Estimating Population and Urban Areas at Risk of Coastal Hazards, 1990-2015: How data choices matter

Kytt MacManus,[1] Deborah Balk,[2,3] Hasim Engin,[1,2] Gordon McGranahan[4], Rya Inman[1]

[1]CIESIN, Columbia University
[2]CUNY Institute for Demographic Research (CIDR), City University of New York
[3]Marxe School of Public and International Affairs, Baruch College, City University of New York
[4]Independent Researcher, Lewes, United Kingdom

Correspondence to: Kytt MacManus (kmacmanu@ciesin.columbia.edu)

Style Definition: Hyperlink: Font color: Hyperlink

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: Not Bold, Font color: Auto

Abstract

The accurate estimation of population living in the Low Elevation Coastal Zone (LECZ), and at heightened risk from sea level rise, is critically important for policy makers and risk managers worldwide. This characterization of potential exposure depends not only on robust representations of coastal elevation and spatial population data, but also of settlements along the urban-rural continuum. The empirical basis for LECZ estimation has improved considerably in the 13 years since it was first estimated that 10% of the world's population, and an even greater share of the urban population, lived in the LECZ ~~(McGranahan et al., 2007b).~~(McGranahan et al., 2007a). Those estimates were constrained in several ways, most notably by a single 10-meter LECZ, but also by a dichotomous urban-rural proxy and population from a single source. This paper updates those initial estimates with newer, improved inputs and provides a range of estimates, along with sensitivity analyses that reveal the importance of understanding the strengths and weaknesses of the underlying data. We estimate that between 750 million to nearly 1.1 billion persons globally, in 2015, live in the ≤10m LECZ, with the variation depending on the elevation and population data sources used. The variations are considerably greater at more disaggregated levels, when finer elevation bands (e.g. the ≤5m LECZ) or differing delineations between urban, quasi-urban and rural populations are considered. Despite these variations, there is general agreement that the LECZ is disproportionately home to urban dwellers, and that the urban population in the LECZ has grown more than urban areas outside the LECZ since 1990. We describe the main results across these new elevation, population, and urban proxy data sources in order to guide future research and improvements to characterizing risk in low elevation coastal zones. DOI: assigned upon completion of data peer-review.

1. INTRODUCTION

Climate change threatens people around the world, but particularly in locations where concentrations of people can be expected to overlap with concentrations of physical hazards resulting from climate change. Low elevation coastal zones (LECZs) are likely to contain a disproportionate and growing share of such locations. Sea level rise and a greater prevalence of extreme weather events are correlates of climate change and heighten the risks of flooding, coastal erosion, groundwater salinization and other hazards in low lying coastal areas ~~(Oppenheimer and Hinkel, 2019).~~(Oppenheimer et al., 2019). People are also more concentrated in coastal areas, and continued urbanization can be expected to increase this concentration, unless urban development patterns change substantially. ~~Flooding~~Urbanisation and ~~other hazards~~ related water abstraction, especially in deltas where natural subsidence is often already occurring, can also contribute to ~~sea~~ subsidence, which adds to and amplifies the hazards associated with sea-level rise and extreme weather events (Nicholls et al., 2021). As such, quantifying urbanization within the LECZ is both important to monitoring the sources and pathways creating these hazards and to identifying increasing concentrations of receptors, who will bear a growing share of burden. The hazards of concern not only threaten human life and wellbeing directly, but also ~~indirectly through damage to~~ have adverse consequences for the homes, businesses and ~~infrastructure~~infrastructures characteristic of built-up areas, as well as to local ecosystems and the services they provide. The principal purpose of delineating a Low Elevation Coastal Zone the (LECZ) is to identify the broad areas in which there are likely to be sub-populations at a heightened and rising risk of exposure to selected hazards being exacerbated by climate change. The hazards of particular concern here are those arising from sea level rise and more extreme weather events, and being aggravated by land subsidence, where that is occuring.

A foundational study of settlement in the LECZ globally ~~(McGranahan et al., 2007b)~~(McGranahan et al., 2007a) found that in the year 2000, coastally contiguous areas of less than ten meters in elevation contained an estimated 10 percent of the world's population and 13 percent of its urban population. That study also contained case studies of China and Bangladesh, suggesting that from 1990-2000, populations in these countries' LECZs were growing faster than outside the LECZ, with urban LECZ populations growing fastest of all. Since that study, a number of new tools and data sets have been developed, allowing for more refined estimates of land areas, built-up areas, and populations in LECZs, and their changing urban-rural compositions over a number of years (1990 - 2015).

Accurate estimates of populations living in the LECZ depend on robust representations of coastal elevations ~~(Gesch, 2018; Hinkel et al., 2014; Lichter et al., 2010) and population at a fine resolution (Mondal and Tatem, 2012; Leyk et al., 2019).~~(Gesch, 2018; Hinkel et al., 2014; Lichter et al., 2010, Hooijer and Vernimmen, 2021) and population at a fine resolution (Mondal and Tatem, 2012; Leyk et al., 2019a). Estimates of *urban* population and land in the LECZ require additional data on the spatial extent and density of urban areas, ideally encompassing a full urban-rural continuum of settlements ~~(Dijkstra et al., 2020; OECD and European Commission, 2020).~~(Dijkstra et al., 2020; OECD

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

and European Commission, 2020). While the empirical basis for such estimates has improved considerably (cf. excellent review in ~~McMichael et al., 2020)~~(McMichael et al., 2020)) since the first analysis ~~(McGranahan et al., 2007b),~~(McGranahan et al., 2007a), there has also been a proliferation of new internationally available data sources containing a wide range of estimates of the population in LECZs. For example, a recent study finds that, "New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding" ~~(Kulp and Strauss, 2019).~~(Kulp and Strauss, 2019). To improve decision-making there needs to be a better understanding of the strengths and limitations of each data set, the applications it is best suited for, and why estimates vary so widely. This study fills that gap by constructing new estimates of population and land area found along the urban-rural continuum within the LECZ, based on four elevation data sources, four population data sources, and four urban proxy data sources, each with their own strengths and weaknesses, all designed to be internationally comparable and substantially improved in the past decade ~~(Leyk et al., 2019; Gesch, 2018).~~(Leyk et al., 2019a; Gesch, 2018).

Improvements in the spatial resolution of these data sets also allows for a more fine-grained analysis of potential exposure: within the ≤10m LECZ and along an urban-rural continuum. Using four elevation sources, we first constructed the LECZs: ≤5m above sea level, 5-10m above sea level (which can be added together to form a ≤10m LECZ), and a category of higher coastal areas and non-coastal areas of any elevation. We then estimated the population of these LECZs, disaggregated by settlement type, based on an array of population sources and urban proxy data sets. The four elevation data sets obtain their estimates through a variety of different sensors, which in one case (CoastalDEM) is combined with statistical modeling. The four population data sets use different approaches to mapping and disaggregating population, and the four data sets representing the urban-rural continuum use a variety of different underlying data sources, such as satellite imagery of built-up areas or night-time lights, and different modelling approaches (some with population criteria, others without), to categorize the level of urbanization of settlements. Defining an urban-rural continuum, largely in contrast to defining population, requires researchers to make decisions that reflect the best available knowledge and expert judgements, but are at some level necessarily arbitrary, or may be more suitable for some research questions than others.

The primary focus of this paper is on methodology and includes a sensitivity analysis in order to compare the many sources of population, urban area delineation, and digital elevation models used to construct the LECZs and estimate at-risk populations. The sensitivity analysis reveals similarities and differences in each of the data sources that we considered for population, urban proxy, and elevation as well as indicates how the results would change if the measures along an urban-rural continuum were defined with a different indicator (such as night-time lights rather than built-up area), or if the thresholds or boundaries of the definitions were adjusted.

We begin in the section on Data and Methods with summaries of input sources and continue with a detailed description of how the LECZs were constructed, and how populations and land areas were tabulated. This is followed by a Results section which includes a series of zonal summaries of gridded population data categorized by LECZ and along an urban-rural continuum for three time points (1990, 2000, 2015) and the Sensitivity Analysis. Finally, there is a discussion of fitness for use along with conclusions and future research needs. Accompanying this paper are tabular data on ~~country-level summaries~~country-level summaries as well as spatial data, where redistribution is permissible and a python notebook which provides an algorithm to produce LECZs from elevation and coastline data ~~(Center for International Earth Science Information Network - CIESIN - Columbia University, and CUNY Institute for Demographic Research - CIDR - City University of New York, 2021).~~(Center for International Earth Science Information Network - CIESIN - Columbia University, and CUNY Institute for Demographic Research - CIDR - City University of New York, 2021).

2. DATA AND METHODS

The basic method used here to quantify potential exposure to sea level rise is based on fairly straightforward spatial summaries (zonal statistics), but depends on substantial improvements to and suitable conditioning of underlying data sets, which we describe in detail below. There have been many advances in earth observation, population censuses, and scientific computing capacity since the original LECZ Urban-Rural Population Estimates, v1 (1990, 1995, 2000) data set ~~(McGranahan et al., 2007a)~~(McGranahan et al., 2007b) was constructed. In this section, we provide an overview of the various input data sets, including a discussion of the uncertainties. Such uncertainties may have considerable impacts on the ultimate estimates of persons, particularly when stratified along an urban-rural continuum, in the LECZ, and must therefore be carefully understood ~~(Gesch, 2018; Hawker et al., 2019; Mondal and Tatem, 2012;~~

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Pattern: Clear, Highlight

Formatted: Pattern: Clear, Highlight

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

~~Lichter et al., 2010; Leyk et al., 2019).~~(Gesch, 2018; Hawker et al., 2019; Uuemaa et al, 2020; Mondal and Tatem, 2012; Lichter et al., 2010; Leyk et al., 2019a). Since it has been shown that accuracy of the digital elevation models (DEMs) – upon which the LECZs are based – is highly sensitive to local conditions, including land cover, it is therefore sensible to evaluate a variety of DEMs that can be used to estimate population and land area in the LECZ. In this section, we describe the data strengths and limitations, along with the conditioning, transformations, and processing required to generate LECZs and accompanying population and land area estimates along an urban-rural continuum. At the end of each type of data, we choose a 'core data set' to form the basis of comparison with all others and to simplify the presentation of results. (Estimates available on all combinations of the data sets are available in the dataset.) Rationales for our selection of these core data sets are given.

Formatted: Font: 12 pt, Font color: Black

2.1 Data on Elevation

For elevation data to construct the LECZ, we considered the sources as described in Table 1 and shown in Fig. 1. These data are all freely available at 3 arc second horizontal resolution or higher, but some have restrictions on usage and redistribution of derived data products. There is general agreement that the newer data products have made substantial improvements in vertical accuracy since the early releases of the Shuttle Radar Topography Mission (SRTM) data~~(Gesch, 2018; Hawker et al., 2019).~~ (Gesch, 2018; Hawker et al., 2019; Uuemaa et al, 2020). We selected four data sets for use here based largely on recent studies, such as that by Gesch (2018), which finds that only some of the global DEMs are suitable for delineating the LECZ ≤10m elevation at or above the 68% confidence level, (including TanDEM-X, CoastalDEM, NASADEM, AW3D30, and MERIT DEM) whereas other data sets (such as SRTM) are not. (SRTM nonetheless is included here in order to compare to previous work.) Despite compelling interests from the policy arena, it should be noted that the implication of Gesch's 2018 study is that delineating LECZs in finer increments than 10m is subject to great uncertainty. (Hooijer and Vernimmen's (2021) recent study uses a Digital Terrain Model to generate LiDAR data rather than those primarily used here, to estimate land area and population up to 2m above mean sea level, but use of this welcomed addition to the literature will be deferred to future research.)

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Black

Major improvements of these data notwithstanding, we discuss below considerations specific to the coastal zone (such as detection of and correction for mangrove forests) and to urban areas (where manmade structures may bias measurement.) The 2019 study by Hawker et al. identifies the types of land cover that each of these data sets best detects or is prone to misinterpret (see graphical abstract and Table 4 in ~~Hawker et al., 2019).~~(Hawker et al., 2019)). While urban environments are evaluated in these recent studies, urban is just another land cover class. To our knowledge, no targeted or multi-criteria evaluation of these global ~~elevations~~elevation data sets in the urban setting has been made (but see related analysis of the built environment in cities ~~(Esch et al., 2020)).~~, which leverages TanDEM-X (Esch et al., 2020)). Furthermore, case studies reveal that the data sets which perform best globally, on average, may not necessarily be the most accurate in a given location or under particular geographic conditions ~~(Minderhoud et al., 2018, 2019).~~(Minderhoud et al., 2018, 2019). Notably these global DEMs tend to do a comparably poor job of capturing low-lying elevation in small island states ~~(Taupo and Noy, 2016; Taupo et al., 2018; Yamano et al., 2007; Lewis, 1989).~~(Taupo and Noy, 2016; Taupo et al., 2018; Yamano et al., 2007; Lewis, 1989). Once again, there are important policy implications here.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

The scientific community of modelers that produce the many relevant data sets and models to predict sea levels, tides, storm surges and other coastal flooding is large, growing and vibrant. Future LECZ estimates stand to be substantially improved by their current efforts. In this work, we have used global elevation data sets to construct LECZs using a simple but inclusive approach. Global hydrodynamics-based models could potentially provide a better basis for identifying some of the relevant hazards, including most notably those caused by flooding. However, at the time of this writing the results of global models that account for the fuller and complex set of factors at and connected to the seacoast are not available. ~~Local studies (Schumann and Bates, 2018; Orton et al., 2020)~~Local studies (Schumann and Bates, 2018; Orton et al., 2020) suggest clearly that the hydrodynamics of the coastal zone are complex and nuanced, perhaps even more so in urban areas where impervious surface, underground infrastructure (sewage and subway systems), and other modifications to the landscape (including accumulations of uncollected solid waste) may impact inland flows and drainage. New work on coastal storm surges and tidal heights ~~(Muis et al., 2020; Arns et al., 2020) and empirical flood events (Tellman et al., 2020)~~(Muis et al., 2020; Arns et al., 2020) and empirical flood events (Tellman et al., 2021) make new avenues of research possible in the coming years, but currently hydrodynamic modeling has been restricted to certain locations or events.

Table 1. Elevation data sets used in the construction of the Low Elevation Coastal Zones (LECZ)

Source Data Set on Elevation	Abbreviation	Input Spatial Resolution	Acquisition (or modelling) Period	Paper Reference
Shuttle Radar Topography Mission Elevation Low Elevation Coastal Zones	SRTM	3 arc second	February 11-22, 2000	(ISciences, 2003)(ISciences, 2003)
MERIT DEM: Multi-Error-Removed Improved-Terrain DEM	MERIT	3 arc second	see SRTM and AW3D30	(Yamazaki et al., 2017)(Yamazaki et al., 2017)
TerraSAR-X add-on for Digital Elevation Measurement 90m	TanDEM-X	3 arc second	2010-2015	(Wessel et al., 2018)(Wessel et al., 2018)
CoastalDEM90	CoastalDEM	3 arc second	see SRTM and AW3D30, plus LiDAR from 2003-2009, 2015	(Kulp and Strauss, 2018)(Kulp and Strauss, 2018)
ALOS World 3D - 30m Digital Surface Model	AW3D30	1 arc second	2006-2011	(Tadono et al., 2014)(Tadono et al., 2014)

NB: ALOS 30m global DEM is used as a supplement in CoastalDEM90 at latitudes north of 60N and south of 56S. It is also used as an input data set in the construction of MERIT DEM.

2.1.1 SRTM

The NASA Shuttle Radar Topography Mission (SRTM) set the standard for characterizing global elevations, but the SRTM data products, now nearly 20 years old, have been widely understood to have limitations in some critical areas. There is, for instance, high Root Mean Square Error (RMSE) in the elevation estimates of mangrove forests (Gesch, 2018).(Gesch, 2018). These vertical errors (termed tree-height bias) are particularly problematic in some low-lying vegetated areas (such as in populous southeastern Bangladesh). A coastally contiguous derivation of SRTM was produced in 2003 by ISCIENCES, Ltd. and it was these data which were used in the first LECZ study (McGranahan et al., 2007b), and the NASA SEDAC update (Center for International Earth Science and Information Network - CIESIN - Columbia University, 2013).(McGranahan et al., 2007a), and the NASA SEDAC update (Center for International Earth Science and Information Network - CIESIN - Columbia University, 2013). We include the same data here exclusively for the purpose of comparison with the original study.

2.1.2 MERIT DEM

The Multi-Error-Removed Improved-Terrain DEM (MERIT), "separated absolute bias, stripe noise, speckle noise and tree height bias using multiple satellite data sets and filtering techniques" to improve on the SRTM baseline (Yamazaki et al., 2017).(Yamazaki et al., 2017). The Japanese Aerospace Exploration Agency (JAXA) produces the ALOS Global Digital Surface Model (DSM) "ALOS World 3D - 30m" (AW3D30), which along with SRTM3 v2.1 was used as primary data in the construction of MERIT. In the land cover classes of short vegetation and forested areas, MERIT performs well when compared to locally available LiDAR data (Hawker et al., 2019).(Hawker et al., 2019). Compared

Formatted ... (repeated)

to TanDEM-X, which Hawker et al. (2019) find to be of comparable overall accuracy, MERIT has a marginally higher Margin of Error (ME) (1.09 m), but lower Mean Absolute Error (MAE) (1.69 m) and RMSE (2.32 m). If the RMSE metric is the only metric considered, MERIT is the most accurate Global DEM across landcover types. Uuemaa et al, 2020 found that MERIT peformed well in accuracy tests, despite having a coarser resolution than SRTM. Despite these many improvements, Gesch (2018) notes that MERIT has an RMSE of about 3m globally, which has implications for producing LECZs in finer increments below 10m.

Formatted: Font: 12 pt, Font color: Black

2.1.3 TanDEM-X 90

TerraSAR-X add-on for Digital Elevation Measurement (TanDEM-X) departs from an SRTM baseline and provides a new synthetic-aperture radar (SAR) interferometry-based estimate of elevations globally ~~(Wessel et al., 2018).~~(Wessel et al., 2018). TanDEM-X has been shown to be the most accurate global DEM in some landcover categories (bare, shrubland, sparse vegetation, urban) ~~(Hawker et al., 2019),~~(Hawker et al., 2019), and at the 95% confidence level, Gesch (2018) finds that only TanDEM-X is suitable for delineating the LECZ below 10 meters. Wessel et al. (2018) found TanDEM-X to have biases in areas of rugged terrain, where there is heterogeneity in the landscape/landcover and elevation, and additional analyses have revealed that while TanDEM-X is highly accurate in flat to slightly undulating terrains, it tends to overestimate elevation when used in areas characterized by more sharply uneven terrain ~~(Bhardwaj, 2019).~~(Bhardwaj, 2019). Higher resolution versions of TanDEM-X (0.4 arc second and 1 arc second) are available through proposal and service fee for scientific use, but were not utilized in this study. While no study has closely examined the vertical accuracy of these DEMs globally for urban areas, Hawker and colleagues' (2019) analysis does suggest that TanDEM-X may be well-suited to capturing elevation in core urban areas where high-rise buildings are common (For specific cities, some analysts have used TanDEM-X for urban elevation mapping ~~(Rossi and Gernhardt, 2013) to construct urban extents (Esch et al., 2012, 2013)).~~(Rossi and Gernhardt, 2013) to construct urban extents (Esch et al., 2012, 2013)). However, TanDEM-X's restrictive licensing limits dissemination and replicability, making it less useful for many purposes.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.1.4 CoastalDEM90 and ALOS World 3D

CoastalDEM utilizes neural networks with an array of spatial covariates to improve on SRTM (see Fig.1 in Kulp & Strauss 2018). CoastalDEM90 is made available for research free of charge and a higher resolution (1 arc second) version of CoastalDEM is available with a licensing fee. AW3D30 was used as supplemental data for CoastalDEM in latitudes north of 60N and south of 56S as was done by the data authors. CoastalDEM is produced from a 23-dimensional vertical error regression analysis using variables including neighborhood elevation values, pop density, land slope and local SRTM deviations from ICESat altitude observations and vegetation cover indices. Importantly, since one of the covariates is population density from LandScan 2010 (resampled to 1 arc second), estimation of population exposure in the LECZ is complicated since population itself was used to determine elevation values. Studies by the data producers show that in the Caribbean Basin, the CoastalDEM data provides greater vertical accuracy than other data sets, including SRTM data and AW3D30 data set which both overestimate by more than 2m on average ~~(Climate Central, 2018).~~(Central, 2018). However, Gesch (2018) notes that globally CoastalDEM, like MERIT, has an RMSE of about 3m, implying a need for caution when using it to delineate LECZs in increments finer than 10m.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.1.5 Core data choice - Elevation

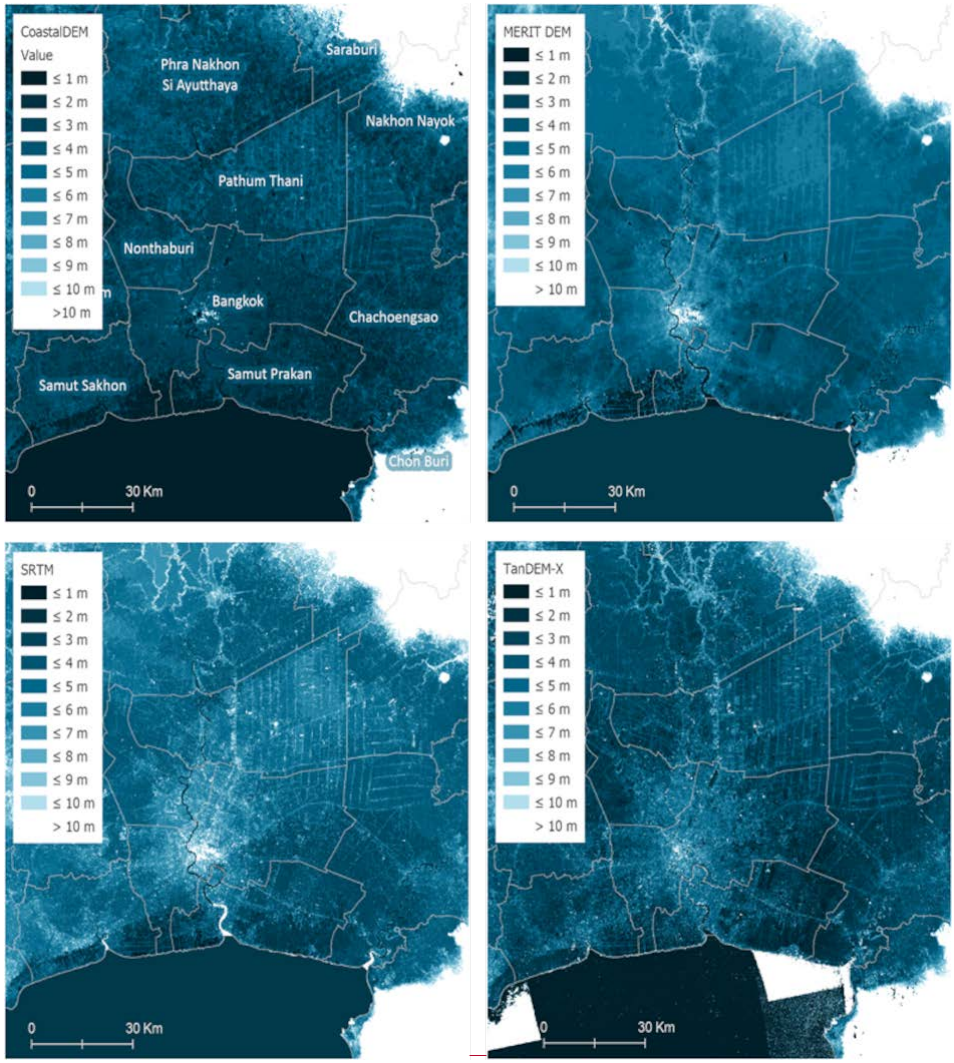
Hawker et al., (2019) evaluated the accuracy of global DEMs by land cover type and found that both TanDEM-X and MERIT outperform SRTM across all categories, and that MERIT achieves greater accuracy than TanDEM-X in short vegetation and forested land cover classes. CoastalDEM performs well in the Caribbean Basin (~~Climate Central~~Strauss and Kulp, 2018), but globally has a similar RMSE to MERIT. According to Gesch (2018), TanDEM-X, with an RMSE of 1.69m, can be used to delineate the ≤10m LECZ with the greatest confidence, but MERIT and CoastalDEM, which have RMSEs ~3m, can be used, albeit with somewhat less confidence (see Table 5 in Gesch, 2018). However, as is shown below, the precision of TanDEM-X results in a highly varied landscape, with raised roadways clearly identified at higher elevations than surrounding land. This results in wide areas of TanDEM-X losing their direct connectivity to the coast according to image segmentation (region grouping) methods, and therefore removes them from the LECZ, which requires coastal contiguity. Additional research on the presence of natural (wetlands, floodplains), and manmade (raised berms, buildings) barriers, as well as connections (sewer systems, stormwater management systems/culverts, estuaries, and other water channels) is needed in order to improve the TanDEM-X-based LECZ.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

We have selected MERIT as the core elevation data set. This is because of the complications with identifying coastal connections in TanDEM-X, and since MERIT is the only elevation data set considered which is both accurate ~~(Gesch, 2018; Hawker et al., 2019)~~(Gesch, 2018; Hawker et al., 2019; Uuemaa et al, 2020), and has wide dissemination rights (open for use both non-commercially and commercially, so that any data we create from it can be widely and openly distributed as well, both in spatial and tabular format (Yamazaki et al., 2017)).



Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted: Default Paragraph Font, Font color: Auto
Formatted: Font color: Auto
Field Code Changed
Formatted: Font: 12 pt, Font color: Black

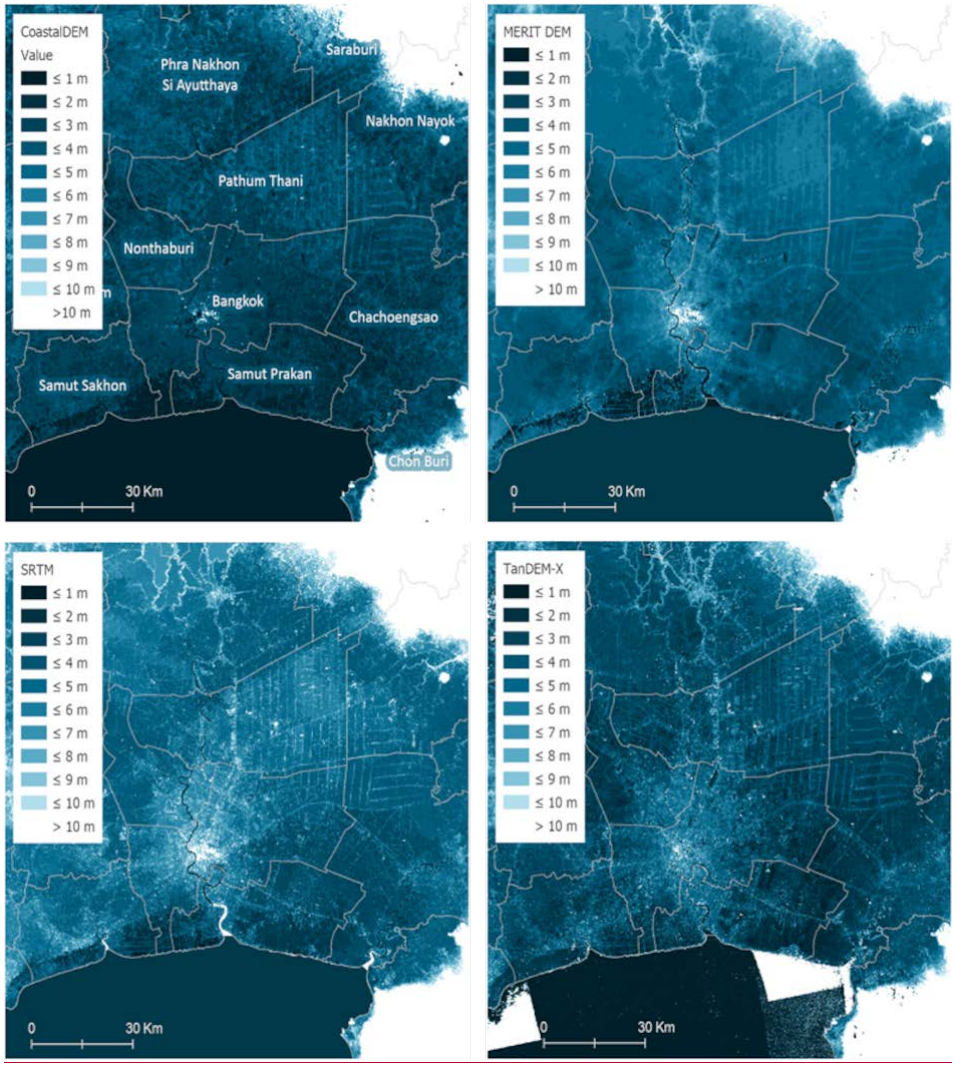


Figure 1. Elevation source data for constructing Low Elevation Coastal Zones (LECZ), Bangkok and surrounding areas, Thailand. Note that the darkest blue indicates ocean and gray boundaries indicate first-order administrative boundaries.

2.2 Data on Population

For population data, we compared several data sets from the leading producers of global population data, listed in Table 2 and visualized for Bangkok in Fig. 2. ~~Mondal and Tatem, 2012, as well as Lichter et al., 2010,~~Mondal and Tatem, 2012, as well as Lichter et al., 2010, compared several gridded population data sets and recommended that studies utilizing a particular data set should acknowledge how the inherent uncertainties of the underlying input data and methods are likely to impact conclusions. A recent and thorough review by ~~Leyk et al., 2019~~Leyk et al., 2019a discussed the nature and source of these uncertainties at great length. Characteristics such as the relative resolution of

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

underlying input vector data sets, the selection and relative accuracy of spatial covariates for dasymetric maps, and the currency of population estimates all have major impacts on grid level population counts. For a full description of each of these, including strengths and weaknesses, please see Leyk et al. ~~2019~~2019a, and Table 2 for the selection of data sets we use herein.

Table 2. Population data sets used to estimate persons living in the Low Elevation Coastal Zones (LECZ)

Source Data Set on Population	**Abbreviation**	**Input Spatial Resolution**	**Temporal Resolution**	**Paper Reference**
~~Global Human Settlement Layer -~~ Gridded Population ~~Grid r2019a~~of the World, version. 4.11	~~GHS-POP~~ GPW	~~9~~30 arc sec	1990, 2000, 2015	~~(Florczyk et al., 2019)~~(Doxsey-Whitfield et al., 2015)
~~Gridded~~Global Human Settlement Layer - Population ~~of the World, version. 4.11~~Grid r2019a	~~GPW v4.11~~GHS-POP	~~30~~9 arc sec	1990, 2000, 2015	~~(Doxsey-Whitfield et al., 2015)~~ (Florczyk et al., 2019)
~~LandScan 2015~~WorldPop Global High Resolution Population ~~Data Set~~Denominators	~~LandScan~~WorldPop	~~30~~3 arc sec	2000, 2015	~~(Bright and Coleman, 2001; Bright et al., 2016)~~(Lloyd et al., 2019)
~~WorldPop Global~~LandScan 2015 High Resolution Population ~~Denominators~~Data Set	~~WorldPop~~LandScan	~~3~~30 arc sec	2000, 2015	~~(Lloyd et al., 2019)~~ (Bright and Coleman, 2001; Bright et al., 2016)

2.2.1 GPW

Gridded Population of the World version 4.11 (GPW) is a minimally-modeled 30 arc second horizontal resolution data set which uses source data from the 2010 round of international censuses. Because GPW uses a uniform allocation across space to distribute census-based populations across the smallest areas for which population estimates were made available, the accuracy of its estimates at a pixel level is very closely linked to the relative resolution of input vector data. The layer "Mean Administrative Unit Area" in the GPW data collection provides an indicator of the relative size of input geographies (Administrative areas refer here loosely to the units in which census data are reported and may include enumeration areas, which typically are statistical rather than administrative, as well as truly administrative areas used for reporting) and is used here along with the GPW UN-WPP Adjusted Population Counts data set, which adjusts census reported national totals to estimates from the UN World Population Prospects ~~(Doxsey-Whitfield et al., 2015; United Nations, 2018).~~(Doxsey-Whitfield et al., 2015; United Nations, 2018). When overlaying a population grid with irregularly-shaped and variably-sized zones (such as a narrow LECZ in some locations), a uniform allocation method of coarse underlying data will sometimes lead to misestimation ~~(Mondal and Tatem, 2012; Balk, 2009).~~(Mondal and Tatem, 2012; Balk, 2009). However, an important advantage of the uniform allocation approach is that these data do not include additional spatial layers which could themselves be the source of errors and uncertainties.

2.2.2 GHS-POP

The Global Human Settlement Population Grid r2019a (GHS-POP) is derived from GPW inputs and the Global Human Settlement Layer (GHSL) built-up data set (GHS-BUILT ~~(Pesaresi et al., 2016).~~(Pesaresi et al., 2016), also discussed below) to improve the horizontal resolution and positional accuracy of free and open population data ~~(Freire et al., 2016).~~(Freire et al., 2016). A distinguishing characteristic of GHS-POP is the use of the GHS-BUILT time series, which was derived from satellite observations from the LandSat program's long history of earth observations. GHS-POP data uses a dasymetric mapping approach at a 9 arc second horizontal resolution to reallocate GPW census inputs based on the percentage built-up, as defined by GHS-BUILT ~~(Freire et al., 2016).~~(Freire et al., 2016). In this approach, population from large, sparsely populated administrative units is moved to the detected built-up areas rather than being assumed to be evenly distributed throughout the entire polygon: reallocation of population occurs in proportion to the distribution of built-up area (within a given cell), otherwise areal weighting is applied (see Fig. 2 in Freire et al., 2016). Because GHS-POP relies on GHS-BUILT, for which detection in sparse and rural areas is lacking

~~(Leyk et al., 2018).~~(Leyk et al., 2018), it may tend to over-concentrate population into built-up ~~-~~areas, overestimating the number of urban residents (depending on how urban areas themselves are delineated). GHS-POP has been shown in recent studies to produce the most accurate pixel-level population estimates when compared to local data for some locations, especially in urban areas ~~(Archila Bustos et al., 2020; Calka and Bielecka, 2020).~~(Archila Bustos et al., 2020; Calka and Bielecka, 2020)

Formatted: Font: 12 pt, Font color: Black

2.2.3 WorldPop

WorldPop Global High Resolution Population Denominators (WorldPop) also uses census-based population inputs from GPW to produce estimates for 2000 and 2015 (it does not include 1990). Its disaggregation approach uses country-specific machine learning (ML)-based dasymetric models which employ random forest classifications to disaggregate population on the basis of a variety of spatial covariate layers such as slope, impervious surface, night-time lights and others ~~(Lloyd et al., 2019; Gaughan et al., 2015).~~(Lloyd et al., 2019; Gaughan et al., 2015). WorldPop produces population estimates at a 3 arc second horizontal resolution, which is the same resolution as the input elevation data. Importantly, the covariate data used to delineate WorldPop estimates ~~are static for~~ include elevation as one of the ~~year that they were collected (even though~~ weighting factors, and some ~~represent~~ time-varying ~~characteristics, like~~(or modelled to be time-varying) data sets (GHSL, the ~~night-time lights), and therefore the spatial distributions of population estimates are also static. Despite this,~~Global Urban Footprint (Esch et al., 2017), and ESA land cover.) WorldPop has been shown to produce accurate disaggregations in many locations (see for example, ~~Bai et al., 2018; Chen et al., 2020; Mohanty and Simonovic, 2021)~~Bai et al., 2018; Chen et al., 2020; Mohanty and Simonovic, 2021) and is widely used particularly in health applications. Like GPW, WorldPop is highly transparent in the methods and underlying inputs used in its creation.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

2.2.4 LandScan

Oak Ridge National Laboratory's LandScan data set uses input population data from a variety of sources, (including censuses, surveys, work and school registers, and other sources), and a ML-based dasymetric model (~~including use of~~using a wide-variety of covariates, including elevation and slope) to produce annual population estimates for 2000 and 2015 (it does not include 1990). It deviates from the other population data sets in that it aims to measure ambient population – that is, population distribution averaged over a 24-hour period, rather than census *de jure* measures linked to usual residence. LandScan Global is a 30 arc second population surface which is not directly comparable year over year since methodologies are updated with each release ~~(Bright and Coleman, 2001; Bright et al., 2016; Rose and Bright, 2014).~~(Bright and Coleman, 2001; Bright et al., 2016; Rose and Bright, 2014). Thus LandScan is not suitable for use as a time-series. The ML model used by LandScan is proprietary, so the efficacy of covariate sources cannot be evaluated, and LandScan is not free and publicly available for non-commercial and commercial use. Nevertheless, LandScan is a spatial population data set that is often used in policy-making ~~(Leyk et al., 2019)~~(Leyk et al., 2019a) and has produced accurate disaggregations in many locations. For applications requiring ambient rather than *de jure* population estimates, LandScan is suitable.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

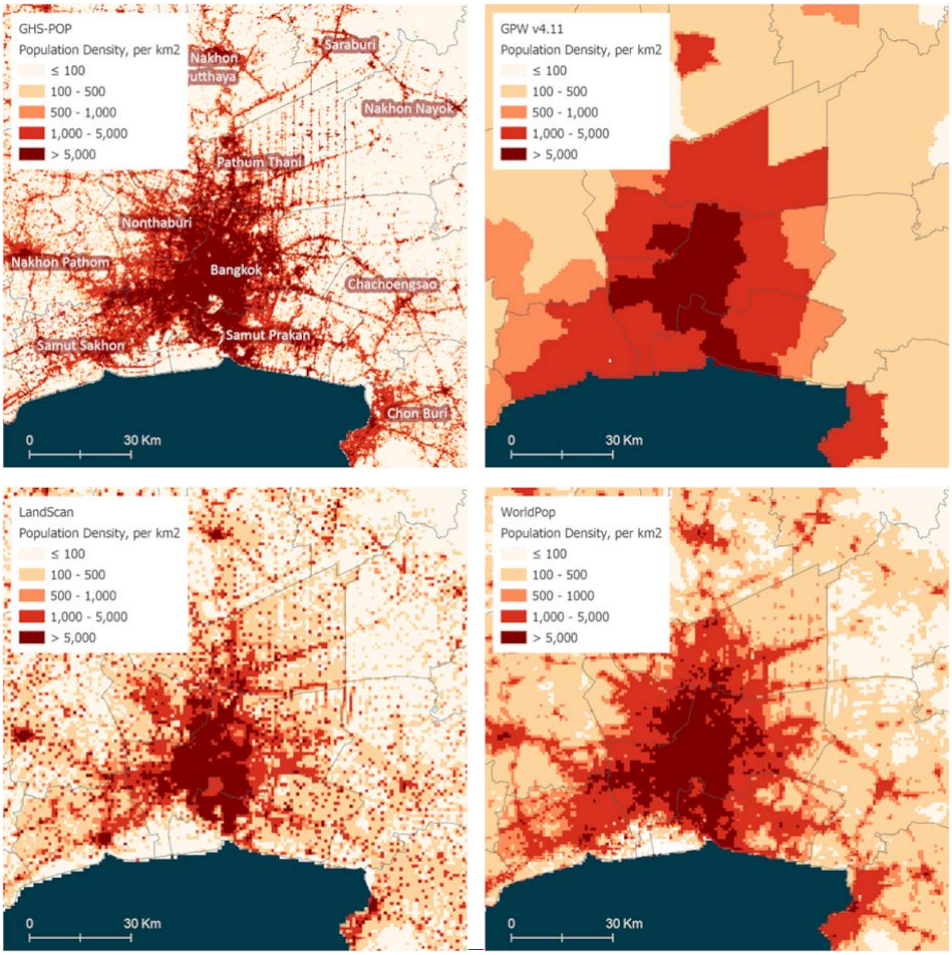
2.2.5 Core data choice - Population

Unlike physical data (elevation), the evaluation of the accuracy of gridded population data is complicated by the unavailability of baseline population estimates. Estimates from census and survey sources are static, and human mobility makes it difficult to validate those sources. ~~Leyk et al., 2019~~Leyk et al., 2019a discuss the strengths and weaknesses of global population data sets in great detail, and help data users select the best data for their specific use. For our analysis, being able to construct estimates of population in the LECZ for a 25-year interval (from 1990-2015) was important, in order to evaluate population change in different settlement types. ~~Therefore,~~In this study, we chose GHS-POP as our core population data set because ~~GHS-POP is~~it does not use elevation to reallocate population (as do LandScan and WorldPop), and represents the ~~only data representing a true~~longest time series ~~in regards to the underlying spatial structure~~(back to 1990) in that it uses built-up data for each of its target years to allocate population (rather than interpolating or extrapolating estimates of population based only on growth rates, as in GPW) , and was acceptable in other regards as mentioned above~~, it was chosen as our core population data set~~.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black


GHS-POP
Population Density, per km2
≤ 100
100 - 500
500 - 1,000
1,000 - 5,000
> 5,000
Saraburi
Nakhon Nayok
Pathum Thani
Nonthaburi
Nakhon Pathom
Bangkok
Chachoengsao
Samut Prakan
Samut Sakhon
Chon Buri
0
30 Km
GPW v4.11
Population Density, per km2
≤ 100
100 - 500
500 - 1,000
1,000 - 5,000
> 5,000
0
30 Km
LandScan
Population Density, per km2
≤ 100
100 - 500
500 - 1,000
1,000 - 5,000
> 5,000
0
30 Km
WorldPop
Population Density, per km2
≤ 100
100 - 500
500 - 1000
1,000 - 5,000
> 5,000
0
30 Km

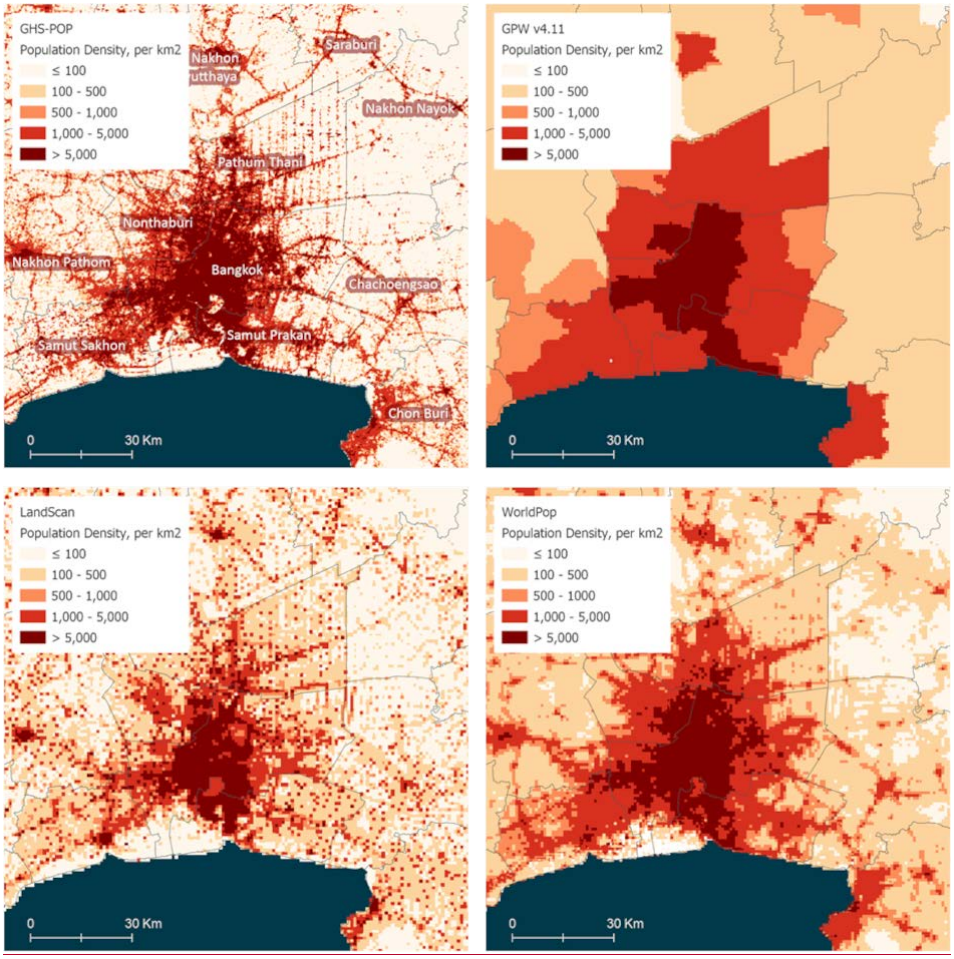


Figure 2. Population source data, Bangkok and surrounding areas, Thailand, 2015. Note that the dark blue indicates ocean and gray boundaries indicate first-order administrative boundaries. (These images show the processed 9-arc second data rather than the higher resolution inputs, given that the resolution differences are hard to detect in this visualization.)

2.3 Data on Urban Proxy

There is no authoritative source of data to delineate urban areas in an internationally comparable manner. Indeed, national statistical agencies and different disciplines have different ways to measure urbanization ~~(Buettner, 2015; Uchiyama and Mori, 2017).~~(Buettner, 2015; Uchiyama and Mori, 2017). Yet since the Global Rural Urban Mapping Project ~~(Balk, 2009; Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2021)~~(Balk, 2009; Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2021) – the first-ever global, spatial rendering of urban areas and the data set used in McGranahan et al. (~~2007~~2007a) – as with the above advances in elevation data and population models, there have been many advancements in data, models and methods which have led to many new data sets that aim to ~~to~~ capture settlements across the urban-rural continuum. All of these various new efforts to capture urban areas use proxy data sets – such

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

as night-time lights (dLIGHT, GRUMP) or built-up area (GHS-BUILT, GHS-SMOD) – that measure conceptually different dimensions of urban. The data sets selected for this study are listed in Table 3. Two of the data sets we consider represent physical processes whose spatial concentration is closely related to urban settlement (dLIGHT, GHS-BUILT) while two are more heavily modelled with the goal of urban classification (GRUMP, GHS-SMOD). All of the new underlying inputs (not including GRUMP) can be expressed as continuous data, which is important for representing a fuller urban-rural continuum ~~(Dorélien et al., 2013).~~(Dorélien et al., 2013), here we classify the urban proxy inputs into three large classes, described in ~~the methods section below~~Table 4.

Table 3. Urban Proxy data sets used to estimate persons living in the Low Elevation Coastal Zones (LECZ)

Source Data Set on Urban Proxies	**Abbreviation**	**Input Spatial Resolution**	**Temporal Resolution***	**Paper Reference**
Global Rural Urban Mapping Project - Urban Extents Grid v1	GRUMP	30 arc second	1994-95	~~(Balk, 2009)~~(Balk, 2009)
VIIRS Plus DMSP Change in Lights (VIIRS+DMSP dLIGHT), v1 (1992, 2002, 2013)	dLIGHT	15 arc second	1992, 2002, 2013	~~(Small and CIESIN- Columbia University, 2020)~~(Small and CIESIN- Columbia University, 2020)
Global Human Settlement-Built-Up Grid r2018a	GHS-BUILT	9 arc second	1975, 1990, 2000, 2014*	~~(Corbane et al., 2018)~~(Corbane et al., 2018)
Global Human Settlement - "Degree of Urbanization" model Grid r2019a v2	GHS-SMOD	30 arc second	1990, 2000, 2015	~~(Pesaresi et al., 2016)~~(Pesaresi et al., 2016)

*The temporal resolution represents acquisition years of the underlying satellite data, for GRUMP, dLight and GHS-BUILT. GHS-BUILT creates 'epochs', that is, the period by which observation(s) was (were) made, in contrast, VIIRS and the night-time lights on which GRUMP was based represent observations from a more narrow temporal range (such as one year). The temporal resolution of GHS-SMOD, being based on GPW and GHS-BUILT, indicates the specified target output year of the variables in question.

2.3.1 GRUMP

The Global Rural Urban Mapping Project (GRUMP) Urban Extents Grid v1 was the first gridded global data product to delineate urban areas ~~(Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2021).~~(Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2021). This was accomplished through the use of stable-city (Nighttime) Lights observations from the Defense Meteorological Satellite Program Operational Line Scanner (DMSP-OLS) circa 1995 ~~(Elvidge et al., 1999; Small et al., 2005)~~(Elvidge et al., 1999; Small et al., 2005) and confirmed by the presence of a named settlement above a certain population size (5,000 persons, where the data collection permitted). To address gaps in DMSP-OLS, these data were supplemented by "alternate sources (e.g. Tactical Pilotage Charts), or approximated by circles whose sizes were given by population–area relationships calibrated (through a regression analysis) on existing data" ~~(Balk, 2009).~~(Balk, 2009). GRUMP is distributed at a 30 arc second horizontal resolution. While GRUMP has been widely used, because its urban footprint is based on the stable-city lights known for their blooming quality (Small et al., 2005), it is well known to be an inclusive measure of urban extents. Furthermore, the spatial extent of the urban area represents a simple dichotomy: urban or rural. For this reason, GRUMP (like the original SRTM-based LECZ) is only included here for the purpose of comparison with the original study.

2.3.2 dLIGHT

Because night-time lights have been shown to be a good proxy for economic activity ~~(Henderson et al., 2012; Donaldson and Storeygard, 2016; Ghosh et al., 2009)~~(Henderson et al., 2012; Donaldson and Storeygard, 2016; Ghosh et al., 2009) and because the spatial concentration of economic activity is associated with urban location, night-time lights data products continue to be a valuable data source as an urban proxy ~~(Hu et al., 2020).~~(Hu et al., 2020). The VIIRS Plus DMSP Change in Lights, v1 (dLIGHT) data set depicts the relative luminosity in stable lights areas (as determined by VIIRS annual composites for the year 2015) for the years 1992, 2002, and 2013 respectively ~~(Small and Center For International Earth Science Information Network-CIESIN-Columbia University, 2020).~~(Small and

Center For International Earth Science Information Network-CIESIN-Columbia University, 2020). The dLIGHT data set combines nighttime lights imagery from DMSP-OLS with a stable night light composite from Suomi National Polar-orbiting Partnership (NPP) Visible Infrared Imaging Suite (VIIRS) Day-Night Band in a 15 arc second horizontal resolution grid. While dLIGHT makes great improvements in resolution and accuracy over DMSP-OLS, it represents relative changes in brightness of the DMSP-OLS sensor, and is constrained to lit areas based on the 2015 VIIRS data ~~(Small et al., 2018a).~~(Small et al., 2018a). Like some of the data sets used here, dLIGHT is a new data product and has not been used extensively with other data sets indicating the spatial extents (and change thereof) of urban areas. While there has been continued improvement to reduce gas flares from the underlying data products, because gas flares are not expected to be associated with urban areas ~~(Elvidge et al., 2009),~~(Elvidge et al., 2009), it is also understood that not all economic or human activity is accompanied by light sources, particularly in poor economies or in particular land covers (such as deserts) ~~(Stokes and Seto, 2019).~~(Stokes and Seto, 2019).

Formatted: Font: 12 pt, Font color: Black

2.3.3 GHS-BUILT

Another approach to urban representation comes from a land-use perspective. Early work in this area classified pixels from moderate resolution satellite-data products detecting vegetation (e.g. Normalized Difference Vegetation Index (NDVI)) that were typically heterogeneous and residual, and thus did not represent specific land-use classes found outside urban localities ~~(Potere et al., 2009; Schneider et al., 2010):~~(Potere et al., 2009; Schneider et al., 2010): in other words, this approach did not directly detect structures or features typically concentrated spatially in urban areas (buildings, roads, etc). Developments in ML approaches and related modelling have led to a new class of derived products that explicitly classify built features commonly concentrated in urban settings. Recent global-scale efforts include NASA SEDAC's Global Man-made Impervious Surface (GMIS), the European Commissions' Joint Research Center's GHSL, and the German Aerospace Center's World Settlement Footprint (WSF) projects ~~(Esch et al., 2018; Marconcini et al., 2020).~~(Esch et al., 2018; Marconcini et al., 2020). Here we use two products from the GHSL suite. GHS-BUILT represents estimations of built-up presence as derived from LandSat image collections. Built-up estimates are provided for the epochs 1975, 1990, 2000, and 2014 ~~(Florczyk et al., 2019).~~(Florczyk et al., 2019). At its core are more than 40,000 LandSat scenes which have been processed in a consistent manner across countries and over time using advanced machine learning algorithms ~~(Pesaresi et al., 2016).~~(Pesaresi et al., 2016). The 1 arc second data are binary, indicating either the presence or absence of a built structure and are aggregated to 3 arc second to represent the fraction of built-up land in each pixel. This data set has been cross-validated or analyzed with census-designated classes of urbanization in the recent studies of the U.S ~~(Balk et al., 2018; Leyk et al., 2018).~~(Balk et al., 2018; Leyk et al., 2018), and generally confirmed the accuracy of the GHSL algorithms except in very sparsely settled rural regions (Leyk et al., 2018). One feature of this data set that some would consider to be a disadvantage is that once a location is detected as built-up, that location remains built and while it can become more built-up it cannot become un-built. Similarly, in the version of GHS-BUILT used here all built structures are treated equally: future versions of this data set will distinguish industrial built structure from other types.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

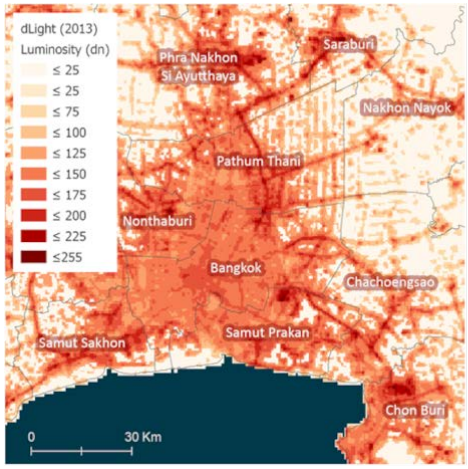
Formatted: Font: 12 pt, Font color: Black

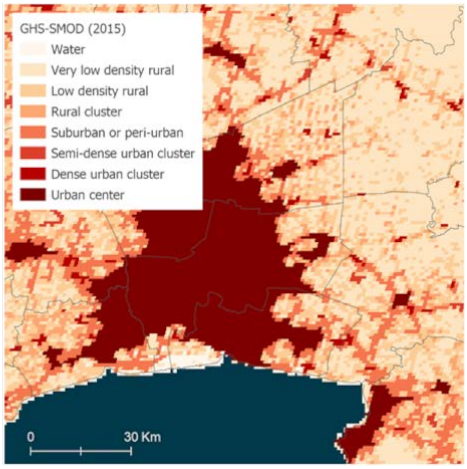
2.3.4 GHS-SMOD Degree of Urbanization (DoU)

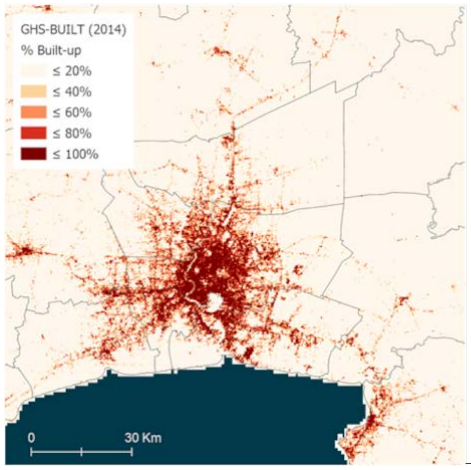
The Global Human Settlement - "Degree of urbanization" model Grid r2019a v2 (GHS-SMOD) delineates settlement types by modeling population size, and population and built-up area densities from GHS-POP and GHS-BUILT to construct a "degree of urbanization" grid (Florczyk et al., 2019). This modelled surface uses built-up area (GHS-BUILT) along with population data (GHS-POP) and a set of density and proximity criteria to classify population and land area into seven classes (plus a category for inland open water) along an urban-rural continuum. This new data product has not yet been cross-validated in the peer-reviewed literature, but such studies are underway and it has already been used in policy applications ~~(Henderson et al., 2012; OECD and European Commission, 2020; Colenbrander et al., 2019).~~(Henderson et al., 2012; OECD and European Commission, 2020; Colenbrander et al., 2019). The Degree of Urbanization methodology has been endorsed by the UN Statistical Commission as a means of identifying areas as being urban to different degrees ~~(Dijkstra et al., 2020, 2019; OECD and European Commission, 2020).~~(Dijkstra et al., 2020, 2019; OECD and European Commission, 2020).

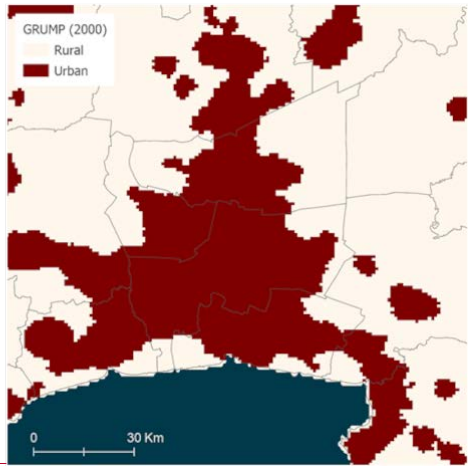
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


dLight (2013)
Luminosity (dn)
≤ 25
≤ 25
≤ 75
≤ 100
≤ 125
≤ 150
≤ 175
≤ 200
≤ 225
≤255
Phra Nakhon Si Ayutthaya
Saraburi
Nakhon Nayok
Pathum Thani
Nonthaburi
Bangkok
Chachoengsao
Samut Prakan
Samut Sakhon
Chon Buri
0
30 Km


GHS-SMOD (2015)
Water
Very low density rural
Low density rural
Rural cluster
Suburban or peri-urban
Semi-dense urban cluster
Dense urban cluster
Urban center
0
30 Km


GHS-BUILT (2014)
% Built-up
≤ 20%
≤ 40%
≤ 60%
≤ 80%
≤ 100%
0
30 Km


GRUMP (2000)
Rural
Urban
0
30 Km

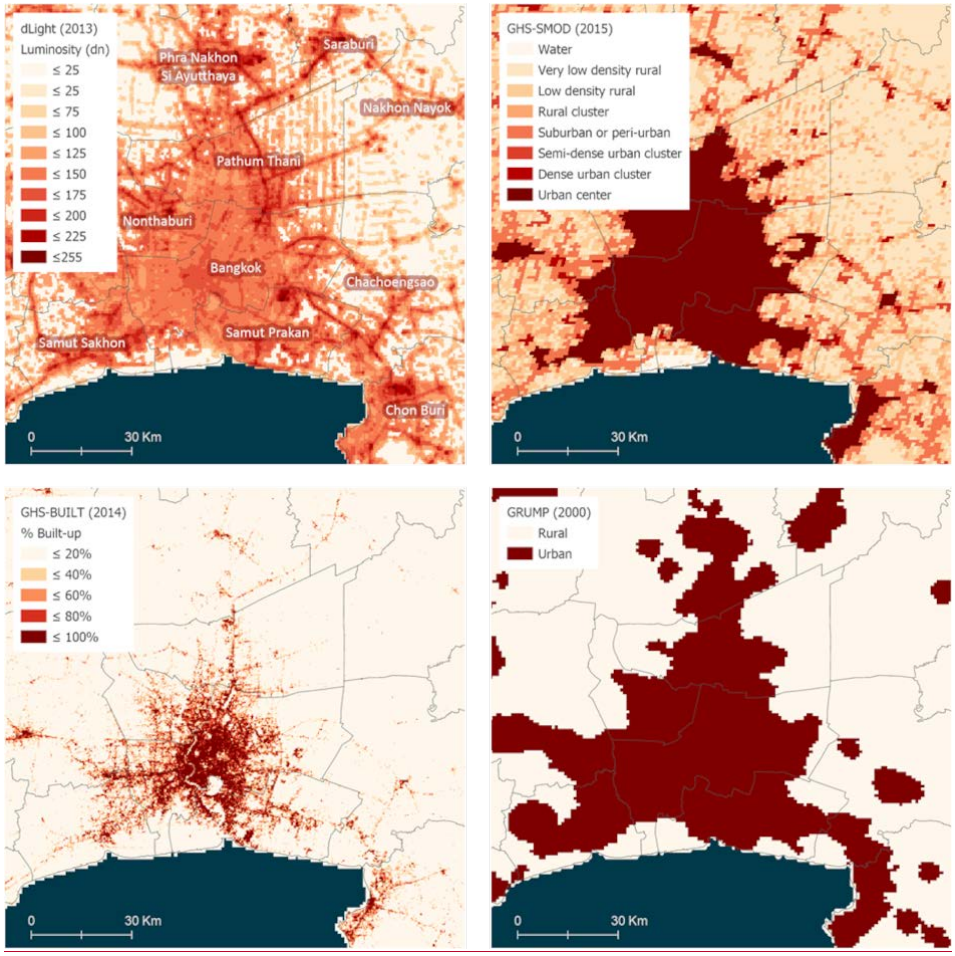


Figure 3. Urban proxy source data, Bangkok and surrounding areas, Thailand. Note that the dark blue indicates ocean and gray boundaries indicate first-order administrative boundaries.

2.3.5 Core data choice - Urban Proxy

The choice of a core data set to delineate urban areas was based on three criteria: availability of time-series, consistency with the population data, and intentionality to capture a continuum of urban locations. Of the four data sets included, only GHS-SMOD (Dijkstra et al., 2020) and GRUMP (Balk et al., 2005) claim by design to represent urban extents. Between these, we select GHS-SMOD as the core data set for this analysis as it allows us to consider change over time (and allows for the longest temporal comparison). GHS-SMOD classifies grids cells into an urban-rural continuum based directly on GHS-POP and GHS-BUILT data for each epoch, which makes population and urban proxy data spatially consistent. Nevertheless, given that validation of the GHS-SMOD is only just under way, we reduced the seven native classes (GHS-SMOD Level 2, Florczyk et al., 2019) to three broad classes (Level 1) as described below.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.4 Other data sets

In addition to the elevation, population, and urban proxy data sets described in the preceding section, a number of ancillary data sets were also considered.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.4.1 National Identifier Grid

~~The GPW collection (Center for International Earth Science Information Network - CIESIN - Columbia University, 2018)~~The GPW collection (Center for International Earth Science Information Network - CIESIN - Columbia University, 2018) includes an ancillary National Identifier Grid (NID) which we have used to represent the extents of countries and territories in this analysis in order to construct summary statistics for these units. GPW is the only one of the population data sets that includes this information. (None of the elevation or urban proxy data sets include this information.) The horizontal resolution of the NID is 30 arc seconds.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.4.2 Area Grids

The land area grid from the GPW Land and Water Area data set ~~(Center for International Earth Science Information Network - CIESIN - Columbia University, 2018)~~(Center for International Earth Science Information Network - CIESIN - Columbia University, 2018) forms the basis of the land area estimates in this study. The land area grid is a surface which accounts for the reduction in the underlying area of regular rectangular grid cells as they approach the poles. This allows for accurate area measurements without requiring the use of an Equal Area projection.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Additionally, a Mean Administrative Unit Area raster is part of the GPW collection's data quality indicators data set ~~(Center for International Earth Science Information Network - CIESIN - Columbia University, 2018).~~(Center for International Earth Science Information Network - CIESIN - Columbia University, 2018). It represents the nominal resolution of input vector geographies which were matched to census population estimates prior to gridding. Since GPW population counts and density data are created with a uniform allocation method, the Mean Administrative Unit Area raster is essential for understanding the precision and accuracy of pixel level population estimates across and within countries.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.4.3 Built-up Density

GHS-BUILT is the building block of the GHSL data collection; it is a multitemporal information layer on built-up presence as derived from LandSat image collections (GLS1975, GLS1990, GLS2000, and ad-hoc Landsat 8 collection 2013/2014~~)(Corbane et al., 2018)).~~)(Corbane et al., 2018)). In addition to forming the basis for GHS-POP and GHS-SMOD, the GHS-BUILT data set at its core provides information on the density (sometimes referred to as intensity) or percentage of land that is developed with built-structures. Measured as whether a 3 arc second pixel is made up of more built surfaces than not, then aggregated to 9 arc second to represent the percentage of cell that is built, the fraction "built-up" ranges from 0-100. (GHSL measures area, not volume such as the vertical dimension of built-up areas or cities.) Elsewhere these data have been used to describe change in the extent or footprint of the urban environment ~~(Balk et al., 2018)~~(Balk et al., 2018) and as an indicator for urbanization of land area ~~(Liu and Balk, 2020; Pinchoff et al., 2020; Gao and O'Neill, 2020).~~(Liu and Balk, 2020; Pinchoff et al., 2020; Gao and O'Neill, 2020). We use it independently to evaluate how much land in the LECZ is built-up.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

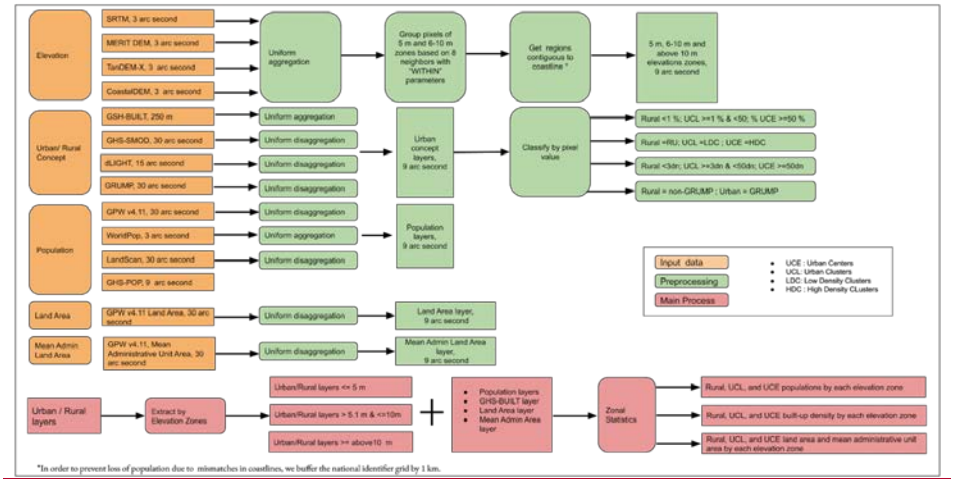
Formatted: Font: 12 pt, Font color: Black

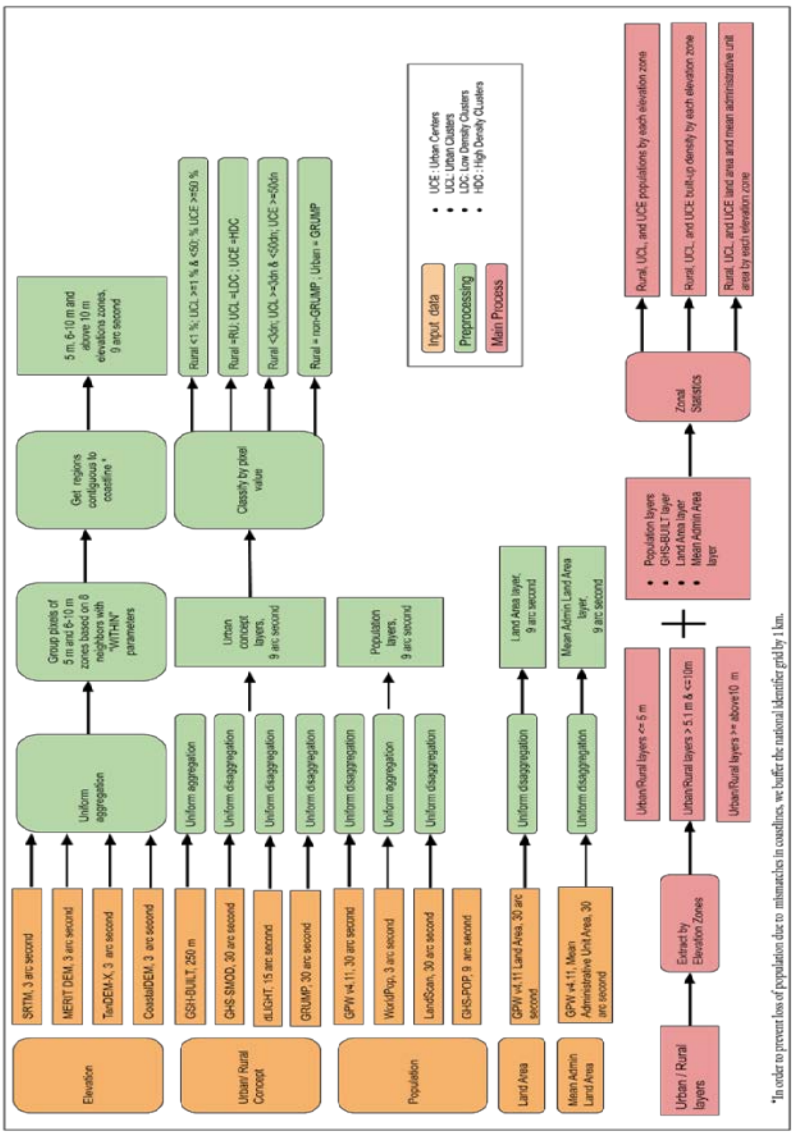
2.5 Methods

As shown in Fig. 4, the methodologies used to produce the new data layers for this study (i.e., LECZ and urban-rural classifications) followed this sequence: first, elevation data were preprocessed into common frameworks and subset by country boundaries as defined by the GPW NID. Then, areas contiguous to the coast were identified to create LECZs that are classified as up to 5m, 5-10m and above 10m. Next, urban proxy data sets were conditioned into common thematic classes (classified broadly as urban, quasi-urban and rural), harmonized into a common horizontal resolution and subset by the NID. Similarly, population data sets were harmonized into a common horizontal resolution and subset. Area data from the GPW land area grid and Mean Administrative Unit area grid, along with built-up percentages from GHS-BUILT were also harmonized and subset. Finally, using spatial overlays and zonal statistics, we constructed estimates of population, land area and built-up density that are summarized by country, urban class, and LECZ. The methodology is depicted in the flow chart (Fig. 4) and described in detail below.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


Elevation
SRTM, 3 arc second
MERIT DEM, 3 arc second
TanDEM-X, 3 arc second
CoastalDEM, 3 arc second
Uniform aggregation
Group pixels of 5 m and 6-10 m zones based on 8 neighbors with "WITHIN" parameters
Get regions contiguous to coastline *
5 m, 6-10 m and above 10 m elevations zones, 9 arc second
Urban/ Rural Concept
GSH-BUILT, 250 m
GHS-SMOD, 30 arc second
dLIGHT, 15 arc second
GRUMP, 30 arc second
Uniform aggregation
Uniform disaggregation
Uniform disaggregation
Uniform disaggregation
Urban concept layers, 9 arc second
Classify by pixel value
Rural <1 %; UCL >=1 % & <50; % UCE >=50 %
Rural =RU; UCL =LDC ; UCE =HDC
Rural <3dn; UCL >=3dn & <50dn; UCE >=50dn
Rural = non-GRUMP ; Urban = GRUMP
Population
GPW v4.11, 30 arc second
WorldPop, 3 arc second
LandScan, 30 arc second
GHS-POP, 9 arc second
Uniform disaggregation
Uniform aggregation
Uniform disaggregation
Population layers, 9 arc second
Input data
Preprocessing
Main Process
UCE : Urban Centers
UCL: Urban Clusters
LDC: Low Density Clusters
HDC : High Density CLusters
Land Area
GPW v4.11 Land Area, 30 arc second
Uniform disaggregation
Land Area layer, 9 arc second
Mean Admin Land Area
GPW v4.11, Mean Administrative Unit Area, 30 arc second
Uniform disaggregation
Mean Admin Land Area layer, 9 arc second
Urban / Rural layers
Extract by Elevation Zones
Urban/Rural layers <= 5 m
Urban/Rural layers > 5.1 m & <=10m
Urban/Rural layers >= above10 m
Population layers
GHS-BUILT layer
Land Area layer
Mean Admin Area layer
Zonal Statistics
Rural, UCL, and UCE populations by each elevation zone
Rural, UCL, and UCE built-up density by each elevation zone
Rural, UCL, and UCE land area and mean administrative unit area by each elevation zone
*In order to prevent loss of population due to mismatches in coastlines, we buffer the national identifier grid by 1 km.



505 **Figure 4. Flow chart describing methods used to construct population and land area along an urban continuum in Low Elevation Coastal Zones (LECZs)**

2.5.1 Elevation

The LECZ are constructed from elevation data with one main rule applied to it: contiguity to coast-line. We construct two zones – ~~0~~ below 5 meters (including 5.0 and where below 0 is rounded up to 0) and 5-10 (including 10.0) meters

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

contiguous to coast – and compare these with all other areas within a country, that is, those areas above 10 meters (or at or below 10 meters, but not contiguous to coastline). The ≤10m LECZ is constructed by combining the ≤5m and 5-10m zones.

Elevation data from four sources were used, each projected to WGS84 horizontal coordinate system with EGM96 geoid heights: ~~MERIT, SRTM, TanDEM-X, CoastalDEM (see Table 1).~~MERIT, SRTM, TanDEM-X, CoastalDEM (see Table 1). In vertical terms, these elevation data models aim to set zero elevation at mean sea level using global datums with local variation. Out of the 4 DEMs evaluated, 3 of them (SRTM, MERIT, CoastalDEM) were referenced to the EGM96 Vertical Coordinate System (EPSG:5773). Only TanDEM-X was not. TanDEM-X 90 elevations are referenced to the WGS84 (G1150) ellipsoid (EPSG:4979). Therefore, TanDEM-X was converted from its native WGS84 ellipsoidal heights to EGM96 geoid heights using the GDAL Warp tool. Each of these high resolution DEMs were obtained from data distributions which followed regular, but unique tiling schemes. Tiling of high resolution raster data is often necessary to control for file size and usability (e.g. memory footprint), with the cost of complicating global scale analyses when different data sets use their own schemes. Therefore, each of the DEMs were preprocessed into country units to enable the ultimate goal of country scale analyses, and to harmonize the objects to be processed apart from their unique tiling schemes. This was accomplished by loading the elevation tiles into an ESRI File Geodatabase Mosaic Data set, which includes vector layers (footprints) of the input raster extents that identify the file name and location of each input.

Next, a python script was used to clip the vectorized raster footprints by country boundaries extracted from the NID. This created country level layers with attributes (file names and locations) from intersecting footprints for each of the elevation sources which were used to isolate a subset list of elevation tiles belonging to a given country. Those subset lists were then mosaicked into country specific DEMs using the ArcGIS Mosaic to New Raster tool with the MEAN mosaic method; when a country was completely covered by a single tile, that tile was simply used without need for mosaic. All of the elevation data were then aggregated with the MEAN method of the ArcGIS Aggregate tool to a 9 arc second horizontal resolution. The following additional steps were taken to ensure that the coastlines and coastal regions were adequately identified in our processes.

2.5.2 Determining Coastal Contiguity

Buffering the coastline

There is no international standard for coastlines, and administrative boundary data sets may or may not conform strictly to the physical reality of the coastline ~~(Mcleod et al., 2010).~~(Mcleod et al., 2010). Elevation data sets sometimes include representations of coastlines, but this too may differ between sources: for example, SRTM, MERIT-DEM and TanDEM-X use different implied coastline beyond which elevation is assumed to be zero but CoastalDEM does not. This discordance in the definition of a coastline occurs for many reasons including (1) administrative boundaries that intentionally include water bodies for which there is jurisdiction; (2) coarse scale administrative boundaries that are likely to be imprecise with respect to the physical coastline; and (3) the nature of physical coastlines to change over time (daily, monthly, yearly), which impacts both administrative and elevation sources based on the date of their data capture.

The problem of coastline disagreement is compounded for gridded data, where precise vectorized coastlines are pixelated in accordance with the raster resolution. The NID used in this study to represent coastlines has a native resolution of 30 ArcSeconds which implies some imprecision. However, the NID was used so that zonal statistics of population grids could capture, and not double count, every populated pixel in one and only one country. In this analysis, where the overlay of the administrative data with elevation data at the coastline matters for estimation, alignment between the input spatial layers is paramount. This is particularly true for those small island nations where the majority of their land area is coastal, and therefore mismatches can lead to substantial ~~misestimation~~misestimating of land area and population in the LECZ.

In order to prevent the loss of population due to coastline mismatches, or the loss of LECZ land area when the elevation data source uses a coastline that is seaward of the NID, the NID is buffered by 1 km on a per country basis in order to create an inclusive coastline definition which accounts for imprecision. Examples of these problem areas – and with and without this buffer – are shown in Fig. 5 for the case of Sri Lanka. The inclusive version was utilized in this work.

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

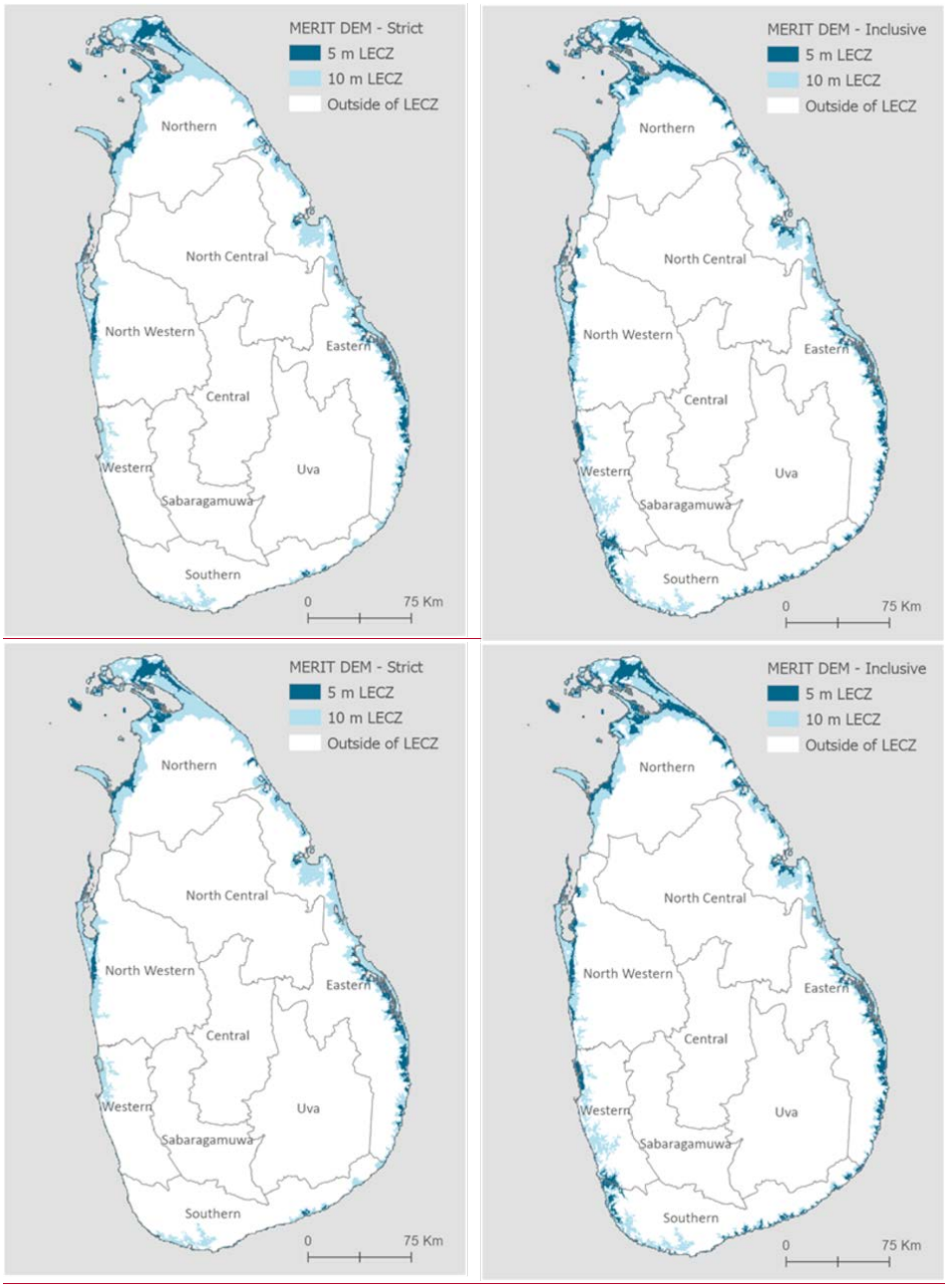

MERIT DEM - Strict
5 m LECZ
10 m LECZ
Outside of LECZ
Northern
North Central
North Western
Eastern
Central
Western
Uva
Sabaragamuwa
Southern
0
75 Km
MERIT DEM - Inclusive
5 m LECZ
10 m LECZ
Outside of LECZ
Northern
North Central
North Western
Eastern
Central
Western
Uva
Sabaragamuwa
Southern
0
75 Km
MERIT DEM - Strict
5 m LECZ
10 m LECZ
Outside of LECZ
Northern
North Central
North Western
Eastern
Central
Western
Uva
Sabaragamuwa
Southern
0
75 Km
MERIT DEM - Inclusive
5 m LECZ
10 m LECZ
Outside of LECZ
Northern
North Central
North Western
Eastern
Central
Western
Uva
Sabaragamuwa
Southern
0
75 Km

Figure 5. MERIT-DEM LECZ constructed strictly and inclusively (with 1km buffer), Sri Lanka.

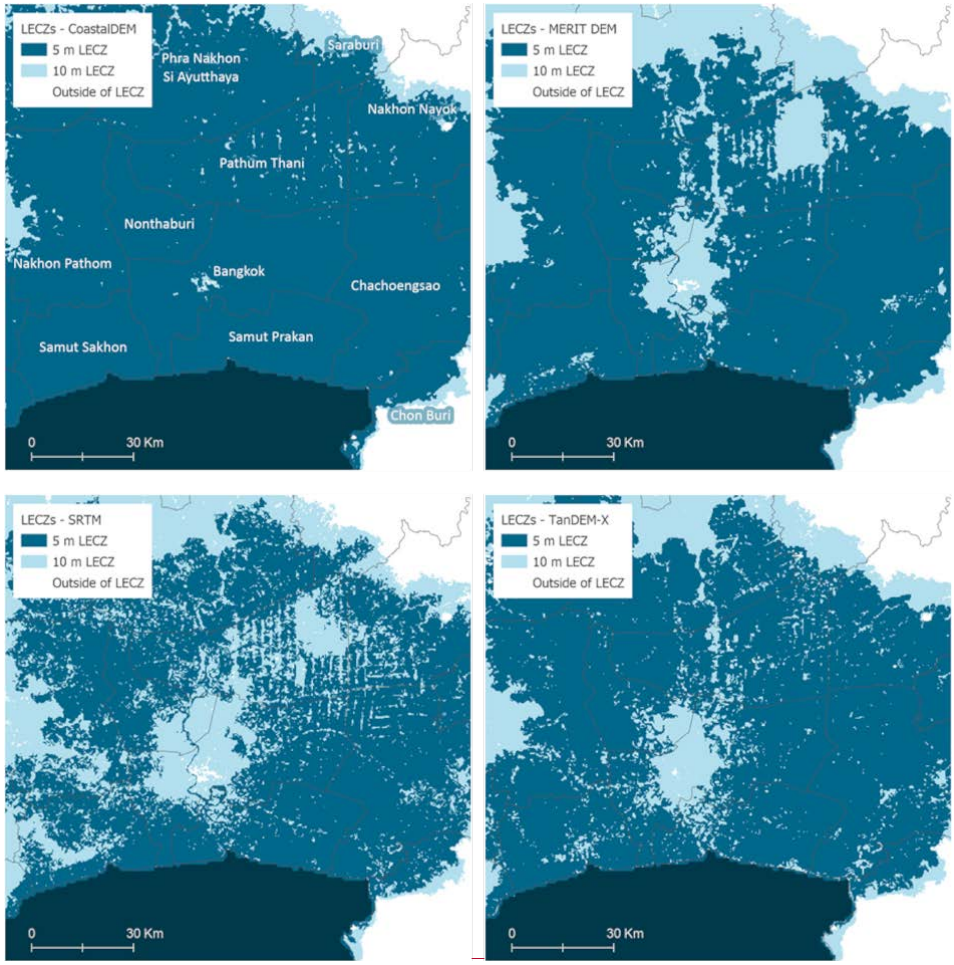
2.5.3 Isolating Coastally Contiguous Regions

The 9 arc second country elevation mosaics for each elevation source were reclassified into integers for the following zones: ≤5m, 5 to 10m and greater than 10m. For example, all continuous values less than or equal to 5 were assigned a value of 5, all values greater than 5 and less than or equal to 10 were assigned a value of 10, and all values greater than 10 or not contiguous to coast below 10m were assigned an arbitrary value of 31. The reclassified images were extracted by attribute into ≤5m, and ≤10m rasters, and were then segmented with the ArcGIS "Region Group" tool with eight neighbors using the WITHIN parameter. Region-grouped images are those where groups of pixels with like values are combined such that each connected group (region) receives its own unique identifier along with a count of the number of pixels within the grouping (for example see cute cat picture in the appendix Fig. B1). In order to isolate coastally contiguous regions, the region-grouped images were converted into polygons and selected by location where each polygon intersected the border of a country as determined from the 1km buffered NID. This effectively isolated all regions connected to administrative boundaries. Since this could potentially include inland areas, each of the files were visually inspected in order to identify spurious lowland areas contiguous with inland country boundaries (although laborious, this quality check was completed within 1-2 days of effort.) When errors were discovered, they were manually removed. The isolated, quality assured regions were then used as extraction masks on the reclassified DEMs and null inland values were coded as above 10m (the corresponding value in our resulting spatial data is coded as 31). The resulting rasters contained coastally contiguous pixels coded into ≤5m and 5-10m LECZs, and a third category representing the area outside of LECZs. Figure 6 shows the final LECZ designations for Bangkok Thailand by elevation source.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black


LECZs - CoastalDEM
5 m LECZ
10 m LECZ
Outside of LECZ
Phra Nakhon Si Ayutthaya
Saraburi
Nakhon Nayok
Pathum Thani
Nonthaburi
Nakhon Pathom
Bangkok
Chachoengsao
Samut Sakhon
Samut Prakan
Chon Buri
0
30 Km
LECZs - MERIT DEM
5 m LECZ
10 m LECZ
Outside of LECZ
0
30 Km
LECZs - SRTM
5 m LECZ
10 m LECZ
Outside of LECZ
0
30 Km
LECZs - TanDEM-X
5 m LECZ
10 m LECZ
Outside of LECZ
0
30 Km

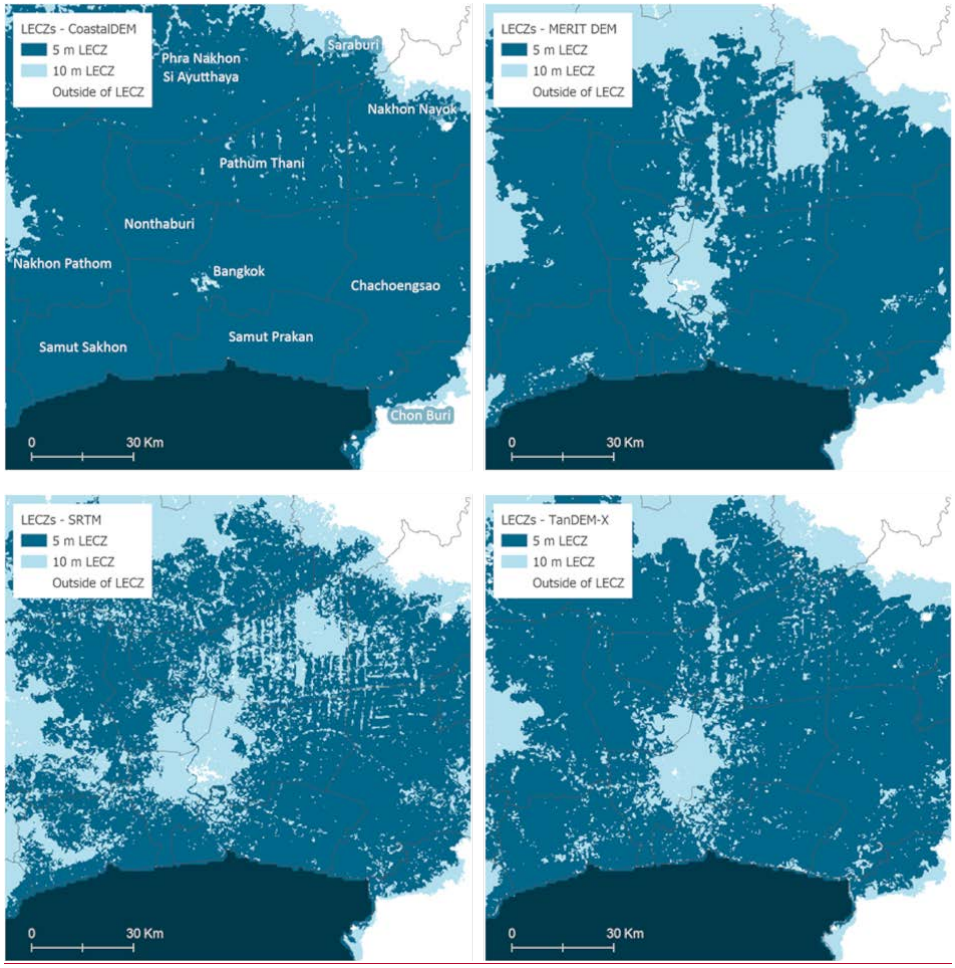


Figure 6. Low Elevation Coastal Zones (LECZ) constructed from source DEMs, Bangkok and surrounding areas, Thailand. Note that the darkest blue indicates ocean and gray boundaries indicate first-order administrative boundaries.

2.5.4 Population

As introduced in Table 2, four population sources were utilized: GHS-POP (1990, 2000, 2015), GPW (1990, 2000, 2015), WorldPop (2000, 2015), and LandScan (2000, 2015). The horizontal resolution of these data sets vary: WorldPop is 3 arc second, GHS-POP is 9 arc second, and both GPW and LandScan are 30 arc second. Therefore, methods for constructing comparable resolution population data sets and subsetting into countries varied as follows: (1) WorldPop was aggregated from 3 arc second to 9 arc second using the ArcGIS Aggregate tool with the SUM method and then subset; (2) GHS-POP was simply subset at its native 9 arc second resolution; and (3) GPW and LandScan were uniformly disaggregated by a factor of 100 (e.g., 1 pixel was divided into 100 pixels, given its 1km resolution) and quality assured to have the same total population before and after the sampling, then subset by country. Population distributions shown in Fig. 2 represent these data processed to 9-arc second (nominally 300m) resolution.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.5.5 Urban Proxy

Official UN urban population statistics ~~(United Nations, 2018)~~(United Nations, 2018) are based on the very wide range of country-specific procedures for classifying areas as urban. This variation in urban definitions presents significant challenges in making international urban comparisons. Further, these statistics do not correspond to a spatial data set, making it impossible to use with spatial data to estimate urban (or rural) populations residing in LECZ. Thus, leveraging recent global efforts (e.g., ~~Dijkstra et al., 2020, 2019; Pesaresi et al., 2019; Florczyk et al., 2019; Small et al., 2018; Corbane et al., 2018; Balk, 2009),~~Dijkstra et al., 2020, 2019; Pesaresi et al., 2019; Florczyk et al., 2019; Small et al., 2018; Corbane et al., 2018; Balk, 2009), and precedent used in McGranahan et al. (~~2007~~2007a), we rely on satellite depictions to distinguish settlements and places along an urban-rural continuum. As in Table 3, four data sources were used: ~~GHS-SMOD; GHS-BUILT, GRUMP; and dLIGHT.~~GHS-SMOD; GHS-BUILT, GRUMP; and dLIGHT. The horizontal resolution of GHS-BUILT is 9 arc second, dLIGHT is 15 arc second, and both GHS-SMOD and GRUMP are 30 arc second. All of these data sets were natively in the WGS84 coordinate system except for GHS-BUILT which is natively in the World Mollweide Equal Area Projection. As with population, these data were conditioned into a common 9 arc second horizontal resolution through uniform upsampling, but using a Nearest Neighbor approach since the underlying data is categorical. GHS-BUILT was also projected into the WGS84 coordinate system with Nearest Neighbor cell assignment at 9 arc second. All of these data sets were subset by country using the NID.

2.5.6 Constructing Classes along an Urban-Rural Continuum

While the underlying urban proxy data (Table 4) are continuous or ordinal along many classes, for the purposes of our summaries here we constructed three simplified, common thematic categories: Urban, Quasi-Urban, and Rural. It was not possible to do this for the GRUMP data set, which includes only two classes, described urban and rural, but which is nonetheless summarized here in order to compare with the benchmark study by ~~McGranahan et al., 2007b,~~McGranahan et al., 2007a, which was the first of its kind to delineate urban population in the LECZ.

It is worth noting that this represents an important improvement from estimates in ~~McGranahan et al., 2007b,~~McGranahan et al., 2007a, and that newer, more recent estimates of global populations in the LECZ ~~(Kulp and Strauss, 2019~~(Kulp and Strauss, 2019 which showcases CoastalDEM), do not stratify by any urban-rural classes. Other studies have highlighted population in case-study cities ~~(Small et al., 2018b; Ahmed et al., 2018; Khan et al., 2019)~~(Small et al., 2018b; Ahmed et al., 2018; Khan et al., 2019) but these are not global in extent; others have focused on types of cities (such as ports, ~~De Sherbinin et al., 2007~~De Sherbinin et al., 2007 or megacities, ~~Nicholls, 1995)~~Nicholls, 1995) at risk.

In creating a globally applicable urban-rural categorization, inserting a quasi-urban category serves to acknowledge that urban-rural is a continuum, and explicitly separates out a hard-to-classify and rapidly evolving, but not especially large, middle range of localities. Historically, emphasis on cities, and large ones at that, has been in part because these localities are populous, but also arguably consistent in some basic aspects of form (largely built-up for example, or with population densities above a given threshold), which makes them easier to identify in imagery ~~(L.Imhoff et al., 1997; Schneider et al., 2010).~~(Imhoff et al., 1997; Schneider et al., 2010). Similarly, areas identified as rural have a largely consistent signature ~~(Doll and Pachauri, 2010).~~(Doll and Pachauri, 2010). Debate arises over the treatment of the heterogeneous collection of places such as small towns, suburbs, and peri-urban settlements. Whether these should be considered urban is open to interpretation, and may not be discernible using features like nighttime lights, population density and built-up area. Many countries, for example, include administrative criteria or use others based on country-specific criteria in their identification of urban areas (United Nations, 2018). While such variation undermines international comparability, it can make the classification more useful locally. Variation also exists, however, among urban-rural allocation procedures designed to be internationally comparable, such as those included here, and remains even when this variation is mitigated somewhat by introducing the category of quasi-urban.

GHS-SMOD was dissolved using the ArcGIS reclassify tool from its native seven classes of settlement (level 2 classification), into three classes: urban, quasi-urban and rural. This type of aggregation is inherent to the GHS-SMOD data set as the level 1 classification structure (Florcyk et al., 2019). GHS-BUILT is made up of estimates of the percentage of built-up in a given 9 arc second pixel. The raw GHS-BUILT data was thresholded into urban (> 50% built-up), quasi-urban (> 3% and ≤ 50% built-up), and rural (≤ 3% built-up) using the ArcGIS Reclassify tool; as 3% built-up is used as a delineation of classes in GHS-SMOD that fall into the quasi-urban class, we used here for the threshold of GHS-BUILT as well. ~~dLIGHT (Small and Center For International Earth Science Information Network-~~

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Pattern: Clear, Highlight

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

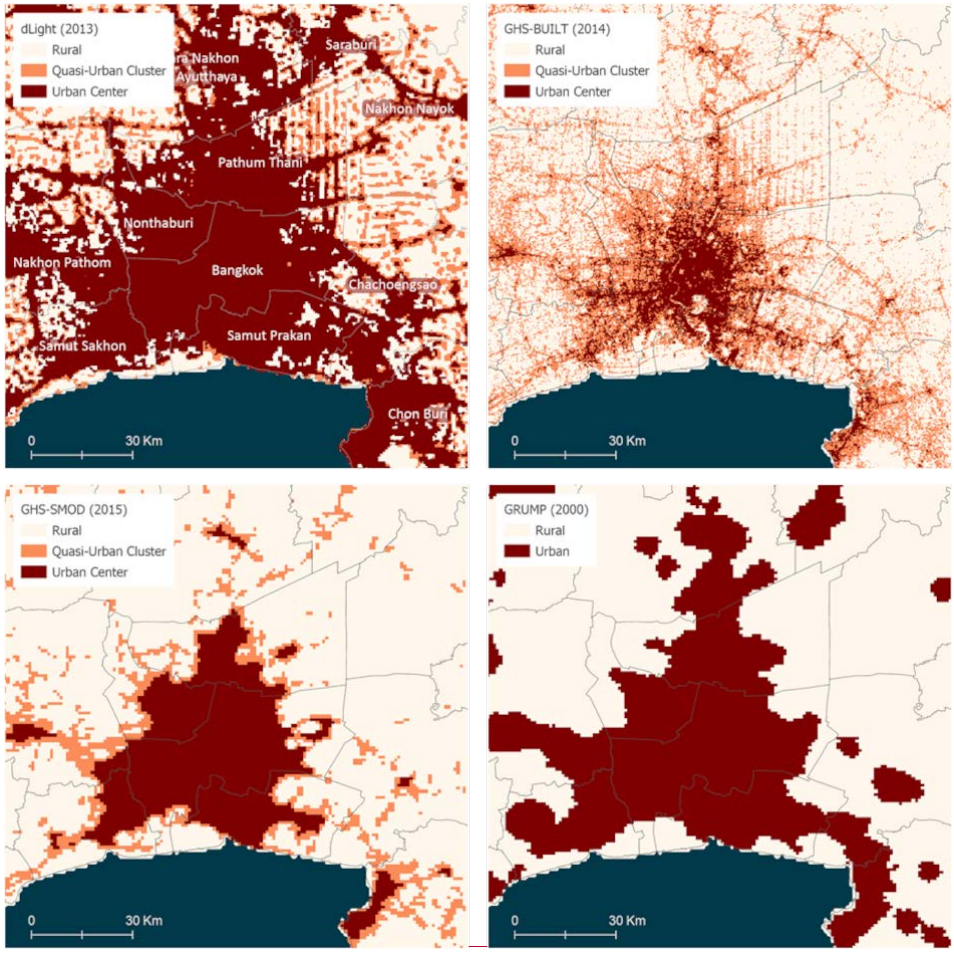
~~CIESIN-Columbia University, 2020)~~dLIGHT (Small and Center For International Earth Science Information Network-CIESIN-Columbia University, 2020) is made up of digital numbers (dn), 0 to 255, which represent the relative luminosity of pixels across the time periods represented in the data set (1992, 2002, 2013). Cross-classifying this with other urban depictions was done by visually comparing dLIGHT with GHS-SMOD and GHS-BUILT to find areas of agreement to guide thresholding. Based on this, the raw dLIGHT data was thresholded into urban (> 100 dn), quasi-urban (> 3 and < 100 dn), and rural (< 3 dn) using the ArcGIS Reclassify tool.

Table 4. Urban Proxy Data Sets: Specifications of Underlying inputs Classification Schema

		Criteria used for each urban proxy data set*		
Short formal description	**~~Intuitive description~~**	**~~GHS-SMOD~~****	**~~GHS-BUILT~~**	**~~dLight~~**
Urban Centers	Cities	A cell that is part of a cluster along with its 4 contiguous and directly adjacent grid cells, in which: ▪ Density ≥ 1,500 inhabitants/km2 within the cell and at least 50,000 inhabitants in cluster; or ▪ 50% built-up surface share on permanent land	> 50% built-up	> 100 dn
Quasi-Urban Clusters	Towns, peri-urban areas, suburbs	A cell that is not urban but is part of a cluster with its 8 adjacent or diagonally contiguous grid cells with: ▪ Density ≥ 300 inhabitants/km2 in the cell and at least 5,000 inhabitants in the cluster; or ▪ > 3% built-up surface share on permanent land	> 3% and ≤ 50% built-up	> 3 and < 100 dn
Rural grid cells	Rural areas	All cells not belonging to Quasi-Urban Clusters or Urban Centers ~~(i.e. generally, density < 300 inhabitants/km2~~ ~~Greater density is possible for rural cells if cells are not part of a cluster with sufficient total population to be classified as quasi-urban cluster or urban center~~	≤ 3% built-up	< 3 dn

*GRUMP is constructed as a dichotomous urban-rural grid. Due to its construction, it is known to include a lot of land area that might be classified as quasi-urban (peri-urban and suburban-type areas extending beyond core urban areas).

** See additional detail on the Degree of Urbanization (GHS-SMOD) construction in Florczyk 2019.


dLight (2013)
Rural
Quasi-Urban Cluster
Urban Center
Saraburi
Nakhon Nayok
Pathum Thani
Nonthaburi
Nakhon Pathom
Bangkok
Chachoengsao
Samut Prakan
Samut Sakhon
Chon Buri
0
30 Km
GHS-BUILT (2014)
Rural
Quasi-Urban Cluster
Urban Center
0
30 Km
GHS-SMOD (2015)
Rural
Quasi-Urban Cluster
Urban Center
0
30 Km
GRUMP (2000)
Rural
Urban
0
30 Km

665

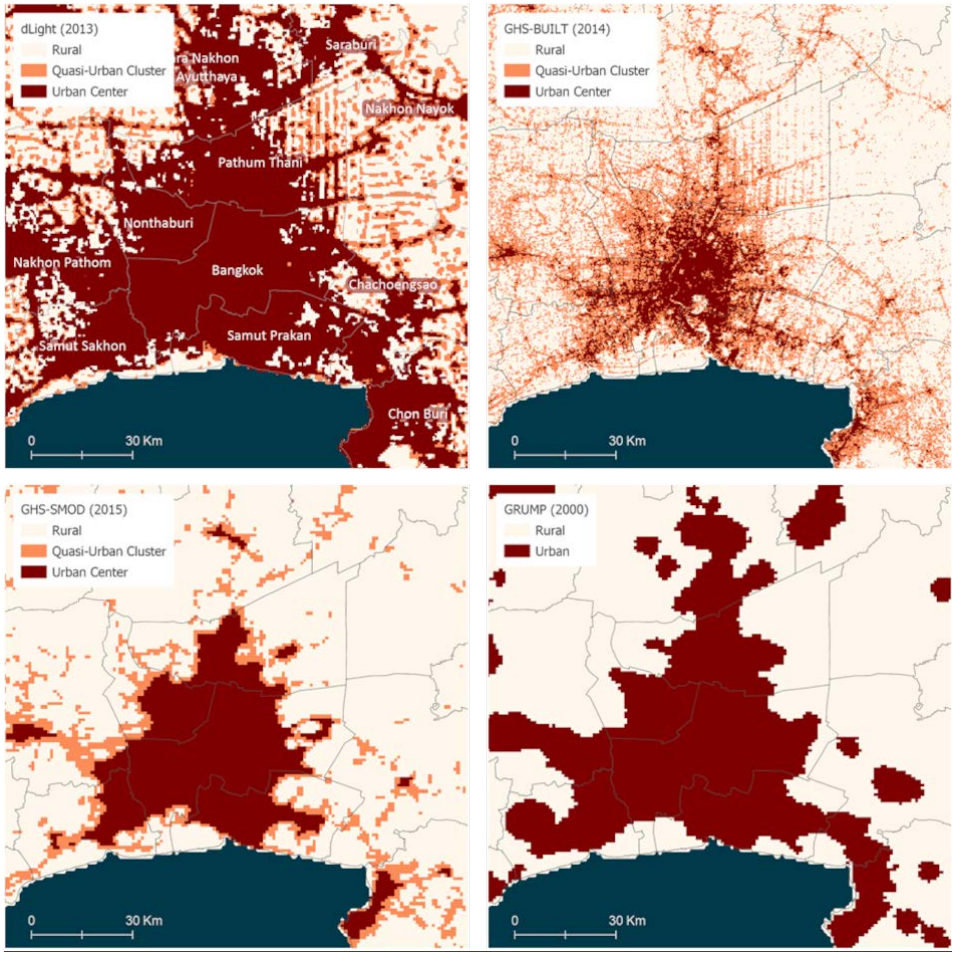


Figure 7. Urban proxy data classified into urban, quasi-urban and rural, Bangkok and surrounding areas, Thailand. Note that the dark blue indicates ocean and gray boundaries indicate first-order administrative boundaries.

2.5.7 Other Data Sets

The GPW land area grid had a native horizontal resolution of 30 arc seconds. It was uniformly upsampled to 9 arc second resolution by a factor of 100 and quality assured to have the same total land area per pixel both before and after the sampling, then it was subset by country. The mean administrative unit area grid also had a native horizontal of 30 arc seconds, but because the values in this grid represent the average size of input population units, there was no need to ~~upsample.~~alter or disaggregate the data values when increasing the cell size resolution. These data were simply resampled at 9 arc second resolution and subset by country. GHS-BUILT was used here not only to discriminate between urban, quasi-urban and rural as a categorical data set, but also to summarize built-up densities as a measure in its own right. It was projected from the World Mollweide projected coordinate system into WGS84 coordinates using Nearest Neighbor at 9 arc seconds, and subset by country.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

2.5.8 Calculating Summary Statistics

We produce estimates for each of the permutations of these 12 sources using the ArcGIS Zonal Statistics as Table tool, by country. A Python script was then used to compile these data into a single master table. These tabular data are summarized for the globe in this section and are available along with spatial data and a python notebook demonstrating how to produce LECZs here.

We used 9 arc seconds as the horizontal resolution of analysis, despite the native resolutions of elevation data ~~nominally~~nominally being 3 arc seconds. The reason for this is in order to leverage GHSL layers, which are the only data sets which have data for 3 points in time, without simply applying growth rates to a single spatial structure. GHSL's native resolution is 9 arc seconds (roughly 300m at the equator).

3. RESULTS

Using our core data sets as described above (MERIT, GHS-POP, and GHS-SMOD), we find that for 2015, 815 million persons globally live in the ≤10m LECZ, with nearly 300 million of those persons living in the higher risk ≤5m zone. About 60% of the population of the LECZ live in locations classified as urban and another 24% live in quasi-urban areas. Outside the LECZ, by way of contrast, the population is only 45% urban, while the share that is quasi-urban is comparable to in the LECZ, at 25%. The finding that the LECZ is disproportionately urban is robust across all data combinations of input data, as shown in Table 5 and the Figures 8-15 below; however, the range of these estimates vary substantially by the choice of data sets. Thus, in the following sensitivity analysis, we aim to understand the differences in these estimates, highlighting areas of agreement as well as divergence, and to draw out the implications where possible~~.~~ (the full range of global estimates by elevation source, population source, and urban proxy are available as summary tables with the data download).

Table 5. Summary of estimates of the global population in the LECZ, by LECZ and urban-rural classes. Core data sets (2015) shown with the range of estimates from other data sets given parenthetically.

	Urban-Rural Classification: Population Counts (in millions), with shares in *italics*						
Elevation	**Urban**	**%**	**Quasi-Urban**	**%**	**Rural**	**%**	**Total**
In LECZ (≤10m)	**487**	*60%*	**198**	*24%*	**130**	*16%*	**815**
	(220-714)		**(116-376)**		**(92-434)**		**(750-1056)**
≤5m LECZ	150	*50%*	85	*28%*	64	*21%*	299
	(55-488)		(55-236)		(45-294)		(276-691)
5-10m LECZ	337	*65%*	113	*22%*	66	*13%*	517
	(94-360)		(53-175)		(43-171)		(301-517)
Out of LECZ	**2,959**	*45%*	**1,651**	*25%*	**1,923**	*29%*	**6,533**
	(1,014-3,423)		**(692-2,670)**		**(1,259-4,587)**		**(6,233-6,596)**
Total	**3,446**	*47%*	**1,849**	*25%*	**2,053**	*28%*	**7,348**
	(1,322-3,950)		**(851-2,962)**		**(1,376-4,944)**		**(7,280-7,348)**

In the discussion that follows, we first review the results for data sets within a given domain (elevation, population, urban) but then use only the core data set when adding dimensions. Results with full permutations are found in the supplement materials.

3.1 Comparing population and land area estimates of LECZ with different Elevation data sets

3.1.1 Land area estimates by LECZ and Elevation Source

Figure 8 shows that for the ≤5m LECZ, CoastalDEM assigns the highest ~~proportion~~percentage of land area, and almost a third more land, than MERIT, SRTM, and TanDEM-X. In the 5-10m LECZ CoastalDEM, SRTM and TanDEM-X all assign the same ~~proportion~~percentage (0.83~~)~~%) of land area, whereas MERIT allocates almost a quarter more (1.02~~)~~.%). As a whole, CoastalDEM estimates the highest total land area falling within the ≤10m LECZ, followed by MERIT, SRTM and TanDEM-X.

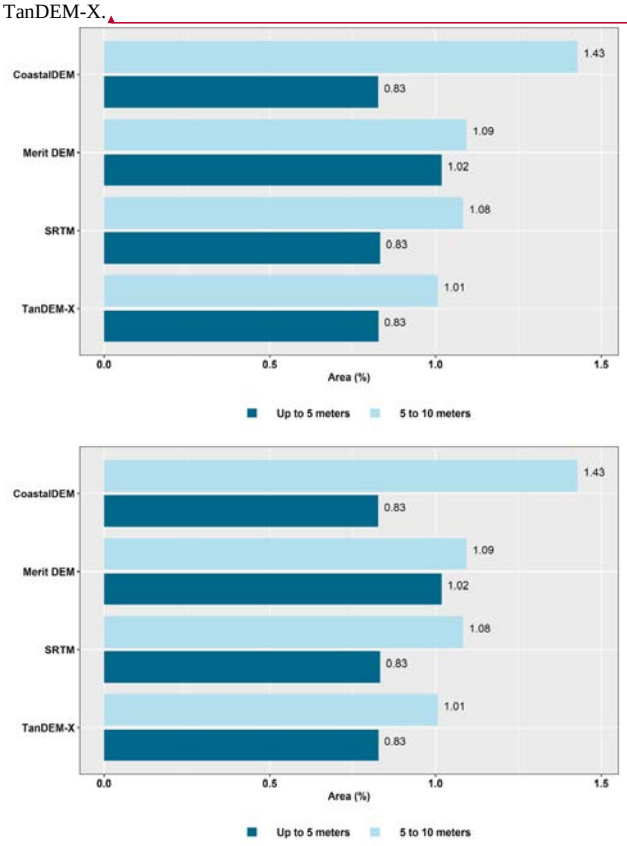


Figure 8. ~~Proportion~~Percentage of total land area in the ≤5m and 5-10m LECZ, by different elevation sources.

3.1.2 Population estimates by LECZ and Elevation Source

Estimates for the global population residing in the LECZ, by different elevation and population data sources, are shown in Fig 9. for 2015 and appendix Fig. B11 for 1990. The shares of population residing in either the ≤5m LECZ or the 5-10m LECZ have increased somewhat in the past 25 years, irrespective of which elevation data source is used to estimate the LECZ or which population estimates are used (only GPW and GHS-POP have estimates for 1990). Depending on the data sources, an additional .25 to .49 percent of the world's population was living in the ≤10m LECZ in 2015 than in 1990, which equates to ~~200~~20,000 – ~~400~~,000 – 40,000,000 more people.

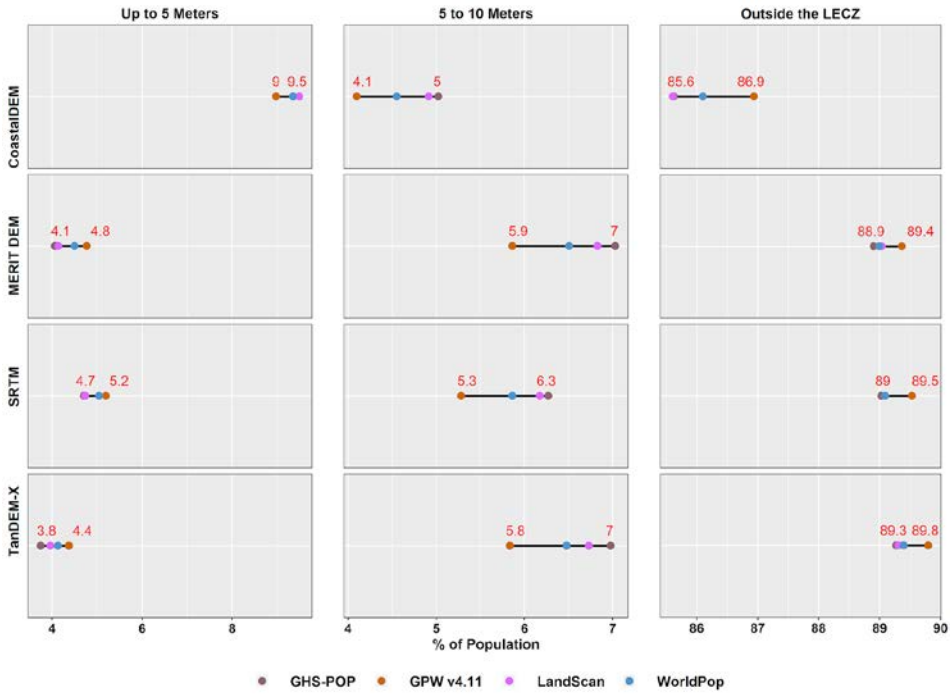
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Figure 9 shows the impact of population data choice on estimating the percentage of people living in LECZs globally in 2015. The relationship is clear to discern in the 5-10m LECZ, where GPW consistently estimates the lowest percentage, WorldPop the second lowest, LandScan the third lowest, and GHS-POP the highest percentage regardless of the elevation source used to define the LECZ.



Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

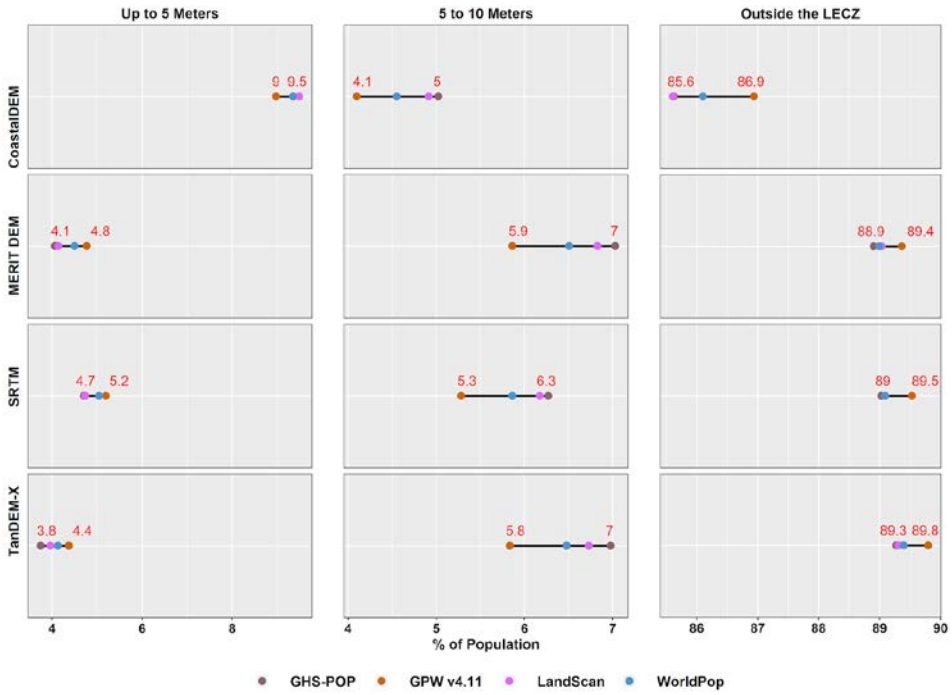


Figure 9. Estimates of Population in different LECZ zones, by elevation and population data sources, 2015.

Based on Fig. 9, it is clear that the selection of an elevation source has a greater impact on the estimation of population (and land area) in the zones than the selection of a population data source itself. The largest difference (in percentage points) between population sources is for areas outside the LECZ: using CoastalDEM for elevation there is 1.3% difference between LandScan and GPW. This is the largest difference in the percentage of population estimated in or outside the LECZ within any single elevation data source. The combined largest difference across all categories of elevation and population is 5.7% when comparing CoastalDEM and TanDEM-X in the ≤5m LECZ, where TanDEM-X estimates 3.8% using GHS-POP, and CoastalDEM estimates 9.5% using LandScan. Nevertheless, the selection of a population data source on its own is significant when considering that a difference of even 1% globally between sources amounts to approximately 80 million people in 2015. Also, since these differences in LECZ shares are not uniform, within some local areas the selection for population data may have considerably more impact.

3.1.3 What is driving the differences?

Considering only GHS-POP 2015 population estimates (without stratifying the urban-rural continuum), by using CoastalDEM, we estimate that 687 million people live in the ≤5m LECZ globally, whereas when the other data sets are used we estimate far fewer – 299 million with MERIT, 346 million with SRTM, and 276 million with TanDEM-X – people live in that same zone. Why is there such a large discrepancy? First and foremost, as indicated in Table 8, the land area ≤5m LECZ is about 40% more in CoastalDEM ~~≤5m LECZ~~ than in the others.

Looking at the end members of this range of estimates (CoastalDEM on the high end, and TanDEM-X on the low end), roughly 80% of the population difference can be found across 11 countries (China, India, Bangladesh, Indonesia, Viet Nam, Japan, Philippines, Egypt, Thailand, United States of America, Brazil), and more than 30% of this

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

difference occurs in a single country, China, where CoastalDEM predicts approximately 184 million people in the ≤5m LECZ, and TanDEM-X predicts approximately 54 million. A closer inspection of the elevation data sets sheds light on how these two data sets vary in their detection of low-lying areas.

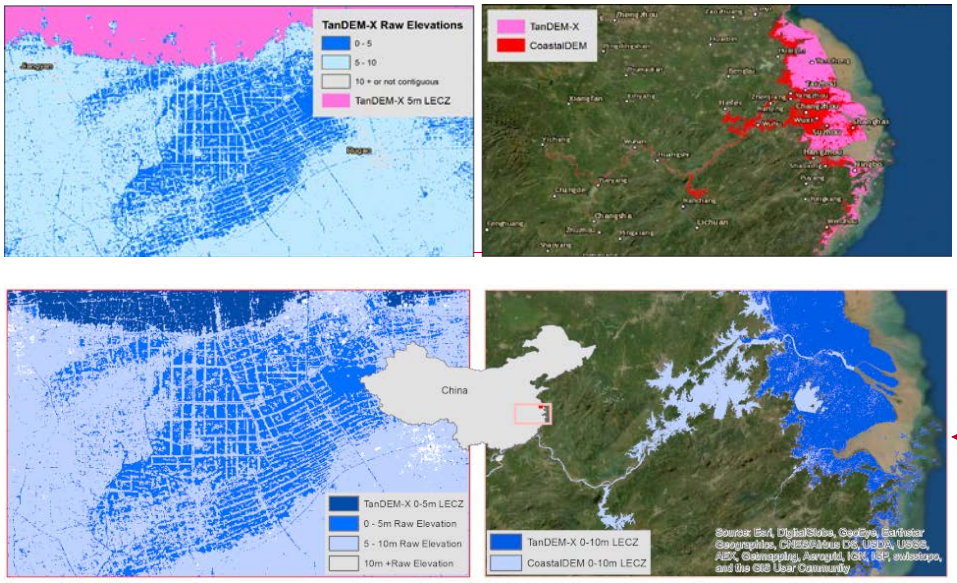


Figure 10. Comparison of ~~coastal contiguity in~~LECZ data sources along the Yangtze River, China for CoastalDEM and TanDEM-X elevation sources. Basemap from Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.

Figure 10 shows differences in the evaluation of coastal contiguity. In the right hand panel, the CoastalDEM elevation data source extends inland up the Yangtze River which leads to identification of low-lying areas near Hefei and Nanchang which are not considered contiguous to coastline according to TanDEM-X (or MERIT/SRTM). Figure 10 also shows in the left hand panel an area near Rugao, China, where a portion of the low lying area is not included in the final ≤5m LECZ based on TanDEM-X because it is not contiguous to the coastline owing only to the fact that it is cut off by a roadway.

CoastalDEM sets all grid cells over inland water to an elevation of zero, therefore when we evaluate coastal contiguity, the ~~banks of rivers and~~contiguous coastal zone extends further inland (e.g., to include more river tributaries ~~are more often captured)~~ than with other data products which have variable elevation values over inland water which sometimes exceed 5m or 10m. This partly explains why more inland areas are captured within the LECZ by CoastalDEM than the other sources. The LECZ based on TanDEM-X produces the smallest estimates of population. Unlike the other elevation data, it detects roads (notably found at higher density in urban settings), and classifies them as being at higher elevation than their surroundings as shown in Fig. 10 above (and it is overall more sensitive to built-structures and other elements of the landscape than the other elevation data sources). This is especially relevant when constructing LECZs because in the evaluation of coastal contiguity, we require direct connectivity to the coastline. Because TanDEM-X classifies roads (and other features) at higher elevation than their surroundings, it effectively creates contiguity barriers and thus smaller ≤5m and ≤10m LECZ zones. Similar phenomena are observed when considering MERIT or SRTM, namely that raw elevation estimates in these sources sometimes produce barriers which prevent coastal contiguity. (Whether these barriers indeed function as higher-elevation impediments to flooding is an open question that local studies may be able to address ~~(Orton et al., 2015, 2020)).~~(Orton et al., 2015, 2020)). The CoastalDEM model produces a more homogenous surface which therefore expands the zone of contiguity to the coast which increases the land area and population estimates within the zone (See appendix Fig. B2).

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

3.2 Comparing population and land area estimates with different Urban Proxy data sets

3.2.1 Land estimates by Urban Classes

Before evaluating the population in the LECZ along the urban-rural continuum, it is helpful to see how the different urban proxy data sets differ in their estimation of land area. Table 6 shows estimates of land area for the years 2000 (so that the comparison to GRUMP can be made) and 2015. The GRUMP data, sometimes criticized for the blooming quality inherited from use of the stable city lights as a key input ~~(Da Costa et al., 2017),~~(Da Costa et al., 2017), can be taken to combine the urban and quasi-urban categories into urban only, at least this is observed when comparing with urban and quasi-urban data not based on city lights. Combining urban and quasi-urban, for year 2000, the results according to dLIGHT are the most inclusive (5.3%) estimates of global land area, followed by GRUMP (2.9%), then GHS-BUILT (1.6%), and finally GHS-SMOD (1.2%). The same general pattern is seen for the year 2015, when GRUMP is omitted; additionally, changes over time in total area and percentages are also detected. These different urban proxies produce somewhat different depictions of land area. However, we find fairly strong agreement in the land area estimated in the urban class between GHS-BUILT and GHS-SMOD. This is not surprising because they share an important underlying data source (GHS-BUILT), but dLIGHT (like GRUMP before it) places more land area in both urban and quasi-urban classes which is also not surprising as both GRUMP, and dLIGHT are based on night-time lights which have known blooming effects.

Table 6. Land area of urban, quasi-urban and rural by urban proxy data sets.

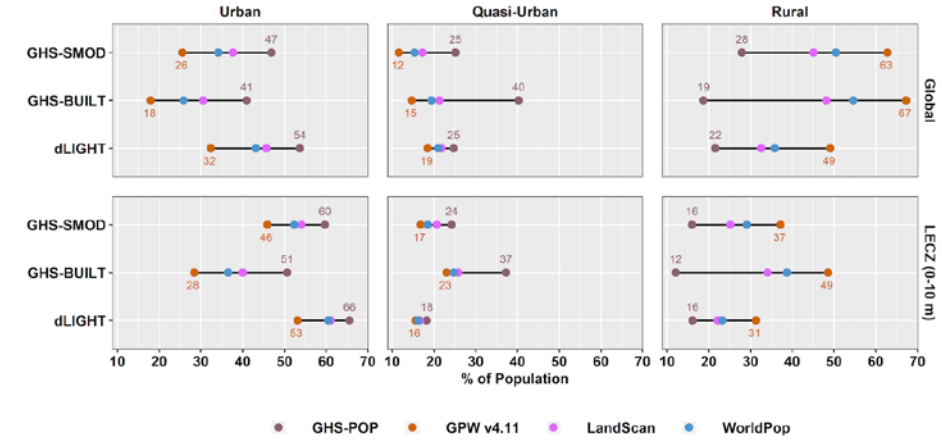
		Urban-Rural Classes, Area (in 1000 km2) and (%)					
		Urban		**Quasi-Urban**		**Rural**	
Year	**Urban Proxy Data Set**	Area	~~%~~(%)	Area	~~%~~(%)	Area	~~%~~(%)
		Globally (all land area)					
2000	dLIGHT	1,430	(1.1)	5,427	(4.2)	123,184	(94.7)
	GHS-BUILT	482	(0.4)	1,591	(1.2)	127,862	(98.4)
	GHS-SMOD	512	(0.4)	1,100	(0.8)	128,429	(98.4)
	GRUMP	3,766	(2.9)			126,265	(97.1)
2015	dLIGHT	2,038	(1.6)	6,313	(4.9)	121,690	(93.6)
	GHS-BUILT	568	(0.4)	2,013	(1.5)	127,355	(98.1)
	GHS-SMOD	639	(0.5)	1,270	(1.0)	128,131	(98.5)
		LECZ (0-10m) only, MERIT-DEM					
2015	dLIGHT	230	(8.4)	290	(10.6)	2,227	(81.1)
	GHS-BUILT	72	(2.6)	186	(6.8)	2,483	(90.6)
	GHS-SMOD	103	(3.8)	154	(5.6)	2,489	(90.6)

3.2.2 Population estimates by Urban Classes

The UN *World Urbanization Prospects* estimates that in 2018, 55% of the world's population lives in urban areas (United Nations, 2018), and whether this estimate is accurate or not ~~(Cohen, 2004)~~(Cohen, 2004) it remains the established benchmark of urban population statistics. Since the UN's estimate is derived from collections of country-specific urban measurements, the open questions are whether globally-consistent and spatially derived estimates are in fact more accurate, and whether or not these agree with the UN's estimates? Without additional information, we cannot evaluate accuracy, but we can characterize whether or not there is agreement.

Using these globally consistent urban proxy data sets, we show in Fig. 11 the ~~proportion~~share of the population that resides in urban, quasi-urban or rural settlements in 2015. The top panel of Fig. 11 shows the variation in the estimates by data source along this continuum. For any given population data set, the total population sums to 100% across the urban, quasi-urban and rural classes. In general, GHS-POP concentrates more people into urban and quasi-urban categories, no matter which urban proxy data set is used, and GPW concentrates more people into the rural category no matter which urban proxy is used. In terms of comparison to the UN estimates, whether the ~~proportion~~percentage of the population estimated to be urban shows agreement with the UN's estimate depends both on which urban proxy *and* which population data set are used. GHS-POP and LandScan place at least 55% of the population (and sometimes, quite a bit more) in urban, and quasi-urban areas regardless of which urban proxy data is used, whereas WorldPop (except when using dLIGHT) and GPW place less than 55% of the population in urban and quasi-urban areas. Use of dLIGHT as an urban proxy data set leads to comparable or higher proportions of the population in urban and quasi-urban areas across all population data sources. Importantly, none of the population data sources, irrespective of the urban proxy data set, place 55% of the population in areas classified as urban only, ~~suggesting that some of the official definitions are drawn from areas that have a more quasi-urban character (such as towns, suburbs, etc).~~however when combined with quasi-urban do approach 55%. Similarly, irrespective of the urban proxy data set, the percentage of the global population in urban and quasi-urban areas has grown substantially since 1990 (see Fig. B12).

Perhaps it is not surprising that estimates of population in rural areas vary more than those in urban areas, because satellite data broadly agrees on the urban category due to its relatively consistent and identifiable morphology. Most notably, the ranges in rural areas are largest when using GHS-BUILT or GHS-SMOD. The ends of these ranges are GPW, which uses no modelling towards settlements (or other attributes), and GHS-POP, in which the population reallocation is dominated by settlements (but not other ancillary features). Additionally, GHS-BUILT produces the widest range of population estimates across the three urban classes; in other words, the GHS-BUILT urban proxy is highly sensitive to the choice of population data.



Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Figure 11. Percent of total population, by urban-rural classes, using different urban proxy and population data sources, globally and in the ≤10m LECZ (using MERIT DEM) 2015.

3.2.3 Population estimates by Urban Classes in LECZs

In comparison to the global distribution, the lower panel of Fig. 11 identifies the population distributions along the urban-rural continuum in the ≤10m LECZ (using MERIT): the denominator in this panel is the total population *in the LECZ*. It is clear that the population is more concentrated in urban areas in the ≤10m LECZ than globally. For instance, using GHS-SMOD as the urban proxy and GHS-POP population data, less than half – 47% of global population – reside in the urban class, in contrast to in the ≤10m LECZ, where 60% of the population live in urban areas. Similarly, the population of LECZ is less rural than the global average. Indeed, compared to the global figures, the urban population shares in the ≤10m LECZ are higher and the rural shares lower for all combinations of population and urban proxy data sets (and for all the elevation data sources, as shown in the summary tables available with the data download). However, the quasi-urban population shares are sometimes higher and sometimes lower in the ≤10m LECZ than globally depending on which population and urban proxy is used. For each of the urban proxies, the estimates of the quasi-urban shares based on the different population data sets are closer to each other within the ≤10m LECZ, though the ordering remains the same as global, with GHS-POP having the highest quasi-urban share, and GPW the lowest (as for urban).

Comparing the upper and lower panels of Fig. 11, it is clear that the range of estimates of the population share for each urban class is narrower in the LECZ than the respective global ranges. This is likely because the LECZ is itself more urban, and urban areas are where the resolution of the underlying census data is finest. (See appendix A1 for a discussion of the role of underlying resolution on the population.) Similarly, all of the urban proxies show less sensitivity to the choice of population data set within the LECZ than they do globally, and both GHS-SMOD and dLIGHT show the least sensitivity in the quasi-urban class both inside the LECZ and globally. Notably, as shown in the lower panel of Fig. B12, the same patterns held in 1990, when the combined urban and quasi-urban population shares in the LECZ exceeded 50% (for all combinations of data sets except one). Whether population shares in the urban and quasi-urban areas of the LECZ have increased more than those shares globally is a question we address below.

Taking a closer look at the distribution of population within the LECZ, the pie charts in Fig. 12a, shown using only MERIT and GHS-SMOD, reveal some other interesting patterns (including information necessary to understanding the fractions of the population in the LECZ, and their associated densities.) (1) The population data sets vary in the total number of persons estimated to reside in the LECZ. GHS-POP places the greatest number of persons in the LECZ (815 million), and GPW the least (781 million), a difference of 35 million persons. (2) Despite differences in the

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

shares of the population estimated to live in the different urban classes (also shown in Fig. 11), the ratios of population living under ~~5~~5m to that living in the 5-10m is relatively consistent across population data sets, with notably greater fractions of the urban population in the LECZ living in the 5-10m zone. (3) Consistent with much different land areas (Table 6), the urban proxy data sets (shown in appendix Fig. B3) reveal different distributions of population in the LECZ.

Figure 12a. Proportion of the population in each urban class (urban, quasi-urban and rural) in the ~~5~~≤5m and 5-10m LECZ, for different Population data sets, 2015. Shown using GHS-SMOD as the urban proxy and ~~-~~MERIT for LECZ delineation.) Labels indicating population count (in 000s) are shown.

Given the comparably high shares of urban population in the LECZ, Fig. 12b shows the *urban (or quasi-urban, or rural, respectively)* population that resides in the ≤10m LECZ as a proportion of the global total in each respective urban-rural class (using MERIT, and GHS-SMOD with full results shown in appendix Fig. B3). Even though GPW estimates the smallest population in the LECZ, and the smallest urban population both globally and in the LECZ, it estimates the highest proportion of total urban population in the ≤10m LECZ: nearly 1 out of every 5 urban dwellers live in a city in the LECZ. This pattern holds no matter which urban proxy data set is used. In contrast, GHS-POP estimates the largest population in the LECZ, and the largest urban fractions globally, but as shown in Fig. 12b, places only 1 out of every 7 urban dwellers in the LECZ. Smaller proportions of the global rural and quasi-urban populations live in the LECZ, but particularly notable is that of the rural population in LECZ, roughly half live in the higher risk ≤5m zone.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

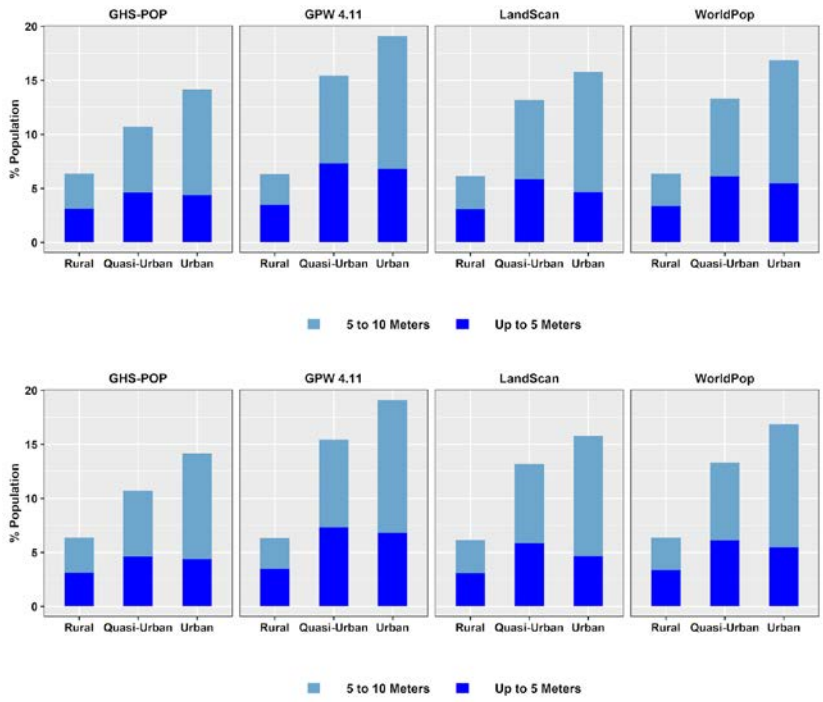


Figure 12b. ~~Proportion~~Percentage of population in each urban class (urban, quasi-urban and rural) in the ≤5m and 5-10m LECZ, by each *respective* urban-rural class, according to different Population ~~and Urban Proxy~~ data sets, 2015. ((GHS-SMOD is used for the urban proxy and MERIT is used for LECZ delineation.)

3.2.4 What is driving the differences?

Differences in the estimates of population – especially within classes along an urban-rural continuum – between urban proxy data sets are largely driven by four factors: (1) the selection of population data source; (2) the underlying satellite data measure and its associated urban construct; (3) the resolution of the underlying sensor; and (4) the thresholds used in the construction of urban classes: choices we have imposed for GHS-BUILT and dLIGHT, and the choice to use JRC's GHS-SMOD level 1 classification (which the user community at large can continue to evaluate and revise.) A fifth consideration that interacts with these factors relates to the underlying and intrinsic resolution of population data in urban areas, an issue that becomes more significant when considering the LECZ because it is disproportionately urban.

As described in Sect. 2.2, the key differences between the population data sources arise from how they allocate population within census based geographic units. GPW distributes population uniformly within these areas, without taking any account of indications that the land is urban (e.g. built-up) or rural (e.g. forested). GHS-POP distributes the same populations on the basis of built structures, in effect concentrating the population in those areas more likely to be classified as urban or quasi-urban. WorldPop also distributes the same population using different indicators, and also likely to concentrate population in more urban locations. LandScan builds on somewhat different sources for its initial population inputs and uses a somewhat different model to distribute the population, but is also designed to distribute more population to land with more urban characteristics. In densely-populated urban areas where the underlying census units tend to be more finely resolved (even in data poor settings), there is likely to be the greatest

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

agreement between the ~~urban, quasi-urban and rural~~ population estimates across the population data sources (this is explained in more detail in appendix A1.) Where there are differences, one would expect GPW population counts to be more rural than the others. The global statistics presented above conform with this expectation. So do those for the ≤10m LECZ, though the differences between the population sources are less, which is probably explained by higher overall densities and higher input resolution in coastal areas. All the population data sources estimate that the share of global urban populations located in the ≤10m LECZ (using MERIT) is higher than the share of global quasi-urban population, which is in turn higher than the share of global rural population. However, GPW has the highest shares of the urban and quasi-urban populations in the ≤10m LECZ, ~~as to be expected~~which is notable given its particularly low urban and quasi-urban populations outside of the zone, and provides additional evidence that the zone is disproportionately urban.

The urban proxy data sets determine which areas are classified as urban, quasi-urban and rural, using different indicators, as well as cut-off points between the classes that are inherently somewhat arbitrary, and help determine the share of land in each class. GHS-BUILT uses a physical (built settlements) model, GHS-SMOD expands it by also using population density, whereas dLIGHT and GRUMP use the detection of night-time lights which combine physical, atmospheric and environmental factors. In terms of application to urban locations, areas that are built are almost always included in the night-lights based products. One of the reasons that the lights-based products are more inclusive is because they contain land-uses that are not built structures but which may be lit (urban parks, roads, etc.., Stokes and Seto, 2019), and that the lights 'bloom' beyond the area where the actual light originates (Small et al., 2005), and while the resolution of the data underlying dLIGHT are much higher and thus reduce concern, this concern persists. While many user communities prefer the more spatially delimited built-construct, others (notably ecologists) prefer lights-based extent because its more expansive nature is better suited to the study of ecosystems, capturing the dynamics of land fragmentation ~~(McDonald et al., 2011).~~(McDonald et al., 2011).

The input horizontal resolution of GHS-BUILT is highly refined at 9 arc second, whereas dLIGHT has a native resolution of 15 arc second, and GRUMP of 30 arc second. The resolution of GHS-SMOD is also 30 arc second, but it is constructed from higher resolution (9 arc second) input data. These differences in resolution impact the classification of areas into urban, quasi-urban and rural since data which originated from a coarser scale is likely to be more inclusive. For example, on the edges of the urban class there are often transitions to quasi-urban which are clearly captured when using high resolution data, but are combined into the urban class at lower resolutions until greater than 50% of a pixel is captured as quasi-urban.

The selection of thresholds that we used for the GHS-BUILT and dLIGHT data sets (as well as the use of GHS-SMOD level 1 classification) is another important factor contributing to the variation in land area estimated to be in each class. The determination of any critical values to differentiate settlement types is somewhat subjective, as evidenced by the wide range of country definitions of settlement types utilized in global censuses (United Nations, 2018). We applied thresholds here based on a limited number of other studies that have evaluated the impact of thresholds on detection ~~(Leyk et al., 2018; Balk et al., 2018; Tong et al., 2018).~~(Leyk et al., 2018; Balk et al., 2018; Tong et al., 2018). GHS-SMOD is a fairly new data product as well that continues to make improvements to its model, in particular on the rules used to create the different urban classes; such as whether to apply a given built-up threshold or not. (Early work applied a 3% threshold, the latest and stable release removes this threshold.) Future work should help to evaluate the usefulness, and sensitivity, of these and other thresholds.

Because estimates of land area differ in the LECZ by urban proxy data set (as well as elevation), the last result section will evaluate differences in population (and built) density.

3.3 Comparing built-up and population density estimates by urban proxy data sets

This section turns from comparing population and land area estimates for different LECZ and urban-rural classes, to examining the related population densities and built-up area density estimates across these same classes. Population densities are simply the population counts divided by the land areas. Urban areas are associated with high population density, and indeed a high population density is often treated as a ~~criteria~~criterion to define urban areas ~~(Solecki et al., 2015; Dijkstra et al., 2020).~~(Solecki et al., 2015; Dijkstra et al., 2020). Having a high proportion of the land built-up is also sometimes treated as a defining feature of urban areas ~~(Melchiorri et al., 2018; Wentz et al., 2014).~~(Melchiorri et al., 2018; Wentz et al., 2014). Moreover, nighttime light is assumed to be associated with where human populations

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

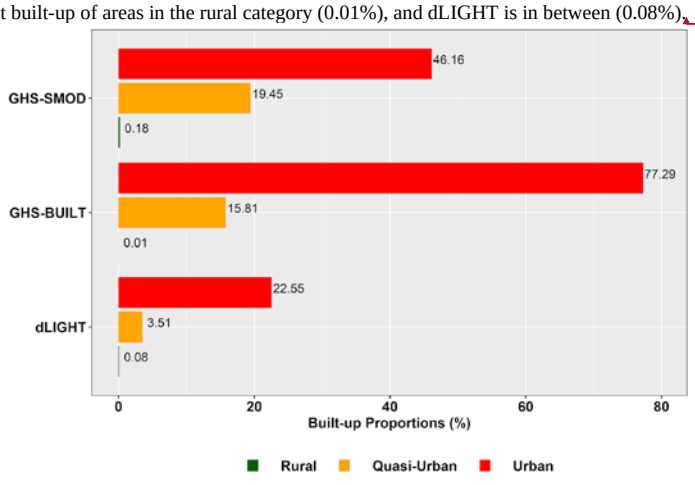
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

and built-up urban areas are located ~~(Wentz et al., 2014; Henderson et al., 2012).~~(Wentz et al., 2014; Henderson et al., 2012).

Along ~~-~~the urban-rural continuum, urban areas can be expected to be the most built-up, and rural areas the least: being built-up is part of how urban areas can be identified (as is fully the case with GHS-BUILT, and partially the case with GHS-SMOD), and being built-up is also generally associated with population density (used as one of several criteria to identify urban area in GHS-SMOD) and lit-up areas (used to identify urban areas with dLIGHT). While one would expect the relationship to be particularly close for the classification based on GHS-BUILT, where one would also expect it to depend on the thresholds (of the respective measure in each urban proxy data set) chosen, with particular thresholds resulting in smaller urban areas yielding higher urban built-up shares in those areas, and thresholds resulting in larger rural areas yielding higher rural built-up shares (and quasi-urban areas having higher built-up shares when urban areas are smaller and rural areas larger). Somewhat similar expectations apply to population density, though in this case the urban proxy data set most likely to pick out high density urban areas and low density rural areas is GHS-SMOD because it uses population density as a ~~criteria~~criterion in the construction of the GHS-SMOD classes. Also, excessively tight thresholds could in principle reduce urban population densities, as city ~~centres~~centers are often dominated by commercial property, exhibiting lower population densities (at least at night).

3.3.1 Built-up and population density by urban class and elevation

Figure 13a shows the global, average built-up percentage for urban, quasi-urban and rural categories across the urban proxy data sets (i.e., this is the average concentration of built-up areas based on GHS-BUILT). GHS-BUILT is a measure of the ~~proportion~~percentage of any given pixel that is considered built-up. The urban class is on average, ~~53.86~~46.16% built-up according to GHS-SMOD data; according to the thresholded GHS-BUILT urban proxy, the urban class has higher concentration of built-up area, at 77.29%; and according to dLIGHT, which produces a more expansive urban area (see Table 6), the urban class has an average built-up percentage of only 22.55%. GHS-SMOD captures areas as rural which are a factor of twenty more built-up (0.21%) as compared to GHS-BUILT, which only includes the least built-up of areas in the rural category (0.01%), and dLIGHT is in between (0.08%).



Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

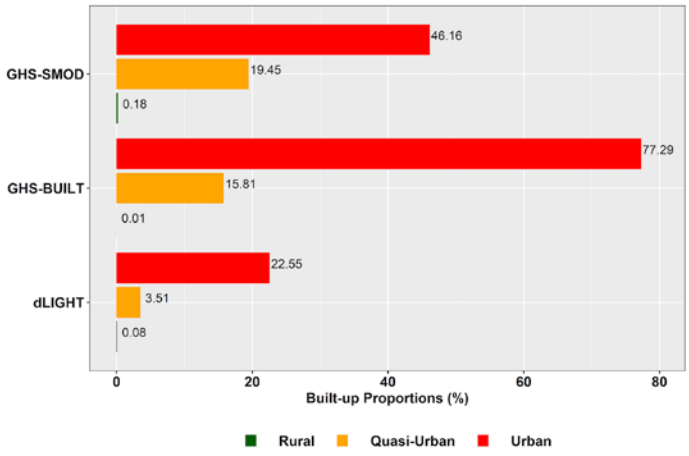
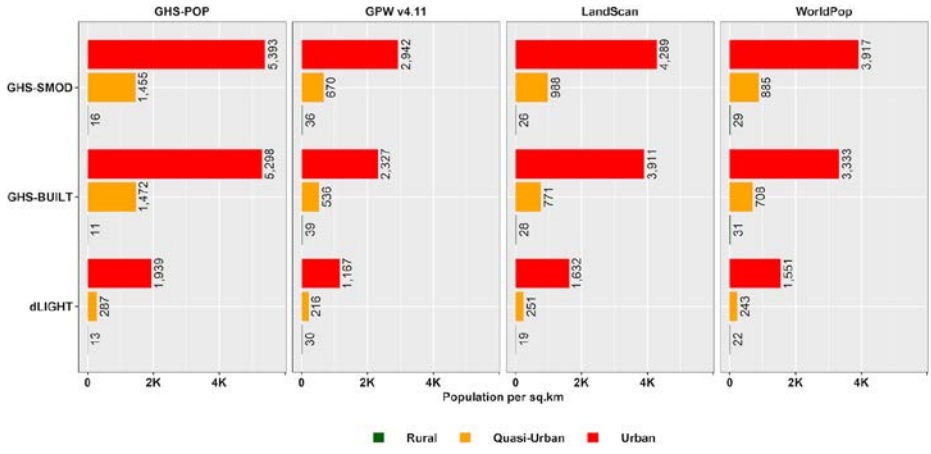


Figure 13a. Built-up density of urban, quasi-urban and rural by different urban proxy data sets, 2015.

Figure 13a can be compared to Fig. 13b, which shows the same classes, but by population rather than built-up densities, for each population data source (along the x-axis) and urban proxy data set (on the y-axis). The highest population densities are found, as expected, in urban areas; and these are several times those of quasi-urban areas, which are in turn several orders of magnitude greater than those in rural areas. This is similar to the built-up area differences, though for GHS-SMOD in particular the population density differences between urban and quasi-urban areas are greater than the differences in the proportion of area built-up. Within a given urban proxy data set, the estimate of population density depends largely on which population data set one uses – and these differences between population data sets are substantial. For example, using GHS-SMOD for our urban proxy data set, the population density in urban areas according to GPW is 2,942 persons/km2 whereas with GHS-POP it is 5,393 persons/km2. In contrast, within a population data set, the difference across the urban proxy data sets are largely comparable between GHS-SMOD and GHS-BUILT, but substantially smaller with dLIGHT. In short, and importantly, estimates of population density depend on one's choice of both the urban proxy data set and the population data set.



Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

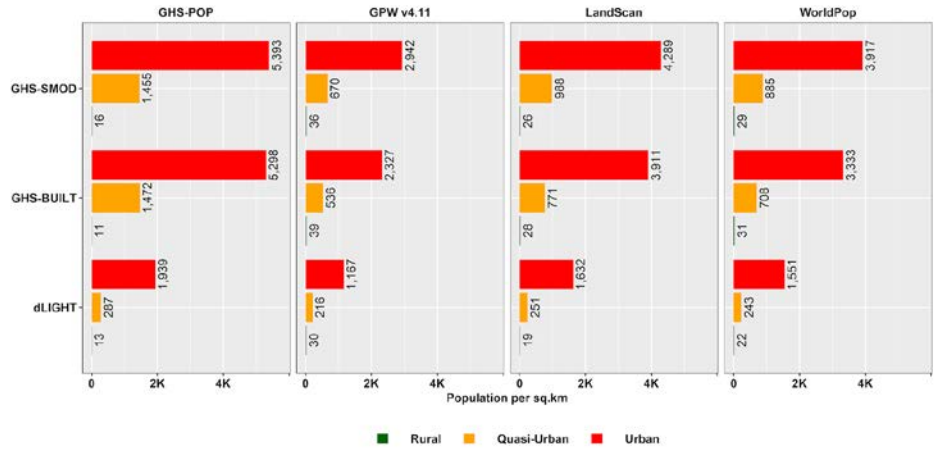


Figure 13b. Population density of urban, quasi-urban and rural areas by urban proxy data sets (y-axis) and population data (x-axis).

Whereas Fig. 13a and Fig. 13b show average global densities, in Fig. 14a and Fig. 14b, we separate out the LECZ classes. (For simplicity, we show the LECZ based on MERIT only, with the full comparison shown in the appendix Figures B4 and B5, noting that differences are small between elevation data sets.) While there are inherent relationships between both built-up share and population density and an area's degree of urbanization (i.e., ~~some of the urban proxy~~ because built-up and population ~~data sets~~density are often taken to be defining features of urban settlement), there are no ~~equivalent~~inherent relationships between either population or built-up density with elevation levels and the LECZ zones. ~~Yet, given~~Nevertheless, the ≤10m LECZ is disproportionately urban ~~nature of the ≤10m LECZ, we expect to find a~~, with a tendency to have higher population and built-up densities ~~in the LECZ~~ than outside of ~~it~~the LECZ.

Are urban and quasi-urban areas in the LECZ more built-up than areas outside of the LECZ? The answer in part depends on which urban proxy data set is used, and on differences within the zone itself. Fig. 14a shows that the ≤5m LECZ is less built-up than the 5-10m LECZ, with large differences when using GHS-SMOD or dLIGHT. Notably, the built-up percentages are greater in the 5-10m LECZ in all classes of the urban continuum than areas outside of the LECZ with the exception of GHS-SMOD, where urban areas in the 5-10m zone are quite similar in their built densities (53.5% vs. 55.3% outside the LECZ). GHS-BUILT produces built-up percentages that are almost invariant across the different LECZ zones.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font color: Auto

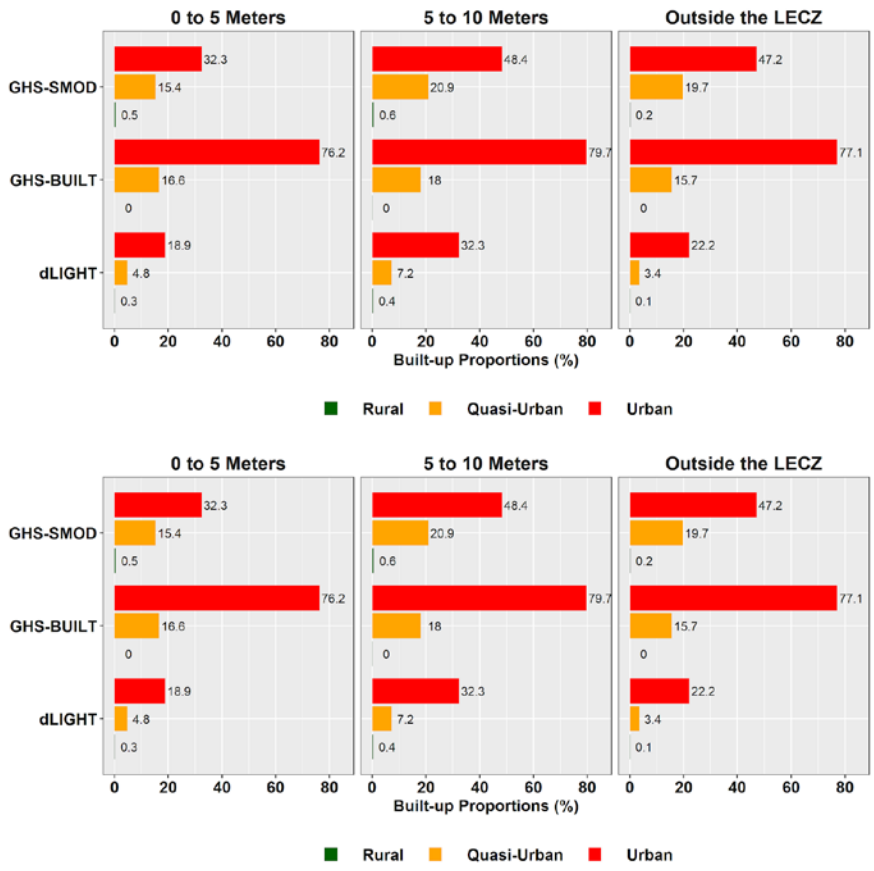
Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black



0 to 5 Meters
5 to 10 Meters
Outside the LECZ
GHS-SMOD
GHS-BUILT
dLIGHT
32.3
15.4
0.5
76.2
16.6
0
18.9
4.8
0.3
48.4
20.9
0.6
79.7
18
0
32.3
7.2
0.4
47.2
19.7
0.2
77.1
15.7
0
22.2
3.4
0.1
0
20
40
60
80
Built-up Proportions (%)
Rural
Quasi-Urban
Urban

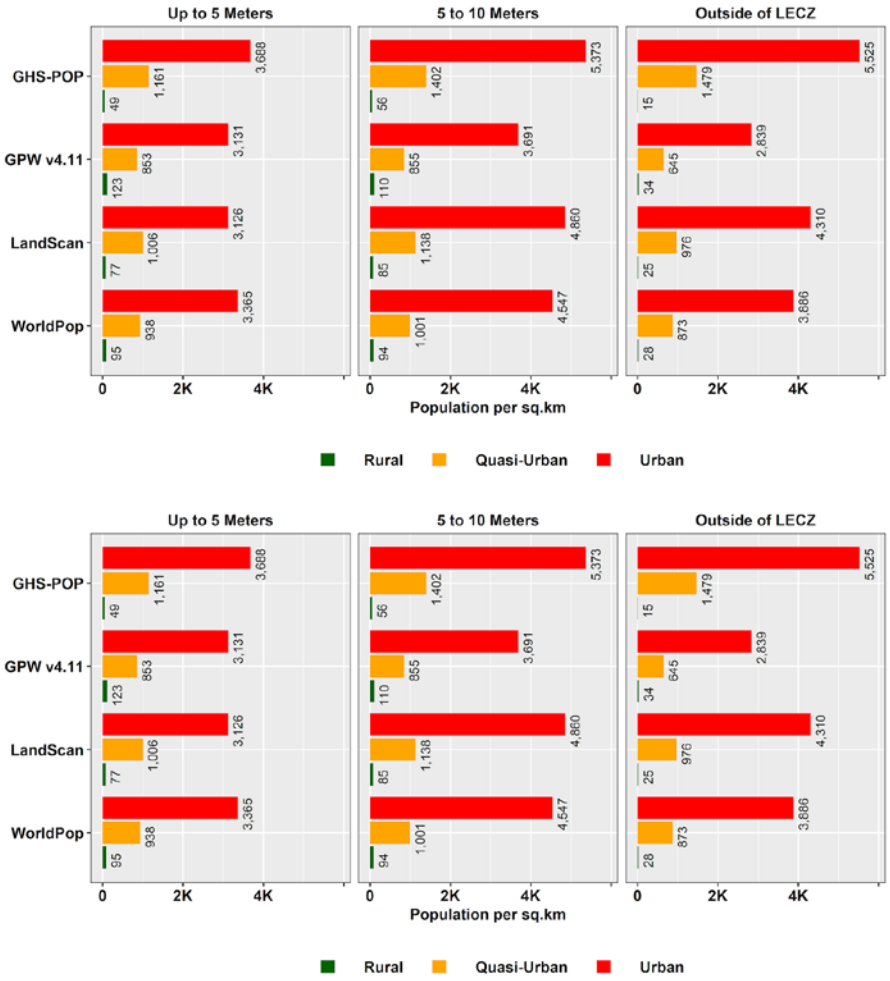
Figure 14a. Built-up density (%) of urban, quasi-urban and rural areas by urban proxy data set, using MERIT elevation source.

Like Fig. 14a, Fig. 14b asks how population density varies within the LECZ and compares to areas outside of the LECZ (results shown only for GHS-SMOD with other urban proxy data found in appendix Fig. B5). Population densities are lowest in the ≤5m LECZ, along the urban continuum (with the exception of GPW where population densities in the ≤5m zone are higher than outside of the LECZ). All of the population densities in the 5-10m LECZ are higher than those outside of the LECZ ~~regardless of which population source is used~~except for GHS-POP, where the density in the 5-10m LECZ is only slightly less than outside of the LECZ.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black



Up to 5 Meters
5 to 10 Meters
Outside of LECZ
GHS-POP
GPW v4.11
LandScan
WorldPop
3,688
1,161
49
3,131
853
123
3,126
1,006
77
3,365
938
95
5,373
1,402
56
3,691
855
110
4,860
1,138
85
4,547
1,001
94
5,525
1,479
15
2,839
645
34
4,310
976
25
3,886
873
28
0
2K
4K
Population per sq.km
Rural
Quasi-Urban
Urban

Figure 14b. Population density of urban, quasi-urban and rural areas using GHS-SMOD, by LECZ (using MERIT), according to different population data sets, 2015.

3.4 Change over time: Population growth in the LECZ along the Urban-Rural Continuum

In 1990, according to GHS-POP and GHS-SMOD 570 million persons lived in the ≤10m LECZ; 25 years later, that population has grown by another 245 million persons. As Fig. 15 shows, while the population has grown everywhere in the LECZ – that is, in urban, quasi-urban and rural areas – urban areas have experienced the greatest increases. The share of the global urban population living in the LECZ has grown from 13% in 1990 to 14.2% in 2015, whereas the shares of the respective quasi-urban and rural areas have declined – presumably, in part because some areas that were classified as quasi-urban or rural in 1990 have transformed to urban during this period. Also notable is that the

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

proportionate change in this 25-year period is greatest in the urban areas in the ≤5m LECZ (a pattern that holds across elevation data sets).

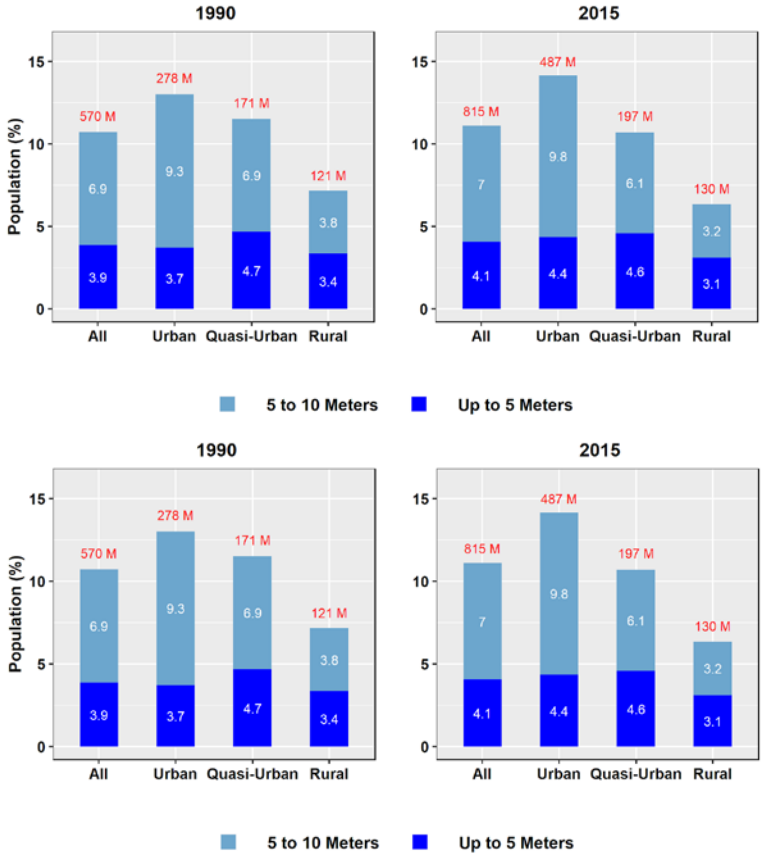


Figure 15. Respective share of population in LECZ, by LECZ and urban-rural classes, 1990 and 2015. Core data sets used (Merit DEM for LECZ; GHS-SMOD for urban continuum classification and GHS-POP for population).

Has the urban population in the LECZs grown more than the urban population overall? The answer to that is yes. Table 7 shows the shares of the respective urban, quasi-urban, or rural population living in the ≤10m LECZ in 1990 and 2015, and their change over the 25-year period: In 1990, 11.4% or the urban population — one out of every urban 7.7 person – lived in the LECZ; by 2015, 14.1% – one out of every 7.1 – urban person lived there, using GHS-POP, GHS-SMOD, and MERIT, whereas the percentage of the respective populations living in quasi-urban and rural areas has fallen in these 25 years. The population living in urban areas in the LECZ has grown considerably more – ranging from 67% increase to more than a doubling, depending on which data set one uses – than in urban areas outside the LECZ. While populations have grown in quasi-urban and rural areas, according to most data combinations – they have grown much less than in these areas outside the LECZ. While the levels of these shares differ somewhat across urban proxy and population data sets, the main message is unambiguously consistent: urban population has grown more in the LECZ than outside of it.

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Table 7. Estimates of the global population, and percentage change, in the LECZ, by urban-rural classes (based on different urban proxy and population data sets), 1990-2015.

Urban/Rural Classification		% Population Residing in <10m LECZ				% change in Population in LECZ from 1990-2015			
		2015		1990		<10m LECZ		Outside LECZ	
Classification	Urban Proxy Data	GHS-Pop	GPW	GHS-Pop	GPW	GHS-Pop	GPW	GHS-Pop	GPW
Urban	dLIGHT	13.5%	17.5%	11.4%	14.0%	99.9%	104.3%	63.3%	57.4%
	GHS-BUILT	13.7%	16.8%	13.0%	15.3%	67.3%	84.9%	57.2%	65.9%
	GHS-SMOD	14.1%	19.1%	13.0%	17.3%	74.9%	84.7%	58.7%	63.1%
Quasi-Urban	dLIGHT	8.2%	9.0%	12.4%	14.5%	0.8%	-1.9%	59.7%	68.7%
	GHS-BUILT	10.3%	16.7%	10.7%	16.7%	29.8%	54.6%	35.5%	55.3%
	GHS-SMOD	10.7%	15.4%	11.5%	16.5%	16.1%	30.2%	26.1%	41.4%
Rural	dLIGHT	8.3%	6.8%	8.8%	7.2%	-15.1%	14.3%	-9.5%	20.8%
	GHS-BUILT	7.1%	7.7%	7.3%	8.0%	10.6%	24.5%	13.5%	29.6%
	GHS-SMOD	6.3%	6.3%	7.2%	6.9%	7.7%	18.0%	22.9%	30.1%

Note: MERIT-DEM used as elevation data for LECZ.

4. FITNESS FOR USE & DIRECTIONS FOR FUTURE RESEARCH

When modelling three different phenomena, all imperfectly, and then combining them, it is prudent to reflect on limitations (e.g., accuracy or uncertainty) and future usages of these data, including many that go well beyond what ~~we've~~we have described in this analysis. Following the example found in a recent review of global population grids (Leyk et al. ~~2019~~2019a), here we pose some questions that users of the data sets produced herein may use in order to describe fitness for, and possible limitations, in use. As we have made evident above, it is important to get under the hood of any given data set to understand what data went into its construction along with important modelling assumptions.

Importantly, estimates of population in the LECZ, along an urban continuum, are sensitive to the choice of data sets. For some uses, it will be important to be explicit about how sensitive conclusions or recommendations are to the choice of data sets, and results from multiple sets should be combined to provide the needed sensitivity analysis. There are, however, significant disadvantages in having to manipulate and present the results of multiple data sets, and there will be times when it is more appropriate to select only one population data set, one urban-rural data set and one elevation data set. In either case, it is important to recognize the strengths and weaknesses of the different data sets, and not simply to pick the ones that support ~~favoured~~favored recommendations or conclusions. Some of the questions whose answers may help determine both which and how many data sets are appropriate to particular uses are discussed below.

Formatted ... (repeated 100 times)
Formatted Table ...

4.1 How does spatial resolution impact your analysis? Both horizontal and vertical resolutions are of utmost importance when trying to characterize at-risk populations in LECZs. Vertical resolution and corresponding uncertainties of elevation data sets (measured by RMSE etc.) vary by location and land cover type, and therefore must be examined carefully when used in local or regional scale studies. If local elevation data (e.g. LiDAR) is not available, then users should consider evaluating multiple global DEMs to identify areas of disagreement and related uncertainties. Particular small and dynamic geographies are more susceptible to misrepresentation arising from vertical and horizontal resolution issues, and co-registration of overlaying data sets. This is especially important for small islands geographies in global data products ~~(Taupo et al., 2018; Taupo and Noy, 2016; Yamano et al., 2007; Lewis, 1989),~~(Taupo et al., 2018; Taupo and Noy, 2016; Yamano et al., 2007; Lewis, 1989), for ~~deltatic~~deltaic geographies (Minderhoud et al., 2018, Tessler et al., 2015; Syvitski et al., 2009) and in coastal areas and cities more generally (Nicolls et al., 2021, Erkens et al. 2015; Kaneko and Toyota, 2010),) where subsidence may be occurring at higher rates than previously thought ~~(Minderhoud et al., 2018),~~ and in urban areas where building heights may impact the accuracy of elevation measurements ~~(Pesaresi et al., 2021),~~(Pesaresi et al., 2021).

Although the horizontal resolution (e.g., cell size) of global elevation data is generally uniform at approximately 100m (nominally at the equator), when combining with other data users must account for integration complexity. The population and urban proxy data sets used in this study include a range of horizontal resolutions from 100m to 1km (nominally at the equator), and 2019~~).~~. However, simply selecting the finest resolution data does not imply greater accuracy, since fine resolution disaggregation depends on models which themselves are uncertain based on their inputs and modeling approach. Aggregating all data layers to a common ~~–~~coarser scale can reduce uncertainties in the population estimation, but at the cost of also reducing the resolution of elevation models (which causes information loss.)

In this study, we selected 3 arc seconds as the common horizontal resolution, which is coarser than the native 1 arc second resolution of global DEMs. This choice has the advantage of smoother coastal contiguity surfaces in producing LECZs, and harmonization to the full GHSL time series, but the disadvantage of not using the source resolution present in global DEMs. The choice of which population data set to use depends in part on the type of area being studied. In general, we expected and found convergence of population estimates in larger urban areas. This is explained by the fact that underlying census geographies (ie., administrative boundaries) are usually the smallest in large urban areas, which therefore reduces the need for complex allocation models. Users interested in largely rural or quasi-urban localities would be advised to compare among the population and urban proxy data sets since the models are more likely to diverge, as well as to utilize any local data available.

The recent study by Hooijer and Vernimmen (2021) produces estimates of population and land area up to 2m above mean sea-level. While they demonstrate greater vertical accuracy than are reported in the elevation data sets we used as the basis of our LECZs, their data are a much coarser spatial resolution (nominally at 5km), and they used only one population data set (GPW) in their estimation rather than a multiple data to generate a range of possible exposures. Future work can evaluate these data and determine how to integrate with ours in order to benefit from their gains in the vertical resolution.

4.2 Can these data be used to observe changes over time? If comparability over time is important to your analysis, some combinations of these data may be objectively better than others. To be clear, all the elevation data used to produce the LECZs represent only one time point (circa 2000 for the SRTM based measures, or 2015 for TanDEM-X). It is an open question as to what the ideal observation period is for identifying changes in elevation, which is the main data set used here to delineate LECZs. ~~Future climate change is~~Future climate change, and an increasing understanding of the rate of subsidence in coastal and deltaic megacities (Zoccarato et al. 2018, Kaneko & Toyota, 2011; Syvitski et al., 2009), are likely to shorten the periodicity for which we want such observations collected and made available in a regular way, like national censuses. Yet there is no international or collection of national organizations with this mandate at present (the TanDEM-X website describes plans by the German Space Agency to reacquire elevation data as of 2019 and produce a “Change DEM”, but as of present that data is not available for analysis).

Some population and the urban proxy data sets ~~user~~represent multiple points in time, but users must exercise caution when using them as if they were a spatially-precise time series. The LandScan population data specifically advises users to not consider annual estimates as a comparable time series since they are based on different methodologies

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Auto

Formatted: Font color: Auto

~~(Rose and Bright, 2014).~~(Rose and Bright, 2014). GPW ~~and WorldPop produce~~(2000-2020, in 5-year time steps) produces population estimates for multiple points in time by using subnational growth rates ~~(~~at the finest scale at which those data are available and applied to the full~~-~~-resolution 2010 round census geographies (which represent an even finer resolution in most cases), but ~~their underlying~~the spatial structure is nonetheless based only on 2010 round census geographies~~.~~ and therefore does not vary. GHS-POP ~~also utilizes~~(1975, 1990, 2000, 2015) and WorldPop (2000-2020, annually) utilize those same 2010 geographies, but the finer distribution of population is modeled using built-up ~~estimates~~(and nighttime lights and landcover as well in the case of WorldPop) data unique to each epoch, which therefore ~~varies~~vary the spatial structure of estimates over time. Each of these data sets makes assumptions about representing change over time and space, and, particularly for processes that operate at a local scale (such as the population distribution associated with urbanization), users would be advised to consider the implications of these assumptions for their analysis.

Of the urban proxy data, GRUMP is from a single temporal range (circa 1995), but dLIGHT (1992, 2002, 2013), GHS-SMOD (1975, 1990, 2000, 2015), and thresholded GHS-BUILT (1975, 1990, 2000, 2015) are all variable over time.

4.3 How important is transparency in underlying assumptions of source data sets? With the promulgation of more elevation, population, and urban proxy data sets, users must consider how to meaningfully combine these data to characterize population and land area estimates in the LECZ, and be aware of the interdependence of their choices. Transparency in the underlying methodologies of these data is important in order to avoid confirmation bias (or what has been termed policy-based evidence) in the pursuit of better decision making (through evidence-based policy). The producers of data used in this analysis are transparent in that they have published peer-reviewed articles on their data sets, however, due to the complexity of the modeling and that articles are often written for technical and discipline-specific audiences, some of the assumptions made in the process ~~are obscured~~remain obscure to end users or non-experts.

Above we have reviewed the relevant modelling methods and ancillary data inputs to produce the various data sets, here we will address issues as they relate to use when being combined. Models are used because the underlying data are inadequate to some degree for the purpose at hand. Yet, many of the modelled data sets in all three areas (population, urban proxy and elevation) are endogenous to some degree. Models depend not only on use of inputs – some of which could produce circular reasoning in results (as addressed above) but also on assumptions of how to model the constructs at hand. Those assumptions should be understood by downstream users of data products as well, but are often left implicit.

For instance, the CoastalDEM elevation surface is the only elevation data set among the four here that uses information other than elevation in its model: among many data sets, it uses LandScan population data as an explanatory variable to predict elevations ~~(Kulp and Strauss, 2018).~~(Kulp and Strauss, 2018). This makes population partially endogenous to the CoastalDEM elevation surface, and treating the resulting elevation data as explaining the spatial distribution of population risks circularities, as the model used to help estimate CoastalDEM elevations already contained population as an explanatory variable.

Of the population data sets, GPW is the only one that does not use covariate layers in its population model and GHS-POP only uses GHS-BUILT. WorldPop uses SRTM Elevation data, GHS-BUILT, and VIIRS Nighttime Lights (although they are stronger predictors in certain countries than others) to delineate a population surface, which creates a similar endogeneity problem to that of CoastalDEM and LandScan. LandScan's list of covariate layers are not publicly documented. Although GPW avoids any endogeneity problems, the uniform allocation results in a higher share of population allocated to rural areas. GHS-POP uses only one covariate layer, GHS-BUILT, which leads to a higher share of population in urban areas, and also endogeneity when being used with the urban proxy data layers also rendered from GHSL products. Since only one covariate layer is used, the potential for bias is more transparent than when highly complex or unspecified models are used. The reason that more complex population distribution models are used is to provide more precise estimates at the pixel level, but this comes at the cost of introducing ~~unrecognised~~unrecognized or poorly understood endogeneity problems. Users must see these endogeneity vs. precision concerns as trade-offs and consider the nuances when selecting which data to use. Transparency is vital in supporting reasoned decisions in that regard.

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Of the urban proxy data used in this study, all of the data sets are based on varying assumptions on how to best represent urban areas along a continuum – which we simplified to three categories of urban, quasi-urban, or rural (as described in Table 4). It is an open question as to whether such renderings require data representing both the population and land perspective of urban areas, in part because there is no agreement on what defines an urban center or settlement. If one adopts a demographic perspective and treats population concentration as the ~~the~~ defining feature of urban areas, this relationship between population concentration and areas identified as urban is not an endogeneity problem but an explicit assumption, though there is still the risk that the spatial population data sets may, for example, overestimate population concentration on built-up land, and lead to the misspecification of just how urban areas are. In contrast if one adopts a (physical) geographic perspective ~~-~~and treats built-up land use as the defining feature of urban areas, then the same tendency to overestimate population concentration on built-up land would not create errors of urban misspecification, but of misrepresentation of the relationship between population concentration and the degree to which the land was urban.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Of the data sets we use here, two use both population and settlement proxies together (GHS-SMOD and GRUMP), but no other inputs (unlike the complex population models), and two are based on just the physical urban footprints – built-up or lights. But regardless, in order to generate urban proxies, some sets of thresholds and other rules were applied to construct the three classes here. GHS-SMOD and GRUMP are complex data integration projects that downstream users cannot easily reimplement but GHS-BUILT and dLIGHT, which do not include population-based criteria, are easy for spatial data users to use as they wish. The selection of use-appropriate thresholds and more complex criteria for dLIGHT and GHS-BUILT data (as alternatives to what we have done here) is something which users may wish to do in order ~~-~~to more fully optimize use of those data. Moreover, the assumptions across all of the urban proxy data used here are global assumptions, whereas there is strong reason to believe that locally-adaptive models (by levels of economic development, biome, or other characteristics) could produce more precise and optimized results. Users should experiment in this regard whenever possible depending on their use case and study area.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

The sensitivity analysis here shows a consistent relationship between GPW and GHS-POP forming the end members of the array of possible populations both within and outside of the LECZ, with GPW dispersing population uniformly on the low end resulting in a larger rural share, and GHS-POP concentrating population into built-up areas on the high end resulting in larger urban shares. Where the underlying census data are at high resolution (typically, at administrative level four, five, or six, but this depends on the geographic size of the country), we found high agreement across population data products in areas classified as urban, and across elevation data sets. While much effort has gone in improving the resolution at which census data is collected and made available (United Nations, 2014), as more censuses implement and distribute high-resolution data (e.g., for enumeration areas or settlement points) the need for modelling will dissipate (or be needed only in special use-cases, like remote areas); this is relevant for the spatial distribution of both population and urban areas ~~(Champion and Hugo, 2004).~~(Champion and Hugo, 2004). To the extent that the research community engages with national statistical offices, reiteration of this need remains a priority.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Similarly, future versions of GHS-BUILT that distinguish between industrial and other types of structures will be an improvement for those using these data (or derived products like GHS-POP and GHS-SMOD) to distribute population spatially, and to identify urban areas. Particularly if nighttime population concentration (as is the construct for all population data sets here except LandScan) is meant to be an independent indicator of an area being urban, alongside being built-up (as intended in GHS-SMOD), it would represent an important improvement to avoid population allocation procedures that shift population to non-residential built-up areas. It could also improve estimates of nighttime coastal population to avoid non-residential port development, for example, from being implicitly assumed to be residential, and having populations allocated to them.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

4.4 How important is the accessibility of the data sets to other potential users? A separate but related issue to that of transparency is whether or not one's analysis requires data that can be accessed by others or whether it has large user-restrictions or fees (since for publicly available data, unrestricted use is part of transparency in that they foster replication and comparison). We make all of our results publicly available in tabular form. However, as discussed above, we can only redistribute our spatial data layer for the LECZ based on MERIT, as the other sources have restrictions on redistribution~~.~~ (SRTM can also be distributed but is not here due to its known limitations.) Apart from whether a data set can be redistributed, some data sets are freely available and others have fees. TanDEM-X and Coastal-DEM among the elevation data sets, and LandScan among the population data sets, are freely available for

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

research usages, but not other commercial or operational ones. All of the data sets used here as urban proxies are publicly available. Creative Common and Open Database Licenses are increasingly used, and previously for-fee and restricted data sets are becoming more open, thus users are encouraged to check with data providers for updates.

4.5 How important is consistency with international, national or disciplinary norms and usage? Much progress has been made in the past two decades in the spatial rendering of population, urban locations, and elevation by a global community of researchers. That has been accompanied by a critical lens of usage and discussion among data producers, notably among the population data producers (Leyk et al., ~~2019~~2019a). The coupling of population and land-use based data to describe urban location and population is the most novel of the three data types we use here and therefore the one requiring the most scrutiny. Importantly, only GHS-SMOD and GRUMP explicitly aim to locate urban areas and both of these depart from the long-accepted, aspatial standard by which global estimates of urban population are estimated (United Nations, 2018).

As the goal here is to create globally coherent data sets, international comparability is critical. For local uses, there will often be more relevant and/or accurate data, including on the elevations needed to identify the LECZ, the spatial distribution of population and built-up land, and the locations of more or less urban areas. Until recently, as indicated above, international data on urban populations have been based on national definitions of urban area, which vary widely, even if they tend to coincide with more densely populated and built-up areas, more likely to contain the sort of structures, infrastructures and institutions that planners and others associate with urban settlements. For this report, we have used more internationally comparable methods of identifying the urban-rural continuum. As noted above, one of these - GHS-SMOD - has been ~~recognised~~recognized by the United Nations Statistical Commission (UNSC, 2020) as a means of helping countries identify the degrees to which areas are urban, and thus to provide a valuable complement to or even eventually input to international urban population series based on national definitions. It is to be expected, however, that appropriate data choice will continue to depend not only on disciplinary and national norms, but on the relative priority given to international comparability versus local relevance, and to consistency with the ever expanding and improving data sets from sources such as satellite imagery and other evolving technologies. Using a refined measure of urban locations rather than a simple dichotomy is important given that the bulk of future population growth will take place predominantly in the cities and towns of Asia, Africa and Latin America, and thus understanding cities of different sizes, their characteristics and relationships to one another, is increasingly important and these new data and methods make it easier to do so ~~(Dorélien et al., 2013; Tacoli, 1998; Menashe-Oren and Bocquier, 2021).~~(Dorélien et al., 2013; Tacoli, 1998; Menashe-Oren and Bocquier, 2021). Thus, understanding what data and criteria are used to construct a continuum of urban classes is only likely to gain in relevance. How any given user chooses or not to specify a continuum will depend on a given usage.

Since the construction of the first LECZ ~~(McGranahan et al., 2007),~~(McGranahan et al., 2007a), others have adopted the basic methodology to create improved data for more local areas ~~(Reimann et al., 2018; Vafeidis et al., 2019; Hauer et al., 2020),~~(Reimann et al., 2018; Vafeidis et al., 2019; Hauer et al., 2020), as a basis for forecasting future exposure ~~(Reimann et al., 2018; Neumann et al., 2015),~~(Reimann et al., 2018; Neumann et al., 2015), and at finer elevation bands~~(Lichter et al., 2010)).~~ (Lichter et al., 2010)). But the international norms and data available to model coastal flooding and sea level rise are presently emerging ~~(Nicholls et al., 2021; Muis et al., 2020; Tellman et al., 2020; Vafeidis et al., 2019; Haigh et al., 2020; Kopp et al., 2015; Tebaldi et al., 2012),~~(Nicholls et al., 2021; Muis et al., 2020; Tellman et al., 2021; Vafeidis et al., 2019; Haigh et al., 2020; Kopp et al., 2015; Tebaldi et al., 2012), and to date mostly depend on local data.

5. DATA AVAILABILITY

Only some of the underlying data sets used here are licensed for derivation and redistribution. Links in Tables 1-3 point to the data originator. ~~Here~~Here we disseminate:

- Table of results as indicated (population and land area in LECZ by urban-rural classes, by all elevation, population and urban proxy data sets), by country, continent and year.
- A spatial layer of ~~0-~~<5m and 5-10m LECZs based on MERIT at 300m horizontal resolution in the WGS84 coordinate system.

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

This is a preliminary open data release, pending peer review of the data and associated journal articles. Following the peer review process, data curation will be completed by the NASA Socioeconomic Data and Applications Center (SEDAC) and the data will be disseminated through the SEDAC catalog.

Formatted: Font: 12 pt, Font color: Black

6. CODE AVAILABILITY

Many of the techniques we use here to generate estimates of populations by elevation, population source, and along the urban continuum leverage well-known workflows and geoprocessing tools. ~~The code provided here~~The code provided here focuses on the novel aspect of the work, namely how to produce a LECZ from some DEM and Coastline data for a country or other area of interest. License restrictions on some of the data utilized in this work prevent their redistribution, therefore the sample code utilizes sample data from the open MERIT product and coastline and area of interest data files from GPW, which is also open.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

The sample code is provided as a Python Notebook which utilizes the ESRI arcpy module. Although the ESRI arcpy module is proprietary, analogous tools exist in open source python modules and in R so that this example can help guide users who do not have access to arcpy. Sample input and output data are also included. To run the code to produce new outputs users should update data paths or delete the sample outputs provided.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

7. DISCUSSION & CONCLUSIONS

The analysis in this paper updates and confirms, but also refines and extends, the findings from the ~~McGranahan et al., 2007~~McGranahan et al., 2007a study: In 2015, based on elevation data from MERIT and population data from GHS-POP, over 10% of the world's population – more than 815 million people – lived within 10 meters above sea level, and based on GHS-SMOD, 84% of those people lived in urban centers or quasi-urban clusters. Close to 10 percent (9.4) of the world's land area in the ≤10m LECZ is urban or quasi-urban, compared to 1.5% globally.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

The sensitivity analysis, which incorporates four sources each for elevation, population, and urban proxy, reveals a much wider range of possible estimates than previously noted. Despite the variation in estimates, there is nonetheless consistency among several key findings. There is high agreement in the estimates of the global population in the ≤10m LECZ, regardless of population or elevation data set choice: the four different population data sets place between 10.2 and 11.1 percent of global population in the LECZ as measured by three of the four elevation data sources. Notable here is that two of the elevation sources – SRTM and MERIT – are based on the same underlying inputs (i.e., STRM) whereas TanDEM-X uses a different instrument to detect elevation. (CoastalDEM places between 13.1-14.5 percent of the global population in the ≤10m LECZ making it a notable exception; and while CoastalDEM also uses SRTM as its base, it also uses many ancillary data sets, including LandScan population data, and different modeling assumptions, which explain the difference.)

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

Furthermore, and importantly, the population of the LECZ is disproportionately more urban and less rural than the global population is, on average, by a substantial degree (about 1.25-1.75 times), regardless of which data sets one uses. This does not mean that in any given location or for any particular strata (i.e., in the urban continuum), data set choices do not matter, but the overwhelming pattern of a more urban LECZ is clear. Among the urban proxy data sets, there is substantial agreement in classes at the ends of the continuum – that is, locations that are classified as urban or rural – as distributed in and outside of the LECZ, but locations that are classified as quasi-urban seem to be found in roughly equal proportions in and outside of the LECZ. This may be explained in part by how urban areas are detected – urban centers have a large share of built-up area, with little unplanned vegetation, and are comparably dense (both in terms of structures and population), and the rural areas are largely vegetated or not built-up areas with sparse settlement – so if underlying detection is an issue, it is most likely to be manifest in the quasi-urban class where there is a mixture of built and unbuilt areas. Within the LECZ, most urban, and even quasi-urban area, and population, is found within the 5-10m zone, across the different population and urban proxy data sets. Consistent with this observation, population densities in the 5-10m LECZ are higher than those in the ≤5m LECZ or outside of the LECZ regardless of which population (or urban proxy) data source is used. Finally, from 1990-2015, we find unambiguous evidence that urban population has grown more in the LECZ than outside of it.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

The sensitivity analysis also reveals where input data choices result in very different estimates of population in the LECZ. Differences within the LECZ are most prominent when subdividing the zone into finer bands. The elevation data sets allocate different land areas and population totals in the zones, and result in different population estimates. Notably, CoastalDEM puts about 40% more land area and double the population in the ≤5m zone, despite still having less population and less or equal land in the 5-10m zone than the other elevation sources; it also estimates around 25% more land area overall in the ≤10m LECZ than the other DEMs. MERIT places about 20% more land area in the 5-10m zone than the other data sets, but estimates roughly the same population as TanDEM-X. The different population data sets produce estimates within the ≤10m LECZ that are generally consistent within any given population data set choice: there is about a 1 percentage point difference - or approximate 73 million persons - so by no means trivial, but much less than the differences in population estimates across the elevation data sets, or whether one subdivides the LECZ into finer bands. CoastalDEM stands as an outlier overall, but even choices between the other elevation data sets result in differences of 2% of the global population, which is large. In this regard, population estimates in the LECZ are more sensitive to the choice of elevation data than to the choice of population data.

Similarly, the urban proxy data sets result not only in different depictions of urban and quasi-urban land areas, but also population estimates by urban-rural class (which vary substantially within each urban proxy data set). Importantly, while these differences persist in and out of the LECZ, due to its urban nature the differences are more substantial outside of the LECZ. The lights-based estimates include much more land area in urban or quasi-urban than the built-up area based measures. This could be in part because of the physical nature of night-time lights, which have been shown to bloom (scatter) resulting in larger apparent footprints ~~(Small et al., 2005).~~(Small et al., 2005). The range of population estimates by population source can vary dramatically even within one proxy: by as much as 48% in the rural category and 23% in the urban category depending on which population source is used. Therefore, this sensitivity analysis indicates that the choice of population data set has large impacts on the total estimates for a given settlement class within any given urban proxy. Importantly, regardless of the urban-proxy or population data sets used as the basis for estimation, from 1990-2015, we find unambiguous evidence that urban areas have grown more in the LECZ than outside of it.

Population density measures are often used to proxy aspects of urbanization in studies of climate adaptation ~~(Solecki et al., 2015; Creutzig et al., 2015)~~(Solecki et al., 2015; Creutzig et al., 2015) and thus in this analysis we felt it important to examine the range of both population and built-up densities along the urban continuum in the LECZ. Despite the strong agreement that population density is highest in the 5-10m zone of the LECZ and that rural areas have relatively low built-up and population densities, population and built-up densities estimates vary substantially by data set choice. Population density estimates vary considerably depending on the population *and* urban proxy data ~~-~~used: ~~Globally~~globally, GHS-SMOD produces the highest population densities in all urban classes and closely followed by GHS-BUILT, but these are generally 2-3 times greater than those estimated by dLIGHT. The lights-based measures produce much lower estimates of built-up and population density ~~-~~in the urban class, in large part because this is where they also include the most land area (in the urban and quasi-urban classes). Within population data sources, even for a given urban proxy data set, the average population of urban centers varies by close to 2,000 persons/km^2. These differences across population data sets are substantially smaller within the LECZ, but how one defines the urban continuum still matters.

While population density varies by population and urban proxy data choices, we had only one measure to evaluate built-up densities. Within the LECZ, the GHS-BUILT proxy produces similar estimates of average built-up density in the ≤5m and 5-10m, as well as outside of the LECZ. This is in contrast to estimating higher population densities in the 5-10m zone, where in urban areas, densities range from 3,514 to 6,355 persons/km^2 ~~-~~(see Appendix Fig. B4), regardless of which population source data is used. The levels of built-up densities vary by the choice of urban proxy data set, but when deconstructed by LECZ zone, it is apparent that the ≤5m zone is less built than average across all input data sets. ~~Given that both population and built-density differ by urban proxy data sets, even within the LECZ, we caution users to consider carefully what a given measure means to their analysis.~~

It is clear that variations in the estimates in the LECZ can be explained through examining input data choices, but that is not the only factor which might lead to variations. The methodologies used to summarize those at risk are also important. In our use of the elevation and urban proxy data sets, we made choices that are reflected in the results. Other users of these input data would be free to make different choices and that would result, likely, in different estimates. For example, the delineation of LECZs in our work is dependent on contiguity to coastlines (connectivity),

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Pattern: Clear, Highlight

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

which eliminates spurious low lying inland areas from being misclassified as part of an LECZ. We have found that the conditioning of input DEM data has an impact on this delineation. Specifically, the reason why the use of CoastalDEM results in a more expansive LECZ is precisely because the CoastalDEM model is a smooth surface which is highly connected to coastlines. We did not apply any smoothing to TanDEM-X, and the raw data was more heterogeneous such that grid cells that were both less than 5m and within short distances of coastline, were often surrounded by barriers greater than 5m elevation. Therefore, in our construction of the LECZ, those areas are not considered as contiguous to the coast in the ≤5m zone. The same is true of the ≤10m LECZ, and results in lower estimates of population and land area based on TanDEM-X, even though it is known that TanDEM-X has the lowest RMSE globally of those DEMs we evaluated. Local studies of connectivity – both in urban settings or waterways (such as deltaic areas) – are important areas for future research to improve estimation below 10m~~.~~, and identification of the landward limits of the LECZ. While the coastal contiguity rule is ideal for application to high-resolution data in an urban setting, it should be revisited in future work along with local studies, perhaps those using new methods which model barriers explicitly, to validate ~~the existence of barriers to~~ coastal contiguity and inform our understanding of how they impede or amplify flood risk.

Similarly, with regard to the urban proxy data, the decisions we made to reduce GHS-SMOD into its level 1 classification (urban, quasi-urban, rural), and the subsequent thresholds we applied to GHS-BUILT and dLIGHT to produce those same categories could be changed or refined with other modeling rules, which would in turn alter the estimates. The settlement classes we adopted are not discrete and homogeneous as one might wish to assume, but rather encompass a range of settlement types along a spectrum. Defining urban, quasi-urban and rural requires researchers to make decisions that reflect the best available knowledge and expert judgements, but which are at some level necessarily arbitrary, or may not be the most suitable definitions for certain research questions. Given that the share of the urban population in the LECZ has grown much more so than outside the LECZ, it remains imperative that urban research continues to reflect on the conceptualization and measurement ~~-~~of this dynamic process.

Recent work has been paying greater attention to the socio-demographic issues, such as migration ~~(McMichael et al., 2020; McLeman, 2018; Hauer et al., 2020),~~(McMichael et al., 2020; McLeman, 2018; Hauer et al., 2020), adaptation in ~~(Reimann et al., 2018; Hinkel et al., 2014).~~ (Reimann et al., 2018; Hinkel et al., 2014), and managed retreat and planned relocation from ~~(Solecki and Friedman, 2021; Dannenberg et al., 2019; Geisler and Currens, 2017)~~(Solecki and Friedman, 2021; Dannenberg et al., 2019; Geisler and Currens, 2017) the LECZ. These are welcome additions to understanding populations at risk. New work has also examined aspects of the geography of LECZ, such as flooding in deltaic areas ~~(Edmonds et al., 2020; Minderhoud et al., 2019) and modelling tidal heights (Muis et al., 2020; Taherkhani et al., 2020; Du et al., 2018; Pickering et al., 2017) and multiple stressors (Anderson et al., 2018; De Dominicis et al., 2020; Moftakhari et al., 2017)~~(Edmonds et al., 2020; Minderhoud et al., 2019), modelling tidal heights (Muis et al., 2020; Taherkhani et al., 2020; Du et al., 2018; Pickering et al., 2017), subsidence (Nicholls et al., 2021, Minderhoud et al., 2018, Erkins et al., 2015) and multiple stressors (Anderson et al., 2018; De Dominicis et al., 2020; Moftakhari et al., 2017) so relevant to improving estimates of coastal exposure itself. Along with improvements to understanding DEMs in urban areas (Pesaresi et al., 2021), these represent promising new avenues towards fuller understanding of ~~seward~~seaward hazards to town and city dwellers.

In a data-rich age, we must be careful to reveal our assumptions, to understand uncertainties, and to highlight those things which are not yet well enough known. Headlines tend to highlight boldly stated findings, such as recent claims that the number of people at risk of catastrophic flooding is far greater than previously understood ~~(Kulp and Strauss, 2019; Herscher, n.d.).~~(Kulp and Strauss, 2019; Herscher, 2021). Although such claims may turn out to be true, when it comes to coming up with estimates supporting or denying such claims, the devil is in the detail, and it is important ~~-~~ to avoid an exaggerated impression of scientific debate or a rapidly fluctuating scientific consensus. This work demonstrates the impact of data choices on estimates of population and land area in the LECZ and unambiguously finds that those choices can lead to drastically different understandings of where people live and under what conditions. Improvements are continuous, and often incremental, and while there is considerable agreement on the broader patterns and trends, there is a lot of variation that reflects real uncertainties, and high levels of uncertainty will not be disappearing any time soon. The clearer we can be in articulating those areas of uncertainty, the more effective future research and policy can be.

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

8. APPENDICES

Appendix A

A1 Comparing Population Estimates by Urban Class and Population Data Sets: A closer look

There are large differences in population estimates between population sources in different measurements regardless of which elevation data set is used for the LECZ or which urban proxy data set is used to indicate the classes along the urban continuum. Even though the estimates between population data sources vary when classified by these data strata, there is also a clear pattern: estimates from GPW and GHS-POP are found on the opposite ends with estimates based in WorldPop and LandScan in between. ~~One key explanation for the~~ This variation across population data sources is ~~driven~~explained by a combination of the input resolution of the administrative units of the underlying census data that are made available (by national statistical offices) ~~. Another explanation is the modeling choices (both types and number of ancillary data).~~and modelling choices used to reallocate population within administrative units: Three of the four data sets share the same underlying population data units, but countries vary considerably in their administrative level of the data that is publicly available. When the data are available at a high resolution, variation in the modelling choices (both types and number of ancillary data) makes little difference. But where data are coarse or even of moderate resolution, the modelling choices matter. Figures B6 to B10 help to explain these differences. Figures B6, B7 and B8 focus on understanding the differences in terms of the underlying resolution GPW input units, comparing GHS-POP, WorldPop and LandScan –population estimates to GPW along the urban continuum by different urban proxy data sets (GHS-SMOD, GHS-BUILT, DLIGHT) and LECZ (using Merit DEM only). In Figures B9, B10 and Table C1, we look at differences in population estimates between population sources from a modeling perspective: in these figures, we show an independent "settlement extent" data set as a validation data (available only for some African countries) to compare population data sets used here for Kenya, Mozambique and Nigeria. We select these countries because the population data use different administrative resolutions. The validation settlement extent data set is from the GRID3 project and vector (polygon) data are derived from building footprints ~~(Center For International Earth Science Information Network and Novel-T, 2020).~~(Center For International Earth Science Information Network and Novel-T, 2020). First, buildings are extracted from high resolution imagery, then the building footprints are aggregated into polygons. Imagery that was used for feature extraction were mostly captured between 2017 and 2020. The settlement extent data set has three subclasses: built-up areas (BUA), small settlement areas (SSA) and hamlets. Building count is used for the classification. Settlements that have more than 3,000 buildings (in the aggregated polygon) are classified as BUA and they generally correspond to cities and large towns. Settlements that have buildings count between 50-3000 are classified as SSA, and generally correspond to small towns or villages. Hamlets correspond to individual farm houses or small villages.

In Figures B6, B7 and B8, the three left panels show population by urban continuum globally whereas the three right panels show the same estimates in the LECZ only. GPW is used as a baseline because it is the population input data for both GHS-POP and WorldPop. (As a reminder, GPW uses 2010 round census data.) The country-specific administrative unit levels for LandScan are not indicated in the metadata, but LandScan is still included here for comparison. Administrative level zero is country-level data and level one is first-order administrative units such as states or provinces. Levels four and higher are finely resolved, often representing enumeration areas or the finest geographic unit available in a census (such as blocks in the US). We plot the proportion of the population that falls into each urban continuum class comparing GPW to the other three population data sources. Each dot represents one country. Colors represent the administrative level that was used in the production of GPW.

Color-coded fit lines indicate administrative-level inputs and show homogeneity between countries that have the same level inputs. Fit-lines with higher R^2 statistics indicate that countries with the same administrative-level resolution are similar between population sources in terms of their differences across the urban continuum in their population estimates. The position of the fit lines relative to the diagonal line indicates the amount of difference in population shares between GPW and the other population data sets –across the urban continuum. If a fit line is in the same orientation, and close to the diagonal line, it indicates that there is high agreement between the population data set on the x-axis and GPW in terms of the population share in each respective urban class in the countries that have the same administrative-level resolution.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

In the scatterplots (B6, B7 and B8) we examine not only the input resolution of the population data sets, but also differences in the choice of urban proxy data sets as well as the modelling choices of population data sets (in Figures B9, B10, and Table C1). GHS-SMOD, GHS-BUILT and dLIGHT data sets, used for the classification of urban continuum, are shown Figures B6, B7 and B8 respectively. As shown in the figures, the way we classify the urban continuum affects the population estimate differences across the urban continuum, but it does not explain the whole story. We see the same pattern in terms of population estimates between population sources across all urban proxy data sets. Fit lines are closer to the diagonal lines as we move from GHS-BUILT-based classes to a dLIGHT-based classification. This is mostly due to the differences in area estimates: dLIGHT is more inclusive meaning that it adds more land area in urban centers and quasi urban classes (Table 6), and when the land area increases the agreement between GPW estimates also increases because of GPW's uniform distribution of population. GHS-BUILT is more exclusive in that urban centers and quasi-urban areas have the smallest area (of the urban proxy data sets) because we did not apply a contiguity rule when classifying the urban continuum. We simply applied basic cut points in built-up density (1%, 3% and 50%). This causes some single cells to be either quasi-urban or urban centers. This is much more evident in quasi-urban areas because there are many single cells with built-up density between 3% and 50% in the GHS-BUILT layer. Grid cells with built-up density of 50% or more are most likely part of a large urban agglomeration area. Therefore, in GHS-BUILT based classes, the difference in population estimates follows a somewhat different pattern especially in quasi-urban areas. Does the urban proxy data set affect the population difference across the urban continuum? Yes, but the respective differences between population sources and GPW remains the same; WorldPop and GHS-POP are located at opposite ends of the spectrum, and LandScan is in between. Also, the differences by the resolution of the administrative level remain the same. Countries with higher administrative resolution have smaller population differences across the urban continuum regardless of the urban proxy and population data sets. Therefore, urban proxy data sets affect the population differences across the urban continuum due to the extent or area of the urban classes.

Next we look at input administrative unit resolutions. As shown in the scatterplots, input administrative resolution is important: it explains much of the differences in the estimation of population between population sources. Regardless of the population sources and urban proxy data set, the population difference across the urban continuum is smaller in the countries with higher administrative resolution (i.e., those at level 4 and higher). These countries also have very high R^2 generally. High R^2 in this case means that these countries are alike in terms of the population differences across the urban continuum. Even though countries with lower administrative resolution also have higher population differences (i.e., levels 1 or 2), in most of the combinations of different data sets shown in Figures B6, B7 and B8, countries with level two have larger population differences than level one because countries with level one are mostly geographically small, such as small island countries or city-states. Another interesting pattern related to the administrative resolution is that regardless of the respective level, fit lines are much closer to the diagonal lines in urban areas than in quasi-urban areas. This is because regardless of the country-specific administrative resolution, the delineation of urban areas, in general, is at finer resolutions. We see the same pattern in the ≤10m LECZ. Regardless of administrative resolution, all of the fit lines are slightly closer to the diagonal lines in the LECZ (right panels in the scatterplots), because generally speaking, the ≤10m LECZ is more urbanized than inland areas and therefore, has finer administrative delineations regardless of the overall country specific administrative resolution.

To demonstrate, using Mozambique as an example to make that point clear, the maps in Fig. B9 and Fig. B10 illustrate how these factors (input resolution, urban areas and population models) come together inside vs outside of the LECZ. (These figures also show the reference data set, settlement extents from GRID3 building-footprint data.) GPW uses third-level administrative population data for Mozambique, which means these data are also the baseline population units from which modelling occurs in GHS-POP and ~~Worldpop~~WorldPop. As shown on ~~the~~ map B9, there is a large urban agglomeration around the capital city of Maputo, which itself has variable resolution, including many smaller units (where population is concentrated), and a relatively large portion of it is in the LECZ. However, in general, in Mozambique, inland areas (outside the LECZ) as well as those with low population, as shown in Fig. B10, the administrative resolution is very coarse.

Lastly, we look at the population differences across the urban continuum in terms of downscaling methods that were used between population sources. Even though the way we classify the urban continuum has an effect on the population differences, as shown in the scatterplots and map series, modeling is the main factor that causes the population differences across the urban continuum along with administrative resolution. Even though we do not know what level of administrative units were used in LandScan, based on WorldPop and GHS-POP, we can say that there

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

is an inverse relationship between the level of input administrative resolution and uncertainty ~~–~~in the ~~–~~applied downscaling methods. GPW is spatially imprecise in terms of estimating population because it takes the units as given and uniformly allocates population within a given spatial unit. Imprecision is greater in spatially coarse units in general and those where the population is inherently unevenly distributed (e.g., large desert or rural regions where population may be concentrated but the administrative unit is much larger than that concentrated area).

GHS-POP allocates the population within administrative units using very clear rules based on built-up presence and built-up density, whereas the other two population data sets use more complex models~~, which at the subnational level may be less transparent inherently.~~. Due to the input built-up layer that is used in the GHS-POP model, it allocates population heavily in dense built-up areas, and tends to underestimate population in areas of lower built-up density (such as rural locations). This is more important where the satellite data used to detect built-up are weak, such as in cloud-prone areas in the tropics. This overallocation is greatest if the administrative units are relatively coarse. In WorldPop and LandScan, which use additional ancillary spatial features and modelling parameters, they are able to overcome some of the overallocation issues of GHS-POP. Nevertheless, WorldPop and LandScan produce different spatial distributions of population. In general, we find that WorldPop tends to overestimate rural population and LandScan tends to overestimate urban population. The degree of misestimation is unknowable at a global scale because there is no objective baseline on which to compare them all; but some studies have compared these data sets in particular locations.

In order to clarify the role of differences in the allocating methods between population data sources shown in ~~-~~the Fig. B9 and Fig. B10, we see that GPW distributes population evenly in each admin unit but other population sources ~~-~~allocate population to settled areas. The GHS-POP layer has higher agreement with BUA class in the GRID3 settlement extent layer, but it has a value of zero (that is, estimated to be unpopulated) in SSAs and hamlets. WorldPop also captures large settlements, but it gradually lowers the population as it gets far from dense settled areas. Unlike GHS-POP, all cells are populated, but as shown in the settlement extent layer, most of these areas do not have any settlements. Therefore, WorldPop over estimates rural population and this is why WorldPop estimates are much closer to GPW population estimates in the scatterplots. LandScan falls somewhere between GHS-POP and WorldPop in terms of disaggregating population between dense (urban) and lower density (rural) settled areas. Like WorldPop, the majority of the cells are populated, but the population of unsettled areas is lower in LandScan than WorldPop, and LandScan does not allocate as much population to medium and large settlements as does GHS-POP.

Finally, in Table C1 we compared Kenya, Mozambique and Nigeria ~~-~~in order to quantify the effect of input resolution on uncertainty in the downscaling processes. We overlay population layers with settlement extent and summarize population by GRID3 settlement extent subclasses. In Kenya, where high resolution administrative level 5 data is available, the population shares between settlement extent classes are very close to each other; including population of unsettled areas across population sources as compared to Mozambique and Nigeria, which rely on coarser administrative unit inputs as their base. It is important to note here that GHS-POP also adds population to unsettled areas and as administrative resolution increases, population in unsettled areas also increases. This is due to a rule that is applied in GHS-POP; it disaggregates population evenly when there is no built-up captured in an administrative unit. In this table, we do not account for the average geographic size of the units, and even though in general, administrative level corresponds to average size, there is variation in this pattern. So for example, the agreement among population data sources in level 2 Nigeria is higher than in level 3 Mozambique for places classified as unsettled by GRID3 – this is because on average, the geographic size of the administrative units in Nigeria is indeed smaller in terms of area than in Mozambique.

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Appendix B - Figures

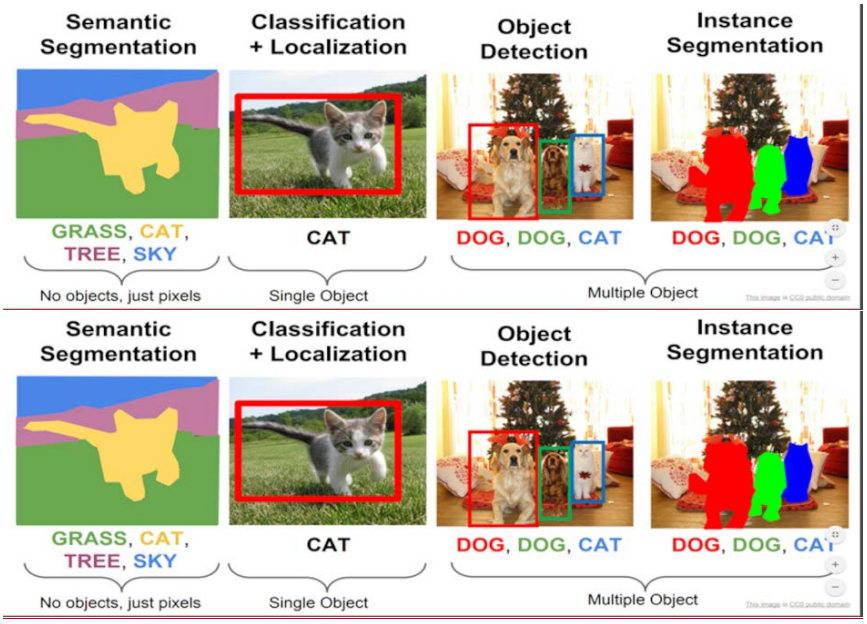
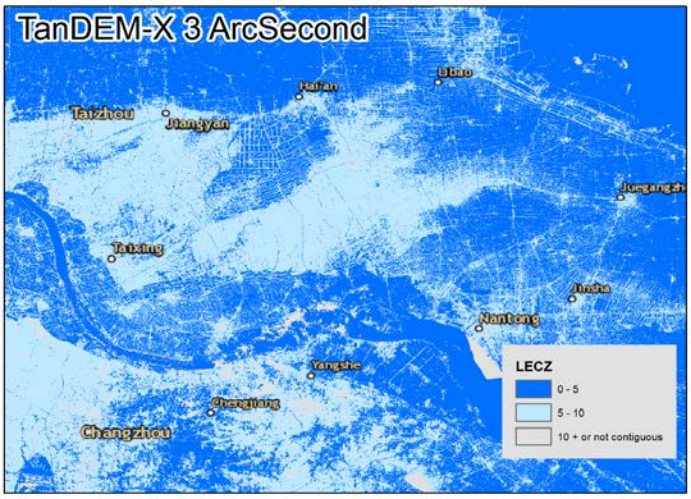
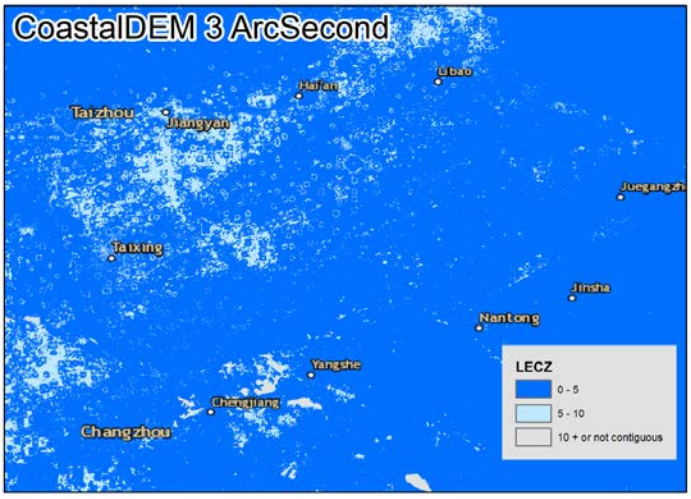


Figure B1. Region Group Concept from https://data-flair.training/blogs/image-segmentation-machine-learning/

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


TanDEM-X 3 ArcSecond
Haian
Lvbao
Taizhou
Jiangyan
Taixing
Jinsha
Nantong
Yangshe
Changzhou
LECZ
0 - 5
5 - 10
10 + or not contiguous


CoastalDEM 3 ArcSecond
Haian
Lvbao
Taizhou
Jiangyan
Taixing
Jinsha
Nantong
Yangshe
Changzhou
LECZ
0 - 5
5 - 10
10 + or not contiguous

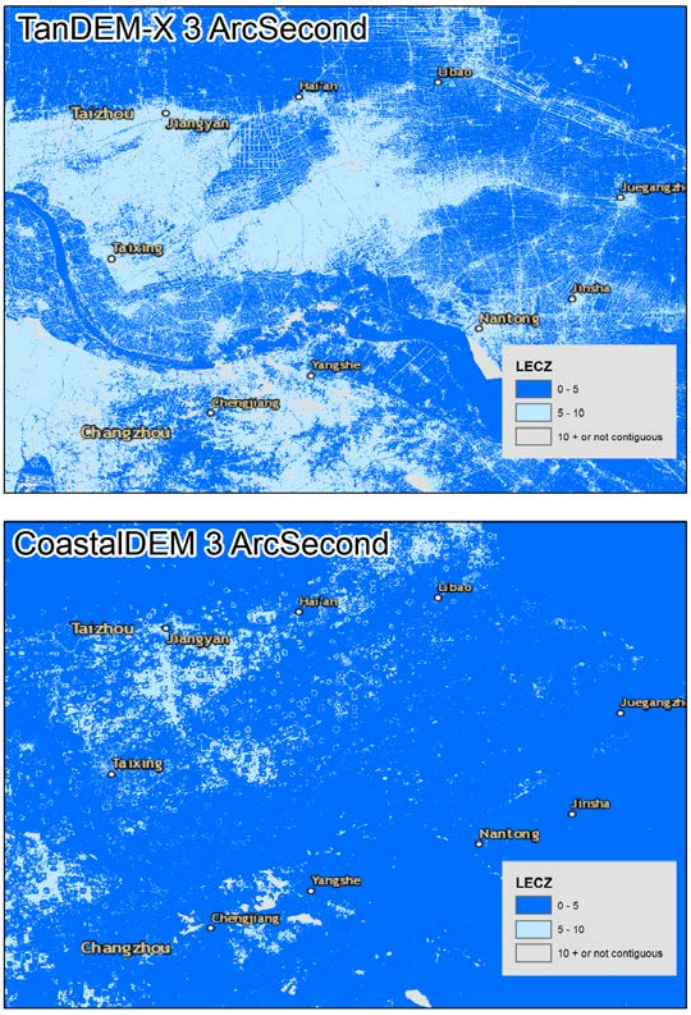


Figure B2. Raw TanDEM-X captures roadways, whereas CoastalDEM seems to capture agricultural land uses.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

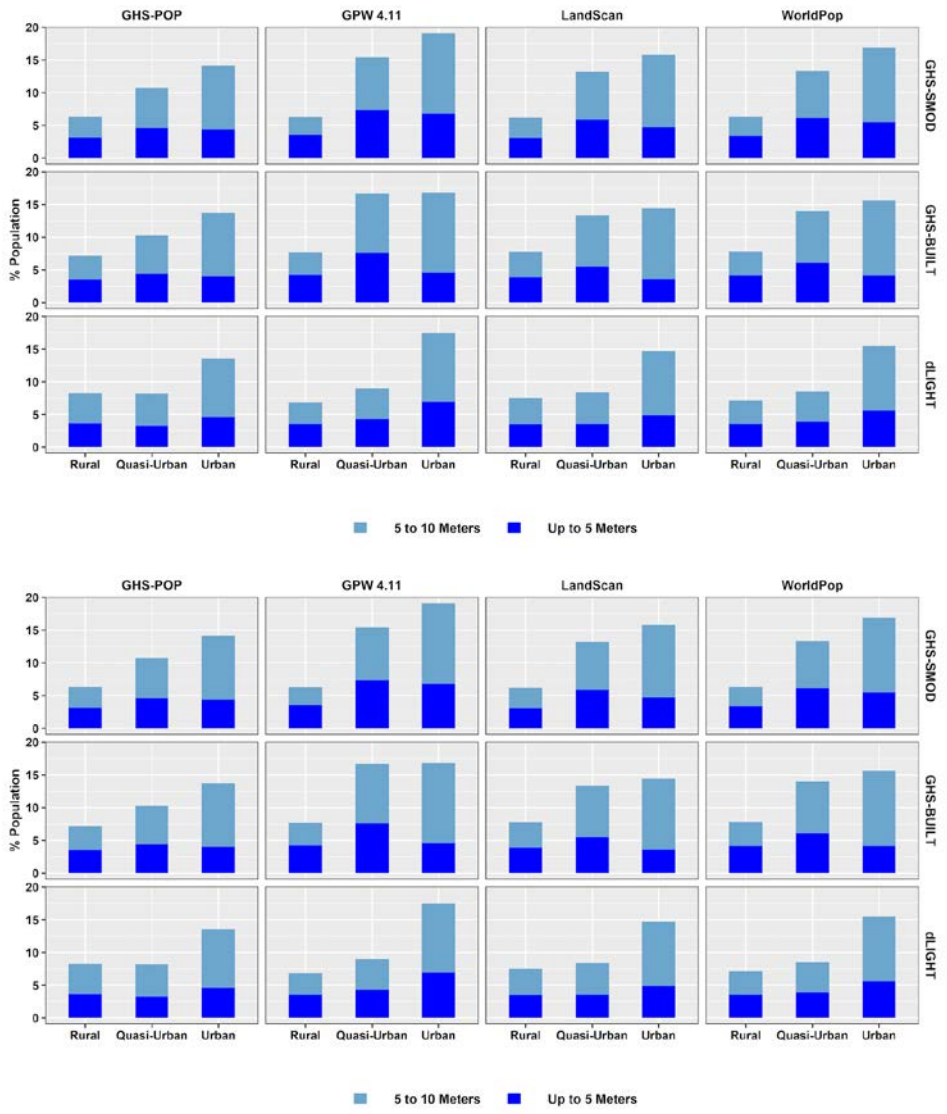
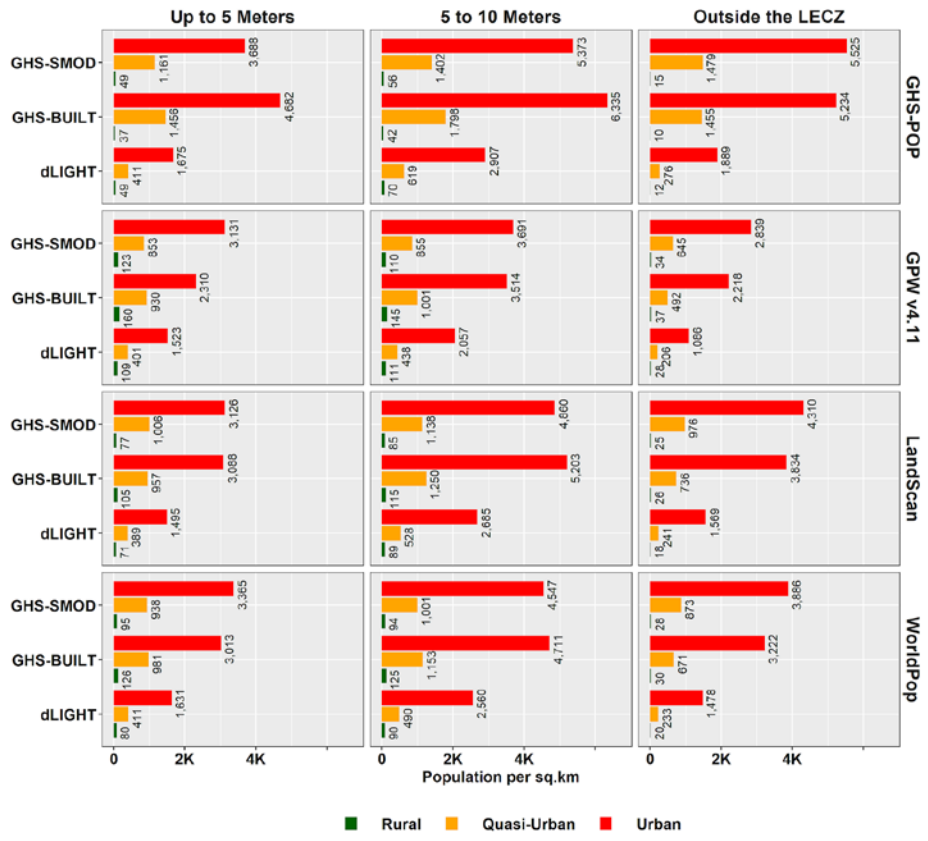


Figure B3. Proportion of the global population in each urban class (urban, quasi-urban and rural) in the ≤5m and 5-10m LECZ, by each *respective* urban-rural class, according to different Population and Urban Proxy -data sets, 2015. (MERIT DEM is used for LECZ delineation.)

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


Up to 5 Meters
5 to 10 Meters
Outside the LECZ
GHS-POP
GPW v4.11
LandScan
WorldPop
GHS-SMOD
GHS-BUILT
dLIGHT
0
2K
4K
Population per sq.km
Rural
Quasi-Urban
Urban

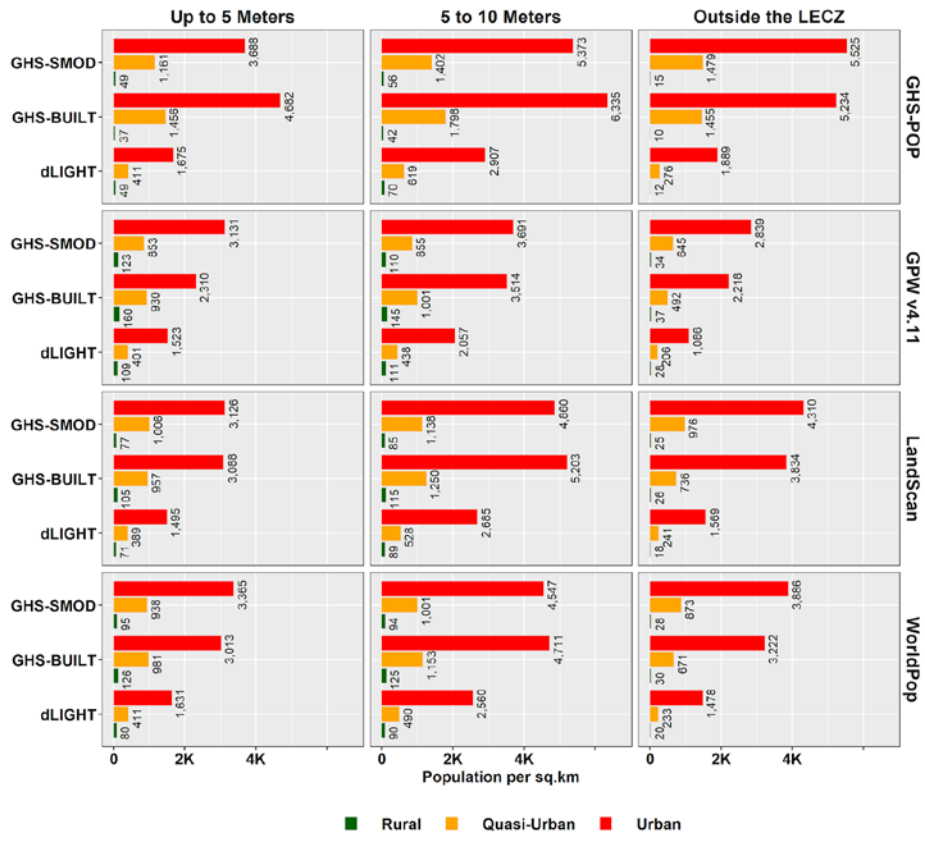
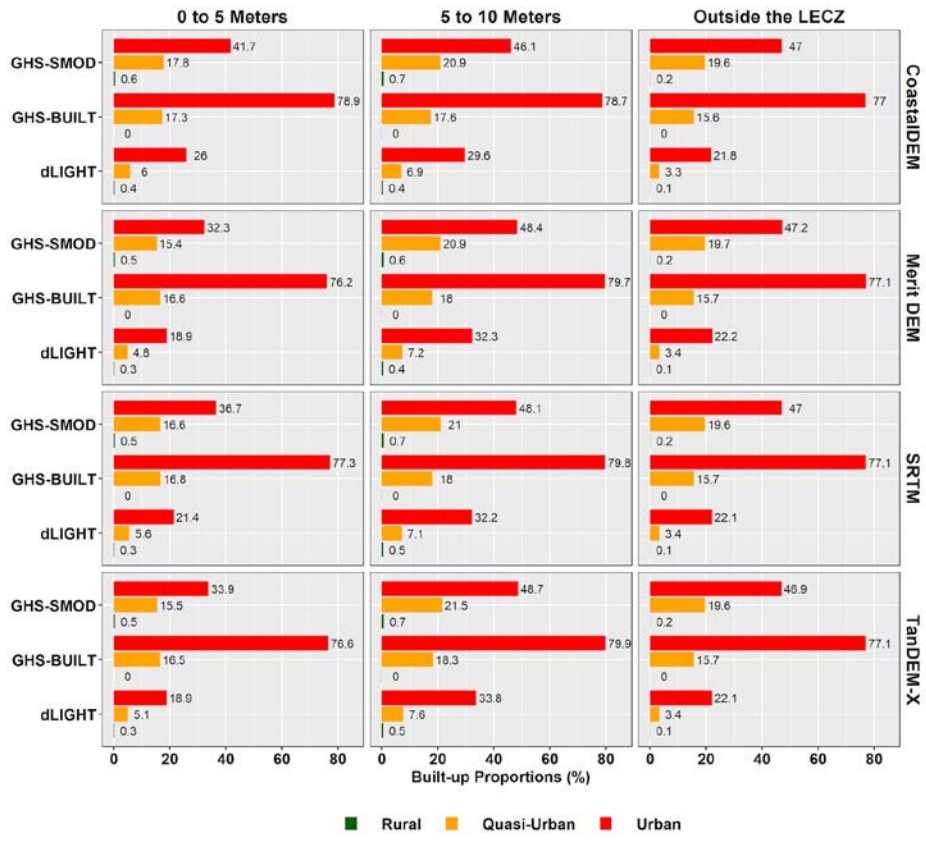


Figure B4. Population density of urban, quasi-urban and rural by urban-rural classes, by LECZ using MERIT DEM.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


0 to 5 Meters
5 to 10 Meters
Outside the LECZ
CoastalDEM
Merit DEM
SRTM
TanDEM-X
GHS-SMOD
GHS-BUILT
dLIGHT
41.7
17.8
0.6
78.9
17.3
0
26
6
0.4
46.1
20.9
0.7
78.7
17.6
0
29.6
6.9
0.4
47
19.6
0.2
77
15.6
0
21.8
3.3
0.1
32.3
15.4
0.5
76.2
16.6
0
18.9
4.8
0.3
48.4
20.9
0.6
79.7
18
0
32.3
7.2
0.4
47.2
19.7
0.2
77.1
15.7
0
22.2
3.4
0.1
36.7
16.6
0.5
77.3
16.8
0
21.4
5.6
0.3
46.1
21
0.7
79.8
18
0
32.2
7.1
0.5
47
19.6
0.2
77.1
15.7
0
22.1
3.4
0.1
33.9
15.5
0.5
76.6
16.5
0
18.9
5.1
0.3
48.7
21.5
0.7
79.9
18.3
0
33.8
7.6
0.5
48.9
19.6
0.2
77.1
15.7
0
22.1
3.4
0.1
0
20
40
60
80
Built-up Proportions (%)
Rural
Quasi-Urban
Urban

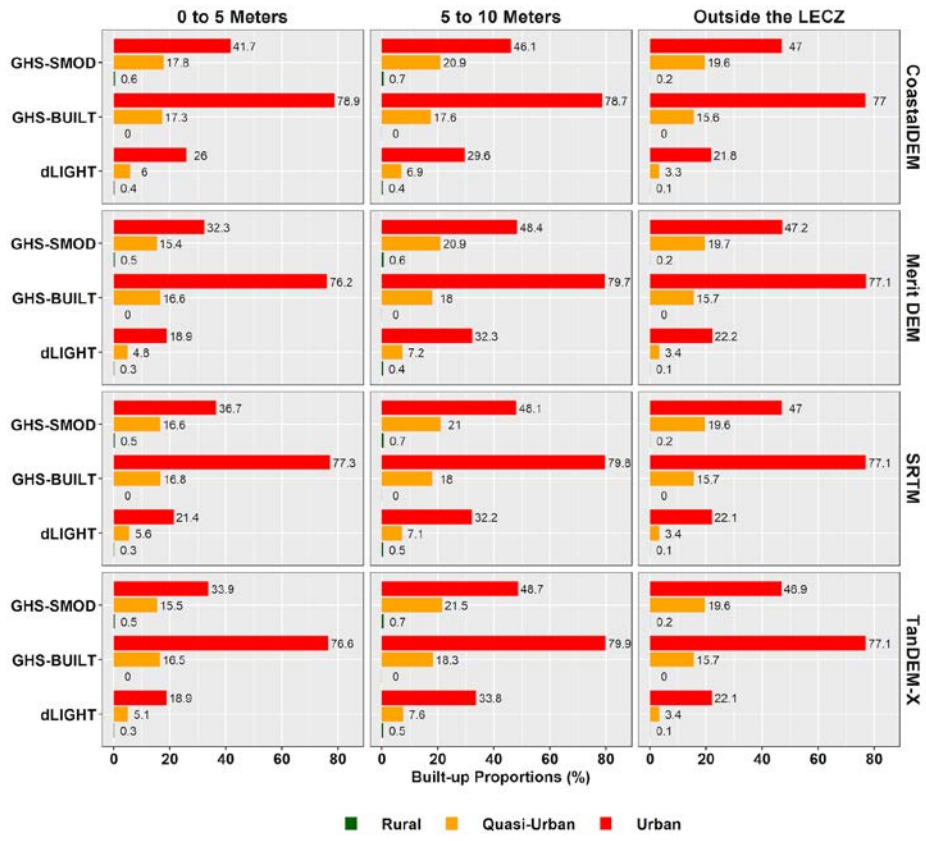


Figure B5. Built-up -density (%) of urban, quasi-urban and rural by urban-rural classes, by urban proxy and election data sets in and outside of the LECZ.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

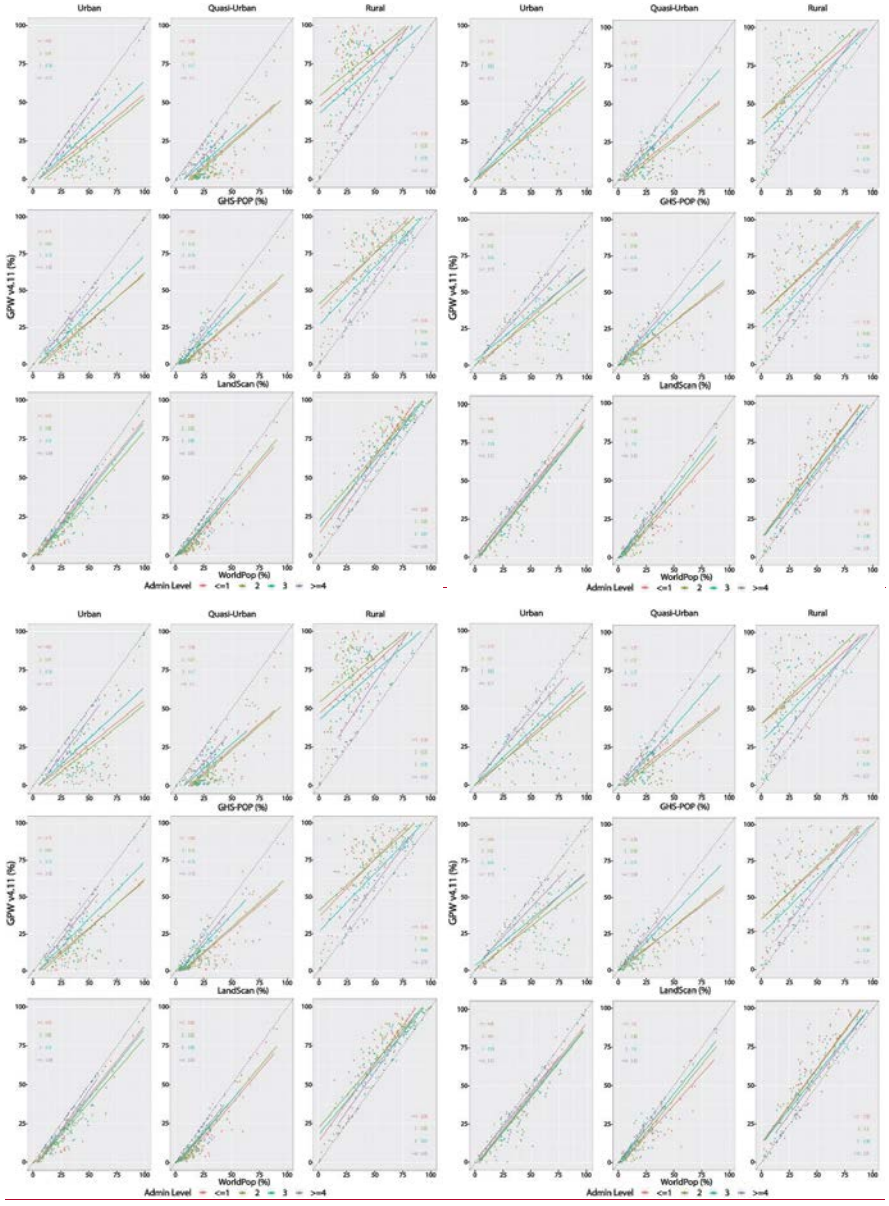
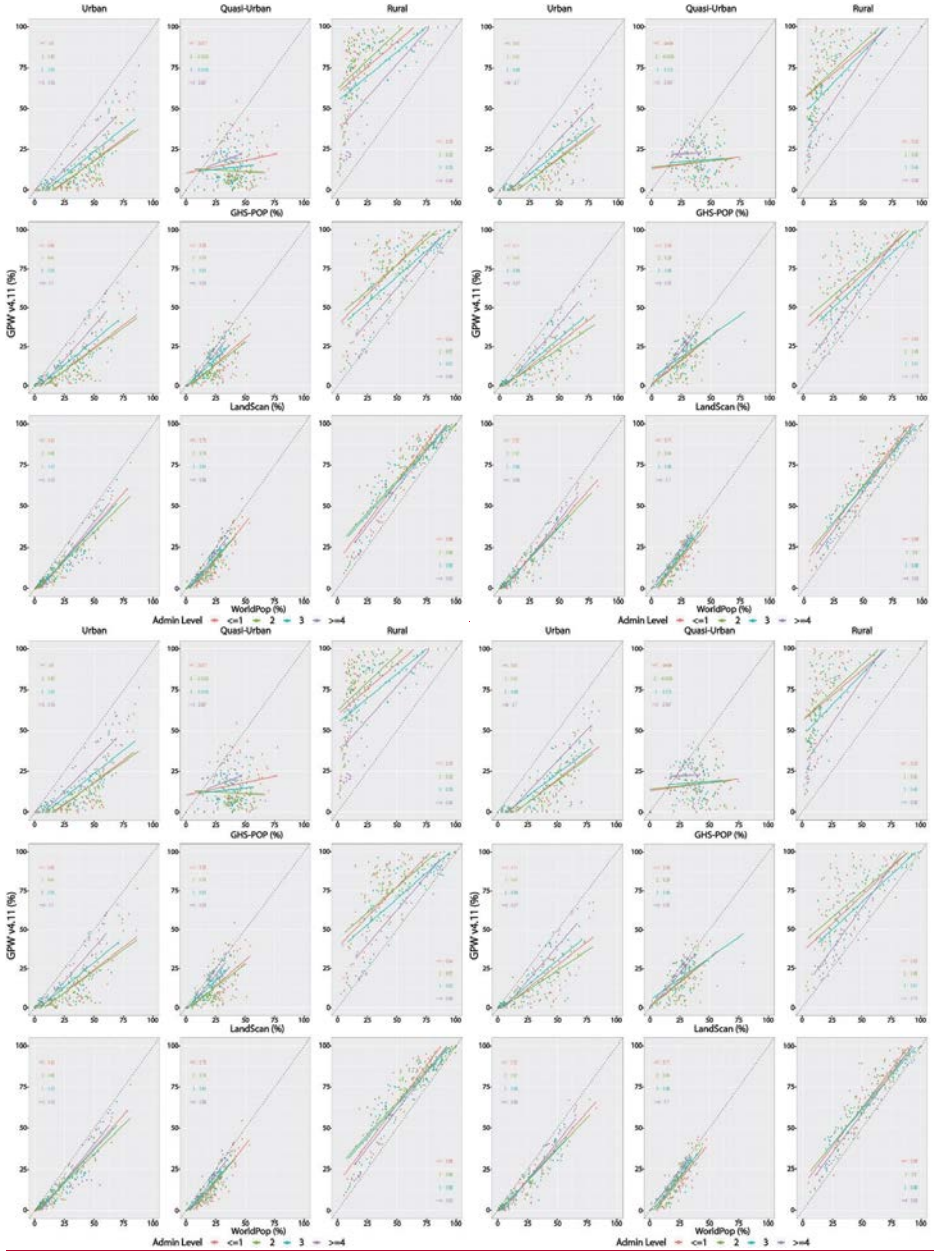


Figure B6. Comparison of Population Data Sources by level of input unit in GPW, by Urban-Rural Classes and Urban proxy data sets (GHS-SMOD). Left panel shows all land areas, the right panel shows area only in the ≤10m LECZ.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

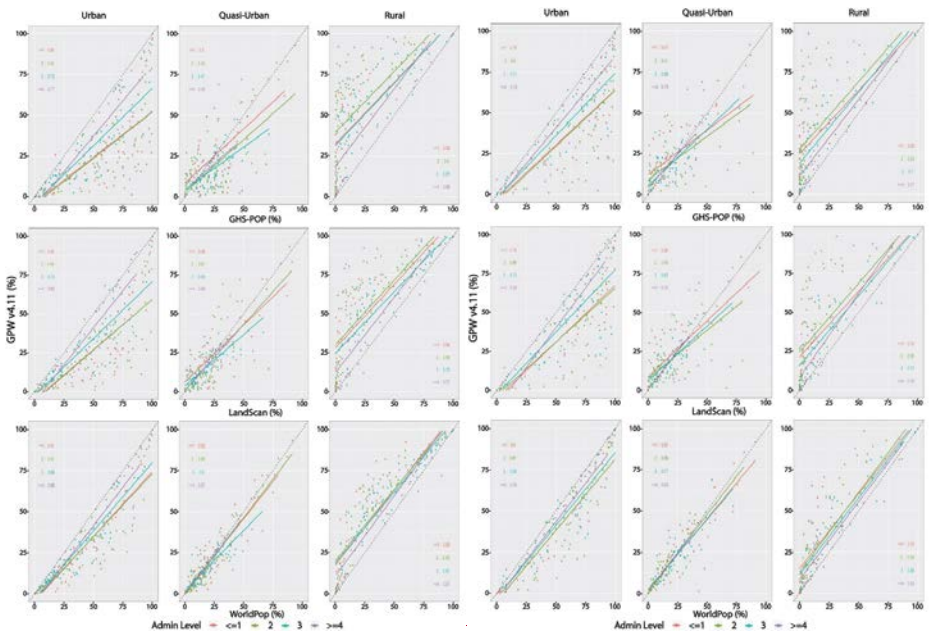
Formatted: Pattern: Clear, Highlight

Formatted: Font: 12 pt, Font color: Black


Urban
Quasi-Urban
Rural
GHS-POP (%)
GPW v4.11 (%)
LandScan (%)
WorldPop (%)
Admin Level
<=1
2
3
>=4

1670

Figure B7. Comparison of Population Data Sources by level of input unit in GPW, by Urban-Rural Classes and Urban proxy data sets (GHS-BUILT). Left panel shows all land areas, the right panel shows only under 10 meters (LECZ).



Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

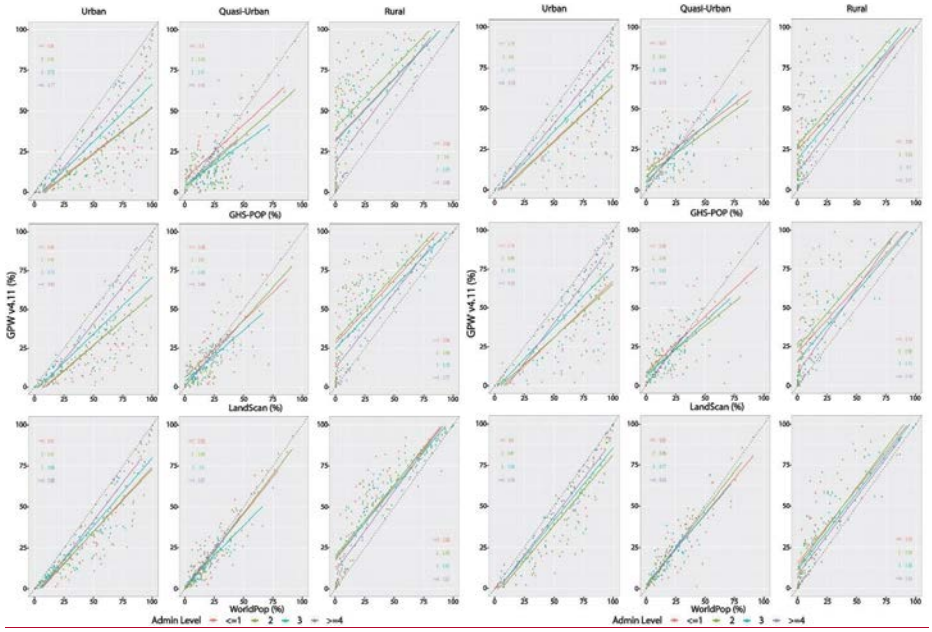
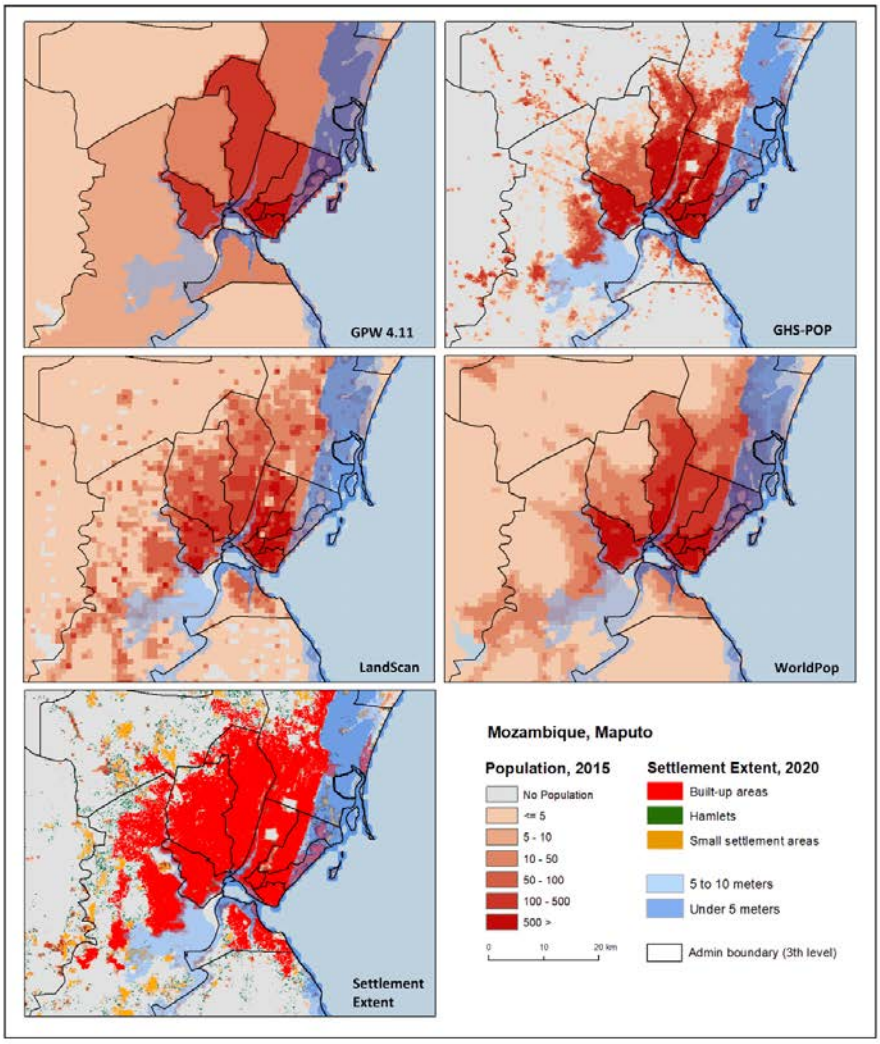


Figure B8. Comparison of Population Data Sources by level of input unit in GPW, by Urban-Rural Classes and Urban Proxy data sets (DLIGHT). Left panel shows all land areas, the right panel shows only under 10 meters (LECZ).

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


GPW 4.11
GHS-POP
LandScan
WorldPop
Settlement Extent
Mozambique, Maputo
Population, 2015
No Population
<= 5
5 - 10
10 - 50
50 - 100
100 - 500
500 >
0
10
20 km
Settlement Extent, 2020
Built-up areas
Hamlets
Small settlement areas
5 to 10 meters
Under 5 meters
Admin boundary (3th level)

1680

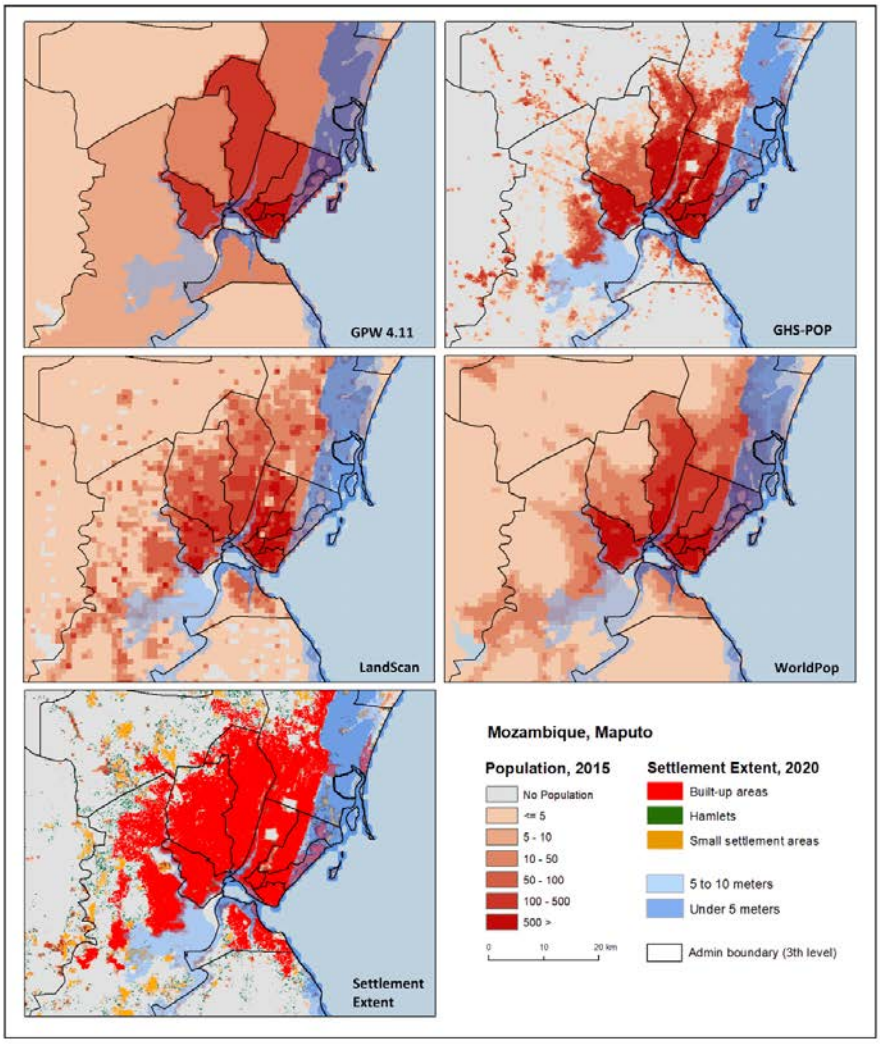
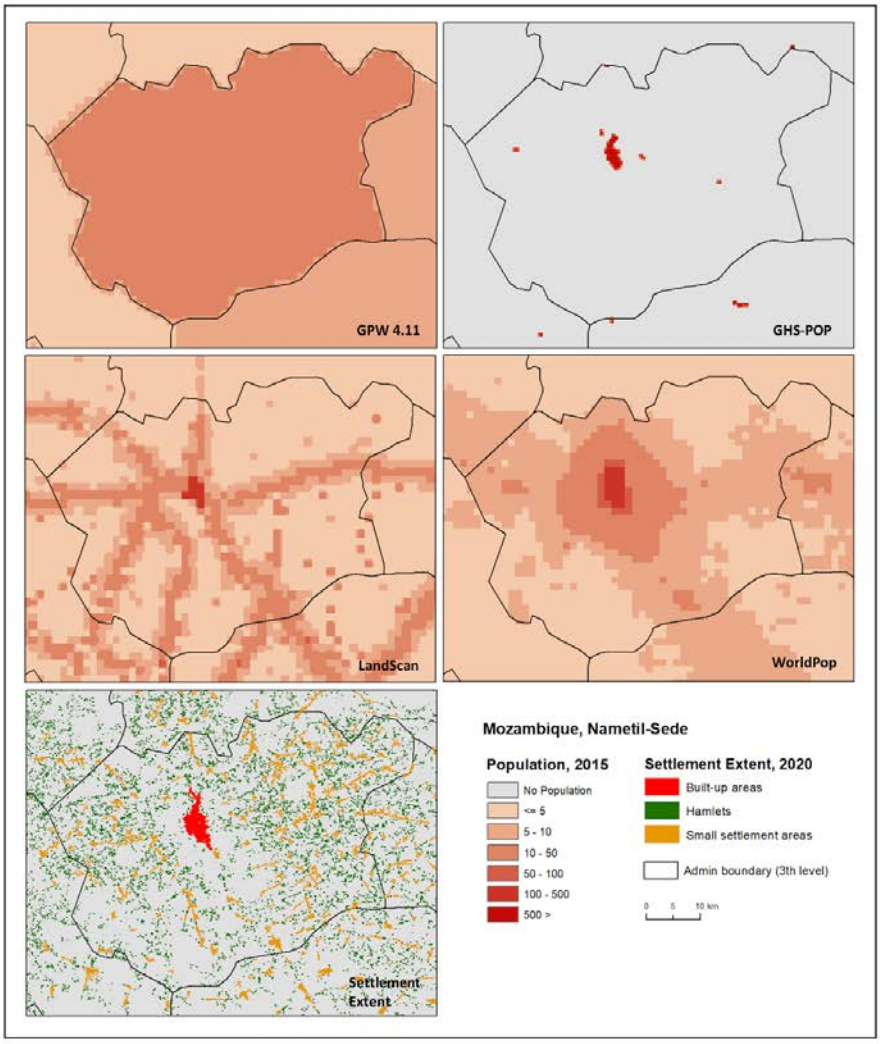


Figure B9. Comparison of population distributions by population data sources shown with administrative boundaries and GRID3 building-footprint derived settlement extent data, Maputo and surrounding region.

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


GPW 4.11
GHS-POP
LandScan
WorldPop
Settlement Extent
Mozambique, Nametil-Sede
Population, 2015
No Population
<= 5
5 - 10
10 - 50
50 - 100
100 - 500
500 >
Settlement Extent, 2020
Built-up areas
Hamlets
Small settlement areas
Admin boundary (3th level)
0 5 10 km

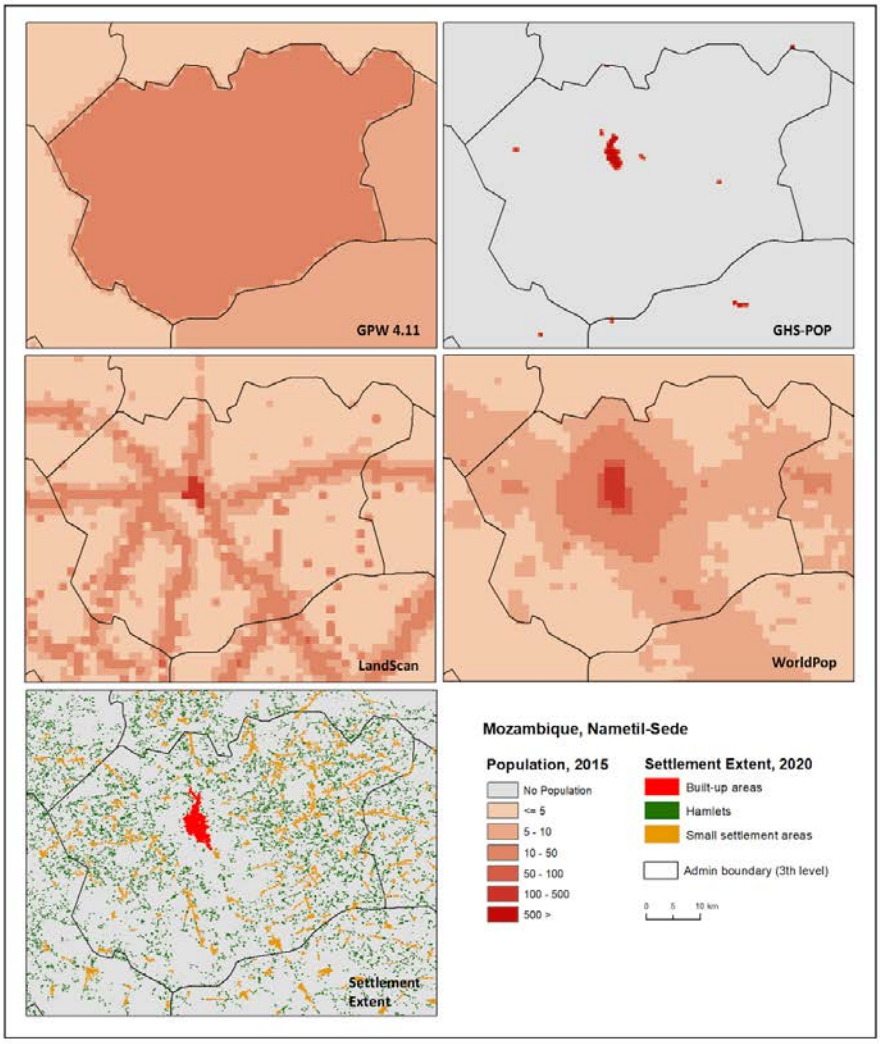
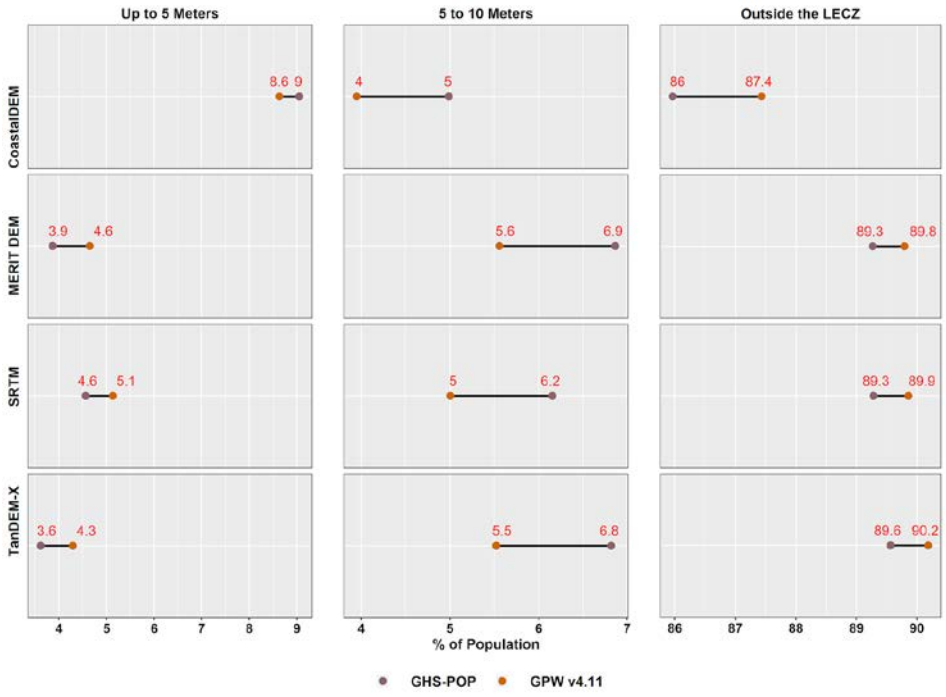


Figure B10. Comparison of population distributions by population data sources shown with administrative boundaries and GRID3 building-footprint derived settlement extent data, Nametil-Sede and surrounding region.

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black


Up to 5 Meters
5 to 10 Meters
Outside the LECZ
CoastalDEM
MERIT DEM
SRTM
TanDEM-X
8.6 9
4 5
86 87.4
3.9 4.6
5.6 6.9
89.3 89.8
4.6 5.1
5 6.2
89.3 89.9
3.6 4.3
5.5 6.8
89.6 90.2
4 5 6 7 8 9
4 5 6 7
86 87 88 89 90
% of Population
GHS-POP
GPW v4.11

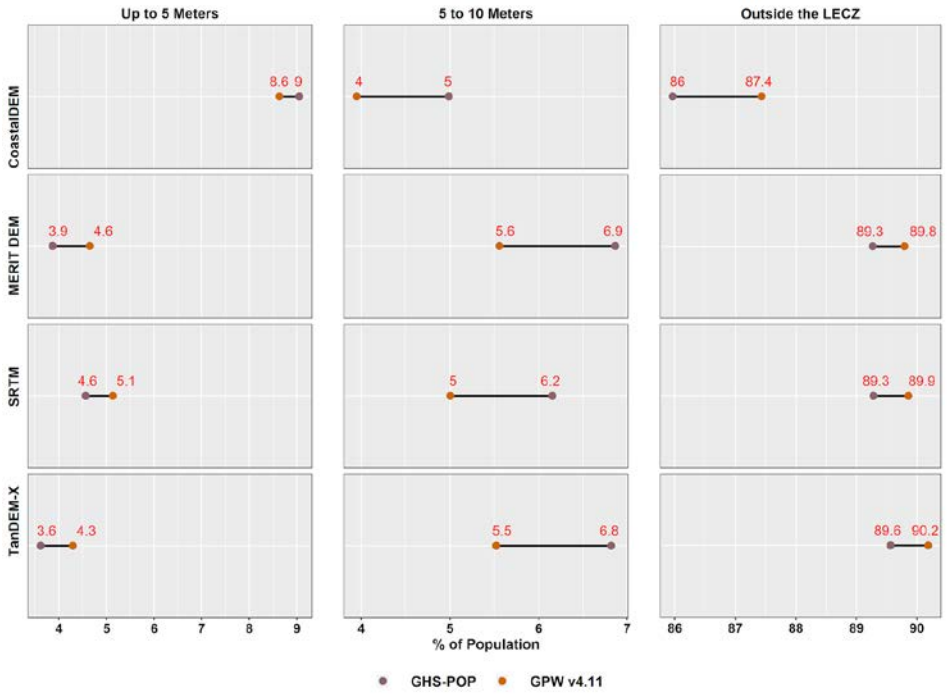


Figure B11. Estimates of Population in different LECZ zones, by elevation and population data sources, 1990.

Formatted: Normal, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

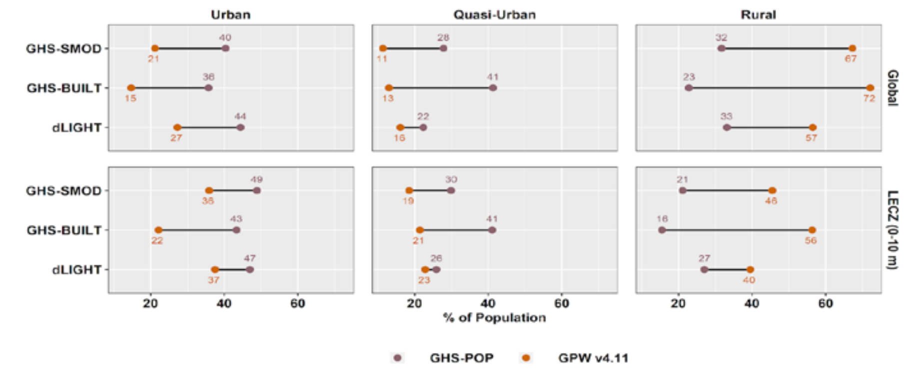


Figure B12. Percent of total population, by urban-rural classes, using different urban proxy and population data sources, globally and in the ≤10m LECZ (using MERIT DEM), 1990.

Appendix C - Tables

Table C1. Comparison of Population Data Sources by level of input unit in GPW, by Urban-Rural Classes and Urban proxy data sets, in the LECZ.

Country	Admin level	Digital-Globe based Settlement Type	WorldPop (%) Total	WorldPop (%) ≤10 Meters	GHS-POP (%) Total	GHS-POP (%) ≤10 Meters	LandScan (%) Total	LandScan (%) ≤10 Meters	GPW (%) Total	GPW (%) ≤10 Meters
Kenya	5	BUA	65	44	69	40	65	58	60	42
		SSA	12	12	13	28	14	17	12	9
		Hamlets	10	16	7	16	10	13	11	17
		No settlement	13	28	11	16	11	12	17	32
Mozambique	3	BUA	26	36	59	53	29	45	15	23
		SSA	12	9	19	14	15	9	10	9
		Hamlets	28	13	9	10	25	13	31	15
		No settlement	34	43	13	24	31	33	44	53
Nigeria	2	BUA	37	41	69	52	57	70	23	32
		SSA	17	10	22	24	27	20	15	7
		Hamlets	17	7	4	5	9	3	21	7
		No settlement	29	42	5	19	7	7	41	41

BUA (Built-up areas): Building count above 3000

SSA (Small settlement areas) : Building count between 50 to 3000

Hamlets-: Building count less than 30

Admin level: ~~Admin resolution~~Level of administrative unit used in GPW

Citation: Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2020. (GRID3). Source of building footprints "Digitize Africa data © 2020 Maxar Technologies, Ecopia.AI". https://doi.org/10.7916/d8-37sa-gy34

GRID3 Mozambique Settlement Extents Version 02, Alpha. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development

9. SUPPLEMENT LINK

10. TEAM LIST

Kytt MacManus, Deborah Balk, Hasim Engin, Gordon McGranahan, Rya Inman

11. AUTHOR CONTRIBUTION

This study was jointly led by KM and DB. It was conceptualized by DB, KM, HE and GM. Data curation was led by KM and HE with support from RI. Analysis was undertaken by DB, KM, HE, and GM. Funding was acquired by DB, GM, and KM. The methodology was developed by KM, DB, HE and GM. DB and KM administered the project. Programming and code was developed and implemented by KM and HE with support from RI. DB, KM and HE supervised the project team. All authors participated in validation, visualization, and writing the paper.

12. COMPETING INTERESTS

N/A

13. DISCLAIMER

An early version of this paper was presented at the 2019 American Geophysical Union Fall Meeting.

Views expressed in this article are not necessarily those of NASA SEDAC, CIESIN, or Columbia University.

This product was made in part utilizing the LandScan 2015 and LandScan 2000 High Resolution ~~global~~Global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy. The United States Government has certain rights in this Data Set. Neither UT-BATTELLE, LLC NOR THE UNITED STATES DEPARTMENT OF ENERGY, NOR ANY OF THEIR EMPLOYEES, MAKES ANY WARRANTY, EXPRESS OR IMPLIED, OR ASSUMES ANY LEGAL LIABILITY OR RESPONSIBILITY FOR THE ACCURACY, COMPLETENESS, OR USEFULNESS OF THE DATA SET.

This is a preliminary open data release, pending peer review of the data and associated journal articles. Following the peer review process, data curation will be completed by the NASA Socioeconomic Data and Applications Center (SEDAC) and the data will be disseminated through the SEDAC catalog.

14. ACKNOWLEDGEMENTS

This work was made possible with funding from the World Resources Institute on "Urban Populations Living in Low-Lying Coastal Areas" to the City University of New York, Institute of Development Studies at the University of Sussex and Columbia University. We thank Sarah Colenbrander, Leah Lazer and ~~Catylne~~Catlyne Haddaoui for their support and engagement especially ~~-~~during the preparation of early estimates produced for the Coalition for Urban Transitions Report *Climate Emergency, Urban Opportunity*. The completion of this work was made possible by additional support from the Socioeconomic Data and Applications Distributed Active Archive Center (DAAC) for the EOS Data and Information Systems (EOSDIS) [#80GSFC18C0111] for Kytt MacManus and Rya Inman and a Dean's Research Award at Marxe School of Public and International Affairs, Baruch College to Deborah Balk. For research assistance, we thank Alexandra Hays, Mairead Milán, Florian Mudekereza, Maria Brandao-Sze and Jessica Miller.

Formatted: Normal
Formatted: Font: 12 pt
Formatted: Font: 12 pt, Font color: Black
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Font: 12 pt, Font color: Black
Formatted: Font: 12 pt, Font color: Black
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Font: 12 pt, Font color: Black
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Font: 12 pt, Font color: Black
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Font: 12 pt, Font color: Black
Formatted: Font: 12 pt, Font color: Black
Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)
Formatted: Font: 12 pt, Font color: Black

15. References

Ahmed, F., Moors, E., Khan, M. S. A., Warner, J., and Terwisscha van Scheltinga, C.: Tipping points in adaptation to urban flooding under climate change and urban growth: The case of the Dhaka megacity, Land Use Policy, 79, 496–506, ~~https://doi.org/10.1016/j.landusepol.2018.05.051~~ https://doi.org/10.1016/j.landusepol.2018.05.051, 2018.

Anderson, T. R., Fletcher, C. H., Barbee, M. M., Romine, B. M., Lemmo, S., and Delevaux, J. M. S.: Modeling multiple sea level rise stresses reveals up to twice the land at risk compared to strictly passive flooding methods, Sci Rep, 8, 14484, ~~https://doi.org/10.1038/s41598-018-32658-x~~ https://doi.org/10.1038/s41598-018-32658-x, 2018.

Archila Bustos, M. F., Hall, O., Niedomysl, T., and Ernstson, U.: A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990–2015, Population and Environment, 42, 255–277, ~~https://doi.org/10.1007/s11111-020-00360-8~~ https://doi.org/10.1007/s11111-020-00360-8, 2020.

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., Niehüser, S., and Jensen, J.: Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts, Nat Commun, 11, 1918, ~~https://doi.org/10.1038/s41467-020-15752-5~~ https://doi.org/10.1038/s41467-020-15752-5, 2020.

Bai, Z., Wang, J., Wang, M., Gao, M., and Sun, J.: Accuracy Assessment of Multi-Source Gridded Population Distribution Datasets in China, ~~10, https://doi.org/10.3390/su10051363~~Sustainability, 10(5), 1363, https://doi.org/10.3390/su10051363, 2018.

Balk, D.: More than a name: why is global urban population mapping a grumpy proposition?, in: Global Mapping of Human Settlement: Experiences, Data Sets, and Prospects, edited by Gamba, P. & Herold, M., Taylor and Francis, Ch. 7: 145–161, 2009.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., and Nelson, A.: The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents, in: International Society for Photogrammetry and Remote Sensing, proceedings of the urban remote sensing conference, 14–16, 2005.

Balk, D., Montgomery, M., McGranahan, G., and Todd, M.: Understanding the impacts of climate change: Linking satellite and other spatial data with population data, Population dynamics and climate change, 206, 2009.

Balk, D., Leyk, S., Jones, B., Montgomery, M. R., and Clark, A.: Understanding urbanization: A study of census and satellite-derived urban classes in the United States, 1990-2010, PLoS one, 13, e0208487, 2018.

Bates, P. D., Dawson, R. J., Hall, J. W., Horritt, M. S., Nicholls, R. J., Wicks, J., and Mohamed Ahmed Ali Mohamed Hassan: Simplified two-dimensional numerical modelling of coastal flooding and example applications, Coastal Engineering, 52, 793–810, ~~https://doi.org/10.1016/j.coastaleng.2005.06.001~~ https://doi.org/10.1016/j.coastaleng.2005.06.001, 2005.

Bhardwaj, A.: Assessment of Vertical Accuracy for TanDEM-X 90 m DEMs in Plain, Moderate, and Rugged Terrain, ~~24, https://doi.org/10.3390/IECG2019-06208~~Multidisciplinary Digitial Publishing Institute Proceedings, 24, 8, https://doi.org/10.3390/IECG2019-06208, 2019.

Bright, E. A. and Coleman, P. R.: LandScan 2000, ~~https://landscan.ornl.gov/~~ [data set], https://landscan.ornl.gov/, 2001.

Bright, E. A., Rose, A. N., and Urban, M. L.: LandScan 2015, ~~https://landscan.ornl.gov/~~ [data set], https://landscan.ornl.gov/, 2016.

Buettner, T.: Urban Estimates and Projections at the United Nations: The Strengths, Weaknesses, and Underpinnings of the World Urbanization Prospects, Spat Demogr, 3, 91–108, ~~https://doi.org/10.1007/s40980-015-0004-2~~ https://doi.org/10.1007/s40980-015-0004-2, 2015.

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Calka, B. and Bielecka, E.: GHS-POP Accuracy Assessment: Poland and Portugal Case Study, ~~12, https://doi.org/10.3390/rs12071105~~Remote Sensing, 12, 1105, https://doi.org/10.3390/rs12071105, 2020.

Center for International Earth Science and Information Network - CIESIN - Columbia University: Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2, ~~https://doi.org/10.7927/H4MW2F2J~~ [data set], https://doi.org/10.7927/H4MW2F2J, 2013.

Center for International Earth Science Information Network - CIESIN - Columbia University, and CUNY Institute for Demographic Research - CIDR - City University of New York. 2021. Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 3. [data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). ~~https://doi.org/10.7927/d1x1-d702.~~https://doi.org/10.7927/d1x1-d702.

Center For International Earth Science Information Network and Novel-T: GRID3 Mozambique Settlement Extents Version 02, Alpha, ~~https://doi.org/10.7916/D8-37SA-GY34~~ [data set], https://doi.org/10.7916/D8-37SA-GY34, 2020.

Center for International Earth Science Information Network - CIESIN - Columbia University: Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11, ~~https://doi.org/10.7927/H4TD9VDP~~ [data set], https://doi.org/10.7927/H4TD9VDP, 2018.

Center for International Earth Science Information Network - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT: Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extent Polygons, Revision 02, ~~https://doi.org/10.7927/np6p-qe61~~ [data set], https://doi.org/10.7927/np6p-qe61, 2021.

~~Central, C.: Sea-Level Rise Threats in the Caribbean., Climate Central, 2018.~~

Champion, A. G. and Hugo, G. (Eds.): New forms of urbanization: beyond the urban-rural dichotomy, Ashgate, Aldershot, Hants, England ; Burlington, VT, 420 pp., ~~2004.~~https://doi.org/10.4324/9781315248073, 2004.

Chen, R., Yan, H., Liu, F., Du, W., and Yang, Y.: Multiple Global Population Datasets: Differences and Spatial Distribution Characteristics, IJGI, 9, 637, ~~https://doi.org/10.3390/ijgi9110637~~ https://doi.org/10.3390/ijgi9110637, 2020.

Cohen, B.: Urban Growth in Developing Countries: A Review of Current Trends and a Caution Regarding Existing Forecasts, World Development, 32, 23–51, ~~https://doi.org/10.1016/j.worlddev.2003.04.008~~ https://doi.org/10.1016/j.worlddev.2003.04.008, 2004.

Colenbrander, S., Lazar, L., Haddaoui, C., Godfrey, N., Lobo, A., Clarkson, H., Huxley, R., Parnell, S., Smith, B., Smith, S., and others: Climate Emergency, Urban Opportunity: The unique and crucial roles of national governments, ~~2019.~~available from https://urbantransitions.global/wp-content/uploads/2019/09/Climate-Emergency-Urban-Opportunity-report.pdf, 2019.

Corbane, C., Florczyk, A., Pesaresi, M., Politis, P., and Syrris, V.: GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A, ~~https://doi.org/10.2905/JRC-GHSL-10007~~ [data set], https://doi.org/10.2905/JRC-GHSL-10007, 2018.

Corbane, C., Pesaresi, M., Kemper, T., Politis, P., Florczyk, A. J., Syrris, V., Melchiorri, M., Sabo, F., and Soille, P.: Automated global delineation of human settlements from 40 years of Landsat satellite data archives, Big Earth Data, 3, 140–169, https://doi.org/10.1080/20964471.2019.1625528, 2019.

Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P., and Seto, K. C.: Global typology of urban energy use and potentials for an urbanization mitigation wedge, Proc Natl Acad Sci USA, 112, 6283–6288, ~~https://doi.org/10.1073/pnas.1315545112~~ https://doi.org/10.1073/pnas.1315545112, 2015.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Da Costa, J. N., Bielecka, E., and Calka, B.: Uncertainty Quantification of the Global Rural-Urban Mapping Project over Polish Census Data, in: Environmental Engineering. Proceedings of the International Conference on Environmental Engineering. ~~ICEE, 1–7,~~ICEE, 1–7, https://doi.org/10.3846/enviro.2017.221, 2017.

Dannenberg, A. L., Frumkin, H., Hess, J. J., and Ebi, K. L.: Managed retreat as a strategy for climate change adaptation in small communities: public health implications, Climatic Change, 153, 1–14, ~~https://doi.org/10.1007/s10584-019-02382-0~~ https://doi.org/10.1007/s10584-019-02382-0, 2019.

De Dominicis, M., Wolf, J., Jevrejeva, S., Zheng, P., and Hu, Z.: Future Interactions Between Sea Level Rise, Tides, and Storm Surges in the World's Largest Urban Area, Geophys. Res. ~~Lett., 47, https://doi.org/10.1029/2020GL087002~~Lett., 47, https://doi.org/10.1029/2020GL087002, 2020.

De Sherbinin, A., Schiller, A., and Pulsipher, A.: The vulnerability of global cities to climate hazards, ~~19, 39–64~~Environment and Urbanization, 19, 39–64, https://doi.org/10.1177%2F0956247807076725, 2007.

Demuzere, M., Hankey, S., Mills, G., Zhang, W., Lu, T., and Bechtel, B.: Combining expert and crowd-sourced training data to map urban form and functions for the continental US, Sci Data, 7, 264, ~~https://doi.org/10.1038/s41597-020-00605-z~~ https://doi.org/10.1038/s41597-020-00605-z, 2020.

Dijkstra, L., Poelman, H., and Veneri, P.: The EU-OECD definition of a functional urban area, ~~2019.~~OECD Regional Development Working Papers, https://doi.org/10.1787/d58cb34d-en, 2019.

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., and Schiavina, M.: Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation, Journal of Urban Economics, 103312, ~~https://doi.org/10.1016/j.jue.2020.103312~~ https://doi.org/10.1016/j.jue.2020.103312, 2020.

Doll, C. N. H. and Pachauri, S.: Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery, Energy Policy, 38, 5661–5670, ~~https://doi.org/10.1016/j.enpol.2010.05.014~~ https://doi.org/10.1016/j.enpol.2010.05.014, 2010.

Donaldson, D. and Storeygard, A.: The View from Above: Applications of Satellite Data in Economics, Journal of Economic Perspectives, 30, 171–198, ~~https://doi.org/10.1257/jep.30.4.171~~ https://doi.org/10.1257/jep.30.4.171, 2016.

Dorélien, A., Balk, D., and Todd, M.: What Is Urban? Comparing a Satellite View with the Demographic and Health Surveys, Population and Development Review, 39, 413–439, ~~https://doi.org/10.1111/j.1728-4457.2013.00610.x~~ https://doi.org/10.1111/j.1728-4457.2013.00610.x, 2013.

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., and Baptista, S. R.: Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4, Papers in Applied Geography, 1, 226–234, ~~https://doi.org/10.1080/23754931.2015.1014272~~ https://doi.org/10.1080/23754931.2015.1014272, 2015.

Du, J., Shen, J., Zhang, Y. J., Ye, F., Liu, Z., Wang, Z., Wang, Y. P., Yu, X., Sisson, M., and Wang, H. V.: Tidal Response to Sea-Level Rise in Different Types of Estuaries: The Importance of Length, Bathymetry, and Geometry: Tidal Response to Sea-level Rise, Geophys. Res. Lett., 45, 227–235, ~~https://doi.org/10.1002/2017GL075963~~ https://doi.org/10.1002/2017GL075963, 2018.

Edmonds, D. A., Caldwell, R. L., Brondizio, E. S., and Siani, S. M. O.: Coastal flooding will disproportionately impact people on river deltas, Nat Commun, 11, 4741, ~~https://doi.org/10.1038/s41467-020-18531-4~~ https://doi.org/10.1038/s41467-020-18531-4, 2020.

Elvidge, C., Ziskin, D., Baugh, K., Tuttle, B., Ghosh, T., Pack, D., Erwin, E., and Zhizhin, M.: A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data, Energies, 2, 595–622, ~~https://doi.org/10.3390/en20300595~~ https://doi.org/10.3390/en20300595, 2009.

Formatted: Font: 12 pt, Font color: Black

Elvidge, C. D., Baugh, K. E., Dietz, J. B., Bland, T., Sutton, P. C., and Kroehl, H. W.: Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements, Remote Sensing of Environment, 68, 77–88, https://doi.org/10.1016/S0034-4257(98)00098-4 https://doi.org/10.1016/S0034-4257(98)00098-4, 1999.

Erkens, G., Bucx, T., Dam, R., de Lange, G., and Lambert, J.: Sinking coastal cities, Proc. IAHS, 372, 189–198, https://doi.org/10.5194/piahs-372-189-2015, 2015.

Esch, T., Taubenböck, H., Roth, A., Heldens, W., Felbier, A., Thiel, M., Schmidt, M., Müller, A., and Dech, S.: TanDEM-X mission—new perspectives for the inventory and monitoring of global settlement patterns, J. Appl. Remote Sens, 6, 061702–1, https://doi.org/10.1117/1.JRS.6.061702 https://doi.org/10.1117/1.JRS.6.061702, 2012.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenbock, H., Muller, A., and Dech, S.: Urban Footprint Processor—Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission, IEEE Geosci. Remote Sensing Lett., 10, 1617–1621, https://doi.org/10.1109/LGRS.2013.2272953 https://doi.org/10.1109/LGRS.2013.2272953, 2013.

Esch, T., Wieke H., and Hirner, A.: "The Global Urban Footprint." In Urban Remote Sensing, pp. 3-14. CRC Press, 2017.

Esch, T., Bachofer, F., Heldens, W., Hirner, A., Marconcini, M., Palacios-Lopez, D., Roth, A., Üreyen, S., Zeidler, J., Dech, S., and Gorelick, N.: Where We Live—A Summary of the Achievements and Planned Evolution of the Global Urban Footprint, Remote Sensing, 10, 895, https://doi.org/10.3390/rs10060895 https://doi.org/10.3390/rs10060895, 2018.

Esch, T., Zeidler, J., Palacios-Lopez, D., Marconcini, M., Roth, A., Mönks, M., Leutner, B., Brzoska, E., Metz-Marconcini, A., Bachofer, F., Loekken, S., and Dech, S.: Towards a Large-Scale 3D Modeling of the Built Environment—Joint Analysis of TanDEM-X, Sentinel-2 and Open Street Map Data, Remote Sensing, 12, 2391, https://doi.org/10.3390/rs12152391 https://doi.org/10.3390/rs12152391, 2020.

European Commission. Directorate General for Regional and Urban Policy. and LSE.: Demographic and health outcomes by degree of urbanisation: perspectives from a new classification of urban areas., Publications Office, LU, 2020.

European Commission. Joint Research Centre.: GHSL data package 2019: public release GHS P2019., Publications Office, LU, 2019.available from https://data.europa.eu/doi/10.2760/290498, 2019.

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., and others: GHSL data package 2019, [data set], 29788, 290498, 2019.

Freire, S., Doxsey-Whitfield, E., MacManus, K., Mills, J., and Pesaresi, M.: Development of new open and free multi-temporal global population grids at 250 m resolution, 250, 2000Paper presented at the 19th AGILE Conference on Geographic Information Science, Helsinki, Finland., 2016.

Gao, J. and O'Neill, B. C.: Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways, 11, 1–12Nature Communications, 11, 1–12, https://doi.org/10.1038/s41467-020-15788-7, 2020.

Gaughan, A. E., Stevens, F. R., Linard, C., Patel, N. N., and Tatem, A. J.: Exploring nationally and regionally defined models for large area population mapping, International Journal of Digital Earth, 8, 989–1006, https://doi.org/10.1080/17538947.2014.965761 https://doi.org/10.1080/17538947.2014.965761, 2015.

Geisler, C. and Currens, B.: Impediments to inland resettlement under conditions of accelerated sea level rise, Land Use Policy, 66, 322–330, https://doi.org/10.1016/j.landusepol.2017.03.029 https://doi.org/10.1016/j.landusepol.2017.03.029, 2017.

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Gesch, D. B.: Best Practices for Elevation-Based Assessments of Sea-Level Rise and Coastal Flooding Exposure, Front. Earth Sci., 6, 230, ~~https://doi.org/10.3389/feart.2018.00230~~ https://doi.org/10.3389/feart.2018.00230, 2018.

Ghosh, T., Sutton, P., Powell, R., Anderson, S., and Elvidge, C. D.: Estimation of Mexico's informal economy using DMSP nighttime lights data, in: 2009 Joint Urban Remote Sensing Event, 2009 Joint Urban Remote Sensing Event, Shanghai, China, 1–10, ~~https://doi.org/10.1109/URS.2009.5137751~~ https://doi.org/10.1109/URS.2009.5137751, 2009.

Haigh, I. D., Pickering, M. D., Green, J. A. M., Arbic, B. K., Arns, A., Dangendorf, S., Hill, D. F., Horsburgh, K., Howard, T., Idier, D., Jay, D. A., Jänicke, L., Lee, S. B., Müller, M., Schindelegger, M., Talke, S. A., Wilmes, S., and Woodworth, P. L.: The Tides They Are A-Changin': A Comprehensive Review of Past and Future Nonastronomical Changes in Tides, Their Driving Mechanisms, and Future Implications, Rev. ~~Geophys., 58, https://doi.org/10.1029/2018RG000636~~Geophys., 58, https://doi.org/10.1029/2018RG000636, 2020.

Hauer, M., Hardy, D., Kulp, S., Mueller, V., Wrathall, D., Clark, P., and Oppenheimer, M.: A Framework for Classifying and Assessing Sea Level Rise Risk, ~~2020a.~~SocArXiv, https://doi.org/10.31235/osf.io/tf6r, 2020a.

Hauer, M. E., Fussell, E., Mueller, V., Burkett, M., Call, M., Abel, K., McLeman, R., and Wrathall, D.: Sea-level rise and human migration, Nat Rev Earth Environ, 1, 28–39, ~~https://doi.org/10.1038/s43017-019-0002-9~~ https://doi.org/10.1038/s43017-019-0002-9, 2020b.

Hawker, L., Neal, J., and Bates, P.: Accuracy assessment of the TanDEM-X 90 Digital Elevation Model for selected floodplain sites, Remote Sensing of Environment, 232, 111319, ~~https://doi.org/10.1016/j.rse.2019.111319~~ https://doi.org/10.1016/j.rse.2019.111319, 2019.

Henderson, J. V., Storeygard, A., and Weil, D. N.: Measuring Economic Growth from Outer Space, American Economic Review, 102, 994–1028, ~~https://doi.org/10.1257/aer.102.2.994~~ https://doi.org/10.1257/aer.102.2.994, 2012.

Henderson, J., Liu, V., Peng, C., and Storeygard, A.: European Commission, Demographic and health outcomes by degree of urbanisation: perspectives from a new classification of urban areas., Directorate General for Regional and Urban Policy Publications Office, available from https://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/demogr_health_urban_en.pdf, 2020.

Herscher, R.: A Looming Disaster: New Data Reveal Where Flood Damage Is An Existential Threat, ~~n.d~~available from https://www.npr.org/2021/02/22/966428165/a-looming-disaster-new-data-reveal-where-flood-damage-is-an-existential-threat, 2021.

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S. J., Marzeion, B., Fettweis, X., Ionescu, C., and Levermann, A.: Coastal flood damage and adaptation costs under 21st century sea-level rise, Proc Natl Acad Sci USA, 111, 3292–3297, ~~https://doi.org/10.1073/pnas.1222469111~~ https://doi.org/10.1073/pnas.1222469111, 2014.

Hooijer, A., Vernimmen, R.: Global LiDAR land elevation data reveal greatest sea-level rise vulnerability in the tropics. Nat Commun 12, 3592. https://doi.org/10.1038/s41467-021-23810-9, 2021.

Hu, X., Qian, Y., Pickett, S. T., and Zhou, W.: Urban mapping needs up-to-date approaches to provide diverse perspectives of current urbanization: A novel attempt to map urban areas with nighttime light data, ~~195,~~ Landscape and Urban Planning, 195, 103709, https://doi.org/10.1016/j.landurbplan.2019.103709, 2020.

Imhoff, M., Lawrence, W. T., Stutzer, D. C., and Elvidge, C. D.: A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area, Remote Sensing of Environment, 61, 361–370, https://doi.org/10.1016/S0034-4257(97)00046-1, 1997.

ISciences: TerraViva! SRTM30 Enhanced Global Map — Elevation/Slope/Aspect~~,~~ [data set], Ann Arbor Michigan, 2003.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Black

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Black

Kaneko, S. & Toyota, T.: Long-Term Urbanization and Land Subsidence in M. Taniguchi, editor, Asian Megacities: An Indicators System Approach. Groundwater and Subsurface Environments: Human Impacts in Asian Coastal Cities. 249-270. 10.1007/978-4-431-53904-9_13, 2011.

Khan, S.A., MacManus, K., Mills, J., Madajewicz, M., and Ramasubramanian, L.: Building Resilience of Urban Ecosystems and Communities to Sea-Level Rise: Jamaica Bay, New York City, in: Handbook of Climate Change Resilience, edited by: Leal Filho, W., Springer International Publishing, Cham, 95–115, https://doi.org/10.1007/978-3-319-93336-8_29, 2020.

Khan, A., Chatterjee, S., Akbari, H., Bhatti, S. S., Dinda, A., Mitra, C., Hong, H., and Doan, Q. V.: Step-wise Land-class Elimination Approach for extracting mixed-type built-up areas of Kolkata megacity, Geocarto International, 34, 504–527, ~~https://doi.org/10.1080/10106049.2017.1408704~~ https://doi.org/10.1080/10106049.2017.1408704, 2019.

Kopp, R. E., Horton, B. P., Kemp, A. C., and Tebaldi, C.: Past and future sea-level rise along the coast of North Carolina, USA, Climatic Change, 132, 693–707, ~~https://doi.org/10.1007/s10584-015-1451-x~~ https://doi.org/10.1007/s10584-015-1451-x, 2015.

Kulp, S. A. and Strauss, B. H.: CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network, Remote Sensing of Environment, 206, 231–239, ~~https://doi.org/10.1016/j.rse.2017.12.026~~ https://doi.org/10.1016/j.rse.2017.12.026, 2018.

Kulp, S. A. and Strauss, B. H.: New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding, Nat Commun, 10, 4844, ~~https://doi.org/10.1038/s41467-019-12808-z~~ https://doi.org/10.1038/s41467-019-12808-z, 2019.

Lewis, J.: Sea level rise: some implications for Tuvalu, Environmentalist, 9, 269–275, https://doi.org/10.1007/BF02241827, 1989.

Leyk, S., Uhl, J. H., Balk, D., and Jones, B.: Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States, Remote Sensing of Environment, 204, 898–917, ~~https://doi.org/10.1016/j.rse.2017.08.035~~ https://doi.org/10.1016/j.rse.2017.08.035, 2018.

~~Leyk, S., Balk, D., Jones, B., Montgomery, M. R., and Engin, H.: The heterogeneity and change in the urban structure of metropolitan areas in the United States, 1990–2010, 6, 1–15, 2019a.~~

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J., and Pesaresi, M.: The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use, Earth Syst. Sci. Data, 11, 1385–1409, https://doi.org/10.5194/essd-11-1385-2019, 2019a.

Leyk, S., Balk, D., Jones, B., Montgomery, M. R., and Engin, H.: The heterogeneity and change in the urban structure of metropolitan areas in the United States, 1990–2010, ~~Data, 11, 1385–1409, https://doi.org/10.5194/essd-11-1385-2019,~~Scientific Data, 6, 1–15, https://doi.org/10.1038/s41597-019-0329-6, 2019b.

Lichter, M., Vafeidis, A. T., and Nicholls, R. J.: Exploring Data-Related Uncertainties in Analyses of Land Area and Population in the "Low-Elevation Coastal Zone" (LECZ), Journal of Coastal Research, 27, 757, ~~https://doi.org/10.2112/JCOASTRES-D-10-00072.1~~ https://doi.org/10.2112/JCOASTRES-D-10-00072.1, 2010.

~~L.Imhoff, M., Lawrence, W. T., Stutzer, D. C., and Elvidge, C. D.: A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area, Remote Sensing of Environment, 61, 361–370, https://doi.org/10.1016/S0034-4257(97)00046-1, 1997.~~

Liu, Z. and Balk, D.: Urbanisation and differential vulnerability to coastal flooding among migrants and nonmigrants in Bangladesh, ~~26, e2334~~Population, Space and Place, 26, e2334, https://doi.org/10.1002/psp.2334, 2020.

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Auto

Formatted: Font color: Auto

Formatted: Normal, Left

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., Gaughan, A. E., Nieves, J. J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A., and Tatem, A. J.: Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets, Big Earth Data, 3, 108–139, ~~https://doi.org/10.1080/20964471.2019.1625151~~ https://doi.org/10.1080/20964471.2019.1625151, 2019.

MacManus, K. and Kugler, T. A.: The influence of statistical inputs on global gridded geospatial datasets, ~~2013.~~available from http://www.nsi.bg/efgs2013/data/uploads/presentations/DAY3_WS3_1_Presentation_MACMANUS_ok.pdf, 2013.

Marconcini, M., Gorelick, N., Metz-Marconcini, A., and Esch, T.: Accurately monitoring urbanization at global scale – the world settlement footprint, IOP Conf. Ser.: Earth Environ. Sci., 509, 012036, ~~https://doi.org/10.1088/1755-1315/509/1/012036~~ https://doi.org/10.1088/1755-1315/509/1/012036, 2020.

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., Todd, M., and Montgomery, M.: Urban growth, climate change, and freshwater availability, PNAS, 108, 6312–6317, https://doi.org/10.1073/pnas.1011615108, 2011.

~~McGranahan, G., Balk, D., and Anderson, B.: Low Elevation Coastal Zone (LECZ) Urban-Rural Population Estimates, Global Rural-Urban Mapping Project (GRUMP), Alpha Version, https://doi.org/10.7927/H4TM782G, 2007a.~~

McGranahan, G., Balk, D., and Anderson, B.: The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones, Environment and Urbanization, 19, 17–37, ~~https://doi.org/10.1177/0956247807076960, 2007b~~ https://doi.org/10.1177/0956247807076960, 2007a.

McGranahan, G., Balk, D., and Anderson, B.: Low Elevation Coastal Zone (LECZ) Urban-Rural Population Estimates, Global Rural-Urban Mapping Project (GRUMP), Alpha Version, https://doi.org/10.7927/H4TM782G, 2007b

McLeman, R.: Migration and displacement risks due to mean sea-level rise, Bulletin for Atomic Scientists, 74, 148–154, https://doi.org/10.1080/00963402.2018.1461951, 2018.

Mcleod, E., Poulter, B., Hinkel, J., Reyes, E., and Salm, R.: Sea-level rise impact models and environmental conservation: A review of models and their applications, Ocean & Coastal Management, 53, 507–517, ~~https://doi.org/10.1016/j.ocecoaman.2010.06.009~~ https://doi.org/10.1016/j.ocecoaman.2010.06.009, 2010a.

Mcleod, E., Hinkel, J., Vafeidis, A. T., Nicholls, R. J., Harvey, N., and Salm, R.: Sea-level rise vulnerability in the countries of the Coral Triangle, ~~5, 207–222,~~Sustain Sci, 5, 207–222, https://doi.org/10.1007/s11625-010-0105-1, 2010b.

McMichael, C., Dasgupta, S., Ayeb-Karlsson, S., and Kelman, I.: A review of estimating population exposure to sea-level rise and the relevance for migration, Environmental Research Letters, 15, 123005, 2020.

Melchiorri, M., Florczyk, A. J., Freire, S., Schiavina, M., Pesaresi, M., and Kemper, T.: Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer, ~~10, 768~~Remote Sensing, 10, 768,https://doi.org/10.3390/rs10050768, 2018.

Menashe-Oren, A. and Bocquier, P.: Urbanisation is no longer driven by migration in low-and middle-income countries (1985-2015), ~~2021.~~Population and Development Review, https://doi.org/10.1111/padr.12407, 2021.

Mesev, V. (Ed.): LandScan: a global population database for estimating populations at risk, in: Remotely-Sensed Cities, CRC Press, 301–314, ~~https://doi.org/10.1201/9781482264678-24~~ https://doi.org/10.1201/9781482264678-24, 2003.

Minderhoud, P. S. J., Coumou, L., Erban, L. E., Middelkoop, H., Stouthamer, E., and Addink, E. A.: The relation between land use and subsidence in the Vietnamese Mekong delta, Science of The Total Environment, 634, 715–726, ~~https://doi.org/10.1016/j.scitotenv.2018.03.372~~ https://doi.org/10.1016/j.scitotenv.2018.03.372, 2018.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font color: Black

Formatted: Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Minderhoud, P. S. J., Coumou, L., Erkens, G., Middelkoop, H., and Stouthamer, E.: Mekong delta much lower than previously assumed in sea-level rise impact assessments, Nat Commun, 10, 3847, ~~https://doi.org/10.1038/s41467-019-11602-1~~ https://doi.org/10.1038/s41467-019-11602-1, 2019.

Misra, P., Avtar, R., and Takeuchi, W.: Comparison of Digital Building Height Models Extracted from AW3D, TanDEM-X, ASTER, and SRTM Digital Surface Models over Yangon City, Remote Sensing, 10, 2008, ~~https://doi.org/10.3390/rs10122008~~ https://doi.org/10.3390/rs10122008, 2018.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., and Matthew, R. A.: Compounding effects of sea level rise and fluvial flooding, Proc Natl Acad Sci USA, 114, 9785–9790, ~~https://doi.org/10.1073/pnas.1620325114~~ https://doi.org/10.1073/pnas.1620325114, 2017.

Mohanty, M. P. and Simonovic, S. P.: Understanding dynamics of population flood exposure in Canada with multiple high-resolution population datasets, Science of The Total Environment, 759, 143559, ~~https://doi.org/10.1016/j.scitotenv.2020.143559~~ https://doi.org/10.1016/j.scitotenv.2020.143559, 2021.

Mondal, P. and Tatem, A. J.: Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding, PLoS ONE, 7, e48191, ~~https://doi.org/10.1371/journal.pone.0048191~~ https://doi.org/10.1371/journal.pone.0048191, 2012.

Muis, S., Apecechea, M. I., Dullaart, J., de Lima Rego, J., Madsen, K. S., Su, J., Yan, K., and Verlaan, M.: A High-Resolution Global Dataset of Extreme Sea Levels, Tides, and Storm Surges, Including Future Projections, Front. Mar. ~~Sci., 7, 263, https://doi.org/10.3389/fmars.2020.00263~~Sci., 7, 263, https://doi.org/10.3389/fmars.2020.00263, 2020.

Nations, U.: Principles and Recommendations for Population and Housing Censuses, United Nations Population Division, DESA, 2014.

Nations, U.: World Urbanization Prospects~~.~~. [data set], United Nations Population Division, DESA, 2018.

Nations, U.: 68% of the World Population Projected to Live in Urban Areas by 2050, Says UN | UN DESA | United Nations Department of Economic and Social Affairs, n.d.

Nations, U.: A recommendation on the method to delineate cities, urban and rural areas for international statistical comparisons, n.d.

Nations, U.: United Nations Expert Group Meeting on Statistical Methodology for Delineating Cities and Rural Areas, January 28-30, 2019, n.d.

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J.: Future coastal population growth and exposure to sea-level rise and coastal flooding-a global assessment, PLos one, 10, e0118571, https://doi.org/10.1371/journal.pone.0131375, 2015.

Nicholls, R. J.: Coastal Megacities and Climate Change, GeoJournal, 37, 369–379, https://doi.org/10.1007/BF00814018, 1995.

Nicholls, R. J., Lincke, D., Hinkel, J., Brown, S., Vafeidis, A. T., Meyssignac, B., Hanson, S. E., Merkens, J.-L., and Fang, J.: A global analysis of subsidence, relative sea-level change and coastal flood exposure, Nature Climage Change, 11, 338–342, https://doi.org/10.1038/s41558-021-00993-z, 2021.

Nowak Da Costa, J., Bielecka, E., and Calka, B.: Uncertainty Quantification of the Global Rural-Urban MappingProject over Polish Census Data, in: ~~Proceedings~~Proceedings of 10th International Conference "Environmental Engineering," Environmental Engineering, Vilnius Gediminas Technical University, Lithuania, ~~https://doi.org/10.3846/enviro.2017.221~~ https://doi.org/10.3846/enviro.2017.221, 2017.

OECD and European Commission: Cities in the World: A New Perspective on Urbanisation, OECD, ~~https://doi.org/10.1787/d0efcbda-en~~ https://doi.org/10.1787/d0efcbda-en, 2020.

Formatted: Font: 12 pt, Font color: Black

Oppenheimer, M. and., Hinkel, J. et al.: Sea Level Rise and Implications for Low Lying Islands, Coasts and Communities Supplementary Material, 2019. http://hdl.handle.net/11554/9280, 2019.

Orton, P., Talke, S., Jay, D., Yin, L., Blumberg, A., Georgas, N., Zhao, H., Roberts, H., and MacManus, K.: Channel Shallowing as Mitigation of Coastal Flooding, JMSE, 3, 654–673, https://doi.org/10.3390/jmse3030654 https://doi.org/10.3390/jmse3030654, 2015.

Orton, P. M., Conticello, F. R., Cioffi, F., Hall, T. M., Georgas, N., Lall, U., Blumberg, A. F., and MacManus, K.: Flood hazard assessment from storm tides, rain and sea level rise for a tidal river estuary, Nat Hazards, 102, 729–757, https://doi.org/10.1007/s11069-018-3251-x https://doi.org/10.1007/s11069-018-3251-x, 2020.

Pesaresi, M.: Principles and Applications of the Global Human Settlement Layer, in: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2047–2050, 2018.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., Syrris, V., and others: Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014, 1–62ISPRS Int. J. Geo-Inf, 1–62, https://doi.org/10.3390/ijgi8020096, 2016.

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., and Maffenini, L.: GHS Settlement Grid, Updated and Refined REGIO Model 2014 in Application to GHS-BUILT R2018A and GHS-POP R2019A, Multitemporal (1975-1990-2000-2015) R2019A, [data set], http://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218, 2019.

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

Pesaresi, M., Corbane, C., Ren, C., and Edward, N.: Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling, 16, e0244478, 2021b.

Pickering, M. D., Horsburgh, K. J., Blundell, J. R., Hirschi, J. J.-M., Nicholls, R. J., Verlaan, M., and Wells, N. C.: The impact of future sea-level rise on the global tides, Continental Shelf Research, 142, 50–68, https://doi.org/10.1016/j.csr.2017.02.004 https://doi.org/10.1016/j.csr.2017.02.004, 2017.

Pinchoff, J., Mills, C. W., and Balk, D.: Urbanization and health: The effects of the built environment on chronic disease risk factors among women in Tanzania, 15, e0241810, 2020.

Potere, D., Schneider, A., Angel, S., and Civco, D. L.: Mapping urban areas on a global scale: which of the eight maps now available is more accurate?, International Journal of Remote Sensing, 30, 6531–6558, https://doi.org/10.1080/01431160903121134 https://doi.org/10.1080/01431160903121134, 2009.

Reimann, L., Vafeidis, A. T., Brown, S., Hinkel, J., and Tol, R. S. J.: Mediterranean UNESCO World Heritage at risk from coastal flooding and erosion due to sea-level rise, Nat Commun, 9, 4161, https://doi.org/10.1038/s41467-018-06645-9 https://doi.org/10.1038/s41467-018-06645-9, 2018a.

Reimann, L., Merkens, J.-L., and Vafeidis, A. T.: Regionalized Shared Socioeconomic Pathways: narratives and spatial population projections for the Mediterranean coastal zone, 18, 235–245Regional Environmental Change, 18, 235–245, https://doi.org/10.1007/s10113-017-1189-2, 2018b.

Rose, A. N. and Bright, E. A.: The LandScan Global Population Distribution Project: current state of the art and prospective innovation, Paper presented at the Population Association of America Annual Meeting, Boston MA, USA, 2014.

Rossi, C. and Gernhardt, S.: Urban DEM generation, analysis and enhancements using TanDEM-X, ISPRS Journal of Photogrammetry and Remote Sensing, 85, 120–131, https://doi.org/10.1016/j.isprsjprs.2013.08.006 https://doi.org/10.1016/j.isprsjprs.2013.08.006, 2013.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Saleem Khan, A., MacManus, K., Mills, J., Madajewicz, M., and Ramasubramanian, L.: Building Resilience of Urban Ecosystems and Communities to Sea Level Rise: Jamaica Bay, New York City, in: Handbook of Climate Change Resilience, edited by: Leal Filho, W., Springer International Publishing, Cham, 95–115, https://doi.org/10.1007/978-3-319-93336-8_29, 2020.

Schneider, A., Friedl, M. A., and Potere, D.: Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions,' Remote Sensing of Environment, 114, 1733–1746, https://doi.org/10.1016/j.rse.2010.03.003 https://doi.org/10.1016/j.rse.2010.03.003, 2010.

Schumann, G. J.-P. and Bates, P. D.: The Need for a High-Accuracy, Open-Access Global DEM, Front. Earth Sci., 6, 225, https://doi.org/10.3389/feart.2018.00225 https://doi.org/10.3389/feart.2018.00225, 2018.

Small, C.: Spatiotemporal Network Evolution of Anthropogenic Night Light 1992-2015, 2020.Physics and Society, https://arxiv.org/abs/2005.12197, 2020.

Small, C. and Center For International Earth Science Information Network-CIESIN-Columbia University: VIIRS Plus DMSP Change in Lights (VIIRS+DMSP dLIGHT), https://doi.org/10.7927/9RYJ-6467) [data set], https://doi.org/10.7927/9RYJ-6467, 2020.

Small, C., Pozzi, F., and Elvidge, C.: Spatial analysis of global urban extent from DMSP-OLS night lights, Remote Sensing of Environment, 96, 277–291, https://doi.org/10.1016/j.rse.2005.02.002 https://doi.org/10.1016/j.rse.2005.02.002, 2005.

Small, C., Sousa, D., Yetman, G., Elvidge, C., and MacManus, K.: Decades of urban growth and development on the Asian megadeltas, Global and Planetary Change, 165, 62–89, https://doi.org/10.1016/j.gloplacha.2018.03.005 https://doi.org/10.1016/j.gloplacha.2018.03.005, 2018a.

Small, C., Okujeni, A., van der Linden, S., and Waske, B.: Remote Sensing of Urban Environments, in: Comprehensive Remote Sensing, Elsevier, 96–127, https://doi.org/10.1016/B978-0-12-409548-9.10380-X https://doi.org/10.1016/B978-0-12-409548-9.10380-X, 2018b.

Solecki, W. and Friedman, E.: At the Water's Edge: Coastal Settlement, Transformative Adaptation, and Well-Being in an Era of Dynamic Climate Risk, 42, 211–232Annual Review of Public Health, 42, 211–232, https://doi.org/10.1146/annurev-publhealth-090419-102302, 2021.

Solecki, W., Seto, K. C., Balk, D., Bigio, A., Boone, C. G., Creutzig, F., Fragkias, M., Lwasa, S., Marcotullio, P., Romero-Lankao, P., and others: A conceptual framework for an urban areas typology to integrate climate change mitigation and adaptation, Urban Climate, 14, 116–137, https://doi.org/10.1016/j.uclim.2015.07.001, 2015.

Stokes, E. C. and Seto, K. C.: Characterizing urban infrastructural transitions for the Sustainable Development Goals using multi-temporal land, population, and nighttime light data, Remote Sensing of Environment, 234, 111430, https://doi.org/10.1016/j.rse.2019.111430 https://doi.org/10.1016/j.rse.2019.111430, 2019.

Strauss, B. H.: Flooded Future: Global Vulnerability to Sea Level Rise Worse than Previously Understood, n.dfrom https://www.climatecentral.org/news/report-flooded-future-global-vulnerability-to-sea-level-rise-worse-than-previously-understood, 2019.

Strauss, B. H, & Kulp, S..: Sea-Level Rise Threats in the Caribbean, Inter-American Development Bank, available from https://sealevel.climatecentral.org/uploads/ssrf/Sea-level-rise-threats-in-the-Caribbean.pdf, 2018.

Syvitski, J., Kettner, A., Overeem, I. et al.: Sinking deltas due to human activities. Nature Geosci 2, 681–686. https://doi.org/10.1038/ngeo629, 2009.

Tacoli, C.: Rural-urban interactions: a guide to the literature, Environment and Urbanization, 10, 147–166, https://doi.org/10.1177%2F095624789801000105, 1998.

Formatted: Font: 12 pt, Font color: Auto

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H.: Precise Global DEM Generation by ALOS PRISM, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II–4, 71–76, https://doi.org/10.5194/isprsannals-II-4-71-2014 https://doi.org/10.5194/isprsannals-II-4-71-2014, 2014.

Taherkhani, M., Vitousek, S., Barnard, P. L., Frazer, N., Anderson, T. R., and Fletcher, C. H.: Sea-level rise exponentially increases coastal flood frequency, Sci Rep, 10, 6466, https://doi.org/10.1038/s41598-020-62188-4 https://doi.org/10.1038/s41598-020-62188-4, 2020.

Taupo, T. and Noy, I.: Disaster Impact on Households in Tuvalu, Diakses dariAccessed from https://www.nzae.org.nz/wpcontent/uploads/2016/10/Tauisi…, Taupo.pdf, 2016.

Taupo, T., Cuffe, H., and Noy, I.: Household vulnerability on the frontline of climate change: the Pacific atoll nation of Tuvalu, Environ Econ Policy Stud, 20, 705–739, https://doi.org/10.1007/s10018-018-0212-2 https://doi.org/10.1007/s10018-018-0212-2, 2018.

Tebaldi, C., Strauss, B. H., and Zervas, C. E.: Modelling sea level rise impacts on storm surges along US coasts, Environmental Research Letters, 7, 014032, 2012.

Tellman, B., Sullivan, J., Kuhn, C., Kettner, A., Doyle, C., Brakenridge, G., Erikson, T., and Slayback, D.: Satellite observations indicate increasing proportion of population exposed to floods, 2020Nature, 596, 80-86, https://doi.org/10.1038/s41586-021-03695-w, 2021.

Tessler, Z.D., Vörösmarty, C.J., Grossberg, M., Gladkova, I., Aizenman, H., Syvitski, J.P. and Foufoula-Georgiou, E.: Profiling risk and sustainability in coastal deltas of the world. Science, 349(6248), 638-643, https://doi.org/10.1126/science.aab3574, 2015.

Tong, L., Hu, S., and Frazier, A. E.: Mixed accuracy of nighttime lights (NTL)-based urban land identification using thresholds: Evidence from a hierarchical analysis in Wuhan Metropolis, China, Applied Geography, 98, 201–214, https://doi.org/10.1016/j.apgeog.2018.07.017 https://doi.org/10.1016/j.apgeog.2018.07.017, 2018.

Uchiyama, Y. and Mori, K.: Methods for specifying spatial boundaries of cities in the world: The impacts of delineation methods on city sustainability indices, Sci. Total Environ., 592, 345–356, https://doi.org/10.1016/j.scitotenv.2017.03.014, 2017.

Uuemaa, E.; Ahi, S.; Montibeller, B.; Muru, M.; Kmoch, A. Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). Remote Sens., 12, 3482. https://doi.org/10.3390/rs12213482, 2020.

Vafeidis, A. T., Schuerch, M., Wolff, C., Spencer, T., Merkens, J. L., Hinkel, J., Lincke, D., Brown, S., and Nicholls, R. J.: Water-level attenuation in global-scale assessments of exposure to coastal flooding: a sensitivity analysis, Nat. Hazards Earth Syst. Sci., 19, 973–984, https://doi.org/10.5194/nhess-19-973-2019Sci., 19, 973–984, https://doi.org/10.5194/nhess-19-973-2019, 2019.

Wentz, E., Anderson, S., Fragkias, M., Netzband, M., Mesev, V., Myint, S., Quattrochi, D., Rahman, A., and Seto, K.: Supporting Global Environmental Change Research: A Review of Trends and Knowledge Gaps in Urban Remote Sensing, Remote Sensing, 6, 3879–3905, https://doi.org/10.3390/rs6053879 https://doi.org/10.3390/rs6053879, 2014.

Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., and Roth, A.: Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data, ISPRS Journal of Photogrammetry and Remote Sensing, 139, 171–182, https://doi.org/10.1016/j.isprsjprs.2018.02.017 https://doi.org/10.1016/j.isprsjprs.2018.02.017, 2018.

Yamano, H., Kayanne, H., Yamaguchi, T., Kuwahara, Y., Yokoki, H., Shimazaki, H., and Chikamori, M.: Atoll island vulnerability to flooding and inundation revealed by historical reconstruction: Fongafale Islet, Funafuti Atoll, Tuvalu, Global and Planetary Change, 57, 407–416, https://doi.org/10.1016/j.gloplacha.2007.02.007 https://doi.org/10.1016/j.gloplacha.2007.02.007, 2007.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Formatted: Normal, Left, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations: Accurate Global Terrain Elevation map, Geophys. Res. Lett., 44, 5844–5853, ~~https://doi.org/10.1002/2017GL072874~~ https://doi.org/10.1002/2017GL072874, 2017.

Zink, M., Bachmann, M., Brautigam, B., Fritz, T., Hajnsek, I., Moreira, A., Wessel, B., and Krieger, G.: TanDEM-X: The New Global DEM Takes Shape, IEEE Geosci. Remote Sens. ~~Mag., 2, 8–23, https://doi.org/10.1109/MGRS.2014.2318895~~Mag., 2, 8–23, https://doi.org/10.1109/MGRS.2014.2318895, 2014.

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black