From Pixels to Planning: Large-scale Mapping of Urban Morphology and Population Distribution with the World Settlement Footprint 3D

Daniela Palacios-Lopez, Elisabeth Brzoska, Mattia Marconcini, Julian Zeidler, Thomas Esch

(Dr. Daniela Palacios-Lopez, German Aerospace Center, Weßling, Germany, daniela.palacioslopez@dlr.de)
(Elisabeth Brzoska, German Aerospace Center, Weßling, Germany, elisabeth.brzoska@dlr.de)
(Dr. Mattia Marconcini, German Aerospace Center, Weßling, Germany, mattia.marconcini@dlr.de)
(Julian Zeidler, German Aerospace Center, Weßling, Germany, julian.zeidler@dlr.de)
(Prof. Dr. Thomas Esch, German Aerospace Center, Weßling, Germany, thomas.esch@dlr.de)

1 ABSTRACT

Urban morphology and human population distribution are two interrelated aspects of our urbanization that play a critical role in shaping the sustainability, resilience and liveability of cities. In recent years, the advent of global datasets with 3D information derived from Earth Observation (EO) technologies has revolutionised our ability to study and analyse these two aspects of urbanisation, providing information that is essential for designing cities that can accommodate the needs of their residents while minimizing their environmental impact.

One such dataset is the novel World Settlement Footprint 3D (WSF3D) produced by the German Aerospace Center (DLR). The WSF3D was the first global dataset providing detailed information of the fraction, area, average height and total volume of buildings, at unprecedented spatial resolution, coverage and consistency. Since its development, researchers from different organizations (e.g. WorldBank, United Nations, WorldPop) have employed the dataset as input data for large-scale studies in urban morphology and population distribution, with a level of detail that was previously impossible.

In this paper we present a selection of WSF3D-driven applications with the objective of demonstrating how the new data can be used to support urban planning and management. First, the WSF3D has been employed to demonstrate how the four layers of the dataset can be used to determine a building's functional use, and how this information can be leveraged to improve large-scale models of population distribution at large-scale. Thereafter, the WSF3D has been used to determine the relationships among building height/volume, population density and income, which can provide insights into the efficient use of space (e.g. crowding vs layering) on the one hand, and shed light into infrastructure disparities and variations, on the other. With that being said, due to the global nature of the WSF3D dataset, the previous analyses were conducted from local to regional scales, which can also help identify opportunities for interventions that can be replicated across different locations.

Overall, with the results of this research, the authors aim to provide planners and policy-makers with valuable insights into usability of the globally available WSF3D dataset. By demonstrating its potential as reliable and robust input data, this study seeks not only to empower evidence-based decision-making, but also to advocate for the widespread adoption of geospatial layers in the implementation of strategies towards sustainable development strategies of the built environment.

2 INTRODUCTION

Understanding of how cities are structured and how populations are distributed is crucial for the development of sustainable, resilient and liveable cities. On the one hand, from a socio-economic perspective, being able to identify areas of high population density and areas with high building density (2D and 3D), can promote efficient planning and utilization of resources like water, energy, transportation and waste management (Chokhachian et al., 2020; Wang et al., 2021b). This allows creating more vibrant cities, which in return, attract business, tourists and investments, generating an economic boost that enables social equality and inclusivity. On the other hand, from an environmental point of view, knowledge on these two factors aids in designing cities that minimize urban sprawl and population crowding, promoting compact, secure and eco-friendly urban centres (Lall et al., 2021; Wang et al., 2021a). For example, with up-to-date and precise information, governments and planners can conduct better hazard and environmental assessments that aim at reducing pollution levels and habitat fragmentation and loss, as well as promoting cities that are more resilient to climate-related impacts like sea-level rise, flooding or extreme heat (Cai et al., 2021; Carpio et al., 2021; Mills et al., 2021; Scheba et al., 2021).

In this framework, with the advancement of Earth Observation (EO) technologies (e.g. remote sensing) and artificial intelligence (e.g. machine and deep learning), the development of geospatial datasets has revolutionised our ability to study and analyse urban morphology and population distribution (Chen et al., 2023) . Global-scale datasets, in particular, allow for cross-regional comparisons that enable researchers and planners to understand similarities and differences among cities and countries, helping identify best practices and solutions that can be applied in different contexts (Wang et al., 2023).

In this paper, we present the novel World Settlement Footprint 3D dataset (WSF3D) produced by the German Aerospace Center (DLR) (Esch et al., 2022), and showcase a series of insightful applications that can be performed with the data. Our research objectives are twofold: Firstly, we emphasize the dataset's potential for enhancing large-scale population modelling by leveraging volume and building use information exclusively from the WSF3D dataset. To achieve this, we explore the capabilities of a machine learning algorithm to categorize buildings into residential and non-residential classes, using the fourWSF3D layers: building area, height, fraction, and volume. Additionally, we provide accuracy results obtained from other population modelling methods, demonstrating the significant improvements attained through the utilization of the new WSF3D. Secondly, we conduct analyses across diverse spatial domains to investigate how the WSF3D dataset can be employed to compare and contrast different urban patterns, such as crowding versus layering across cities, countries, and regions. Moreover, we examine how infrastructure disparities and variations relate to income data. The primary focus here is to demonstrate the capability of combining the WSF3D with other information sources to facilitate the derivation of crucial statistical analyses that can be replicated across various locations.

3 APPLICATIONS BASED ON THE WORLD SETTLEMENT FOOTPRINT 3D

The WSF3D was the first global dataset providing detailed information of the fraction, area, average height and total volume of buildings. The dataset has unprecedented spatial resolution (~90m at the Equator) and is open-and-free for download at https://geoservice.dlr.de. In regards to its processing framework, the WSF3D was generated using a combination of a modified version of the ~10m WSF binary settlement mask derived from S1-S2 data (Marconcini et al., 2021) and ~12m digital elevation data and radar imagery collected by the TanDEM-X mission (Zink et al., 2014). Its production relied on three automatic workflows: one to estimate the building fraction and areas, a second to estimate the mean building height, and a third to estimate the total built-up volume. For a more detailed description of these three modules, please refer to Esch et al. (2022). Figure 1 illustrates a subset of the four main layers that compose the WSF3D.

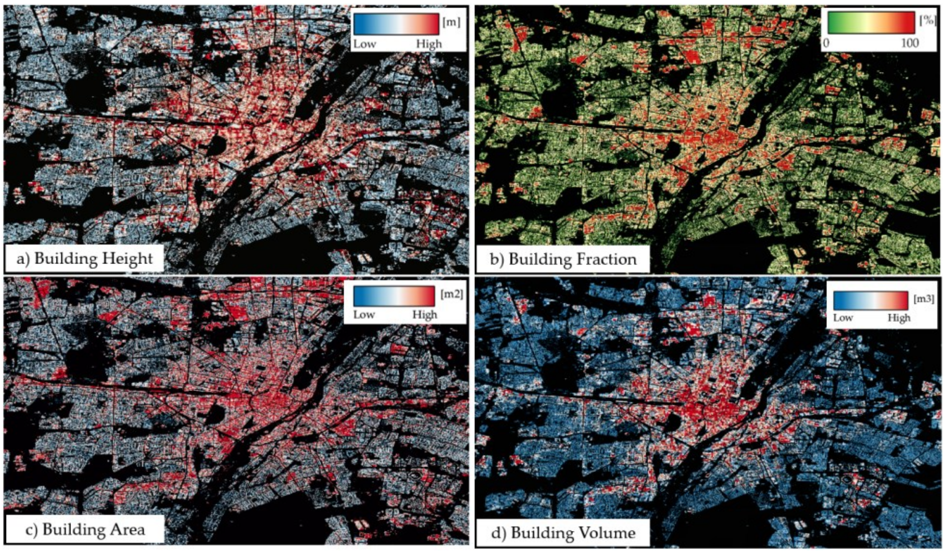

Fig. 1: WSF3D for the city of Munich, Germany. **a)** Building height [m], **b)** Building Fraction [%], c) Building area [m^2] and d) Building Volume [m^3]. Image reference: Palacios-Lopez et al. (2022)

3.1 Large-scale gridded population modelling

Gridded population datasets model human population distributions using a dasymetric modelling approach, in which census counts are disaggregated from administrative units into a reference grid, using different geospatial layers as determinants of allocation. While gridded population datasets have been produced for more than 20 years, recently we have witnessed a significant breakthrough in the field, as current products have begun incorporating 3D and functional use information into their models, thus reducing underestimation errors in rural regions and overestimations in urban areas

In this section we show how the WSF3D can be used to improve the qualitative and quantitative accuracy of residential population estimations without relying on any other data source. The analyses presented here are based on the research of Palacios-Lopez et al. (2022), where the authors apply the same methodology at a Pan-European scale.

First, using a Random Forest (RF) classifier and training data collected from the Urban Atlas 2018 , we show how spatial metrics derived from the WSF3D dataset are sufficient to identify large industrial/commercial areas in the built-up environment. In particular, the use of the mean, median and stdev. of the four WSF3D-layers, are enough to accurately recognise these areas. Using the city of Ljubljana in Slovenia as an example, Figure 2 shows outcome of this classification. At the country level, the classification achieves an overall accuracy (OA): 86.31% and a Kappa coefficient (K) of 0.7. According to the results presented in Palacios-Lopez et al. (2022), at a Pan-European scale the average OA and K can achieve average values of 84.43% and 0.68, respectively.

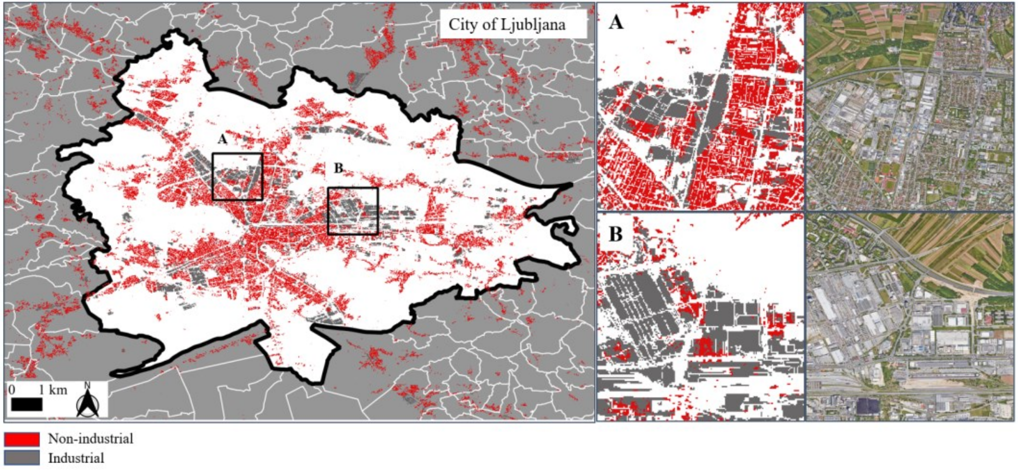

Fig. 2: Industrial/Non-industrial classification based on spatial metrics derived from the WSF3D and RF algorithm.

Using this binary classification, as a second step we evaluate how the accuracy of gridded population maps improves by incorporation building volume and building use information into the modelling framework. We compare our results with methods that are based on general 2D proxies and no building use information, which represent methods that are still largely employed (Leyk et al., 2019). For the city of Ljubljana (NW-area), Figure 4 shows the outcome of population distribution maps produced on the basis of a binary settlement layer (BM), a fraction layer (BF), a volume layer (BV) and volume + use layer (BV-IS).

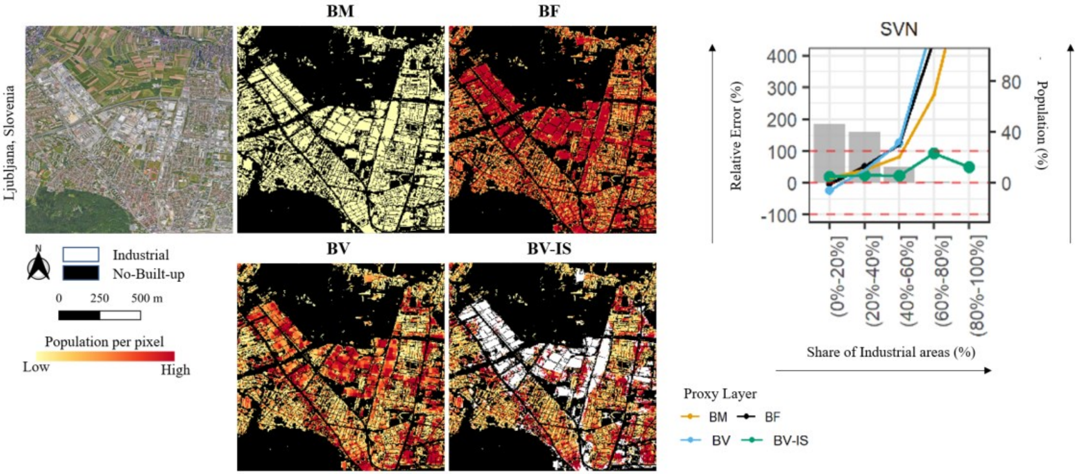

Fig. 3:

As observed, gridded population maps based on the BM proxy layer show homogeneous distributions, with each pixel having the same population. In contrast, maps using BF, BV, and BV-IS proxies exhibit more spatial heterogeneity, following changes in density and volume values. Without settlement use information, BF and BV maps allocate a large population proportion to industrial areas. The BV proxy, however, minimizes this effect, allocating more population to dense non-industrial areas compared to BF, where populations in both dense non-industrial and industrial areas seem balanced. By comparing the relative error produced by the layer, it is clear that the BV-IS delivers the best results. In areas where a large share of industrial areas are found, overestimations are minimized by more than 500%. Comparably, in areas where the share of industrial areas is less (e.g. residential areas), the BV-IS also produced errors closer to 10%.

3.2 Analysis of urban morphology

In the following two subsections we present a set of applications that demonstrate how the WSF3D can be integrated with different data sources, to understand the correlations that might exist between urban morphology and socio-economic development. Analyses are carried-out in two spatial domains:

1. at the country level using the GADM v4.1 polygons and socio-economic data downloaded from the WB development indicators catalogue.
2. at the urban cluster level (globally) using the GHS Urban Clusters, which provide socio-economic information at the level or urban areas,

For each spatial domain the following variables have been collected/used: average building height and total built-up area from the WSF3D, and population, income group, and GDP.

Analysing “Crowding vs Layering”

Figure 4 provides a graphical representation of each country’s position concerning its population (x-axis), average height (y-axis), and built-up area (bubble size), coloured according to their income group. The chart categorizes countries based on their tendencies towards “layering” (top-left) or “crowding” (lower-right) based on their urban development patterns.

Analysing the chart, we observe that high-income (HIC) countries predominantly cluster towards the top-left quadrant, indicating their inclination to construct buildings with greater height and extensive built-up areas compared to lower-income (LIC), lower-middle income (LMIC), and upper-middle-income (UMIC) countries. The behaviour of HIC, coupled with their relatively smaller populations, suggest that they offer more space to their citizens (including industrial and commercial buildings), contributing to a lower population density and -presumably - less crowding.

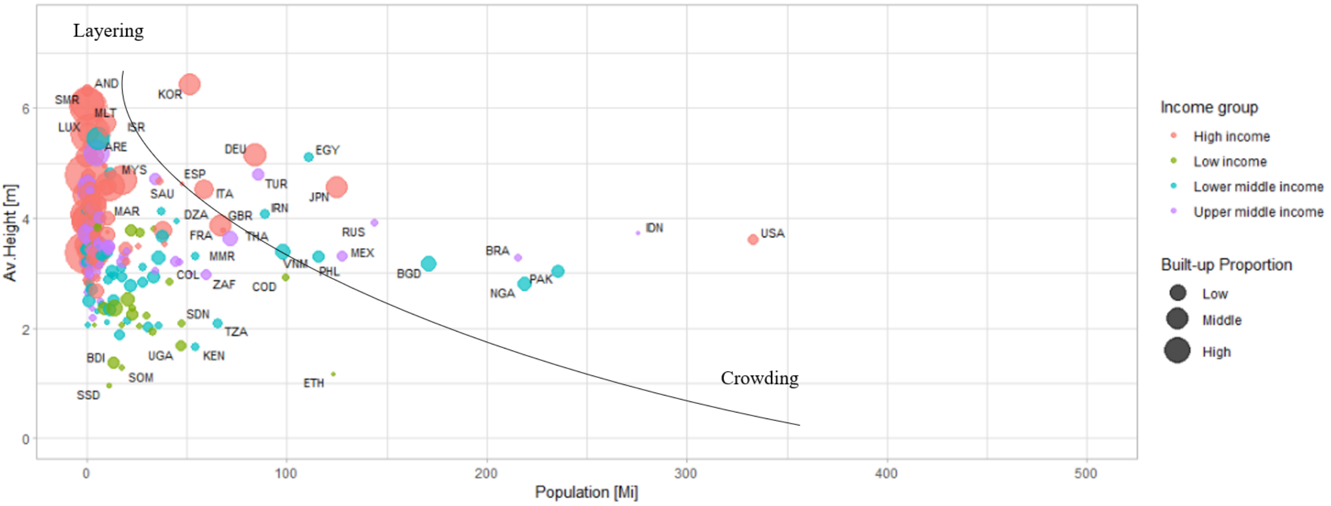

Fig. 4: Country-scale representation of “layering” and “crowding” trends. Total population and income group represent the year 2022.

Comparably, the same data can also be displayed as shown in Figure 5. Here for example, it can be seen that countries that are located in different income groups, but that have similar populations and similar GDP, do not necessarily offer the same space to people. These differences can also be appreciated for countries within the same income group, where the amount of GDP does not necessarily play a role in built-up space. While somehow simple, information like this can help, for example, to shed light into how building policies and financial resources are being implemented to define building space, as well, as well as to understand how possible geographical restrictions including, being an island or having coastal areas, might influence infrastructure development.

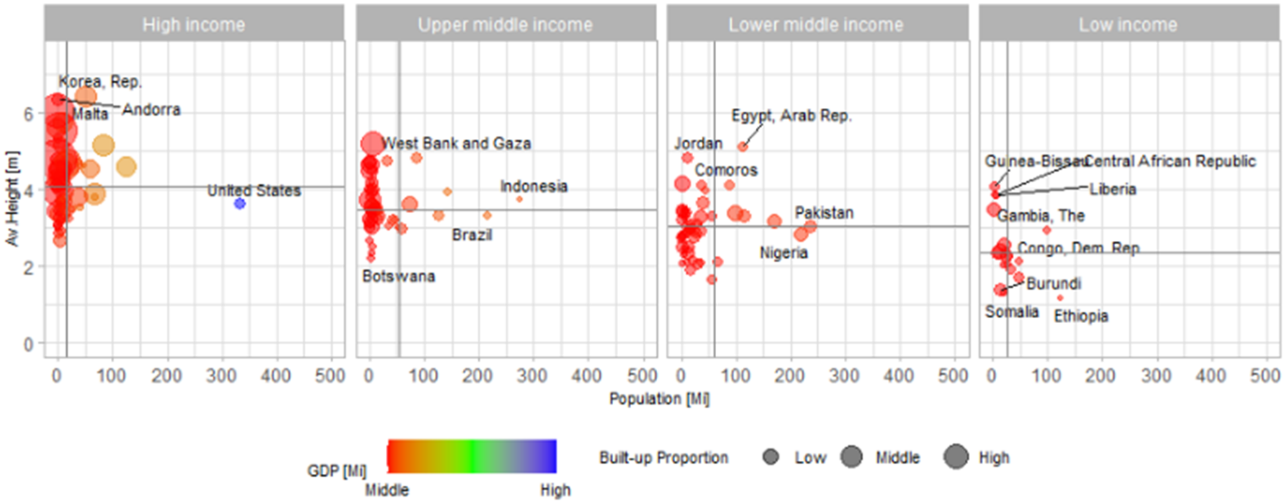

Fig. 5: Country-scale distribution of av. height, built-up area and population with respect GDP and income group.

Analysing "Infrastructure disparities"

While the continental-scale analyses presented in the previous section can be found already useful to define some correlations between morphology and economic development, the use of the WSF3D can also help to analyse infrastructure disparities at the city level. For a number of selected countries, Figure 6 shows the position of different urban clusters in relation to their average height and built-up area proportion.

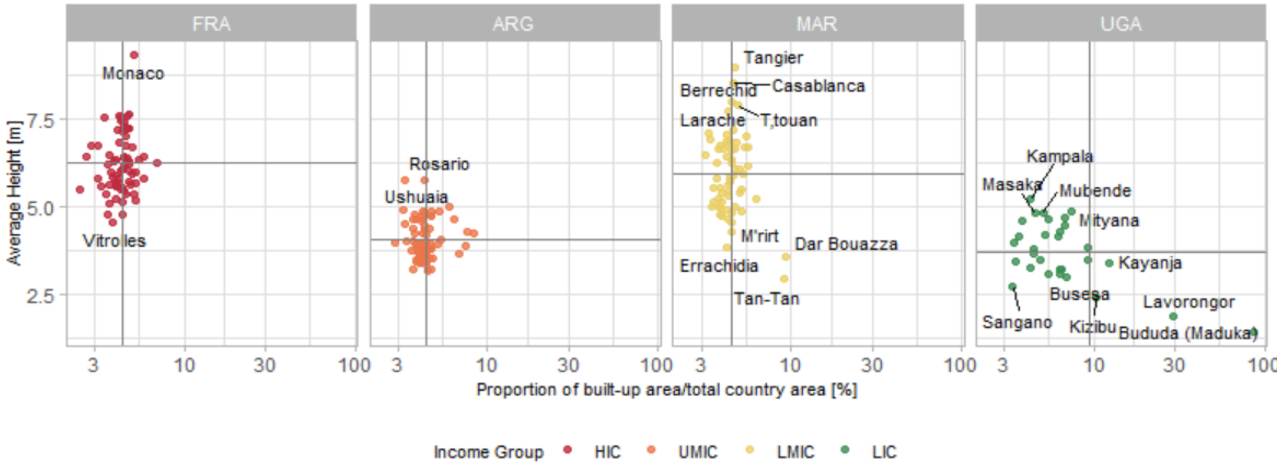

Fig. 6: Urban Cluster distribution in terms of av. height and built-up proportion for France, Argentina, Morocco and Uganda. Each country belongs to a different income group.

Notably, at the country level, France and Morocco display strikingly similar patterns in terms of average height and built-up proportion, as indicated by the horizontal and vertical black lines. This suggests that income levels may not be the sole determining factor influencing infrastructure choices. Surprisingly, Argentina, despite being classified as an upper-middle-income country, exhibits lower average building heights compared to Morocco and appears relatively close to Uganda, which falls into a lower-income group.

In contrast, analysing within-country disparities reveals a different narrative. Within each country, cities exhibit greater variation in their morphological characteristics, as depicted by the spread of points in both the x and y directions. This variation is notably pronounced among Ugandan cities, indicating that local policies and institutional forces play a significant role in shaping urban development patterns within the country

4 CONCLUSION AND DISCUSSION

Accurate knowledge of the 3D characteristics of the built-up environment and population distribution holds significant potential to enhance aspects of sustainability and resilience in urban planning and related decision-making processes. However, a major obstacle in advancing our understanding of these factors on a large scale has been the scarcity of data with sufficient accuracy, spatial detail, and consistency.

This paper introduces and showcases a series of applications that leverage the WSF3D dataset. First, we demonstrate how this data can improve existing population models by integrating volume and functional use information of the built-up environment, leading to more precise population estimates. The key advantage of this approach lies in achieving higher quantitative and qualitative accuracies over binary and density approaches using a single dataset, thus eliminating the technical complexities associated with gathering extensive data. As we continue to develop a multi-temporal WSF3D layer, our outlook is to produce and openly release a global-scale, multi-temporal population dataset with unparalleled accuracy and spatial resolution (approximately 10 meters).

Furthermore, we illustrate how the WSF3D dataset can be employed to comprehend the complex correlations between economic and institutional forces with urban morphology. We highlight how patterns and trends vary depending on the spatial domain of analysis. For example, the observed similarities in urban morphology between countries with different income levels highlight the need to consider various factors beyond income when formulating urban development strategies. Moreover, the substantial variation in urban morphology within a single country underscores the importance of context-specific policies and governance in shaping the built environment. These insights provide valuable guidance for policymakers and urban planners to foster equitable and sustainable development practices, tailoring solutions to address the unique challenges and opportunities presented by different urban contexts.

In conclusion, due to the remarkable versatility and seamless integration of the WSF3D with other data sources, our main goal in presenting these results is to encourage further exploration and utilization of this dataset to address diverse spatial and socio-economic challenges worldwide. By leveraging this valuable resource, researchers and policymakers can make informed decisions that foster sustainable urban development and enhance the overall well-being of communities globally.

5 REFERENCES

Cai, M., et al. (2021). The need for urban form data in spatial modeling of urban carbon emissions in China: A critical review. *Journal of Cleaner Production*, *319*, 128792.

Carpio, A., et al. (2021). Urban form, land use, and cover change and their impact on carbon emissions in the Monterrey Metropolitan area, Mexico. *Urban Climate*, *39*, 100947.

Chen, T.-H. K., et al. (2023). Detecting urban form using remote sensing: spatiotemporal research gaps for sustainable environment and human health. In *Earth Data Analytics for Planetary Health* (pp. 185-217). Springer.

Chokhachian, A., et al. (2020). Urban performance and density: Generative study on interdependencies of urban form and environmental measures. *Sustainable Cities and Society*, *53*, 101952. https://doi.org/https://doi.org/10.1016/j.scs.2019.101952

Esch, T., et al. (2022). World Settlement Footprint 3D-A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, *270*, 112877.

Lall, S. V., et al. (2021). The Evolution of City Form.

Leyk, S., et al. (2019). The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth Syst. Sci. Data*, *11*(3), 1385-1409. https://doi.org/10.5194/essd-11-1385-2019

Marconcini, M., et al. (2021). Understanding current trends in global urbanisation-the world settlement footprint suite. *GI_Forum*, *9*(1), 33-38.

Mills, G., et al. (2021). Characterising urban morphology for urban climate modelling. *Urban Climate Science for Planning Healthy Cities*, 339-354.

Palacios-Lopez, D., et al. (2022). Towards an improved large-scale gridded population dataset: A Pan-European study on the integration of 3D settlement data into population modelling. *Remote Sensing*, *14*(2), 325.

Scheba, A., et al. (2021). Inequality and urban density: Socio-economic drivers of uneven densification in Cape Town. *Environment and Urbanization ASIA*, *12*(1_suppl), S107-S126.

Wang, J., et al. (2023). EO+ Morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning*, *233*, 104691.

Wang, P., et al. (2021a). Sustainability of compact cities: A review of Inter-Building Effect on building energy and solar energy use. *Sustainable Cities and Society*, *72*, 103035.

Wang, W., et al. (2021b). From simulation to data-driven approach: A framework of integrating urban morphology to low-energy urban design. *Renewable Energy*, *179*, 2016-2035.

Zink, M., et al. (2014). TanDEM-X: The New Global DEM Takes Shape. *IEEE Geoscience and Remote Sensing Magazine*, *2*(2), 8-23. https://doi.org/10.1109/MGRS.2014.2318895