INTERNATIONAL MONETARY FUND

Satellite-Based Census of Residential Buildings: Application for Climate Risk Assessment

Andinet Woldemichael and Iyke Maduako

WP/24/204

2024
SEP

WORKING PAPER

WP/24/204

IMF Working Paper
Statistics Department

Satellite-Based Census of Residential Buildings: Application for Climate Risk Assessment

Prepared by Andinet Woldemichael and Iyke Maduako *

Authorized for distribution by Marco Marini
September 2024

ABSTRACT:

Housing represents the largest asset and liability, in the form of mortgages, on most national balance sheet. For most households it is their largest investment, and when mortgages are required also represents the largest component of household debt. It is also directly tied to financial markets, both the mortgage market and insurance sector. Although many countries have a rich set of housing censuses and statistics, others have large data gap in this area and therefore struggle to formulate effective policies. This paper proposes an approach to construct a global census of residential buildings using opensource satellite data. Such a layer can be used to assess the extent these buildings are exposed to climate hazards and how their production and consumption, in turn, affect the climate. The approach we propose could be scaled globally, combining existing layers of building footprints, climate and socioeconomic data. It adds to the ongoing effort of compiling spatially explicit and granular climate indicators to better inform policies. As a case study, we compute selected indicators and estimate the extent of residential properties exposure to riverine flood risk for Kenya.

JEL Classification Numbers:	C89, C81, Q54, R33
Keywords:	Census of buildings; Climate Exposure; Building Footprint
Authors' email addresses:	awoldemichael@imf.org; imakduako@imf.org

* This study is part of the IMF Statistics Department's work to address data gaps related to climate change. The study benefited from a DataDive® event on leveraging opensource satellite imageries to construct a global-scale census of structures, organized in collaboration with DataKind®. The authors wish to thank the organizers and participants of the event. The authors would also like to thank Alessandra Alfieri, Seung Mo Choi, Marco Marini, Niall O'Hanlon, and Jim Tebrake (all IMF Statistics Department) for their guidance and helpful suggestions provided on the preliminary drafts of this paper. We also thank Max Yarmolinsky for his excellent support on econometrics server and Amira Archer-Davies for her assistance. Initial computation of residential properties exposure data for 48 African countries are available from the authors upon request. The authors will also make available the global-scale exposure data on Google Earth Engine.

WORKING PAPER

Satellite-Based Census of Residential Buildings: Application for Climate Risk Assessment

Andinet Woldemichael and Iyke Maduako

Table of Contents

I. Introduction

Housing census, along with population, is one of the foundations of statistical systems that is essential to formulate, implement and monitor policies. The United Nations define housing census as the operation that provides at regular intervals the official counting or benchmarking of all housing stock and their occupants in the territory of a country and in its smallest geographical subterritories.[1] Traditional housing census operation involves a complex process of collecting information from individuals and households across geographic areas simultaneously, making it one of the most complex and costly data collection activity.[2] As a result, most housing censuses are conducted decennially and countries with limited resources and capacity often have outdated censuses. In addition, most housing census are not spatially explicit limiting their application for emerging policy issues, such as climate risk assessment that require information on the geographic location of housing units. This paper examines the potential use of opensource satellite and building footprint data to compile spatially explicit housing (residential building) census and shows its application in climate risk assessment.[3] Using readily available aggregate capital stock data, the paper also introduces a simple technique of assigning monetary value to residential buildings.

The paper focuses on the housing sector because housing is a peculiar asset. It is both investment and a significant part of the consumption basket.[4] It also accounts for the largest share of national wealth making it one of the most important assets in many economies. It is closely linked to the health of the financial systems of countries as booms and busts in the housing markets are major sources of financial crisis across the world. Further, mortgage markets are key transmission channel for monetary policies (IMF, 2008; IMF Global Housing Watch). In addition to its importance in the financial system, housing provides essential social services to a population. It provides security, a place to gather, to entertain, and work. These essential services are key dimensions of economic wellbeing. In democratic systems, spatial distribution of population and housing serve as the basis for distribution of political representation and funding and often constitutionally mandated in advanced and democratic societies (Ericksen and Kadane, 1985).[5] However, up-to-date and spatially explicit housing censuses are lacking in many countries due to cost, capacity, and other factors posing significant challenge in performing timely policy formulation, implementation and monitoring on key policy issues such as climate change.

The housing sector is crucial to both climate mitigation and adaptation, consuming a significant amount of energy in the construction and use of residential buildings. The United Nations Environment Programme (UNEP) reports that in 2015, construction and building operations contributed to 38%, or 13.1 gigatons, of global energy-related CO2 emissions (United Nations Environment Programme, 2021). Moreover, residential properties form an essential part of the capital stock, which includes machinery, equipment, structures, and urban land. IMF data from 2019 estimated the total capital stock at $316,253 billion in international dollars.[6] The built-up area, mainly comprising residential zones, accounts for approximately 83% of the variation in total capital stock (see Figure 1 in Appendix A).[7] The Network for Greening the Financial Sector (NGFS) also

[1] Alternatively, housing census is defined as the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating statistical data pertaining, at a specified time, to all living quarters1 and occupants thereof in a country or in a well delimited part of a country (UN, 1980).
[2] United Nations Statistics Division - Demographic and Social Statistics
[3] "A building is any independent free-standing structure comprising one or more rooms or other spaces, covered by a roof and usually enclosed within external walls or dividing walls that extend from the foundations to the roof." (UN, 1980).
[4] Housing Markets, Financial Stability and the Economy (imf.org)
[5] Why Does the Census Matter? | Council on Foreign Relations (cfr.org).
[6] WhatsNewinIMFInvestmentandCapitalStockDatabase_May2021.pdf. See Figure A.1 in Appendix A.
[7] OECD Data: Built-up area and built-up area change in countries and regions (oecd.org)

recognizes housing or residential properties as Climate Policy Relevant Sectors (CPRS).[8] However, unlike data on machinery, equipment, and urban land area, detailed and timely spatial data on residential buildings are lacking for many countries.

At the UN Climate Change Conferences (COP27 and COP28), countries collectively agreed to increase climate finance, establish a "loss and damage fund" to aid vulnerable developing countries, and committing to transition away from fossil fuels.[9,10] Further, several countries are advancing towards a low-carbon future, enacting policy reforms, and implementing regulations to promote voluntary carbon trading mechanisms.[11] These actions, alongside various policy instruments, are facilitating the energy transition. Much of these efforts are intertwined with the physical structures that constitute our built environment—the buildings we live in and produce; the transportation networks (roads, railways, bridges, ports and airports) that we use to move people and goods; and the utility systems (dams, powerplants, and transmission lines) that we use to produce and supply essential services such as electricity, clean water, and ICT.[12] Residential buildings are primary component of the physical structures or built-environment. However, there are significant data gaps regarding these structures and their interactions with climate—a gap acknowledged by the G20 Data Gap Initiative phase III (DGI-3) recommendations for climate data and statistics.

From a policy perspective, detailed and global-scale data on residential property is crucial for quantifying financial and economic exposure, and vulnerability to climate hazards. It also facilitates policy analysis of transition risks and opportunities. Transition to low-carbon economy implies risks for some and opportunities for others arising from either policy and legal reforms on emissions requirements, innovation and disruptive technologies in energy use, market changes due to changes in demand and access to raw materials, or reputation due to negative consumer sentiment.[13] For instance, new regulations on residential energy use efficiency or retrofitting could require information on building energy use which in turn requires detailed information on location, size, construction materials, age, purpose of buildings, etc. However, such detailed building data have been lacking, particularly in Low Income and Developing Countries (LIDCs) and Emerging Markets (EMs) (Eberenz et al. 2020).

Traditionally, housing data are collected through administrative channels or using housing survey instruments that are time-consuming, costly, and in most cases prone to mismeasurement (Angrist et al., 2021). Although many countries collect such data for form alternative sources and for various reasons—taxation, permits, urban planning/zoning, safety, and regulations—complete housing information is often not available due to data being scattered across different agencies, access may be restricted, or it may not adhere to international standards. With some caveats, leveraging publicly available satellite data could be a cost-effective alternative for national statistical agencies to compile granular data on structures in general and residential buildings, in particular.

This paper presents a methodology to compile a census of residential buildings using open-source building footprint data and shows its application for climate risk analysis. It follows a census approach, which is suited to buildings due to their inherent characteristics: fixed location, construction year, size, shape, height, and

[8] Climate Policy Relevant Sectors | FINEXUS: Center for Financial Networks and Sustainability | UZH
[9] Sharm el-Sheikh Climate Change Conference - November 2022 | UNFCCC.
[10] UN Climate Change Conference - United Arab Emirates | UNFCCC.
[11] Examples include the EU Emissions Trading System (EU ETS).
[12] Basic Information about the Built Environment | US EPA
[13] "Transition risks are those associated with the pace and extent at which an organization manages and adapts to the internal and external pace of change to reduce greenhouse gas. Transitioning requires policy and legal, technology, and market changes to address mitigation and adaptation requirements related to climate change. Depending on the nature, speed, and focus of these changes, transition risks may pose varying levels of financial and reputational risk to organizations. Alternatively, if an organization is a low-carbon emitter and in the renewable energy or climate transition market, they could experience market, technological, and reputational opportunities." Climate Risks and Opportunities Defined | US EPA.

construction materials (including wall and roofing), as well as purpose. More importantly, major space agencies like the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) provide raw and processed satellite imagery to the public at no cost. Similarly, major technology companies provide detailed data about our built environment through applications on personal and business devices, such as navigation maps and business directories. As part of their "Data for Good" initiatives, commercial companies like Google and Microsoft have made such data publicly accessible.[14,15]

Our paper is related to similar studies in the literature which proposed approaches to construct a global model of asset exposure to climate risks. De Bono and Mora (2014) in the Global Assessment Report (2013) is among the early studies which developed a global exposure database utilizing open-source datasets on population and building typology with a spatial resolution of 5 km. This early work focused on compiling exposure data for urban building stock broadly and using low-resolution satellite imageries. Our study extends this research but introduces significant enhancements. Firstly, we employ detailed building footprint data from Google Open Buildings which utilizes Google's advanced Deep-Learning models and high-resolution (~10 meters) satellite imagery to detect and classify buildings. The work takes advantage of high-resolution Sentinel-2 imageries which has substantial improvement over the previous reliance on NASA's Landsat imagery, which, with its 30-meter resolution, misses most building structures. Secondly, we refine building attributes to include ground-level area, height, and classification of buildings as residential or non-residential. Recently, Eberenz et al. (2020) proposed a similar method for compiling global asset exposure data for physical risk assessment, albeit without employing a census approach for identifying building structures. Our research also relates to a recent methodology paper by Doan et al. (2023), which assesses the exposure of populations to extreme weather events and their vulnerability.

The paper contributes to data collection and methodological efforts of national statistical offices with limited resources to leverage open-source satellite and geospatial datasets for compilation of spatially explicit housing census data. It also contributes to the global effort to addressing gaps in climate data and statistics, a key obstacle in tracking the progress in low-carbon transitions and climate adaptation and mitigation policies.[16] It aligns with the Data Gap Initiative phase III (DGI-3) spearheaded by G20 leaders which tasked the IMF to coordinate efforts with the Financial Stability Board and statistical authorities on several data gap recommendations.

The rest of the paper is organized as follows. Section 2 discusses various data sources used to compile exposure layers and provide summary statistics. Section 3 discusses methods and proposed indicators. Section 4 discusses the analytical results and limitations of the proposed indicators, and section 5 concludes the paper.

II. Data and Description

We use processed satellite and geospatial datasets from different sources to compute residential properties exposure layers. These datasets include building footprints, estimated building height and classification,

[14] We use the term "building footprint" as a two-dimensional representation of building outline or polygon (see for example, Wang et al. 2016)).
[15] Google Open Buildings, Microsoft Building Footprint.
[16] Bridging Data Gaps Can Help Tackle the Climate Crisis (imf.org)

gridded population data, urbanization, nighttime light, and dwellings capital stock.[17] We describe each of these datasets in the following.

Google's Open Buildings

The primary dataset that we use to construct residential property layer is Google's Open Buildings dataset.[18] This dataset, developed by Google Open Building researcher team, leverages high-resolution (~50 centimeters) daytime satellite imagery and Deep Learning models for the identification and classification of buildings (Sirko et al., 2021). The coverage is for the Global South and publicly available under the Creative Commons Attribution (CC BY-4.0) license and the Open Data Commons Open Database License (ODbL) v1.0 license. The latest version (v3) of the dataset, released in May 2023, covers about 1.8 billion buildings across continental Africa, South Asia, South-East Asia, Latin America and the Caribbean. The data have information on the building polygon, latitude and longitude of the centroid, area in meters, and confidence levels for each detected building. One of the limitations of Google's building footprint dataset is that it does not have information on building heights and other attributes necessary to compute our proposed indicators of residential properties exposure. To overcome this limitation, we use building height, classification, and other key layers from the European Commission Joint Research Center (EC-JRC) Global Human Settlement Layer (GHSL), enriching detected buildings with more attributes.

Additional building attributes: Height and Classification

The building attributes were derived from the Global Human Settlement Layers (GHSL). The GHSL building height estimates are based on a method developed by Pesaresi et al (2021) and Pesaresi and Politis (2023). The methods utilize Digital Elevation Models (DEMs) sourced from the Advanced Land Observing Satellite (ALOS) Global Digital Surface Model (AW3D30) and the digital topographic database of Earth by the Shuttle Radar Topography Mission. Additionally, they incorporate shadow markers extracted from Sentinel-2 image composites for 2018 to approximate building heights. For this study, we use the average net building height (ANBH) estimates for 2018, which has a resolution of $100mx100m$.[19]

The other key dataset is functional classification of buildings into residential and non-residential. We use GHSL built-up characteristics layer (GHS-BUILT-C) that classifies structures into residential and non-residential at a resolution level of 10 meters. For our analysis, we use the building classification data from 2018 which aligns with the relevant period for our study. Pesaresi and Politis (2023) and European Commission, GHSL Data Package (2023) provide detailed technical descriptions on method of classification.

Gridded Population

To filter buildings within populated and non-populated grid cells and to downscale aggregate dwellings capital stock values to pixel level, we use Gridded Population of the World Version 4 (GPWv4) data from NASA Socioeconomic Data and Applications Center (SEDAC). A notable advantage of the GPWv4 dataset is calibration of estimated population densities to correspond with the United Nations World Population Prospects (UN WPP) country totals, ensuring global consistency and reliability (Center for International Earth Science

[17] Table A.1 presents description of these data sources. Figure A.4 shows the imageries for Nairobi area.
[18] Microsoft's Building Footprint is another opensource building data. It covers about 1.2B buildings across the world. Microsoft also has 174M building height estimates providing 3D polygons for some areas.
[19] Building height estimates are also available at a higher resolution level of 10mx10m. For the purpose of this study, we used the lower resolution layer of 100mx100m.

Information Network – CIESIN, 2018). The data comes at 30 arc-second resolution (~1 km at the equator) and covers the period from 2000 to 2020 in five-year intervals.[20] We use the 2020 global population estimates for our analysis.

Nighttime Light

In addition, we use nighttime light data, which is a good proxy measure of human and economic activities, to proportionally downscale country-level dwellings capital stock values. Specifically, we use the corrected Visible and Infrared Imaging Suite (VIIRS) annual average nighttime light imageries from the Earth Observation Group (EOG) at Colorado Mines. The VIIRS was designed and calibrated to detect electric light from earth's surface.[21] It is one of the most prominent sources of low-light imaging data used to measure human activities (Elvidge et al., 2021). Institutions such as the Earth Observation Group[22] at Colorado Mines and NASA Black Marble[23] process raw satellite images for extraneous artifacts and biases, such as sunlit, moonlit, and cloudy and other light contaminants and provide users ready-to use and high-quality VIIRS DNB composites. These images are cloud and lunar-BRDF-corrected to remove noise from extraneous artifacts and biases. The corrected monthly and annual VIIRS imageries are available since 2014 at a resolution of 15 arc second (around 500 meters at the Equator). For this study, we use annual nighttime light composite for 2020 that corresponds with the GPWv4 gridded population data.

Dwellings Capital Stock

Given the lack of subnational data on property values across many countries, we propose a method to downscale aggregate dwellings capital stock values using nighttime light and population data. We draw upon the Investment and Capital Stock Datasets (ICSD) compiled by the IMF Fiscal Affairs Department (FAD), which provide comprehensive information on public and private capital stock for approximately 170 countries from 1960 to 2019. [24] We specifically use the most recent private fixed capital stock data from the ICSD to estimate dwellings capital stock value at the square meter level.[25] One of the key advantages of using capital stock data is that it is readily available and compiled as part of investment data (De Bono and Mora, 2014). However, due to the lack of specific information on the dwelling component of fixed capital stock for many countries, we use the average share of dwelling in fixed capital formation from countries with available data. These averages are calculated by country income group—Low Income and Developing Countries (LIDCs), Emerging Markets (EMs), and Advanced Economies (AEs).[26] The corresponding average shares of dwellings for the years between 2015 and 2019 in real gross fixed capital stock for LIDCs, EMs, and AEs, respectively, are 0.748,

[20] To cross-examine our analysis, we also use gridded population data from GHSL which has 100m (~3 arcsec) and 1km (30 arcsec) resolutions with global coverage. The data are available for period between 1975 and 2020 in five-years interval and projections into 2025 and 2030 (Schiavina et al., 2023; Freire et al., 2016).
[21] Mounted on Suomi NPP, the VIIRS is an optical spectrum sensor capturing imagery at high spatial resolution ranging from 0.375 to 1.6 km, depending on band. It is equipped with low-light sensor, the Day/Night Band (DNB) and has a global coverage of 3,000-km-wide (Miller et al., 2012). The VIIRS imaging data have several improvements over the older version from DMSP-OLS. These improvements include in-flight calibration, finer resolution, low-light detection, and a much larger Day-Night Band with no saturation, making it one of the preferred nighttime light data sources to study human activities (Miller et al., 2012; Elvidge et al., 2013; Gibson et al., 2021).
[22] VIIRS Nighttime Light (mines.edu)
[23] VIIRS/NPP Lunar BRDF-Adjusted Nighttime Lights Monthly L3 Global 15 arc-second Linear Lat Lon Grid
[24] IMF Infrastructure Governance. InvestmentandCapitalStockDatabaseUserManualandFAQ_May2021.pdf (imf.org).
[25] Investment and Capital Stock - At a Glance - IMF Data
[26] Using data from the Federal Reserve Bank of St Louis (FRED).

0.362, and 0.171.[27] We downscale values to the pixel level based on population and nighttime lights, with the pixel resolution determined by the higher resolution of either the nighttime light or gridded population layers. Pixel level average values are then further downscaled to square meter levels using share of the building area within a pixel as downscaling factor. These methods are described in section III below.

Descriptive Statistics

In this study, we compute residential buildings stock and value data using Google Open Buildings data.[28] Due to the size of the building data and other geospatial layers we use, we demonstrate our approach with Kenya[29] as a case study. We also illustrate how the data could be used as key asset exposure layer for granular climate risk assessment.

Table 1 summarizes some of the key variables. The total number of buildings detected by Google Open Building is 26,405,031. The average ground-level building area is around 50 square meters. The aggregate ground-level building area is 1,324,699,574 square meters and the aggregate height-adjusted building areas is 1,367,850,580 square meters. The average confidence score is 0.78 with minimum and maximum values of 0.65 and 0.99, respectively, and more than 83 percent of detected buildings have confidence scores of 0.7 and above. For each detected building, we extract the average building height from GHSL ANBH layer, which has a pixel size of 100m x 100m. The average net building height for Kenya is 1.29 meters and the maximum is 42.5 meters.[30] In this paper, we assume that buildings within 100m x 100m GHSL ANBH pixel have the same height.

One of the challenges in using GHSL ANBH layers, however, is that about 51 percent of the pixels have zero values, possibly due to model performance, poor building shadows at the time day-time satellite imageries were taken, etc. We therefore impute pixels within which buildings are detected and GHSL ANBH have estimated value of zero with the minimum non-zero ANBH for the country. In the case of Kenya, the minimum non-zero ANBH is 2.5 meters. By dividing imputed building heights by the minimum non-zero height value, we obtain the approximate number of stories/floors for each detected building. We, then compute the estimated total area of the building by multiplying ground-level area by the number of stories, ruling out below ground floors.

The other key set of layers that we use in our analysis is GHSL built-up area classification. We use this information to filter out buildings that are predominantly used for non-residential purposes. As shown in the table, there are three classifications: non-built-up, residential, and non-residential areas. Unfortunately, more than 83 percent of the pixels within which buildings are detected are misclassified as non-built-up area. For this analysis, we only remove pixels that are classified as non-residential and keep detected buildings located in pixels with a human settlement.

[27] Group of countries for which dwelling data was available were Indonesia (EM), Mexico (EM), Argentina (EM), Australia (AE), Japan (AE), Republic of Korea (AE), and United States (AE). Due to lack of data, we used Indonesia's share of dwelling in capital stock, which is the height among EMs, as a proxy for LIDCs.
[28] Building-level residential exposure data in meter square area, population, nighttime light, and downscaled dwellings capital stock value are computed for 48 countries in Africa with more countries to be added. The size of the data is greater than 50Gb.
[29] We selected Kenya as a case study to illustrate the approach for two reasons. First, housing properties survey data is available which will help us estimate and validate property values at next phase of the study. Second, it is one of the low-income countries where detailed residential properties data are lacking.
[30] The other key information that comes with each detected building is the level of confidence associated with each detection. This will help to perform robustness checks in areas where height, population, and building classification are either zero or null.

Table 1. Descriptive Statistics of the Data, Kenya

	Mean	Median	Min	Max
Ground-level area (square meters)	50.19	30.54	2.51	49,630.01
Confidence score	0.78	0.78	0.65	0.99
Buildings with confidence score (%):				
[0.9, 1.0)	2.65			
[0.8, 0.9)	36.14			
[0.7, 0.8)	44.07			
[0.65, 0.7)	17.07			
Average net build height (meters)	1.29	0.00	0.00	41.52
Proportion of zero building heights (%)	50.93			
Built-up area classification (%):				
Non-built-up	83.19			
Residential	16.78			
Non-residential	0.03			
Estimated average # of building stories/floors*	1.02	1.00	1.00	17.0
Population per pixel (2018)—GHSL (100m)	4.42	0.00	0.00	5,329.20
Proportion of zero (%)—GHSL	51.4			
Population (2020) per pixel—GPWv4	1,340.10	502.4	0.00	143,427.70
Proportion of zero (%)—GPWv4	0.00			
Nighttime light radiance per pixel (nW/cm^2/sr)	1.54	0.00	0.00	75.8
Proportion of zero NTL	63.65			
No. of buildings	26,405,031			

Note: *The estimated average number of building stories computed after imputing zero height with minimum non-zero average building height of 2.5 meters for Kenya.

We use the GPWv4 gridded population data which have 1000m x1000m resolution. For comparison, we also considered the GHSL population layer which has a higher resolution of 100mx100m. For instance, the average number of persons within pixels where buildings are detected is 4.42 per 100m x 100m and about 1,340 per 1000m x 1000m, respectively, for GHSL and GPWv4 layers. However, more than half of the GHSL population layer pixels with detected buildings have zero values, whereas GPWv4 pixels within which buildings are detected have all positive values. For these reasons, we use GPWv4 gridded population layer for our downscaling analysis. Finally, we use nighttime light images from EOG in combination with GPWv4 population data to downscale country-level dwellings capital stock value to pixel level. The average nighttime light radiance is 1.54 nW/cm^2/sr, and about 63.65 percent of the pixels within which buildings are detected have zero values.

We encounter additional challenges of extreme values in detected building footprints. As indicated in Table 1, nearly all key datasets have extreme values, potentially impacting downscaling and summary values of indicators. This issue arises from structure like greenhouse sheds, industrial parks, and schools, which, due to large rooftops, are mistakenly classified as residential buildings or non-built areas by the GHSL.[31] We use a simple statistical rule of thumb to identify and remove such outliers which tend to be on the right tail of the distribution. Specifically, we remove outliers if $\left|\frac{x_i-\tilde{x}}{sd(x)}\right| > 3$, where x_i the value of pixel i, $\tilde{x}$ is the median and

[31] Image A.1 in Appendix A shows a satellite and street view of a Greenhouse shed with large roof. Similarly, image A.2 shows a building with large roof in an industrial park. Similarly, we encounter outliers in gridded population data as well as nighttime light luminosity. However, we keep these pixels which are more likely to be densely populated.

$sd(x)$ is the standard deviation. Furthermore, we exclude buildings smaller than 10 square meters and those within GPWv4 pixels (1 square km) and with fewer than four people which is the average household size in Kenya[32]. After filtering out outliers, the total number of buildings dropped from 26.41 million to 22.04 million.

III. Methods and Proposed Indicators

Developing new statistical indicators requires consensus on principles and conceptual framework. As part of the effort of developing statistical indicators to measure development goals, the United Nations Inter-Agency and Expert Group on MDG Indicators (IAEG-MDG) suggested key considerations in selecting indicators (UN, 2013). These include adherence to the principles of indicator selection, statistical criteria, and other considerations. In this study, we propose the following residential properties exposure indicators that have three main dimensions of quantity, density, and monetary value:

i) Quantity indicators: square meters of residential properties per geographic unit (square km, grid cell, administrative boundary, etc.).[33]

ii) Value indicators: value of residential properties per geographic unit (square km, grid cell, administrative boundary, etc.).

iii) Density indicators: square meters of residential properties per person; national share of residential properties per geographic unit (square km, grid cell, administrative boundary, etc.); and national share of residential properties value per a geographic unit (square km, grid cell, administrative boundary, etc.).

These indicators align with most recommended criteria for effective indicators, including the principle that indicators should be directly linked to the objectives and targets under consideration. They also fulfill statistical criteria such as relevance, methodological soundness, measurability, and clarity (UN, 2013). More indicators could also be derived from the underlying data.

Downscaling Aggregate Dwellings Capital Stock Values

Compiling residential property indicators from open-source data presents challenges, particularly the lack of detailed, reliable, and consistent information on property value. To overcome this, we propose downscaling aggregate dwelling capital stock values to the square meter level using a downscaling method which builds upon the work of Zhao et al. (2017), Eberenz et al. (2020), and, more recently, Wang and Sun (2022). These studies used a population and nighttime light (LitPop) approach to downscale aggregate figures such as GDP to grid level. This approach is employed in generating gridded datasets that are used to for common climate projection scenarios or models such as the Shared Socioeconomic Pathway (SSP) framework and the Representative Concentration Pathways (RCP). Grubler et al. (2007) developed a similar method to calculate gridded GDP and population datasets, which have been widely used in the United Nations Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) and related climate research. [34] De Bono and Mora (2014), in their study on a global exposure model for disaster risk assessment, also applied this method to downscale aggregate values to the grid cell level, using population as the

[32] Household Size and Composition | Population Division (un.org)
[33] This indicator could be further disaggregated by the number of stories. For instance, square meters of single-story, two-story, three-story, etc., buildings per geographic unit.
[34] SEDAC - MVA - Downscaling (columbia.edu) is part of NASA's Earth Observing System Data and Information System (EOSDIS).

downscaling factor. Following this methodology, we use population and nighttime light data to estimate the value of buildings, downscaling aggregate dwelling capital stock to the grid level.

Adopting the approach of Wang and Sun (2022), let V_c denote country c's dwellings capital stock and V_{ci} is the downscaled value for pixel i, which takes the minimum of resolution of gridded population or nighttime light layers.

$$V_{ci} = \frac{V_c}{\sum_i^N Lit_{ci} \cdot Pop_{ci}} \times Lit_{ci} \cdot Pop_{ci}, \quad (1)$$

where,

$$Lit_{ci} \cdot Pop_{ci} = \begin{cases} Pop_{ci} & if\ Lit_{ci} = 0 \\ Lit_{ci} \cdot Pop_{ci} & if\ Lit_{ci} > 0\ \&\ Pop_{ci} > 0 \\ Lit_{ci} & if\ Pop = 0 \end{cases},$$

$Lit_{ci} \cdot Pop_{ci}$ is the product of nighttime light radiance and population count for pixel i, N is the total number of pixels for country c. One of the issues we encounter in using LitPop approach is its sensitivity to extreme values. As an alternative, we use log transformed values, i.e., $logLit_{ci} \cdot logPop_{ci} = \log(Lit_{ci} + 1) \cdot \log(Pop_{ci} + 1)$, which addresses issues of extreme values and yields a well-behaved distribution compared to the LitPop approach. Figure A.2. shows the kernel distributions of downscaled property values using these two approaches.

From the downscaled pixel level values of residential properties, V_{ci}, we obtain residential property value per square meter as follows:

$$V_{ci}^{sqr.m} = \frac{V_{ci}}{\sum_{j=1}^{n_j} a_{ci,j}}, \quad (2)$$

where, $V_{ci}^{sqr.m}$ is the average residential property value per square meter within pixel i, $a_{ci,j}$ is the area of residential building j, and n_j is the number of residential buildings within pixel i.

Weighted Downscaling

One key limitation of the downscaling approach proposed in equation (1) is it assigns equal value to residential properties in a country regardless of location and other amenities, such as urban vs. rural. Although nighttime light and population density capture the spatial variation property value arising from some of these amenities, the model assigns equal weight of unity to nighttime and population. This presumes that nighttime light and population have equal contribution to the disaggregation function. Following Eberenz et al. (2020), we assign different weights to nighttime light and population as follows:

$$\widetilde{V_{ci}} = \frac{V_c}{\sum_i^N Lit_{ci}^{\alpha} \cdot Pop_{ci}^{\beta}} \times Lit_{ci}^{\alpha} \cdot Pop_{ci}^{\beta}, \quad (3)$$

where $\widetilde{V_{ci}}$ is the downscaled dwellings capital value for pixel i using weighted downscaling, and $\alpha + \beta = 1$ are weights assigned to nighttime light and population, respectively. To address varying weights by urban and rural locations, which are key factors in determining property values, we let α and β to assume different values for urban and rural locations, i.e., $\alpha^u + \beta^u = 1$ for urban and $\alpha^r + \beta^r = 1$ for rural areas. The challenge in implementing equation (3) is that these weighting parameters are unknowns and needs to be calibrated. To this

end, we estimate log-transformed Cobb-Douglas model of residential property value from survey data on nighttime light and population as follows:

$$\log(h_{ij}) = \alpha \log(Lit_{ij}) + \beta \log(Pop_{ij}) + \varepsilon_i, \quad (4)$$

where h_{ij} is residential property value per square meters from survey data for house i in location j, $\alpha + \beta = 1$ are parameters to be estimated, and $\varepsilon_{ij} \sim iid$ is normally distributed error term. We estimate equation (3) for pooled sample as well as split sample for urban and rural areas to estimate respective weighting parameters—α^u and β^u for urban areas, α^r and β^r for rural areas. We use the 2012/2013 Kenya National Housing Survey data[35] that has information on value of residential properties to estimate the weighting parameters. The estimated values of the parameters are $\alpha = 0.9$, $\beta = 0.1$, $\alpha^u = 0.95$, $\beta^u = 0.05$, $\alpha^r = 0.88$, and $\beta^r = 0.12$. These estimates show that within a given grid, nighttime light have higher weight compared to population in determining the values of residential properties in Kenya. The weight for nighttime light is higher in urban areas compared to rural areas. The estimated weights are then plugged in to equation (3) to obtain the weighted downscaled capital value of properties.

Aggregation Level

In climate risk assessment literature, the selection of an aggregation level is determined by various factors, including the nature of the climate hazards and the available data types. Aggregation can be done at census tracts/blocks, districts, region, or geographic fishnet/grid cell levels. For example, the national risk index developed by the US Federal Emergency Management Agency (FEMA) employs a mix of these aggregation levels, tailored to the data sources and specific climate risks under consideration. In this study, we show computed residential property indicators using two aggregation approaches: equally spaced grid cells and level 3 administrative boundaries[36]. Both approaches have advantages and drawbacks. Equally spaced grid cells provide flexibility in performing aggregations at different levels of climate risk, exposure area, and vulnerability analysis scale. We employ the Uber hexagonal global gridded system, utilizing a hierarchical spatial indexing system (H3), which stands out among various open-source spatial indexing libraries.[37]

IV. Results and discussions

Table 2 presents summary statistics of the computed residential properties indicators for Kenya. It shows the estimated areas of residential buildings, values per square meters, and per capita building values obtained by employing unweighted and weighted downscaling methods. The average total building area is 46 square meters. The estimated values vary depending on the specific downscaling method. Generally, the unweighted LitPop approach yields lower average property value per square meters compared to the values obtained using the logLit+LogPop approach. The variations of downscaled values, as measured by (CV = SD/Mean) are, however, are much lower in logLit+logPop approach compared to the LitPop approach, reflecting the

[35] The sample size is 8,884 households.
[36] The lowest available administrative level for Kenya.
[37] H3: Uber's Hexagonal Hierarchical Spatial Index | Uber Blog. H3 | H3 (h3geo.org). Another openly available library is Google's S2 Geometry Library. See About S2 | S2Geometry. The cons of using equally spaced grid cells are demarcations could be arbitrary and not straightforward for subnational regional- or municipality-level policy analysis. Administrative boundaries, on the other hand, are demarcated with settlement, population density, political representation, topology, etc. under consideration. Figure A.3 in Appendix A illustrates H3 grid cells and administrative boundaries for Kenya.

importance of addressing extreme values that potentially inflate standard deviations and averages. As for the distribution of the values using weighted downscaling approach, the averages are slightly higher than the values obtained from the unweighted approach.

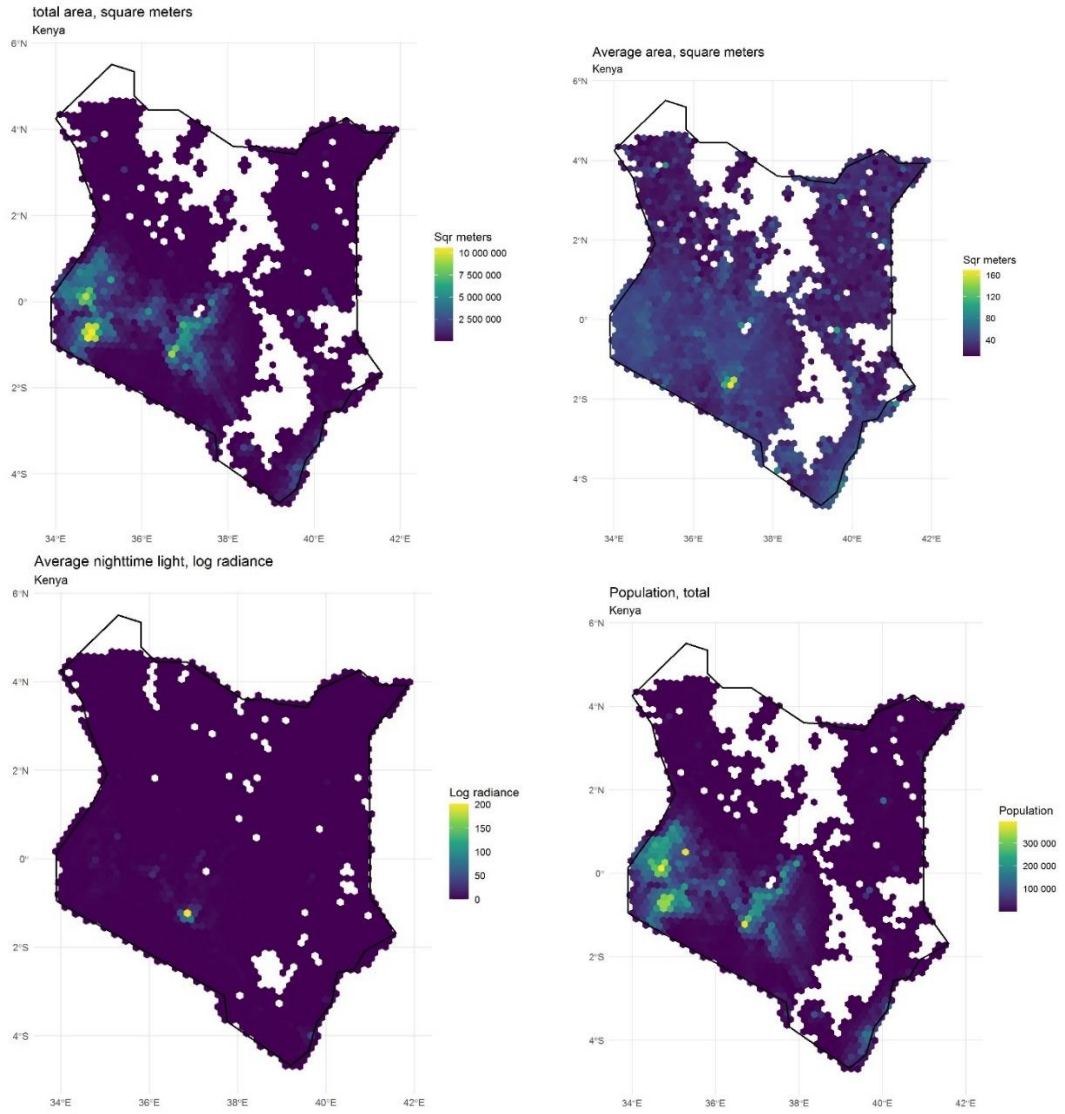
Table 2: Summary of Selected Estimated Residential Property Indicators, Kenya

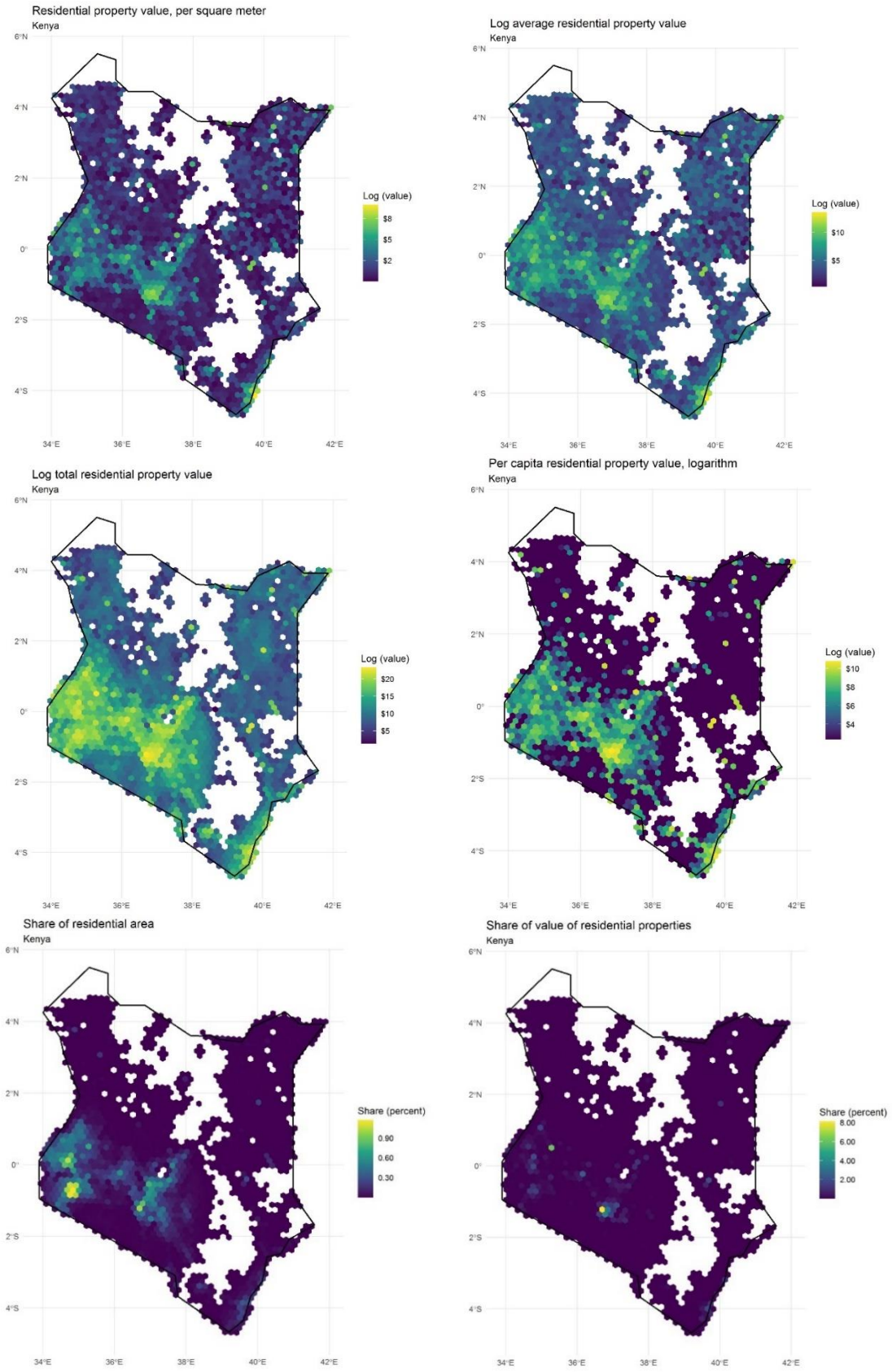
Variable	Mean	Std. Dev.	Min	Max
Building area (sqr meters)	46	(39)	10	2,128
Unweighted downscaling				
Value per sqr meter: LitPop	$102	($723)	$0	$14,170
Value per sqr meter: logLit+logPop	$197	($973)	$2	$97,244
Building value: LitPop	$6,678	($58,041)	$0	$8,209,556
Building Value: logLit+logPop	$6,676	($19,109)	$27	$1,658,925
Weighted downscaling				
Value per sqr meter: LitPop (pooled)	$130	($372)	$0	$659,723
Value per sqr meter: LitPop (urban/rural)	$152	($373)	$0	$385,948
Value per sqr meter: logLit+logPop (pooled)	$207	($1,141)	$2	$114,297
Value per sqr meter: logLit+logPop (urban/rural)	$206	($1,121)	$2	$117,546
Building value: weighted LitPop (pooled)	$6,678	($18,246)	$0	$22,794,341
Building value: weighted LitPop (urban/rural)	$6,677	($12,611)	$6	$12,552,431
Building value: weighted logLit+logPop (pooled)	$6,676	($22,067)	$31	$1,668,971
Building value: weighted logLit+logPop (urban/rural)	$6,676	($21,741)	$31	$1,639,328
No. of buildings	22,044,652			

Note: Values are in constant 2017 international dollars using the GFCF deflators and purchasing power parities taken from the OECD and PWT depending on data availability. See IMF Infrastructure Governance. InvestmentandCapitalStockDatabaseUserManualandFAQ_May2021.pdf (imf.org).

Leveraging the granularity of the data, summary statistics can also be computed at various sub-national levels. Figures 1 show H3 grid-aggregated values of total residential area (log square meters), average area (log square meters), sum of the nighttime light radiance values, population, downscaled dwellings capital stock values in international dollars, and national shares of capital stock values and residential area. For illustration purposes, we use a wider aggregation grid with hex resolution of five that is approximately 252.9 square km.

Figure 1: Spatial Distribution of Residential Properties Indicators for Kenya, by Grid Cell at a Resolution Level of 5

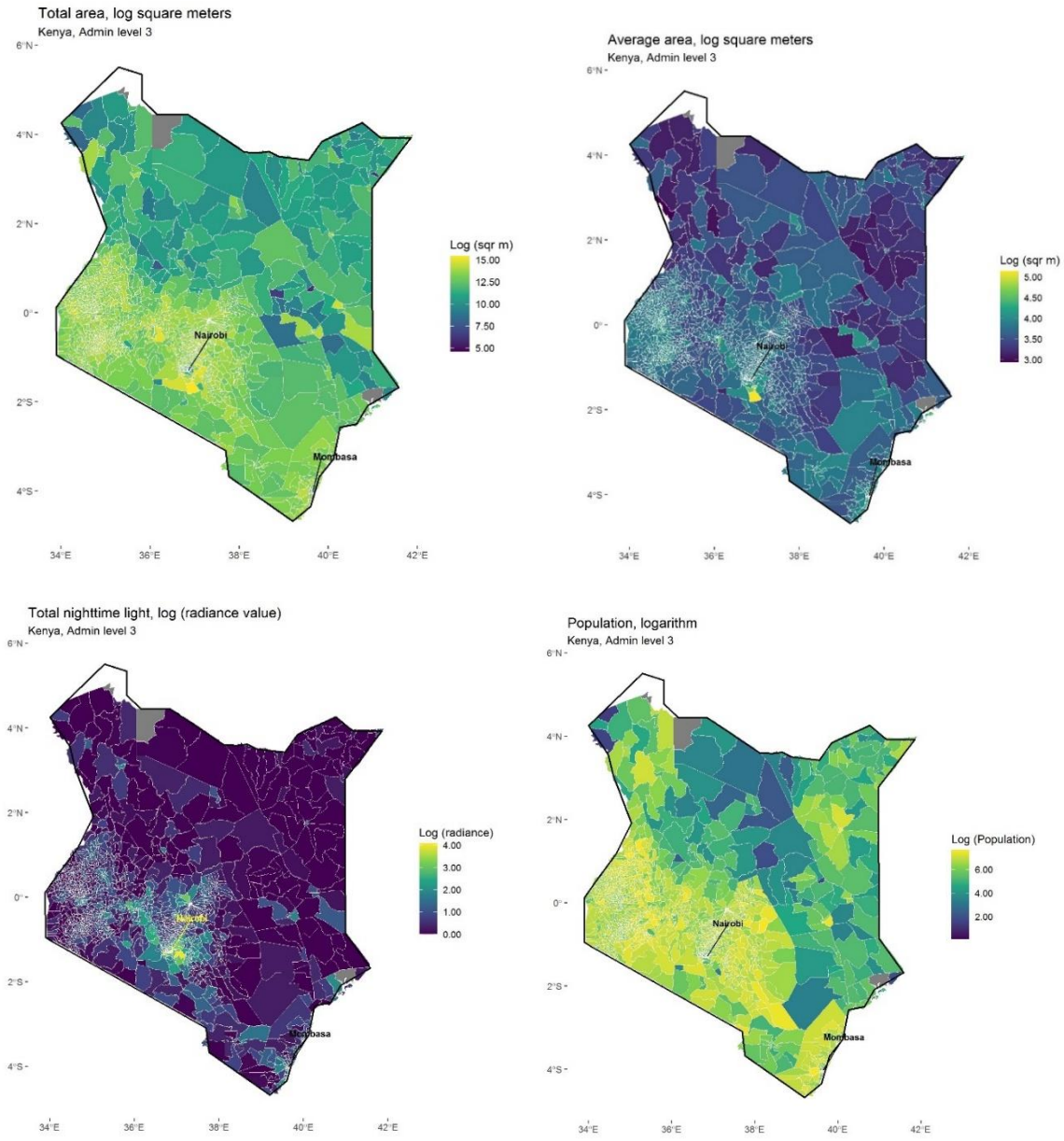


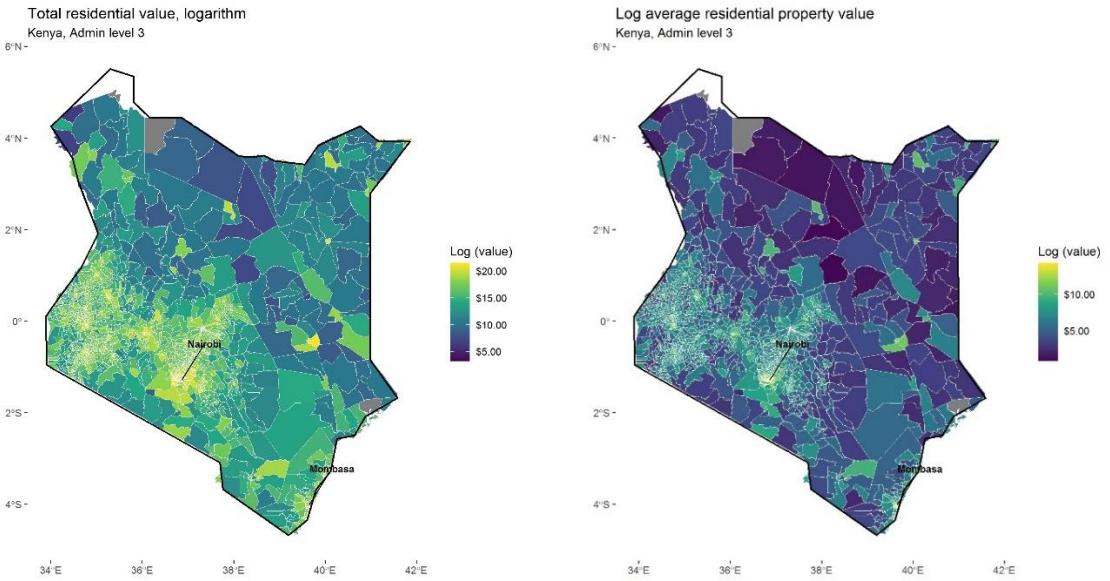


Source: Authors' computation.

Figure 2 shows aggregation of total residential area (in log square meters), average building heights (meters), nighttime radiance value, population, and shares of value of residential properties by level 3 administrative boundaries.

Figure 2: Spatial Distribution of Residential Properties Indicators for Kenya, by Admin Level 3





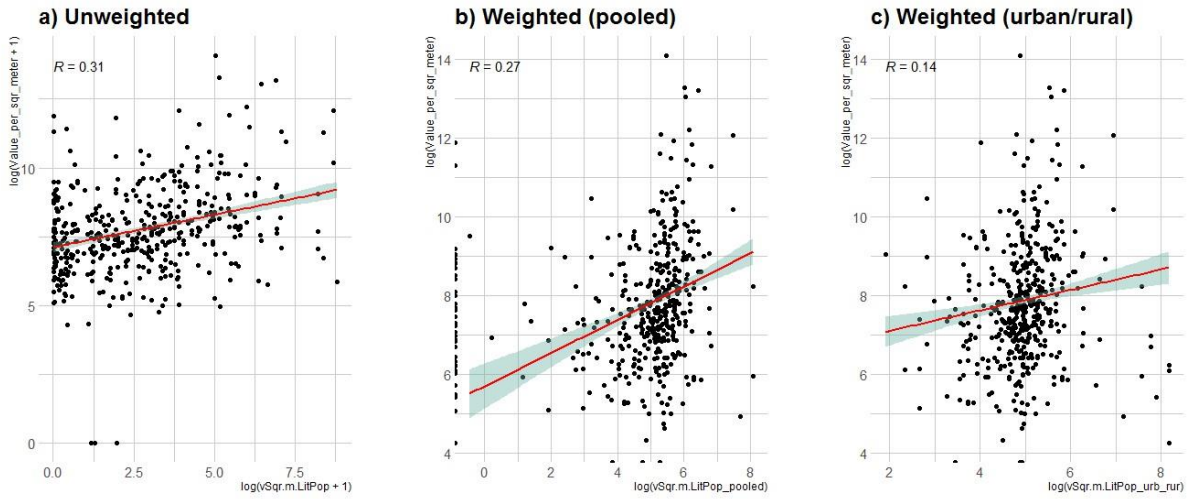
Source: Authors' computation.

Validation Exercise

In this section, we validate the downscaled building values using data from housing survey. Housing survey data comes from the 2012-2013 Kenyan National Housing Survey. The survey was carried out in 44 counties across the country providing a nationally representative sample, covering both urban and rural areas. The total number of observations in the sample is 8,884 with key information on housing units, including area, floor, the amount of money spent to construct or purchase the property and rental value it was rented out. Figure 3 shows scatter plots of housing value from survey and downscaled dwellings capital stock value for unweighted and weighted downscaling approaches. The values averages for a given H3 polygon (at resolution 7) and spatially joined. It shows that the downscaled capital stock values are generally positively correlated with the actual house values collected from housing survey.[38]

Figure 3: Scatter plot housing value from survey and downscaled dwelling capital stock value

[38] The survey values are in Kenyan Shilling. The values of downscaled dwellings capital are in 2017 constant international dollars.



Source: authors' computation.

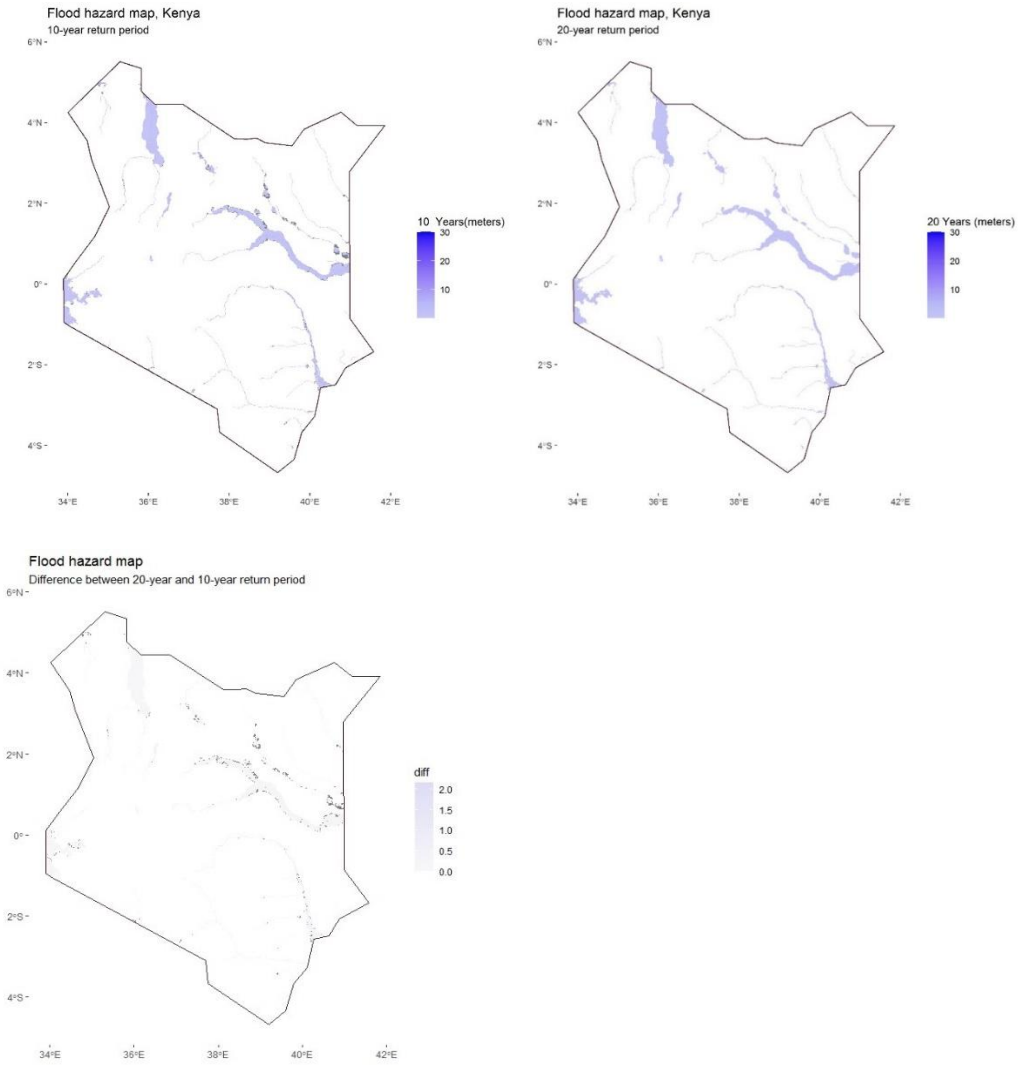
The pearson correlation between the log of house value per square meters and log of unweighted, weighted (pooled), and weighted (urban/rural) downscaled dwellings capital stock value are, respectively, 0.3, 0.27, and 0.14. These underscore that unweighted downscaled dwellings capital stock values have stronger correlation with housing value from survey compared to the weighted downscaling that assign higher weights to nighttime light. These correlations are however not without limitations. One of the limitations in this exercise is that there is a seven year difference between the time the housing survey was conducted (2012/2013) and the capital stock value was obtained (2020). Several factors could change during this period that could potentially change the home values in 2020 resulting in poor correlation. Regardless, we show that the downscaled values are indeed positively correlated.

Exposure to riverine flood risk: an illustration for Kenya

In this section, we demonstrate how residential building exposure data that we compiled above could be used to riverine flood risks using the EC-JRC Global Flood Hazard Map.[39] This layer highlights areas prone to flooding on a global scale for various flood events, presented at a resolution of 30 arcseconds (approximately 1 km). Each grid cell indicates the water depth in meters. Figure 4 illustrates the flood hazard map for return periods of 10 years and 20 years, providing important and granular insights into flood risk (European Commission Joint Research Centre [EC-JRC], 2023). A 10-year flood risk has a 1-in-10 (10%) chance of happening in any one year, whereas a flood hazard with a return period of 20-year has 1-in-20 (5%) chance of happening in any one year.[40]

[39] Another important global flood risk map is the Dartmouth Flood Observatory (DFO). The DFO was established in 1993 and records the extent and temporal distribution of flood events occurring between 2000 and 2018 with each pixel classified as water (in meters) or non-water at 250-meter resolution.
[40] See Floods and Recurrence Intervals | U.S. Geological Survey (usgs.gov) for more discussion of flood recurrence intervals and likelihood of occurrence.

Figure 4: Flood Hazard Map of 10-year and 20-year Return Periods, Kenya

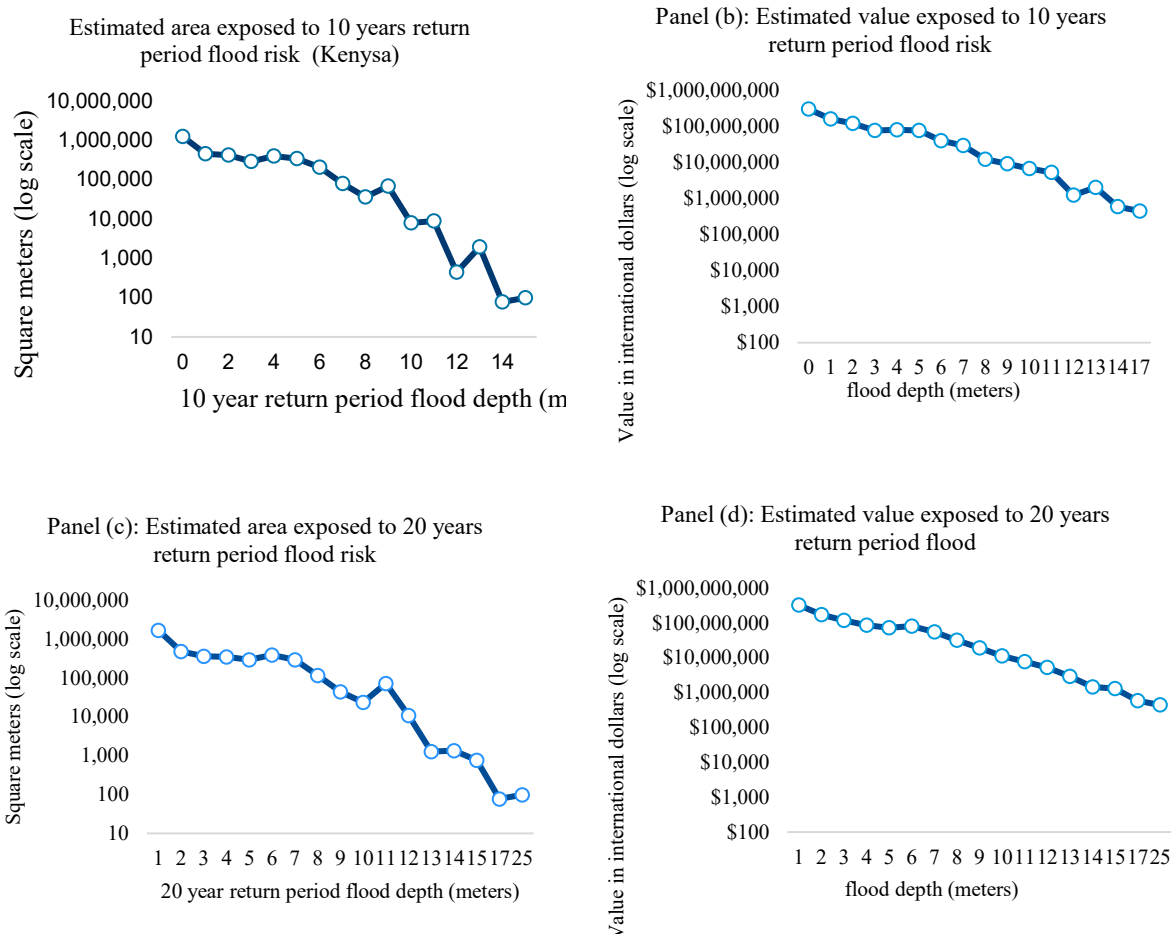


Source: Authors' illustration using Flood Hazard Map of the World (Dottori et al., 2016) at Joint Research Centre Data Catalogue – Flood hazard map of the World – 10-year return per–… - European Commission (europa.eu)

Combining residential properties layer with flood risk map, we calculate the aggregate residential properties and the dollar values that are at risk. For the purpose of this illustration, we use the unweighted downscaled capital stock values. Our calculation shows that the total area of residential properties in 2019 Kenya were 915.1 million square meters that are estimated to be worth 147.2 billion in 2017 constant international dollars (PPP adjusted). Out of these properties, more than 3.5 and 4.1 million square meters are exposed to 10-year return and 20-year return period riverine flood, and the corresponding estimated monetary value of these properties, respectively, are 915.5 million and 1 billion in constant 2017 international dollars. The corresponding real GDP data from the same source was 228 billion in constant 2017 international dollars. Figure 4 panels (a)-(d) shows the quantity and estimated values of residential properties within riverine flood risks of 10-year return and 20-year return periods. These figures indicate that the estimated total area or stock of residential properties within flood risk zones decreases with severity as measured by depth in meters. The same pattern is observed with the total estimated monetary value. These are intuitive results which show that people avoid flood prone

zones when choosing a site to building their houses. However, there are buildings within sever flood risk areas, presumably due to factors, such as urban residential land shortage or poverty.

Figure 4: Quantity and Value of Residential Properties Exposed to Riverine Flood Risks, Kenya



Source: Authors' computation. Monetary values are in constant 2017 international dollars (PPP adjusted). The vertical axis represents of panels (a) and (c) represent the quantity of residential properties in meter square, whereas vertical axis of panel (b) and panel (d) represent estimated values of residential properties in international dollars (2017 constant prices). The horizontal axis represents flood depth in meters.

Given that the focus of this paper is to illustrate approaches on constructing residential buildings exposure layers using existing building footprint and processed satellite data, we do not delve deep into the analysis of exposure to flood risks. However, the computation above illustrates simple approaches to estimate the quantity and value of residential property at a risk of a specific climate risk with greater detail.

Limitations and further discussions

Compiling a residential building census data using various open-source datasets has several limitations. A primary concern is the GHSL's estimated average net building heights, which are zero for a significant proportion of the pixels where buildings were detected. In such instances, we imputed zero values with a minimum non-zero height, approximately 2.5 meters for Kenya. Reliance on such ad-hoc imputations raises potential questions about the accuracy of the GHSL's height estimates. Furthermore, GHSL's building classification sometimes inaccurately labels pixels as neither residential nor non-residential, despite buildings being detected and classified by Google Open Building. To address this, pixels with detected buildings were reclassified as residential, assuming most misclassified structures are in non-urban areas, thus unlikely to be commercial. Additionally, GHSL population estimates are zero for pixels where buildings are detected, while GPWv4 provides non-zero estimates. Also, the absence of detailed information on the dwelling component of gross capital stock for many countries presents a major limitation. Another potential bottleneck in using such layers that are generated using machine learning models involves the propagation of algorithmic biases and model errors. These biases can stem from statistical errors, incomplete data, human bias in labeling training data, or variations in terrain/topology across different regions, potentially exacerbating inaccuracies. For instance, the building footprint data from Google, which utilizes a deep learning algorithm, along with building classifications, estimated building heights, and population data from EC-JRC, all apply various machine learning models with differing performance levels.

V. Conclusion

Many countries have a rich set of housing statistics from which they can formulate, implement, and monitor key policies pertaining to the housing sector. They have access to regular census information, property values from local governments, prices information from real estate boards and property assessment programs. Other countries have a large data gap in this area and therefore struggle to formulate effective policy to help manage this significant national asset and associated risks. This impedes countries' ability to assess financial markets, the supply and demand of housing, and increasingly, the impact climate change is having on the housing stock. This study shows how to leverages opensource satellite data to compile a spatially explicit census of housing or residential buildings.

We follow a census approach using high-precision building footprint data from Google Open Buildings. We also leverage capital stock data to impute monetary value to building level using downscaling technique that uses gridded population and nighttime light as weighting factors. As climate actions take more concrete steps and the need for granular exposure and vulnerability data grows, we show that such data could close existing data gaps and better inform policies. Using Kenya as a country case, we illustrate how the data could be used to perform detailed climate risk analysis on residential buildings exposure to riverine flood risk. The data could also feed into the existing macroeconomic models of the IMF, such as the Debt-Investment-Growth-Natural-Disasters (DIGNAD) model. Given that the lack of spatially explicit and comparable housing exposure data has been recognized as key area of data gap under the G20 Data Gap Initiative phase III (DGI-3) recommendation five, we believe that the research presented in this paper will contribute to the extensive literature assessing the economic impact of climate change.

References

Angrist, N., Goldberg, P. K., & Jolliffe, D. (2021). Why is growth in developing countries so hard to measure? Journal of Economic Perspectives, 35(3), 215-242.

Center for International Earth Science Information Network (CIESIN) Columbia University (2018). Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, Revision 11. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4F47M65. Accessed DAY MONTH YEAR.

De Bono, A., & Mora, M. G. (2014). A global exposure model for disaster risk assessment. International journal of disaster risk reduction, 10, 442-451.

Doan, M. K., Hill, R., Hallegatte, S., Corral, P., Brunckhorst, B., Nguyen, M., ... & Naikal, E. (2023). Counting People Exposed to, Vulnerable to, or at High Risk from Climate Shocks: A Methodology. Policy Research Working Papers; 10619. © World Bank, Washington, DC. http://hdl.handle.net/10986/40685 License: CC BY 3.0 IGO.

Dottori, Francesco, Alfieri, Lorenzo, Salamon, Peter, Bianchi, Alessandra; Feyen, Luc; Hirpa, Feyera (2016): Flood hazard map of the Word - 10-year return period. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-floods-floodmapgl_rp10y-tif

Eberenz, S., Stocker, D., Rsli, T. & Bresch, D. N. Asset exposure data for global physical risk assessment. *Earth System Science Data.* **12**, 817–833, https://doi.org/10.5194/essd-12-817-2020 (2020).

Elvidge, C. D., Baugh, K. E., Zhizhin, M., & Hsu, F. C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. Proceedings of the Asia-Pacific Advanced Network, 35(0), 62.

Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. International journal of remote sensing, 38(21), 5860-5879.

Elvidge, C.D, Zhizhin, M., Ghosh T., Hsu FC, Taneja J. Annual time series of global VIIRS nighttime lights derived from monthly averages:2012 to 2019. Remote Sensing 2021, 13(5), p.922, doi:10.3390/rs13050922.

Ericksen, E. P., & Kadane, J. B. (1985). Estimating the Population in a Census Year 1980 and beyond. *Journal of the American Statistical Association*, *80*(389), 98-109.

Freire S., MacManus K., Pesaresi M., Doxsey-Whitfield E., Mills J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. Geospatial Data in a Changing World; Association of Geographic Information Laboratories in Europe (AGILE), AGILE 2016.

G.R. Brakenridge. Global Active Archive of Large Flood Events. Dartmouth Flood Observatory, University of Colorado, USA. http://floodobservatory.colorado.edu.

Gibson, J., Olivia, S., Boe-Gibson, G., & Li, C. (2021). Which night lights data should we use in economics, and where?. Journal of Development Economics, 149, 102602.

Grübler, A. et al. Regional, national, and spatially explicit scenarios of demographic and economic change based on SRES. Technol. Forecast. Soc. 74, 980–1029 (2007).

Grübler, A.' O'Neill, B., Riahi, K., Chirkov, V., Goujon, A., Kolp, P., ... & Slentoe, E. (2007). Regional, national, and spatially explicit scenarios of demographic and economic change based on SRES. Technological forecasting and social change, 74(7), 980-1029.

IMF (2008), Global Financial Stability Report Durable Financial Stability Getting There from Here. International Monetary Fund.

Miller, S. D., Mills, S. P., Elvidge, C. D., Lindsey, D. T., Lee, T. F., & Hawkins, J. D. (2012). Suomi satellite brings to light a unique frontier of nighttime environmental sensing capabilities. Proceedings of the National Academy of Sciences, 109(39), 15706-15711.

Pesaresi M., Corbane C., Ren C., and Edward N. (2021). Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling. PLOS ONE 16, e0244478. doi.org/10.1371/journal.pone.0244478

Pesaresi, M.; Politis, P. (2023). GHS-BUILT-H R202–A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018).European Commission, Joint Research Centre (JRC).PID: http://data.europa.eu/89h/85005901-3a49-48dd-9d19-6261354f56fe, doi:10.2905/85005901-3A49-48DD-9D19-6261354F56FE

Schiavina M., Freire S., Carioli A., MacManus K. (2023). GHS-POP R202–A - GHS population grid multitemporal (1975-2030).European Commission, Joint Research Centre (JRC) PID: http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe, doi:10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE.

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., ... & Quinn, J. (2021). Continental-scale building detection from high resolution satellite imagery. arXiv preprint arXiv:2107.12283.

UN, 2013. Lessons Learned from MDG Monitoring from A Statistical Perspective. Report of the Task Team on Lessons Learned from MDG Monitoring of the IAEG-MDG.

United Nations Environment Programme (2021). 2021 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector. Nairobi.

United Nations. Statistical Office. (1980). Principles and recommendations for population and housing censuses (No. 67). New York: United Nations, Department of International Economic and Social Affairs, Statistical Office.

Wang, T., & Sun, F. (2022). Global gridded GDP data set consistent with the shared socioeconomic pathways. Scientific Data, 9(1), 221.

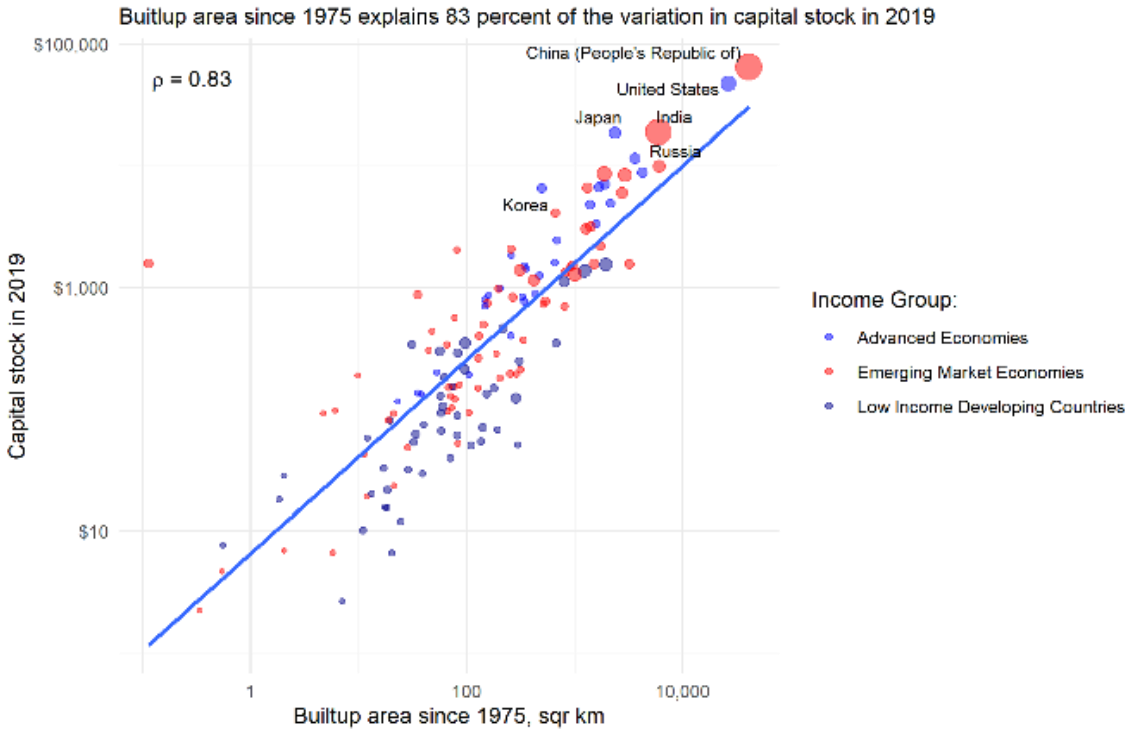
World Bank, NGFS, 2022. Physical Climate Risk Assessment: Practical Lessons for the Development of Climate Scenarios with Extreme Weather Events from Emerging Markets and Developing Economies. Technical Document. September 2022. The World Bank.

Zhao, N., Liu, Y., Cao, G., Samson, E. L. & Zhang, J. Forecasting China's GDP at the pixel level using nighttime lights time series and population images. *Mapping ences & Remote Sensing* **54**, 407–425, https://doi.org/10.1080/15481603.2016.1276705 (2017).

Wang, W., Rivard, H., & Zmeureanu, R. (2006). Floor shape optimization for green building design. *Advanced Engineering Informatics*, *20*(4), 363-378.

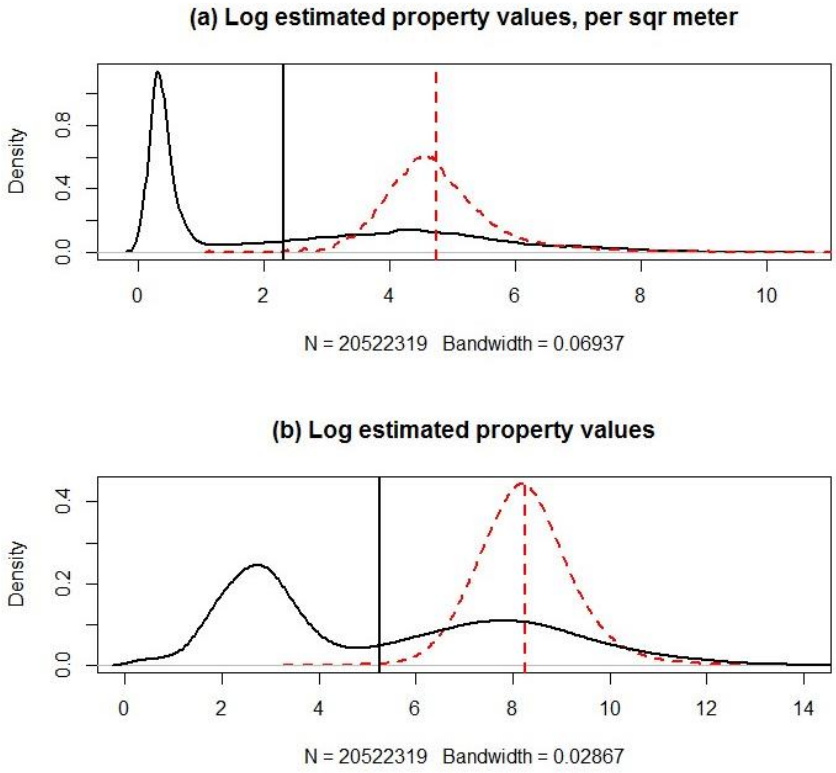
Appendix: Additional Figures

Figure A.1: Built Up Area and Capital Stock



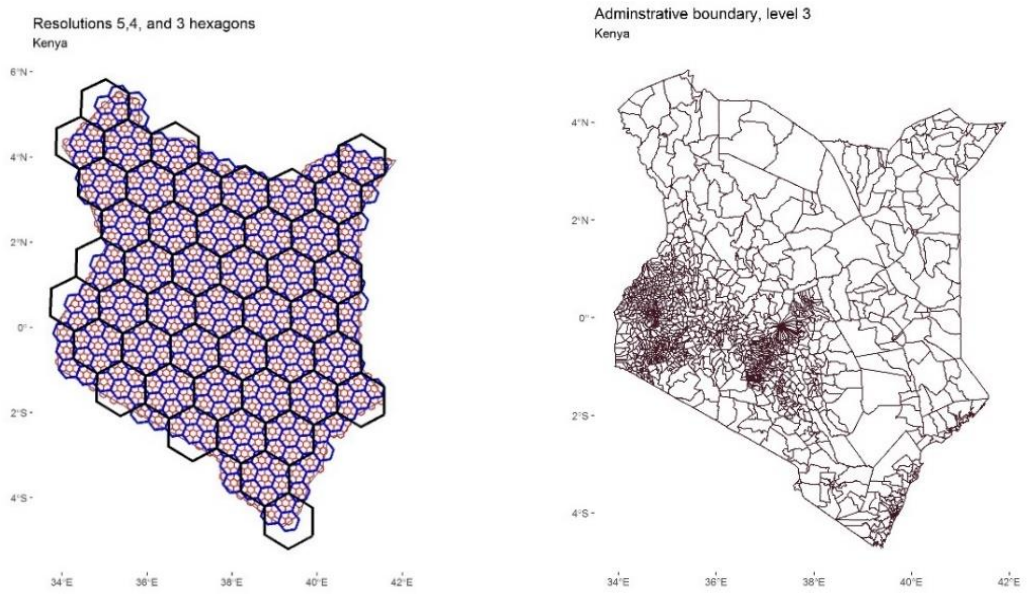
Source: Authors' computation using built-up area data are from OECD (https://stats.oecd.org), capital stock data from IMF (https://data.imf.org).
Note: Capital stock values are in billions of constant 2017 international dollars (ppp adjusted). Bubble size is proportional to population.

Figure A.2: Density of Log Estimated Property Values



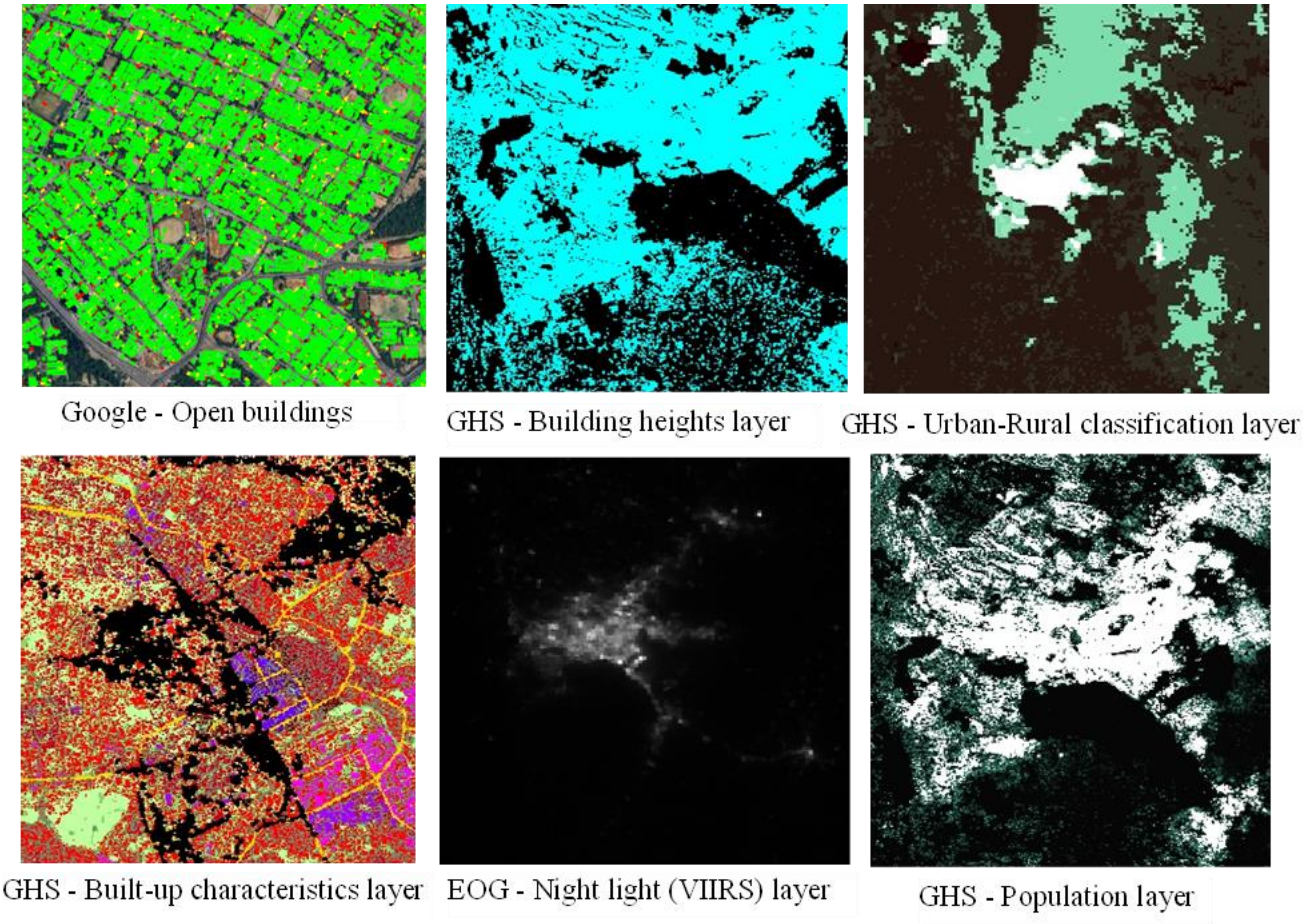
Source: Authors' computation.
Note: solid lines show distribution of computed values using LitPop as downscaling factors. Dotted line shows distributions using log (lit) + log(Pop) as downscaling factors. Vertical lines show mean values.

Figure A.3: Illustration of Hierarchical Grid Cells and Level 3 Admin Boundaries, Kenya



Source: Authors' illustration.
Note: The number of equally spaced H3 grid cells for Kenya are, respectively, 2,212, 315, and 45 for resolution levels of five, four, and three. The highest hex resolution is 0 which has an area of 4,250,547 square km whereas the lowest is 15 which has an area of 0.0000009 square km.

Figure A.4. Key Satellite Datasets and Sources Used for this Study – Extracted from Nairobi Area, Kenya



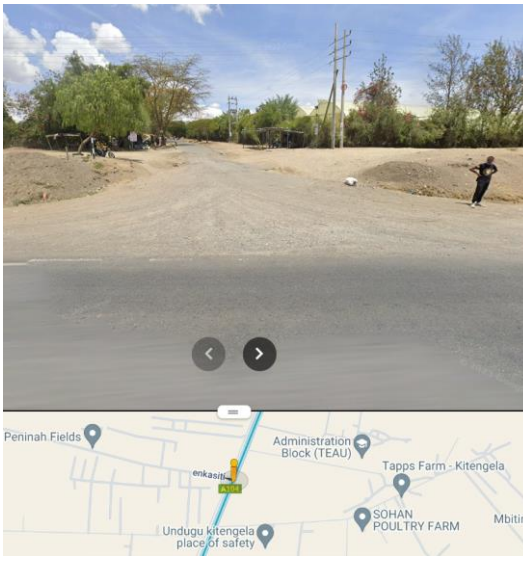
Source: Authors' illustrations.
Note: These images are excerpts of the different datasets used in this study, extracted from the Nairobi urban core. They are approximately from the same locations but at varying scales which comes from the original dataset.

Image A.1: Example of Large, Misclassified Structure: Greenhouse

Satellite view (roof)

Source: Google Open Buildings Open Buildings (research.google).
Note: Longitude and Latitude: 36.92190,-1.56517375 (Pluscode address: 6GCRCWMC+WQJ5)

Google Streetview



Source: Google Map Street View
Note: Longitude and Latitude: 36.92190,-1.56517375 (Pluscode address: 6GCRCWMC+WQJ5)

Image A.2: Example of Large Industrial Structure: Industrial Park

Source: Google Open Buildings Open Buildings (research.google).
Note: Longitude and Latitude: 37.00587-1.44221322 (Plus Code address: 6GCVH254+489H)

Table A.1: Description of Key Data Sources

Data	Source	Coverage	Resolution	Period/Epoch
Building polygon (v3)	Google Open Building	Africa, South Asia, South-East Asia, Latin America and the Caribbean (1.8 billion)	Building level <10m	2023
Building heights	EC-JRC-GHSL	Global	100m	2018
Building classification	EC-JRC-GHSL	Global	10m	2018
Population	EC-JRC-GHSL	Global	100m	2019
UN WPP-Adjusted Population Density, v4.11	SEDAC/ CIESIN	Global	1km	2000-2020 (five years interval)
Urban-Rural classification	GHSL	Global	1km	1975-2030 (five years interval)
Nighttime Light (VIIRS)	EOG	Global	~500m	2013-present
Investment and Capital Stock Datasets (ICSD)	IMF	Global	n.a.	1960-2019
Global Flood Hazard Map	EC-JRC-GHSL	Global	3 arc seconds (~90 m)	-

Note: EC-JRC-GHSL=European Commission Joint Research Center Global Human Settlement Layer. SEDAC/CIESIN = Socioeconomic Data and Applications Center (SEDAC), Center for International Earth Science Information Network (CIESIN). EOG=Earth Observation Group at Colorado School of Mines.