Different Spatiotemporal Patterns in Global Human Population and Built-Up Land

Jing Gao and Brian O’Neill

1Department of Geography and Spatial Sciences, Data Science Institute, University of Delaware, Newark, DE, USA, 2U.S. Department of Energy, Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

Abstract  Population concentration and built-up land expansion are two prominent features of contemporary urbanization. Existing literature on the population aspect of urbanization has mostly focused on national and regional aggregates, and literature on the land development aspect has often relied on spatial case studies of individual cities or their meta-analyses. Using newly-available data, here we conduct the first global-coverage, spatial analysis of the relationship between (changes in) population and built-up land at multiple spatial scales, and compare to existing common beliefs about urbanization based on individual city studies. We find that population and built-up land show distinctly different spatial and temporal patterns (with a global correlation coefficient around 0.6). Contrary to common impressions, our results show that during recent decades, developed and developing regions across the world experienced comparable amounts of built-up land expansion. While meta-analyses have reported that built-up land in urban areas expands globally on average twice as fast as population grows, our results show the global change rates of built-up land and population are similar. Also, most global population, including what national statistics agencies call urban population, reside in areas with low land development levels (which are frequently less than 5% built up). These changes in perspective suggest that urbanization's potential large-scale impacts may need to be re-evaluated, and lead to best-practice recommendations for urbanization modeling and analysis. Especially, the common practice in large-scale earth system modeling of assuming demographically-defined urban population resides in areas with medium to high built-up land development levels should change.

Plain Language Summary  Cities and towns across the world have been expanding (a process known as urbanization) both in terms of population size and land areas. These two expressions of urbanization (i.e., population concentration and built-up land expansion) have often been assumed to be strongly associated. This assumption is influential for understanding people's vulnerability to natural disasters (for example, heatwaves, coastal storms) in urban areas, but has not been tested globally, largely due to the lack of data. Recently new global-coverage maps of population and built-up land became available. Using the newly-available data, we examine the relationship between population and built-up land across the world. We find the correlation between population and built-up land is about 60%, with the two showing some distinctly different spatial and temporal patterns. Contrary to common impressions, we find developed and developing regions across the world experienced similar amounts of built-up land expansion during recent decades. Moreover, most people in the world (including those considered urban people by national agencies, such as census) live in areas with less than 5% built-up land. These changes in perspective suggest that urbanization's potential large-scale impacts may need to be re-evaluated.

1. Introduction

Contemporary urbanization is profoundly transforming our world (e.g., Grimm et al., 2008; Wu, 2014). Rapid concentration of population and massive expansion of built-up land are two of its prominent features (AAAS, 2016). The two phenomena are related: new built-up land development may be demanded by various needs of the growing urban population (e.g., Bierwagen et al., 2010), and the form and functions of built-up land can affect the spatial distribution of population (e.g., Frolking et al., 2013). The spatiotemporal relationship between population and built-up land connects the social and the physical aspects of urbanization, and determines how urbanization interacts with many environmental changes (e.g., Balk & Montgomery, 2015; Hibbard et al., 2010; Sánchez-Rodriguez et al., 2005; Wolff et al., 2020). Understanding
this relationship therefore informs global sustainable development efforts, including international cooperative frameworks like the UN Sustainable Development Goals (SDGs), and helps the calculation of indicators measuring progress toward various sustainability targets, for example, the SDG Indicator 11.3.1 (ratio of land consumption rate to population growth rate) (UN, 2015, 2017).

Despite the apparent association, the relationship between population and built-up land varies across space, time, and scale, and can sometimes be counterintuitive (e.g., Geist et al., 2006; Seto et al., 2010). Existing literature on demographic characteristics of contemporary urbanization has focused more on regional and global aggregate discussions (e.g., national totals) (e.g., Angel et al., 2011; Cohen, 2004; Montgomery, 2008), and literature on urban land conversion, although usually more spatially resolved (e.g., gridded), has mostly been based on case studies of individual cities, stratified samples of individual cities (e.g., Angel et al., 2016; Schneider et al., 2015), or meta-analyses of these studies (e.g., Güneralp et al., 2020; Seto et al., 2011). This literature has provided important understanding and widely-accepted narratives of contemporary urbanization. For example, Asia and developing regions in general have often been considered hotspots of new urban development (e.g., Angel et al., 2016; Schneider et al., 2015); and according to exhaustive meta-analyses (e.g., Güneralp et al., 2020; Seto et al., 2011), urban centers across the world on average expand their land areas twice as fast as their population growth (suggesting a strong increase in urban sprawl).

However, these studies are subjects to selection biases from multiple sources: In existing literature, certain cities and regions (e.g., Houston, Texas) are more frequently studied than others; and small settlement sites are less examined, even though there are many more small towns than large cities in the world and the collective total population growth of smaller settlement sites outpace that of large cities (United Nations, Department of Economic and Social Affairs, Population Division, 2018). Such selection biases of the existing literature hamper the ability of meta-analyses based on small-scale studies to capture the overall spatiotemporal trends of urbanization at global and regional scales. Similarly, although stratified samples of individual cities usually aim to be representative, strata designed for different studies serve different purposes and can be biased for other analyses. For example, a sample stratified to categorize cities according to their population sizes with each category consisting of the same number of cities can help understand how cities of different population sizes compare but would be unlikely to accurately capture global average urbanization trends since the actual frequency distribution of cities with different population sizes is far from uniform.

These selection biases raise the question whether/what common beliefs about urbanization patterns derived from small-scale case studies and their meta-analyses hold at larger scales. Meanwhile, existing literature (including meta-analyses) has not quantified many basic aspects of the relationship between population and built-up land at larger scales (e.g., how spatially correlated are the two variables across the world). Without this kind of knowledge, current spatial modeling outcomes show considerable uncertainties due to their different assumptions about the two variables’ spatial and temporal relationships (Goldewijk & Verburg, 2013).

To address these issues, we take advantage of new, global-coverage, spatially-explicit, multi-epoch observational data (that were not available when meta-analyses were first resorted to for understanding global trends) and examine the spatiotemporal relationship between population and built up land. The global gridded observations enable large-scale testing of commonly-believed urbanization patterns. This spatial time-series data-driven investigation is the first of its kind, can update existing urban theories, and produce globally generalizable and spatially specific knowledge about urbanization.

We ask two overarching questions: (1) How do population and built-up land relate across scales (gridded, national, regional, global) as two prominent aspects of contemporary urbanization? (2) What suggestions can be extended from results of question (1) for socio-environmental analysis and modeling of urbanization? At first glance, question (1) may resemble research questions of the urban scaling literature (e.g., Batty, 2008; Bettencourt, 2013). The key difference is this research focuses on spatially-explicit trends and patterns free of different urban definitions, while the urban scaling literature views each city as a whole without considering variations within a city or across regions.

We compare trends and patterns shown by global-coverage spatial data with existing common beliefs about urbanization derived from small-scale studies, especially: (a) How much has built-up land expanded in various geographic regions across the world (e.g., the developing vs. the developed world) over recent past
Earth’s Future

decades? (b) How do change rates of built-up land expansion and population growth compare across the world? Is built-up land expanding twice as fast as population growth in global-coverage observations? (c) Can population and built-up land substitute each other when defining “urban”? Though the complex relationship between population and built-up land is well noted in urban science, in larger-scale earth system modeling it is common to treat the two variables synonymously when delineating urban areas (e.g., Jones et al., 2018), due to practical considerations like data availability, as well as the common intuitive impression that they are highly correlated considering the Earth’s entire land mass. While the validity of this assumption has seldom been tested, it can strongly affect conclusions of many socio-environmental analyses examining impacts of environmental stresses in urban areas (e.g., urban heat island effects) on the global population. Here, with global-coverage observations, we evaluate (d) what is the typical built-up land development level that urban populations across the world live in? We anticipate our results will confirm some conclusions from existing literature while updating some others.

2. Data

2.1. Remote Sensing Based Built-Up Land Data

For built-up land (a land cover type characterized by man-made structures and materials, e.g., buildings, roads, concrete, asphalt), satellite remote sensing provides globally-consistent direct observations (Potere et al., 2009). Man-made materials and structures have unique physical and thermal properties that can be observed by satellites for mapping built-up land presence on the Earth surface (e.g., Mahtta et al., 2019). While the same remote sensor’s data quality is usually invariant across space and time, current remote sensing based built-up land mapping products are all prone to some known quality issues, for example, lower mapping accuracy in areas with low development levels and mixed pixels (e.g., Theobald, 2005; Wickham et al., 2017). Moreover, the quality of remote sensing based built-up land maps varies across the world, since more training data and knowledge are available for conventionally data-rich regions (e.g., Europe, North America) (e.g., Pesaresi et al., 2015), and also because constructions in the developing world (conventionally data-poor) more frequently use materials different from those common in the developed world (e.g., clay) (e.g., Linard et al., 2013).

We used the Global Human Settlement Layer (GHSL) v.1 (Pesaresi et al., 2015) in this research. Being LandSat-based, it offers the longest time series with fine spatial resolution for global built-up land mapping, which aligns well with the time points of existing spatial population datasets and enables longer-term investigations with population data at the national and above scales (Table 1). Though the abovementioned data quality issues are innate to all remote sensing based land cover products, GHSL’s fine spatial resolution (38 m) limits the scope of the issues. Further, for the time epoch of our spatial analysis (1990–2000), GHSL data are based on the same remote sensor (Landsat Thematic Mapper) with invariant quality, which helps instill confidence in our findings about spatiotemporal trends. The 1975 GHSL data however are based on a different remote sensor more prone to underestimate built-up areas (Landsat Multispectral Scanner). For our aggregate analysis (1975–2014), this systematic bias is likely to inflate the estimated amount of built-up expansion for the time epoch 1975–1990. When interpreting results, we kept these data characteristics in mind and only drew robust conclusions in spite of them.

| Name            | Observed variable | Spatial resolution | Time epochs of observations |
|-----------------|-------------------|--------------------|----------------------------|
| WUP 2018        | (Urban) population| Nations            | once every 5 years since 1950 |
| GRUMP v.1       | Population        | 30" (roughly 1 km) | 1990, 1995, 2000            |
| GPW v.3         | Population        | 2.5' (roughly 4.5 km) | 1990, 1995, 2000            |
| GPW v.4         | Population        | 30"                | 2000, 2005, 2010            |
| GHSL v.1        | Land cover        | 38 m               | 1975, 1990, 2000, 2014      |

*GPW v.4 is not used in this research, but it is a commonly-used dataset discussed in Section 2.2.
2.2. Gridded Spatial Population Data

For gridded spatial population, all current datasets (e.g., those listed in Table 1) are modeled results, as population is surveyed/observed by administrative units. The underlying models’ assumptions and biases are a major source of uncertainty for our analysis. Some popular spatial population data are modeled using remote sensing based built-up land maps as input, for example, LandScan (Bhaduri et al., 2007; Rose et al., 2019) and WorldPop (Lloyd et al., 2017), and hence are not suitable for this study due to the endogeneity (Leyk et al., 2019). To minimize uncertainty, datasets with simpler underlying models are preferred for this research, because that way the gridded population data are as close to direct observations and as independent of built-up land observations as possible.

A recent review of large-scale gridded population data recommended the Gridded Population of the World (GPW) for correlation/relationship analyses because it avoids endogeneity (Leyk et al., 2019). GPW uniformly distributes populations observed at the finest spatial units available within each unit across the world. It can be considered a direct observation with minimal processing, but homogeneity within census units is a strong assumption and can introduce bias to spatially-explicit analyses at the gridded scale. Also, to align with GHSL’s time epochs (Table 1), we used GPW v.3 (CIESIN, 2005) rather than the newest version GPW v.4 (CIESIN, 2016) in this research, though GPW v.3 is based on many fewer spatial units (less than 20,000 globally) than GPW v.4 (millions). Using the older version therefore would mean, for some parts of the world, larger land areas are assumed to have homogeneous population counts. In both versions, the size of the finest census units available varies across the world and can be large in data-poor regions (e.g., Africa). This adds uncertainties to our conclusions, especially those about data-poor regions. Such uncertainties are difficult to quantify, due to the lack of high-quality, spatially-detailed observational data, which is why models were needed to produce gridded population at the first place.

Given potential underlying biases, we took a sensitivity-range approach by repeating all analyses requiring gridded population in this research on two datasets: GPW v.3 and the Global Rural-Urban Mapping Project (GRUMP) v.1 (CIESIN, 2011). GRUMP is also based on census data at the finest spatial units available. It aimed to improve issues of GPW’s homogeneity assumption by using simple models incorporating human settlement records and the U.S. Department of Defense’s Nighttime Light data, which is remote sensing based built-up land information and introduces endogeneity. Though simple, GRUMP’s underlying model has its own biases (e.g., assuming dispersed spatial distributions for rural populations; low ability to detect areas with low or no electrification) (e.g., Bai et al., 2018; Bustos et al., 2020). Despite these limitations, out of all currently available gridded population datasets, GPW and GRUMP are the two closest to direct observations independent of built-up land observations. Moreover, by repeating all analyses on both datasets, our results (e.g., Table 3) come with a sensitivity range, and if the range is narrow, conclusions can be drawn with some confidence. For example, one can imagine the correlation between population and built-up land reported using GPW v.3 is likely lower than the unattainable “true” value due to the spatial homogeneity assumption, and the correlation reported using GRUMP is likely higher due to the endogeneity. If these two correlation coefficients are similar, they together roughly indicate the “true” level of the correlation between population and built-up land. In our analyses, we found this usually the case. For that reason, below we report primarily the GRUMP-based numerical results and only present GPW v.3 results if they show noticeable differences.

2.3. Aggregate National Population Data

To further confine potential uncertainties in our results from gridded population data, for analyses at national and above spatial scales, we used aggregate observations of national total and urban population counts (based on country-specific socioeconomic urban definitions), the latest World Urbanization Prospects (WUP) (United Nations, Department of Economic and Social Affairs, Population Division, 2018). We qualitatively account for available knowledge about underlying data quality, when interpreting our results and inferring their implications for socio-environmental analysis and modeling of urbanization.
3. Methods

3.1. Defining Urban

For this analysis, the multi-faceted nature of defining “urban” must be considered: “urban” areas may be places with higher population density, larger built-up land share, less agricultural employment, more service infrastructure, etc., in contrast to “rural” (e.g., Mahtta et al., 2019; United Nations, Department of Economic and Social Affairs, Population Division, 2018). As a result, the geographic boundary between urban and rural areas is somewhat arbitrary (e.g., Frey & Zimmer, 2001; McGranahan et al., 2005). Some researchers therefore argued for mapping the spatial continuum of “urbanness” (e.g., Small et al., 2011; Uhl et al., 2020). Although this approach avoids drawing a somewhat arbitrary line, it does not reduce the variety of potential criteria associated with the many dimensions of “urban.” These criteria are usually supported by two data sources: (i) national socioeconomic statistics using country-specific urban definitions that greatly vary across countries (Buettner, 2015), for example, national censuses, and (ii) remote sensing based land cover maps with more uniform and spatially-resolved definitions, for example, GHSL. Urban definitions based on different criteria may best serve different analytical purposes but can lead to substantially different numerical results (EU, 2016).

Since urbanization is an integrated dynamic of both population and land cover land use change, we acknowledge that both country-specific demographics-based urban definitions and remote-sensing-derived land-surface-based urban definitions will remain widely used, and account for both types of definitions by identifying numerical links between them as well as quantitative relationships between the two aspects of urbanization.

3.2. Preparing Spatial Data

At gridded scales, we examined two variables: built-up land fraction, and population density (people/sq km), and used 1/8 degree (roughly 14 km at the Equator) as our primary spatial resolution. We used GHSL (which maps built-up land as presence/absence at 38 m) to calculate built-up land fraction as the ratio of built-up pixels to all land pixels within each 1/8-degree grid. We derived two versions of 1/8-degree population density (one version based on GRUMP, and the other based on GPW) by respectively averaging the finer-resolution population densities reported by the two datasets. The GRUMP-based and the GPW-based population density layers are each used with the GHSL-based built-up land fraction, to generate two versions of grid-level analyses. Using fraction and density variables (instead of the typical binary variable of built-up or not) allows us to test/use a range of spatial resolutions (including very coarse ones, e.g., 2-degree) without losing the fine precision of our base datasets (Table 1; e.g., 38 m for GHSL).

Two aspects of the 1/8-degree spatial resolution choice are worth explaining:

1. The spatial granularity: Generally, as the spatial resolution coarsens, the correlation coefficient between two spatial variables increases, which is commonly known as the scale effect of the Modifiable Areal Unit Problem (MAUP) (a common source of statistical bias in spatial analysis; Openshaw, 1981). The trend could naturally arise due to the fact that higher-level spatial aggregation filters out small-scale (and often noisy) signals at very fine spatial resolution, and hence increases the statistical strength of identified correlation between variables. Balancing the needs to map spatial details, filter out small-scale noisy signals, and capture meaningful robust relationships, we implemented a sensitivity test to investigate the correlation between gridded population and built-up land at a range of commonly-used spatial resolutions for global studies (30°, 1/8-degree, 1/2-degree, 1-degree, and 2-degree), and examined how the correlation coefficient changed. As will be shown in results (Section 4.1), 1/8 degree appears an ideal choice for characterizing the population-built-up-land relationship, since it is a good tradeoff between offering a fine spatial resolution and smoothing out excessive, noisy details. When data allow, some analyses in this research were repeated at various spatial resolutions, and we found the results are generally qualitatively similar to the 1/8-degree results, hence below we only present these results when they show important differences from the 1/8-degree results.

2. The coordinate system: Using a spatial resolution based on geographic coordinates (unit: latitude and longitude degrees) rather than projected coordinates (unit: km, miles, etc.) is somewhat unusual for urban analyses. Conventionally, equal-area map projections (i.e., map projections that do not distort area...
measures) are preferred (Slocum et al., 2009), to minimize analytical uncertainty introduced by the fact that land areas within geographic coordinate based grid cells (e.g., 1/8-degree grids) are not the same across the world and are larger at higher latitudes. But, grids in geographic rather than projected coordinate systems are common for current global earth system and environmental models, and large-scale socio-environmental analyses (e.g., Jones et al., 2018) – many of our base datasets were developed for grids based on geographic coordinates (Table 1). 1/8 degree therefore is an appropriate balance among different conventions of all relevant fields, and our use of fraction and density variables attenuates potential impacts on uncertainties in our conclusions. As a sensitivity/validation test, we calculated the correlation between gridded population density and built-up land fraction using an equal-area map projection (Mollweide) at 1 and 14 km, and compared them with corresponding results at 30° and 1/8-degree, respectively. The results (as will be shown in Section 4.1) suggest that, for density and fraction variables, geographic and project coordinates based spatial resolutions lead to similar conclusions about inter-variable relationships.

At the 1/8-degree gridded scale, we analyzed two time points (1990, 2000) when GHSL’s and GPW v3’s time epochs align (Table 1).

3.3. Preparing Aggregate Data

At aggregate scales, we analyzed the world as 227 countries and nine regions (Figure 1). Aggregate totals of built-up land were sums of grid-cell totals, which were estimated as the product of grid-cell built-up land fraction and grid-cell total land areas. Aggregate totals of (urban) populations were extracted from WUP’s national-level observations (United Nations, Department of Economic and Social Affairs, Population Division, 2018).

Aligning the time epochs of relevant datasets, we analyzed four time points (1975, 1990, 2000, 2014/5) at aggregate scales (Table 1).

3.4. Integrated Data Analysis

To characterize the spatial and temporal relationship of population and built-up land across scales, we analyzed many descriptive statistics, including histograms, scatterplots, Q-Q plots, trend lines, thematic maps, correlation coefficients, and semivariograms. A semivariogram is a graph showing how dissimilarity between observations of a geospatial variable changes as the distance between the observations changes. Semivariograms measure spatial heterogeneity-autocorrelation patterns of geographical phenomena that can be conceptualized as continuous fields, for example, whether spatial clustering is present, how large an average cluster may be, etc. Based on these results, we infer best-practice suggestions for socio-environmental analysis and large-scale modeling of urbanization.

Figure 1. World region definitions in this research.
To compare the trends and patterns we identify from global-coverage spatial data with those reported by meta-analyses of existing literature, we explicitly examined the four aspects highlighted in the Introduction: (a) We compared aggregate total amounts of built-up land expansion over recent past decades among various geographic regions and countries, across the developing and the developed world. (b) We compared change rates of built-up land expansion and population growth at multiple scales (gridded, national, regional, and global) across the world. (c) To understand the difference between spatially-explicit “urban” definitions based on population and built-up land characteristics, we identified two sets of spatial urban definitions for global countries. For every nation, we searched for a threshold of gridded built-up land fraction, so that the sum of populations from grid cells with a built-up land fraction greater than the threshold would equal to the country’s total urban population according to WUP. For every nation, we also searched for a threshold of gridded population density that would make the sum of populations in grids with higher population densities match the WUP national urban population total. Different thresholds were found for different countries. We then mapped and compared the national built-up land fraction thresholds, the national population density thresholds, and the urban grids defined using the two sets of thresholds, to examine whether population and built-up land can be used as proxies for each other in urban definitions and analyses. (d) To investigate the distribution of the global population across different land development levels, we derived from our gridded data the amount of population living in every decile between 0-1 built-up land fraction at national and global scales.

Together the analyses (with repeated runs on alternative datasets, spatial resolutions, and parameter settings) resulted in more than 60 GB output in formats of spatial and tabular data. It is ineffective (if not impractical) to present all results. While all our conclusions are supported by multiple pieces of results, below for each conclusion, we present the most direct and the easiest-to-interpret supporting results for clarity and conciseness.

4. Results and Discussion

4.1. Comparing Spatial Patterns of Population and Built-Up Land

Globally, the correlation between gridded population density and built-up land fraction is moderately strong (roughly 0.6; Table 2). That is, about 40% of global spatial variations in population or built-up land cannot be statistically accounted for by the other variable. This relationship varies somewhat across regions (ranging from 0.6 to 0.8; Table 3), with the developed world (e.g., Europe, North America) showing higher correlation than the developing world (e.g., Asia, Africa), suggesting that the population-built-up-land relationship is more heterogeneous in the developing world. This contrast between the developed and the developing regions roots in the known challenge of producing global built-up land maps using remote sensing data. As discussed in Section 2.1, constructions in the developing world may use different materials from those common in the developed world, and cannot be mapped as well by algorithms primarily based on information from the developed world (e.g., Linard et al., 2013; Pesaresi et al., 2015). Our quantitative analysis and the known data quality pattern both indicate that urban theories about the developed world might not apply to the developing world as they become more developed in the future.

Different regions also exhibit different spatial patterns for the two variables: (i) Developing regions show smaller clusters of similar built-up land fraction values (i.e., more fragmented land development patterns) on average than developed regions (Figure 2: see that the trend lines for the Asian regions plateau at a shorter distance than those for North America and Europe). (ii) The two Asian regions show similar built-up

| Geographic coordinate system (latitude/longitude) | Equal-area map projection (Mollweide projection) |
|-----------------------------------------------|-----------------------------------------------|
| 30°   | 1/8 degree | 1/2 degree | 1 degree | 2 degree | 1 km (compare w/30°) | 14 km (compare w/1/8 degree) |
| 1990  | 0.444    | 0.622     | 0.664    | 0.669    | 0.642             | 0.442     | 0.629     |
| 2000  | 0.436    | 0.611     | 0.647    | 0.659    | 0.629             | 0.433     | 0.612     |
land patterns but quite different population patterns (Figure 2: compare the trend line shapes), with South and Southeast Asia showing more spatial contrast (i.e., bigger urban-rural differences) in population density than East and Central Asia (Figure 2: see the higher plateau). (iii) Europe and North America show different development styles: compact (Europe) versus expansive (North America), demonstrated by Europe’s higher spatial contrast (i.e., bigger urban-rural differences) in both population and built-up land than North America (Figure 2: see the higher plateaus).

Moreover, globally, the highest values in population density versus built-up land fraction do not even cluster in the same geographic regions: the highest population density values are mostly in South Asia and East Asia (Figure 3a), while for built-up land fraction, Europe, parts of North America and East Asia stand out (Figure 3b).

These demonstrate that population and built-up land show distinctly different spatial patterns across the world. The common practice in large-scale earth system modeling, treating gridded population and built-up land synonymously when delineating urban areas, can mislead conclusions about spatial trends across scales.

Our results are stable (generally within ±0.1 range), regardless of whether the grid cells with zero values were included, whether a geographic coordinate system or an equal-area map projection was used (Table 2), and whether GPW or GRUMP was used for spatial population (Table 3). As discussed in Section 2.2, it is reasonable to hypothesize that the correlation between population and built-up land reported using GPW v.3 is lower than the unattainable “true” value, and the correlation reported using GRUMP is higher. Our results (Table 3) validate this hypothesis – the GRUMP-based correlation coefficients are always higher than the GPW-based coefficients. Further, since the two versions of the correlation coefficients are similar, they

| Grid-Cell (1/8 Degree) Correlation by Regions: Population Density Versus Built-Up Land Fraction in 2000 |
|---|---|---|---|
|  | GRUMP versus GHSL | GPW versus GHSL |
| North America | 0.798 | 0.778 |
| South & Central America | 0.754 | 0.733 |
| Europe | 0.801 | 0.780 |
| Russia | 0.731 | 0.700 |
| North Africa & Middle East | 0.633 | 0.521 |
| Africa | 0.661 | 0.593 |
| East & Central Asia | 0.687 | 0.608 |
| South & Southeast Asia | 0.612 | 0.497 |
| Oceania | 0.769 | 0.758 |

Figure 2. Semivariograms of population density and built-up land fraction in 2000 for selected world regions. (A semivariogram describes how the dissimilarity between the values of a given geographic attribute at two locations changes, as the distance between the two locations increases. The horizontal axis is the distance between two locations. The vertical axis is the dissimilarity between data values separated by that distance, measured as the squared difference between the two data values. Though degrees are not ideal units for measuring distance, re-projecting raster data introduces location errors of its own. Here both variables under investigation are standardized, and the range examined is only up to 1.8 degree; hence, the concern that grid cell sizes scale with latitude in degree-based raster data is low. The units of the dissimilarity measure are squares of the underlying variables, that is, \((\text{people/sq km})^2\) for population density and \((\text{fraction})^2\) for built-up land fraction. Each blue square point is the average of all pairs of locations (including all grid cells within a geographic region) separated by a distance falling in a set distance range. The horizontal axis is binned into regularly spaced distance ranges. The trend line is the least-squares fit of the commonly-used Stable Model to the blue square points.)
together roughly indicate the “true” level of the correlation between population and built-up land, and give confidence to our conclusions.

Our sensitivity analyses also revealed that, the correlation coefficient between population and built-up land increases as the spatial resolution changes from 30” to 1/8 degree, and stays stable as the spatial resolution further coarsens (Table 2), which suggests that excessive, stochastic spatial variations exist at the 30” resolution and weaken the strength of statistically identifiable association between population and built-up land. The noisy signals can be smoothed out by aggregating to the 1/8 degree resolution, while further aggregation does not offer additional benefits. Hence, 1/8 degree is a good balance between mapping as fine-grained spatial information as possible and smoothing out excessive, noisy details.

4.2. Comparing Temporal Trends of Population and Built-Up Land

Population and built-up land change differently over time: Generally, in observational data, population can increase or decrease, while built-up land either expands or stays the same (Figure 4). This relates to the fact that developed land is often not reverted to undeveloped even after being repurposed (e.g., Pesaresi et al., 2015).

Contrary to the common impression that the developing world has been the global hotspot of new urbanization (e.g., Angel et al., 2016; Schneider et al., 2015), we found that, over recent past decades, the amounts of new built-up land development are comparable in developed versus developing world (e.g., comparing changes in North America and Africa for the same time periods in Figures 5 and 6). The fact that, in the developed world, both regional total and per capita amounts of built-up land keep increasing (Figure 5), indicates that these already highly-urbanized, demographically-stabilized countries have not stopped urban expansion and are becoming more urbanized. This finding is robust from limitations of gridded population data (as discussed in Section 2), since national population observations were used for describing regional patterns.

While the per capita built-up land amount appears to be a good indicator of economic development level (Figure 5), its wide range of values among developed regions suggest a potentially wide range of uncertainties about the planet’s urban future. How much new built-up land may be constructed in years to come will depend not only on what urbanization trajectories the currently-developing regions follow, but also how much urban area the already-developed regions will add. Impacts of urbanization (e.g., those

Figure 3. Global maps of (a) population density (ranging 0–51,217 people/sq km) and (b) built-up land fraction (ranging 0–1) in 2000. (Colors show relative highs and lows within each dataset, and cannot be compared between datasets.) (Enlarged views of selected regions are in Figure S1 in Supporting information.).
on agricultural land, biodiversity, economy, and consumption behavior) will continue to transform developed-world societies and are likely of no less magnitude than those in the developing world.

Meanwhile, when using relative measures (i.e., change rates) instead of absolute measures (i.e., amounts of changes), a different story emerges, where developing regions show faster built-up land change rates than developed regions (Figure 6). The similar absolute amounts of change corresponded to noticeably different change rates, because the base amount of existing built-up land in the developing world is much less than the developed world. These results demonstrate that different measures emphasize different aspects of urbanization and can considerably affect our perception of urbanization. We therefore recommend examining both relative and absolute measures in socio-environmental analysis and modeling of urbanization for better understanding the dynamics and the impacts of urbanization.

![Figure 4.](image)

**Figure 4.** Global maps of changes (1990–2000) in (a) population density (GRUMP-based) and (b) built-up land fraction (GHSL-based). (Colors show relative highs and lows within each dataset.) (Enlarged views of selected regions are in Figure S2 in Supporting information.).

![Figure 5.](image)

**Figure 5.** Changes in total (a) and per capita (b) amounts of built-up land by regions. (Underlying data are in Table S2 in Supporting information.).
The examination of change rates also showed that, for most world regions, built-up land expands faster than population growth, especially for Europe and Russia. This pattern relates to the process of urbanization where urban population grows faster than total population and demands more new built-up land development. However, at the global scale, the change rates of the two variables are similar (last set of bars in Figure 6c), substantially different from what meta-analyses of individual city studies have reported (i.e., built-up land expands twice as fast as population grows suggesting a strong increase in urban sprawl) (Güneralp et al., 2020; Seto et al., 2011). This suggests that the global ratio of land consumption rate to population growth rate (i.e., the SDG Indicator 11.3.1 measuring global communities’ progress toward sustainable human settlement planning and management, UN, 2017) may be substantially different from previously thought, and requires re-evaluation of potential impacts of global urbanization – positive (e.g., driving economic value creation) and negative (e.g., causing biodiversity loss), especially now that we know significant amount of urban expansion continues in the developed world. Meanwhile, the result confirms...
Table 4
Grid-Cell Correlation by Regions: Change in Population Density Versus Change in Built-Up Land Fraction (1990–2000)

| Region                | Change in pop density versus change in built-up land fraction |
|-----------------------|-------------------------------------------------------------|
| North America         | 0.494                                                       |
| South & Central America | 0.486                                                      |
| Europe                | 0.224                                                       |
| Russia                | 0.075                                                       |
| North Africa & Middle East | 0.484                                                      |
| Africa                | 0.429                                                       |
| East & Central Asia   | 0.363                                                       |
| South & Southeast Asia | 0.279                                                      |
| Oceania               | 0.357                                                       |

limitations of using sample urban areas for studying broad global urbanization trends and underscores the necessity of using global-coverage data to avoid selection bias.

At the gridded scale, the global correlation between changes in built-up land fraction and population density is 0.314. The correlation coefficient varies across regions, but is below 0.5 for all world regions (Table 4). Especially, in Russia, built-up land expanded despite the fact that population showed little change in most areas and declined overall; a similar trend also occurred in Europe (Table 4 and Figure 6). These results confirm the known notion that urbanization is a local process and varies over space and time, suggesting that simple predictive models assuming constant relationships between population and built-up land (e.g., a constant per capita amount of built-up land) across large land areas is problematic. Instead, global predictive urbanization models should account for observed spatial and temporal variations in the relationship while covering the whole world (e.g., Gao & O'Neill, 2019, 2020).

4.3. Comparing Spatial Urban Definitions Based on Population Versus Built-Up Land

The two “urban” grid maps, generated by thresholding population density and built-up land fraction while matching to national socioeconomic statistics, both show where cities (loosely defined) are (Figure S3); however, the two definitions can lead to very different numerical results for estimating the amount of “urban” land or people. For example, out of the 997022 global land grid cells, 3.4% were labeled “urban” by both definitions, an additional 3.9% by only the built-up-land-fraction thresholds, and another 1.2% by only the population-density thresholds. More differences between the two urban grid maps were seen for large countries (e.g., India), suggesting that sub-national spatial variations must be accounted for in large-scale socio-environmental analysis and modeling linking population and built-up land.

A somewhat unexpected finding of this analysis is how small-valued the built-up land fraction thresholds had to be, in order to match the census-based national urban population totals (Table S1): For most countries across the world (203 out of 227 countries), the thresholds are less than 0.05 (Figure 7b). Repeating the analysis on 30° instead of 1/8-degree grids led to similar results (in fact, for 104 countries, the 30° thresholds are lower than the 1/8-degree thresholds). Although existing literature has well established that not all urban areas are densely-developed (e.g., Angel et al., 2011; Hansen et al., 2005; Uhl et al., 2020) and has typically been using 0.2 as the built-up land fraction threshold for defining urban areas (e.g., Mahtta et al., 2019), our results further show that most global populations (including what national statistics agencies call urban populations) reside in areas with low land development level that are frequently less than 5% built-up (i.e., built-up land fraction less than 0.05). The population distributions across various land development levels for different countries (Figure 8) also verify this general pattern. It is especially pronounced for developing countries, while consistent across the world: Globally, in 2000, 18% of the global population resided in areas with more than 20% built-up land, while national statistics agencies/WUP labeled 47% of the global population urban.

If we follow the convention to define “urban” as grid cells with built-up land fractions ≥0.2, the global urban population in 2000 is 1.1 billion using GRUMP population at 1/8 degree, 0.9 billion using GPW population, and 1.3 billion using 30° grids. Even when a 50% margin of error is allowed (considering limitations of current gridded population data), the built-up-land-based urban population estimate is ≤1.9 billion, quite different from WUP’s estimate using country-specific census-based “urban” definitions of 2.9 billion. Clearly, “urban” definitions based on population versus built-up land characteristics cannot be used as proxies for each other. For example, whether 1.1 or 2.9 billion people are susceptible to substantial urban heat island effects (associated with dense land development) could make a determining difference for global communities’ adaption to the warming climate. Hence, choosing an appropriate urban definition requires aligning the definition’s emphasis with the analysis’ ultimate purpose: existing national statistics derived “urban”
definitions focus on demographic and economic characteristics, and remote sensing based built-up land maps can more meaningfully map the physical urban environment. This result also indicates that the common practice in large-scale earth system modeling of assuming demographically-defined urban population

Figure 7. National thresholds defining urban grids in 2000 so that sums of populations in urban grids match WUP national urban population totals: (a) population-density (people/sq km) thresholds; (b) built-up-fraction thresholds.

Figure 8. Total population counts in grid cells with different built-up land fraction levels: (a) U.S.A., (b) U.K., (c) China, (d) India, (e) global.
always resides in areas with medium to high built-up land development levels must change, and that the large-scale trends of urbanization’s potential impacts may need to be re-evaluated.

5. Conclusions

In the context of urbanization, we systematically analyzed the relationship between spatial and temporal patterns of population and built-up land at gridded, national, regional, and global scales. Below we organize our conclusions as a set of key findings about contemporary urbanization, and a set of suggestions for urban-related socio-environmental analyses and modeling.

For urbanization across the world, we found the following spatiotemporal patterns:

1. Population density and built-up land fraction show distinctly different spatial and temporal trends (with a global correlation coefficient around 0.6).
2. Different from common impressions, we found the developed and the developing regions experienced comparable amounts of built-up land expansion over recent decades, and both regions can anticipate a wide range of uncertainties about their urban futures.
3. Globally, built-up land and population on average change at similar rates. This finding is substantially different from what meta-analyses of previous literature have reported (that built-up land expands twice as fast as population grows globally on average).
4. Also somewhat surprising, we found that most people across the world (including what national statistics agencies call urban populations) reside in areas with low land development levels that are frequently less than 5% built up. Although existing literature has established that not all urban areas are densely-developed landscapes filled with skyscrapers, 5% is well below the commonly used built-up land threshold for defining urban (20%).

For large-scale socio-environmental analyses and modeling of urbanization, we recommend the following best-practice suggestions:

1. The common practice in large-scale earth system modeling of assuming demographically-defined urban populations always reside in areas with medium to high built-up land development levels is problematic and can lead to large numerical differences in analysis and modeling results.
   If defining “urban” can be avoided, it is advisable to do so, considering the ambiguity among different “urban” definitions. If unavoidable, we recommend carefully choosing a definition that best aligns with the ultimate purpose of the analysis. Generally, existing “urban” definitions derived from national statistics are based more on demographic than landscape characteristics, while built-up land maps based on remote sensing can meaningfully depict the physical urban environment.
   This updated perspective about how “urban” should be defined in urban-related socio-environmental analyses, especially those involving large-scale earth system modeling, suggests that the general trends and patterns of how urbanization interacts with global environmental change need to be re-evaluated.
2. For urbanization modeling, our results suggest that how urbanization is measured (e.g., change amounts vs. change rates) may substantially affect the performance of predictive models. We therefore recommend experimenting with different measures (including both relative and absolute metrics) to identify robust statistical relationships as the basis for model building.
3. We found 1/8 degree an ideal spatial resolution for characterizing the relationship between population and built-up land across the world, which balances the need to show fine-granularity spatial information while smoothing out excessive, noisy details. The geographic coordinate based resolution is able to robustly maintain the spatial precision of finer-resolution underlying data, when used for fraction and density variables (e.g., built-up land fraction, population density).
4. We recommend for large-scale, long-term models to account for the spatial and temporal variations in the urbanization process, for example, modeling large countries with multiple subnational models.
5. Our findings highlight the value of using global-coverage spatiotemporal data for studying large-scale urbanization trends, and demonstrate the limitations of using sample urban areas for these studies.
Conflict of Interest
The authors declare no conflicts of interest relevant to this study.

Data Availability Statement
Datasets for this research (Table 1) are publicly downloadable at their respective referenced citations: United Nations, Department of Economic and Social Affairs, Population Division 2018, CIESIN 2011, CIESIN 2005, and Pesaresi et al., 2015.

References
AAAS (American Association for the Advancement of Science). (2016). Rise of the city. Science, 352(6288), 906–907. https://doi.org/10.1126/science.352.6288.906
Angel, S., Blasi, A. M., Parenti, J., Lamson-Hall, P., Sánchez, N. G., Civco, D. L., et al. (2016). Atlas of urban expansion—2016 edition volume I: Areas and densities. (pp. 50–1227). Choice Reviews Online. https://doi.org/10.5860/CHOICE
Angel, S., Parenti, J., Civco, D. L., Blasi, A., & Potere, D. (2011). The dimensions of global urban expansion: Estimates and projections for all cities of 100,000 or more: 2000–2050. https://doi.org/10.1073/pnas.1104009108
Bai, Z., Wang, J., Wang, M., Gao, M., & Sun, J. (2018). Accuracy assessment of multi-source gridded population distribution datasets in China. Sustainability, 10(5), 1363. https://doi.org/10.3390/su10051363
Balk, D. L., & Montgomery, M. R. (2015). Guest editorial: "Spatializing demography for the urban future". Spatial Demography, 3(2), 59–62. https://doi.org/10.1007/s40980-015-0017-x
Batty, M. (2008). The size, scale, and shape of cities. Science, 319(5864), 769–771. https://doi.org/10.1126/science.1151419
Bettencourt, L. M. (2013). The origins of scaling in cities. Science, 340(6139), 1438–1441. https://doi.org/10.1126/science.1235823
Bhaduri, B., Bright, E., Coleman, P., & Urban, M. (2007). LandScan USA: A high resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal, 69, 103–117. https://doi.org/10.1007/s10708-007-9105-9
Bierwagen, B. G., Theobald, D. M., Pyke, C. R., Choate, A., Groth, P., Thomas, J. V., & Morefield, P. (2010). National housing and impervious surface scenarios for integrated climate impact assessments. Proceedings of the National Academy of Sciences, 107(49), 20887–20892. https://doi.org/10.1073/pnas.1002961107
Buetner, T. (2015). Urban estimates and projections at the United Nations: The strengths, weaknesses, and underpinnings of the world urbanization prospects. Spatial Demography, 3(2), 91–108. https://doi.org/10.1007/s40980-015-0004-2
Bustos, M. F. A., Hall, O., Niedomysl, T., & Ernstson, U. (2020). A pixel level evaluation of five multi-temporal global gridded population datasets: A case study in Sweden, 1990–2015. (pp. 1–23). Population and Environment. CIESIN (2011). Global Rural-Urban Mapping Project (GRUMP), version 1. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from http://sedac.ciesin.columbia.edu/data/collection/grump-v1
CIESIN (2016). GPW, version 4. NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from http://sedac.ciesin.columbia.edu/data/collection/gpw-v3
CIESIN (Center for International Earth Science Information Network, Columbia University). (2005). Gridded Population of the World (GPW), version 3. NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from http://sedac.ciesin.columbia.edu/data/collection/gpw-v3
Cohen, B. (2004). Urban growth in developing countries: A review of current trends and a caution regarding existing forecasts. World Development, 32(1), 23–31. https://doi.org/10.1016/j.worlddev.2003.04.008
European Union. (2016). The State of European Cities 2016: Cities leading the way to a better future. Publications Office of the European Union. https://doi.org/10.2776/770065
Frey, W. H., & Zimmer, Z. (2001). Defining the city. In R. Paddison (Ed.), Handbook of urban studies, (pp. 14–35). SAGE Publications.
Frolking, S., Milliman, T., Seto, K. C., & Friedl, M. A. (2013). A global fingerprint of macro-scale changes in urban structure from 1999 to 2009. Environmental Research Letters, 8(2), 024004. https://doi.org/10.1088/1748-9326/8/2/024004
Gao, J., & O’Neill, B. C. (2019). Data-driven spatial modeling of global long-term urban land development: The SELECT model. Environmental Modelling & Software, 119, 458–471. https://doi.org/10.1016/j.envsoft.2019.06.015
Gao, J., & O’Neill, B. C. (2020). Mapping global urban land for the 21st century with data-driven simulations and shared socioeconomic pathways. Nature Communications, 11(1), 1–12. https://doi.org/10.1038/s41467-020-15788-7
Geist, H., Mcconnell, W., Lambin, E. F., Moran, E., Alves, D., & Rudel, T. (2006). Causes and trajectories of land-use/cover change. In E. F. Lambin, & H. Geist (Eds.), Land-use and land-cover change: Local processes and global impacts, (pp. 41–70). Springer.
Goldewijk, K. K., & Verburg, P. H. (2011). Uncertainties in global-scale reconstructions of historical land use: An illustration using the HYDE data set. Landscape Ecology, 26(5), 861–877. https://doi.org/10.1007/s10980-013-9877-x
Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. Science, 319(5864), 756–760. https://doi.org/10.1126/science.1150195
Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A., & Seto, K. C. (2020). Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis. Environmental Research Letters, 15(4), 044015. https://doi.org/10.1088/1748-9326/ab6669
Hansen, A. J., Knight, R. L., Marzluff, J. M., Powell, S., Brown, K., Gude, P. H., & Jones, K. (2005). Effects of exurban development on biodiversity: Patterns, mechanisms, and research needs. Ecological Applications, 15(6), 1893–1905. https://doi.org/10.1890/05-5221
Hillbärd, K., Jannetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Beets, B., et al. (2010). Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling. International Journal of Climatology, 30(13), 2118–2128. https://doi.org/10.1002/joc.2150
Jones, B., Tebaldi, C., O’Neill, B. C., Oleson, K., & Gao, J. (2018). Avoiding population exposure to heat-related extremes: Demographic change vs climate change. Climatic Change, 146(3–4), 423–437. https://doi.org/10.1007/s10584-017-2133-7
Leyk, S., Gaughan, A. E., Adano, S. B., Sherburnin, A. D., Balk, D., Freire, S., et al. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. Earth System Science Data, 11(3), 1385–1409. https://doi.org/10.5194/essd-11-1385-2019
