6	2.4	3.1
56	6236.0	Yurga	Russia	2.0	0.6	3.1
57	7668.0	Bei ian	China	3.0	1.6	3.2
58	5654.0	Perm	Russia	3.1	3.1	3.2
59	6258.0	Kemerovo	Russia	2.1	0.8	3.2
60	5622.0	Kultaevo	Russia	3.1	3.1	3.2

Table A.6: Sample of 60 cities with the highest "urban areas" area in 2015 [in 9000 cities database]. Column CIPF: Cropland, irrigated or post-flooding

	eFUA_ID	eFUA_name	Cntry_name	Cropland, rainfed	CIPF	Grassland	Urban areas	Bare areas
1	7244.0	New York	UnitedStates	4928	0	18207	74141	731
2	1642.0	Los Angeles	UnitedStates	2880	0	10687	71621	228
3	5129.0	Tokyo	Japan	16892	24531	11	63258	6
4	6654.0	Chicago	UnitedStates	70520	0	11548	63097	323
5	5543.0	Dallas	UnitedStates	30508	0	150163	47746	100
6	5511.0	Houston	UnitedStates	19857	0	71072	46382	453
7	7704.0	Beijing	China	26104	2047	2751	35628	5
8	5288.0	London	UnitedKingdom	29778	0	26960	35008	285
9	6451.0	Atlanta	UnitedStates	1545	0	40578	34866	106
10	6442.0	Miami	UnitedStates	16763	0	906	34635	167
11	6948.0	Detroit	UnitedStates	22912	0	12200	33757	403
12	4532.0	Nagoya	Japan	10972	10592	25	33585	20
13	3956.0	Osaka [Kyoto]	Japan	16178	5289	29	33340	42
14	4897.0	Jakarta	Indonesia	8387	120	0	32735	0
15	2262.0	Moscow	Russia	39566	0	2150	32661	0
16	7165.0	Philadelphia	UnitedStates	27810	0	23328	29945	382
17	4510.0	Seattle	UnitedStates	5227	0	1064	29691	270
18	2702.0	Paris	France	52477	0	934	27990	4
19	4211.0	Phoenix	UnitedStates	14971	0	973	27772	47
20	10097.0	Shanghai	China	2489	36998	929	27442	14
21	2626.0	Toronto	Canada	64923	0	563	27414	0
22	536.0	Dortmund	Germany	41672	0	7088	25925	1
23	1266.0	Melbourne	Australia	4255	214	18100	25221	30
24	7068.0	Washington D.C.	UnitedStates	3308	0	13455	24698	281
25	4215.0	Buenos Aires	Argentina	12974	0	2187	23653	3
26	10153.0	Jieyang	China	8621	10443	128	22590	0
27	1255.0	San Jose	UnitedStates	1351	0	5820	22455	27
28	6433.0	Minneapolis [Saint Paul]	UnitedStates	47064	0	40260	22396	228
29	6567.0	São Paulo	Brazil	2896	0	140	22150	8
30	6237.0	Tampa	UnitedStates	2219	0	5850	20842	44
31	9711.0	Guangzhou	China	5274	12219	1173	19328	5
32	5006.0	Mexico City	Mexico	21495	0	814	18920	127
33	3337.0	Johannesburg	SouthAfrica	2007	272	13586	18183	106
34	6720.0	Indianapolis	UnitedStates	69034	0	5165	18068	6
35	1830.0	Sydney	Australia	29	570	0	17970	4
36	6487.0	Orlando	UnitedStates	3992	0	16127	17952	150
37	94.0	Seoul	SouthKorea	34946	0	1172	17910	2161
38	2204.0	Tijuana	UnitedStates	1192	0	4351	17637	49
39	1600.0	Milan	Italy	18232	4068	45	17460	907
40	6192.0	St. Louis	UnitedStates	29983	0	9456	17161	146
41	3485.0	Montreal	Canada	22293	0	56	17158	8
42	9963.0	Suzhou	China	3900	29067	2011	16826	3
43	5961.0	Kansas City	UnitedStates	21822	0	46929	16617	23
44	4946.0	Berlin	Germany	17670	0	3547	16555	4
45	7466.0	Delhi [New Delhi]	India	7990	39105	155	16462	0
46	1894.0	Bangkok	Thailand	11926	15746	1125	16141	2
47	5332.0	Denver	UnitedStates	24739	0	13912	14555	82
48	7312.0	Providence	UnitedStates	194	0	2106	14435	85
49	7347.0	Boston	UnitedStates	116	0	855	14373	45
50	4444.0	Portland	UnitedStates	22283	0	6125	14227	52
51	10078.0	Hangzhou	China	4470	22423	972	14091	2
52	6990.0	Cleveland	UnitedStates	5228	0	12299	13985	63
53	4333.0	Naples	Italy	5199	1027	686	13654	1379
54	6961.0	Rio de Janeiro	Brazil	322	0	264	13448	1
55	1637.0	Istanbul	Turkey	5970	351	462	13082	973
56	2355.0	Brisbane	Australia	44	270	228	13053	0
57	750.0	Saint Petersburg	Russia	17508	0	20	12992	123
58	360.0	Perth	Australia	1786	307	905	12950	24
59	6887.0	Columbus	UnitedStates	78950	0	9852	12898	113
60	2663.0	Katowice	Poland	26598	0	3962	12819	0

Bibliography

[1] Manuel Aristarán, Mike Tigas, and Jeremy B. Merrill. Tabula: Extract tables from pdfs, 2021. Retrieved 31 July 2021, from https://tabula.technology/.

[2] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec 2011.

[3] Artessa Niccola D.Saldivar-Sali. A global typology of cities: Classification tree analysis of urban resource consumption, 2010.

[4] Sean Gillies et al. Shapely: manipulation and analysis of geometric objects, 2007–.

[5] Sean Gillies et al. Rasterio: geospatial raster i/o for Python programmers, 2013–.

[6] Miguel Grinberg. *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2018.

[7] Mila. Hoornweg, Daniel; Freire. *Building Sustainability in an Urbanizing World : A Data Compendium for the World's 100 Largest Urban Areas. Urban Development Series Knowledge Papers*. World Bank, Washington, DC. © World Bank., 2013. License: CC BY 3.0 IGO.

[8] K Jordahl. Geopandas: Python tools for geographic data. *URL: https://github.com/geopandas/geopandas*, 2014.

[9] KPMG. City typology as the basis for policy . `https://carbonn.org/fileadmin/user_upload/Resources/KPMG_Comparison_Study_Sust_Cities_web.pdf`, 2005. Accessed: 2021-03-20.

[10] L. Maffenini, M. Schiavina, S. Freire, M. Melchiorri, Pesaresi M., and T Kemper. Ghs-pop2g user guide. 2020. ISBN 978-92-76-22414-3, doi:10.2760/500887, JRC121485.

[11] Pranab K. Roy Chowdhury, Budhendra L. Bhaduri, and Jacob J. McKee. Estimating urban areas: New insights from very high-resolution human settlement data. *Remote Sensing Applications: Society and Environment*, 10:93–103, 2018.

[12] Marcello Schiavina, Ana Moreno-Monroy, Luca Maffenini, and Paolo Veneri. Ghs-fua r2019a - ghs functional urban areas, derived from ghs-ucdb r2019a (2015).

[13] Wikipedia. PDF Split and Merge — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=PDF%20Split%20and%20Merge&oldid=1028519467`, 2021. [Online; accessed 31-July-2021].

Alphabetical Index