Developing a pathway to improve large-scale gridded population modelling based on the World Settlement Footprint suite

DISSERTATION

Zur Erlangung

des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

des Fachbereichs Mathematik/Informatik

der Universität Osnabrück

Vorgelegt von

Daniela Palacios Lopez

Prüfer der Dissertation:

Prof. Dr. Peter Reinartz, Universität Osnabrück, DEU

Prof. Dr. Andrew J. Tatem, University of Southampton, UK

Tag der mündlichen Prüfung: 24. November.2022

"The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research"

—Paul Dirac

Acknowledgments

This PhD dissertation would not have been possible without the help and support of many people and organisations to whom I would like to express my deepest gratitude.

First and foremost, I am extremely grateful to Prof. Dr. Peter Reinartz from the German Aerospace Centre (DLR) for giving me the opportunity to undertake my PhD studies under his guidance and supervision, and for allowing me the freedom to carry out my work with utmost independence. I am thankful for his academic support and insightful feedback on all my research articles, and of course, on the final manuscript. I am also thankful for his valuable observations made in every academic presentation done during my PhD studies.

At the same time, I am deeply indebted to my second supervisor Prof. Dr. Andrew J. Tatem (WorldPop & University of Southampton) for sharing his extensive expertise in the field of population modelling. I am particularly thankful for his knowledge-based comments on how to address and focus on some of the most important research gaps, and for his invaluable ideas on how to turn my research into important and practical contributions to the field.

Besides my supervisors, I would also like to extend my deepest gratitude to my mentors Dr. Thomas Esch (DLR) and Dr. Mattia Marconcini (DLR), as the successful completion of my PhD research would not have been possible without the exemplary work and efforts they have placed on the development of the World Settlement Footprint suite. I am truly in debt with them, for giving me the opportunity to participate and contribute to this impressive project, but most importantly, I am thankful for their profound belief in my abilities, their constructive criticism towards my work and their unparalleled support which without a doubt have made me a better researcher.

I must also thank the amazing group of people of WorldPop, the University of Southampton, the Centre for International Earth Science Information Network (CIESIN), the University of Columbia and NASA Socioeconomic Data and Applications Centre (SEDAC) for providing me with a great amount of assistance in terms of data and practical suggestions. An enormous 'thank you!' goes in particular to Dr. Alessandro Sorichetta, Dr. Jeremiah J. Nieves, Kytt MacManus and Gregory Yetman, with whom I hope to have future and ongoing collaborations. Here, I would also like to acknowledge the people working with the German Academic Exchange Service (DAAD), not only for funding my PhD studies, but for their friendly guidance towards all aspects of my residency in Germany.

Furthermore, it has been a great pleasure to pursue my PhD research at the German Remote Sensing Data Centre (DFD) of the German Aerospace Centre (DLR), where I would like to particularly extend my deepest gratitude to Prof. Dr. Claudia Künzer and Prof. Dr. Stefan Dech for fostering a culture of fair and valuable research with global applications. Within the department of Land Surface Dynamics, special thanks also go to all my colleagues in the team of Smart Cities and Spatial Development, in particular to my dear friends Julian Zeidler, Dr. Anne Marconcini-Metz, Dr. Benjamin Leutner, Elizabeth Brzoska, and former members including Dr. Wieke Heldens, Dr. Felix Bachofer, Milena Mönks and Tejas Bhagwat, who greatly contributed to my work by exchanging their

knowledge in different methodological frameworks and by proof reading my research articles. Within DLR, I would also wish to thank my fellow PhD colleagues with whom I exchanged valuable ideas during coffee breaks, summer walks and extensive online chatting during the COVID-19 pandemic, including: Soner Üreyen, Mariel Dirscherl, Chaonan Ji, Dr. Ya-Lun Tsai, Aiyim Orynbaikyzy, Thorsten Höser, Dr. Zhongyang Hu, Dr. Emmanuel Da Ponte and Dr. Celia Baumhoer.

Finally, I cannot begin to express my gratitude to my parents Luis Hector Palacios Hartweg, and Nelly Lopez Monje, my brother Luis Eduardo Palacios Lopez , my husband Manuel Kröger and my friends Barbara, Andie and Shipra. To my family, I have to say that your love and endless support have and will always be the pillars of my success. To my husband, I want to say that without your extensive knowledge on proper writing and priceless motivational coaching, the research presented here would not have had the same high quality, and to my lovely friends, all my gratitude for keeping me going even in the hardest of times. This PhD thesis is dedicated to you all.

Daniela Palacios Lopez

Munich, Germany

English Summary

The field of large-scale human population modelling has emerged as a response to the increasing demands for actionable, consistent and comparable population data needed to support a large number of sustainable development applications. Nowadays, Earth Observation-derived, top-down large-scale gridded population datasets that describe the extent and spatial distribution of the human population as continuous surfaces (rasters), are openly accepted by many governments and institutions around the world, who use them as an alternative source of information to complement/supplement conventional census/estimate-based population data.

Given the wide range of applications where gridded population products are being employed, research performed to improve the accuracy and spatial resolution of large-scale population models has become of utmost importance. For the last decade, the scientific community has constantly leveraged the increasing availability of Earth Observation data, the improvements made on Remote Sensing techniques and the developments made on the field of Machine Learning, to produce large-scale gridded population datasets with higher usability and reliability. For example, some of the most accurate and spatially explicit products available at a global scale are produced mainly by harnessing the inclusion of remotely-sensed derived proxy layers with improved thematic and spatial resolution, especially those describing the characteristics of the built-up environment such as built-area layers and building footprint datasets, respectively.

However, notwithstanding these advancements, a systematic literature review undertaken within this PhD research has revealed that existing top-down large-scale population models still suffer from a number of limitations that affect the final accuracy and usability of their corresponding derived population datasets. In particular, it has been identified that existing models used to produce large-scale gridded population maps are still affected by a) the quality and recency (currentness or age of the data) of the underlying census/estimate-based population data on the one hand; and by b) the still low spatial resolution of the geospatial proxies used for disaggregation, c) the persistent inaccuracy in identifying populated areas from remotely-sensed data, and d) the lack of information on the functional use and 3D characteristics of the built-up environment, on the other hand. Overall, it has been concluded that if some of these limitations still exist in the field, it is because data and methods that can help overcome these issues at very local–scales (e.g. national models) are yet not available or transferable to large-scale applications (e.g. continental or global models).

In this context, this PhD thesis explores the capabilities and effectiveness of the new World Settlement Footprint (WSF) suite in the production of a large-scale gridded population model that allows improving the accuracy and spatial resolution of end-user population datasets. In detail, it presents a methodological framework that explores how and if each of the WSF-layers, namely the WSF2015, the WSF2019 and the WSF3D can overcome the limitations listed above, in particular limitations b), c) and d). Thereinafter, each WSF-layer was evaluated in terms of 1) their ability to improve population estimates compared to binary-dasymetric models, 2) their ability to produce consistent and comparable accuracies across large territorial extents, 3) their ability to produce accurate population estimates acting as single proxies for population modelling, 4) their ability to

reduce the technical complexities of multi-layer weighed-dasymetric models, 5) their ability to discriminate large industrial areas using a simple and spatially transferable machine learning approach, and finally 6) their ability to improve population estimates through the integration of volume and settlement use information.

Within this methodological framework, a comprehensive set of spatial and statistical analyses were designed to evaluate the uncertainties of large a number of population distribution maps at local, national and continental scale. For a robust assessment, population models were produced at varying currencies, qualities and spatial scales of the input census-based population data, with the purpose of analysing how the differences in the level of spatial granularity of the available administrative boundaries and the variability in the morphology of built-up landscapes influence the accuracy of each WSF layer.

Overall, the main findings of this PhD thesis demonstrate that the WSF-layers are capable of tackling some of the main limitations identified in the field of large-scale population modelling. First, the independent weighting framework provided by the non-binary WSF-layers allowed outperforming the mapping accuracies of widely employed binary-dasymetric models and reduce the technical complexities of (multi-layer) weighted dasymetric models. Second, as single proxy layers used for dasymetric disaggregation, each WSF-layer was also capable of delivering consistent and systematic accuracies across large territorial extents (e.g. continent and region-wide); where the robustness of each layer was consistent under varying qualities and spatial resolutions of the input population data. Finally, spatial metrics derived solely from the WSF3D dataset showed to be extremely effective at classifying the built-up environment into industrial and non-industrial land-uses, which ultimately, allowed incorporating for the first time ever, settlement use and settlement volume information into large-scale models of population disaggregation.

In view of these promising results, the main contributions of this PhD research can be summarised as follows:

1. Quantitative and qualitative demonstration of how employing the WSF-suite for population modelling overcomes some of the most prominent limitations in the field.
2. First in-depth quality assessment aimed at evaluating the effectiveness and suitability of each WSF-dataset as proxy layer for large-scale population modelling.
3. Design and implementation of the Settlement Size Complexity (SSC) index as a robust metric to evaluate the uncertainty of population models based on built-up area layers.
4. Improving understanding on the "fitness for use" of large-scale gridded population datasets.
5. The development of a highly accurate, semi-automatic and globally transferable method for the identification of industrial and non-industrial areas using only the WSF3D dataset in combination with a Machine Learning approach.

6. First time delivery of large-scale population datasets produced on the basis of the WSF-layers, to serve as actionable data for a large number of ongoing-projects.

Deutsche Kurzfassung

Der Forschungsbereich der großflächigen Bevölkerungsmodellierung hat sich als Folge des zunehmenden Bedarfs an verwertbaren, konsistenten und vergleichbaren Bevölkerungsdaten entwickelt, die für eine Vielzahl von Anwendungen im Kontext der nachhaltigen Entwicklung benötigt werden. Heutzutage werden aus Erdbeobachtungsdaten abgeleitete, großflächige Bevölkerungsdaten im Rasterformat, die das Ausmaß und die räumliche Verteilung der menschlichen Bevölkerung als kontinuierliche Flächen beschreiben, von vielen Regierungen und Institutionen auf der ganzen Welt offen akzeptiert. Diese nutzen die Bevölkerungsdaten als zusätzliche Informationsquelle, da sie die konventionellen, auf Volkszählungen sowie Schätzungen basierenden Daten auf wertvolle Weise ergänzen.

In diesem Zusammenhang und angesichts des breiten Spektrums von Anwendungen, bei denen gerasterte Bevölkerungsprodukte zum Einsatz kommen, sind Forschungsarbeiten zur Verbesserung der Genauigkeit und räumlichen Auflösung großflächiger Bevölkerungsmodelle von größter Bedeutung. In den letzten zehn Jahren haben Wissenschaftler zunehmende Verfügbarkeit von Erdbeobachtungsdaten, die Verbesserungen der Fernerkundungsmethoden und - in jüngerer Zeit - die Entwicklungen bei den Algorithmen des maschinellen Lernens genutzt, um großflächige Bevölkerungsdatensätze mit einer höheren Nutzbarkeit und Zuverlässigkeit als je zuvor zu erstellen. Heute werden einige der genauesten und räumlich explizitesten Produkte, die auf globaler Ebene verfügbar sind, durch die Einbeziehung von aus der Fernerkundung abgeleiteter Indikatoren mit verbesserter thematischer und räumlicher Auflösung erstellt. Diese umfassen insbesondere solche, die urbane Strukturen beschreiben, wie z. B. Datensätze zu bebauten Flächen oder Gebäudeumrisse.

Eine im Rahmen dieser Doktorarbeit durchgeführte systematische Literaturrecherche hat jedoch ergeben, dass die bestehenden großflächigen Bevölkerungsmodelle nach wie vor einige Einschränkungen aufweisen, die die Genauigkeit und Verwendbarkeit der entsprechend abgeleiteten Bevölkerungsdatensätze stark beeinträchtigen. Insbesondere wurde festgestellt, dass bestehende Techniken zur Erstellung großflächiger Bevölkerungsschätzungen nach wie vor durch die Qualität und Aktualität der zugrundeliegenden, auf Volkszählungen sowie Schätzungen basierenden Bevölkerungsdaten beeinträchtigt werden. Andererseits haben die nach wie vor geringe räumliche Auflösung der für die Disaggregation verwendeten räumlichen Indikatoren, die resultierenden Ungenauigkeiten bei der Klassifikation von besiedelten Gebieten anhand von Fernerkundungsdaten, der Mangel an Informationen über die funktionale Nutzung sowie die fehlende vertikale Information der bebauten Umgebung negative Auswirkungen auf die modellierten Bevölkerungsdaten . Insgesamt wurde der Schluss gezogen, dass einige dieser Einschränkungen weiterhin bestehen und dies darauf zurückzuführen ist, dass Daten oder Methoden, die dazu beitragen können, diese Probleme auf sehr lokaler oder feiner Ebene zu überwinden oder auf das Fehlen, noch nicht verfügbar oder in großem Maßstab übertragbar sind (z. B. kontinentale oder globale Ebene).

Um diese Herausforderungen zu bewältigen, untersucht diese Doktorarbeit die Möglichkeiten und die Effektivität der neuen World Settlement Footprint (WSF) Produkte bei der Erstellung eines großflächigen und rasterbasierten Bevölkerungsmodells, das die Verbesserung der Genauigkeit und der räumlichen Auflösung bestehender Bevölkerungsdatensätze ermöglicht. Im Detail wird ein methodisches Framework vorgestellt, das die Effektivität jeder WSF-Layer, nämlich des WSF2015, des WSF2019 und des WSF3D, in Bezug auf ihre Fähigkeit, 1) Bevölkerungsschätzungen im Vergleich zu binär-dasymetrischen Modellen zu verbessern, 2) konsistente und vergleichbare Genauigkeiten über große territoriale Ausdehnungen zu erzeugen 3) genaue Bevölkerungsschätzungen zu erstellen, die als einzelne Näherungswerte für die Bevölkerungsmodellierung dienen, 4) die technische Komplexität von gewichteten-dasymetrischen Modellen zu verringern, 5) große Industriegebiete mit Hilfe eines einfachen und räumlich übertragbaren maschinellen Lernansatzes zu unterscheiden, und schließlich 6) Bevölkerungsschätzungen durch die Integration von Informationen über Volumen und Siedlungsnutzung zu verbessern.

Innerhalb dieses methodischen Frameworks wurden umfassende räumliche und statistische Analysen durchgeführt, um die Unsicherheiten einer großen Anzahl von Bevölkerungsverteilungskarten auf lokaler, nationaler und kontinentaler Ebene zu bewerten. Für eine robuste Bewertung wurden Bevölkerungsmodelle mit unterschiedlicher Aktualität, Qualitäten und räumlichen Maßstäben der eingegebenen Bevölkerungsdaten erstellt, um zu analysieren, wie die Unterschiede in der räumlichen Auflösung der verfügbaren Verwaltungsgrenzen sowie die Variabilität in der Morphologie bebauter Landschaften die Genauigkeit der einzelnen WSF-Layer beeinflussen.

Die wichtigsten Erkenntnisse dieser Doktorarbeit zeigen, dass die WSF-Layer in der Lage sind, die aufgezeigten Beschränkungen im Bereich der großflächigen Bevölkerungsmodellierung zu überwinden. Erstens konnte durch die unabhängige Gewichtung, die durch die nicht-binären WSF-Layer ermöglicht wird, einerseits die Zuordnungsgenauigkeit der häufig verwendeten binär-dasymetrischen Modelle übertroffen und andererseits die technische Komplexität der (mehrschichtigen) gewichteten dasymetrischen Modelle verringert werden. Zweitens war jeder WSF-Layer als einzelner Proxy, der für die dasymetrische Disaggregation verwendet wird, auch in der Lage, konsistente und systematische Genauigkeiten über den gesamten Raum hinweg zu liefern, wobei die Robustheit jedes Layers konsistent war. Diese wurde unter verschiedenen Qualitäten und räumlichen Auflösungen der eingegebenen Bevölkerungsdaten getestet. Schließlich erwiesen sich räumliche Metriken, die ausschließlich aus den WSF3D-Datensätzen abgeleitet wurden, als äußerst effektiv bei der Klassifizierung von Informationen über die Siedlungsnutzung im gesamten Raum, was letztlich die Einbeziehung von Informationen über die Siedlungsnutzung und das Siedlungsvolumen - zum ersten Mal überhaupt - in groß angelegte Modelle der Bevölkerungsdisaggregation ermöglichte.

In Anbetracht dieser vielversprechenden Ergebnisse können die wichtigsten Beiträge dieser Doktorarbeit wie folgt zusammengefasst werden:

1. Quantitative und qualitative Demonstration, wie der Einsatz der WSF-Produkte für die Bevölkerungsmodellierung die bestehenden Beschränkungen auf diesem Gebiet überwindet.
2. Erste eingehende interne Qualitätsbewertung, die darauf abzielt, die Effektivität und Eignung jedes WSF-Datensatzes als Proxy-Layer für die großflächige Bevölkerungsmodellierung zu evaluieren.
3. Entwurf und Implementierung des Siedlungsgrößenkomplexitätsindex (SSC) als robuste Metrik zur Bewertung der Unsicherheit von Bevölkerungsmodellen.
4. Verbesserung des Verständnisses der „fitness for use“ von groß angelegten gerasterten Bevölkerungsdatensätzen.
5. Die Entwicklung einer hochpräzisen, halbautomatischen und weltweit übertragbaren Methode zur Identifizierung von Industrie- und Nicht-Industriegebieten unter ausschließlicher Verwendung des WSF3D-Datensatzes in Kombination mit Algorithmen aus dem Bereich des Maschinellen Lernens.
6. Erstmalige Lieferung von Bevölkerungsdatensätzen auf der Basis der WSF-Layer, die als verwertbare Daten für eine Vielzahl von laufenden Projekten dienen.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

AI	Artificial Intelligence
ASR	Average Spatial Resolution
BA	Building Area
BC	Building Coverage
BF	Building Fraction
BH	Building Height
BV	Building Volume
BV-IS	Building Volume minus Industrial Use
CCI	Climate Change Initiative
CIESIN	Centre of International Earth Science Information Network
COV	Coefficient of Variation
COVID-19	Corona Virus 2019
DCW	Digital Chart of the World
DEM	Digital Elevation Model
DFD	German Remote Sensing Data Centre
DLR	German Aerospace Centre
DMSP-OLS	Defence Meteorological Program-Operational Line-Scan System
EEA	European Environment Agency
EO	Earth Observation
ESA	European Space Agency
FAO	Food and Agriculture Organisation
FDRR	Framework for Disaster and Risk Reduction
FUAs	Functional Urban Areas
GAIA	Global Annual Impervious Area
GEE	Google Earth Engine
GHSL	Global Human Settlement Layer
GHS-POP	Global Human Settlement Population
GIS	Geographical Information Systems
GISA	Global Impervious Surface
GLCC	Global Land Cover Characterisation
GRUMP	Global Rural-Urban Mapping Project

GPW	Gridded Population of the World
GUF	Global Urban Footprint
HRSL	High Resolution Settlement Layer
ISO	International Organisation for Standardisation
JRC	Joint Research Centre
K	Kappa Coefficient
LiDAR	Light Detection and Ranging
LU/LC	Land Use and Land Cover
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAUP	Modifiable Areal Unit Problem
MNDWI	Modified Normalised Difference Water Index
MODIS	Moderate Resolution Imaging Spectroradiometer
NDBI	Normalised Difference Build-Up Index
NDGB	Normalised Difference Green Blue
NDRB	Normalised Difference Red Blue
NDMIR	Normalised Difference Middle Infrared
NDVI	Normalised Difference Vegetation Index
NLCD	National Land Cover Database
nMAE	Normalised Mean Absolute Error
OA	Overall Accuracy
OSM	Open Street Maps
PA	Producers' Accuracy
PIS	Percent of Impervious Surface
RBF	Radial Basis Function
REE	Relative Estimation Error
RF	Random Forest
RMSE	Root Mean Square Error
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SEADRIF	Southeast Asia Disaster Risk Insurance Facility
SDGs	Sustainable Development Goals
SSC	Settlement Size Complexity

SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
TDX-DEM	TanDEM-Digital Elevation Model
UN	United Nations
UNCTAD	United Nations Conference on Trade and Development
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNICEF	United Nations Children's Fund
UNFPA	United Nations Population Fund
UNPD	United Nations Population Division
USA	United States of America
US	Users Accuracy
VHR	Very High Resolution
VIIRS	Visible Infra-red Imaging Radiometer Suite
WFP	World Food Programme
WGS1984	World Geographic System 1984
WHO	World Health Organisation
WSF	World Settlement Footprint
WSF2015	World Settlement Footprint 2015
WSF2019	World Settlement Footprint 2019
WSF3D	World Settlement Footprint Three-Dimensional

Chapter 1

1. Introduction

1.1 The importance of improving large-scale gridded population models

When the main objective of this PhD research was conceptualised at the end of the year 2017, the world population was reaching the 7.6 billion mark. Four years later, this number has reached the 8.0 billion mark, which indicates that on average, 100 million people were added to the global population per year. According to the United Nations (UN) (UN, 2019; UNFPA, 2021), if the world population should continue to grow at its present rate, the number of humans on Earth could reach the 9.8 billion mark by the year 2050, and increase up to 11.2 billion by end of the century.

Naturally, while population growth and other population dynamics such as urbanisation, migration and population aging can pose challenges as well as opportunities for a given country, over the past decade, abrupt changes in these population processes have mainly acted as a break on social, economic and environmental development. This effect has been more pronounced in the world's least developed countries, where the current policies and economies are not well established to deal with the rapid population changes (Twinoburyo et al., 2019; UN, 2018). For example, owing to the increasing unplanned urbanisation, severe climate change and land degradation has led nearly 39% of the Asia and Pacific region exposed desertification, which in return, has increased food insecurity and hunger. Today, about 351 million people residing in these regions are estimated to be undernourished, which is about 51% of the total amount at a global scale (FAO et al., 2021). Comparably, in most Sub-Saharan African countries, services such as education, health care, electricity, water and decent network infrastructures are severely over stretched as a result of the rapid population and urbanisation growth rates reported every year (~2.4% and ~3.4%, respectively) (Tuholske et al., 2019). Here, approximately 60% of the Sub-African youth between ages fifteen and seventeen are not in school (UNESCO, 2021), 76% of the population do not have access to safe drinking water (UNESCO, 2019) and around 490 million people in live in extreme poverty (UNCTAD, 2021)

Under these circumstances, in an era when we are trying to achieve a global sustainable and inclusive future, the human population has to be considered as one the most important numbers in the sustainability equation (Rosling et al., 2018). The reason for this is that the varying patterns of population growth, composition (e.g. size, density, age, gender, ethnicity, race, income) and distribution of population, influence the patterns of consumption, production, employment and income at global, regional, national and local scales (Aguirre, 2002; National Research Council, 1994). This means, that efforts that remain disassociated from the knowledge and policy options linked to population processes are, and will be destined to fail, as all measures aiming at eliminating poverty,

hunger, land degradation, water pollution, climate change, etc., can only be successfully implemented if they take into account the size, composition and distribution of the populations they target (Herrmann et al., 2012).

In this context, to promote more sustainable pathways of global development and to implement policies that integrate population dynamics into development plans; governments, organisations, policy-makers, and researchers alike, need to have access to timely and reliable population data. Existing international frameworks for development, including the Sustainable Development Goals (SDGs) (UN, 2022b), the Sendai Framework for Disaster and Risk Reduction (Sendai FDRR) (UNDRR, 2015), the Paris Agreement (UN-Climate Change, 2022) and the UN-New Urban Agenda (UN-Habitat, 2022), for example, rely greatly on the availability of population data which are used 1) as denominator in calculating different metrics and indices, 2) as a primary resource to support decision-making, and 3) as main input to design, implement and fine-tune policies aiming at global sustainability (Qiu et al., 2022; Sankoh, 2017; UN, 2021).

For most countries, the traditional form of collecting population data is through national population or housing censuses where information on the number of people and their main characteristics are collected approximately every 10 years. According to the UN, to this day censuses are the most accurate and rich source of population data, as they gather information using the lowest geographical divisions (e.g. household or building level), covering small areas up to the national and international scales (UN-Statistics Division, 2022). In this framework, however, when population totals are made openly available to the public, they are typically aggregated to large administrative units (e.g. census blocks, neighbourhoods, municipalities, etc.) to protect the privacy of the citizens. While this aggregation process is quite standard, it limits the usability of the data, especially in the context of different analytical purposes. For example, census-based population data are collected at different time intervals and are made available using different administrative units among and within countries which makes it difficult to compare population distributions in a consistent and methodological way (Wardrop et al., 2018). At the same time, the administrative units used to aggregate population totals do not correlate with any other geographical phenomena, restricting the integration of census-based population data with other geospatial datasets. Finally, as population counts are aggregated from building level to coarser units, there is a loss of spatial detail which affects the overall accuracy of any subsequent analyses where the data is used.

On this basis, to be of value population data has to be spatially explicit and comparable within and across countries (UN, 2021). Standardised and precise information on where and in what density humans live is essential not only to implement location-based policies, but also to develop local-to-global initiatives that consider populations located in rural and urban settings across the world (POPGRID, 2021). Here, spatially explicit and globally-standardised population data provide the foundation for investigating geographic variations in policy-performance, thus allowing answering questions such as: *Where are sustainable policies over- or underperforming?* or *Does performance vary across space (e.g. between and within countries or regions)*?

To produce more spatially explicit and globally comparable population data, the scientific community has increasingly invested in two main overarching efforts: 1) the development of initiatives to collect, harmonise, and temporally-adjust census-based

population counts (e.g. production of estimates and projections) and cartographic administrative boundaries at a global scale (Doxsey-Whitfield et al., 2015; Freire et al., 2018) and 2) the development of methods to create alternative, global-scale population datasets that provide more precise spatial representations of the population distribution at moderate-to-high-spatial resolutions (Balk et al., 2006). For the last thirty years, these two efforts have been combined with the increasing advances in Remote Sensing (RS), Geographical Information Systems (GIS) and Artificial Intelligence (AI) techniques, allowing the creation and proliferation of georeferenced data products known as "top-down large-scale gridded population datasets".

Concisely, large-scale gridded population datasets provide estimates of the total population as a continue surface or raster at near-global to global scale (Leyk et al., 2019). They are produced using different top-down disaggregation techniques and auxiliary data, in which global census-based population counts, estimates or projections are redistributed from vector-format administrative boundaries to raster pixels of a given spatial resolution. Population allocation is commonly based on a weighting layer that restricts or calculates how many people are allocated per pixel, where the pixel-weights are extracted through simple or complex statistical approaches that investigate the relationships between population densities and different geophysical variables or geospatial proxies (e.g. built-up areas, distance to main roads, services or amenities, elevation, climate, etc.).

As such, top-down gridded population datasets do no replace census/estimate-based population data but rather supplement or complement the information by improving its spatial resolution and interpretation capabilities (Anderson et al., 2017). In its raster format, population data can be more easily integrated with other global gridded datasets such as environmental, economic, or agricultural geodatasets, allowing a deeper understanding of human-environment interrelationships from an increased spatial perspective. At the same time, the data can be aggregated to arbitrary spatial units such as hazard zones, climate zones, risk areas, etc., thus facilitating spatial and statistical analyses. Furthermore, gridded population data provide consistent and comparable information across space, allowing implementing cross-comparison analyses within and across regions (Wardrop et al., 2018).

Due to their remarkable advantages, nowadays large-scale gridded population datasets are greatly accepted by researchers, governments and institutions all over world, who consider them as invaluable sources of population data (Allen et al., 2021). State-of-the-art products, including the Global Rural-Urban Mapping Project (GRUPM) (CIESIN, 2011), the Gridded Population of the World (GPWv4.11) (Doxsey-Whitfield et al., 2015), the LandScan dataset (Bhaduri et al., 2007; Dobson et al., 2000) , the Global Human Settlement Population layers (GHS-POP) (Freire et al., 2016), the WorldPop datasets (Stevens et al., 2015b) and the High-Resolution Settlement Layer (HRSL) (Tiecke et al., 2017) are widely used to support a large variety of research areas, including public health applications (España et al., 2018; Fries et al., 2021; Hay et al., 2005a), public security (Galway et al., 2012), urban planning and characterisation (Amoah et al., 2018; Dhewantara et al., 2018; Serrano Giné et al., 2016; Tuholske et al., 2019), accessibility analyses (Ajisegiri et al., 2019; Linard et al., 2012; Sorichetta et al., 2016), poverty mapping (Barbier & Hochard, 2018; Elvidge et al., 2009; Noor et al., 2008), hazard and environmental

risk assessments (Brown et al., 2018; Calka et al., 2017; Maas et al., 2019; Maroko et al., 2019; Mohanty & Simonovic, 2021; Smith et al., 2019; Tellman et al., 2021), and humanitarian relief applications (Kellenberger et al., 2021). They also support the monitoring and implementation of the SDGs (Kavvada et al., 2020; Kuffer et al., 2020; Qiu et al., 2022; Tuholske et al., 2021), and most recently, with the emergence of the Corona-Virus (COVID-19) pandemic, these datasets have further demonstrated their utility by providing estimates of people exposed or at risk of transmission due to crowding (Rader et al., 2020), as well as to measure assess to vaccines (Rader et al., 2021).

With that being said, the sensitivity of these applications highlights how important it is for large-scale gridded population datasets to be as accurate as possible. These datasets serve not only as empirical evidence, but also as a critical component to better target and allocate financial resources towards vulnerable populations (Aubrecht et al., 2013), where inaccuracies in the estimates can prevent governments and organisations to reach those in more need. Under the UN-motto of "Leave No One Behind" (UN, 2022a), accurate gridded population datasets ensure including populations located even in most remote and reclusive areas of the world.

In this regard, among the many advances done in the field of large-scale top-down gridded population modelling, one of the most notable sources of improvement has been the use of Earth Observation (EO) satellite-based geospatial layers, which through the years, have become increasingly more precise and accurate (Leyk et al., 2019). Here, the development of highly accurate datasets describing the extent and spatial distribution of human settlements and building footprints has played a crucial role, as these proxy layers in particular, have proven to be "the single most highly predictive indicators of human habitation" (Nieves et al., 2017; Reed et al., 2018; Stevens et al., 2020). State-of-the-art built-area layers such as the Global Urban Footprint (GUF) (Esch et al., 2017), the Global Human Settlement Layer (GHSL) (Pesaresi et al., 2016), the WorldPop growth built-up models (Nieves et al., 2020b) and building patterns (Nieves et al., 2020a), the Ecopia/Maxar (Maxar Technologies, 2020), Microsoft (Heris et al., 2020) and Google building footprints (Sirko et al., 2021) have help refined large-scale top-down gridded population models, by improving the identification of human settlements, particularly in rural areas.

Regrettably, even though major qualitative and quantitative improvements have been reported through the years, contemporary research aimed at assessing the accuracy of state-of-the-art large-scale top-down gridded population datasets has revealed that a series of limitations still affect the usability and accuracy of all existing products. These limitations are not unique to any particular existing product, but rather limitations that affect the underlying population models that are used/employed to produce them. Overall, the main challenges that need to be addressed to produce datasets with higher accuracy and usability can be summarised as follows:

1. **Improved spatial resolution:** Most of the currently available large-scale top-down gridded population datasets are produced at spatial resolutions of 100m, 250m and 1km at the Equator adhering the original or modified spatial scale of the proxies used for disaggregation (Lloyd et al., 2019). Different studies deem this as sub-optimal, first because the data cannot be directly integrated with other geospatial layer of higher spatial granularity leading to analytical challenges (Calka & Bielecka, 2020; Smith et al., 2019), second, because population counts extracted for very local areas result in

highly inaccurate counts (Thomson et al., 2021a), and third, because the coarse spatial resolution affect the identification of all potential populated areas (Marconcini et al., 2020). As such, to refine the spatial granularity of future datasets, there is a need to develop population models that can leverage and handle the inclusion of global-scale proxies with higher spatial resolutions (e.g. 10m, 12m or 30m at the Equator).

2. **Improved comparability and replicability:** In general, population models that employ multiple proxies for population disaggregation have shown to have higher qualitative and quantitative accuracies than those models that employ no ancillary data, or just one proxy (Stevens et al., 2015b; Su et al., 2010). In the context of large-scale top-down population modelling, however, these models can show local quality variations, which are difficult to account for in the framework of cross-comparison analyses (Schug et al., 2021). Furthermore, due to the employment of a large number of proxy layers, multi-layer models suffer from endogeneity issues as well as transferability restrictions in space and time (Balk et al., 2006). Therefore, future research should focus on the development of population models capable of delivering systematic patterns of quality across space, leveraging global proxies with comparable spatio-temporal quality. Optimally, these models should also allow for replicability and transferability, and rely on proxies that provide a direct physical relationship with population densities without statistical modelling (e.g. built-up densities).

3. **Integration of use and 3D information of the built-up environment:** State-of-the-art large-scale top-down gridded population datasets are produced with population models that do not integrate the functional (residential vs non-residential) and three-dimensional (3D) information (e.g. volume or floor-space) on the built-up environment. This limitation has led to large error of underestimations in highly dense urban areas with high-rise buildings (Thomson et al., 2021a) , and large errors of overestimation in non-residential areas (e.g. industrial and commercial centres) (Palacios-Lopez et al., 2019; Palacios-Lopez et al., 2021) which affects the accuracy of sub-sequent analyses. While meaningful evidence exists on the advantages of integrating such information into population models (Biljecki et al., 2016; Grippa et al., 2018; Huang et al., 2021b; Schug et al., 2021; Shang et al., 2021; Ural et al., 2011), current research mainly focuses on local to regional-scale analyses due to the lack of geospatial datasets with global-scale coverage. Therefore, future research should focus on the development of global-scale proxies that describe the functional and 3D-characteristics of the built-up environment, as well as on exploring how these datasets can be efficiently used to refine large-scale population estimates.

4. **Improved quality, recency and scale of the input census-based population data counts and boundaries:** National population and housing censuses are considered by far the most reliable source of population data. However, in countries which are mainly located in low-income regions or in areas of conflict, censuses have not been collected some times for more than 40 years (Wardrop et al., 2018). This means that for a given number of countries, top-down population datasets are being modelled with outdated or incomplete data; shortcomings that propagate to derived population grids and the applications where they are used (Kuffer et al., 2022) . In this context, while

reinforcing the economic, political and statistical capacities of countries should remain a major priority to collect up-to-date population counts and boundaries, future research should also focus on exploring and reinforcing alternative methods of producing population estimates. In particular, investigating "bottom-up" population modelling approaches should be of interest, which are methods being developed to bridge this gap at local and national scales (Darin et al., 2021a; Schug et al., 2021; Weber et al., 2018).

1.2 Research Motivation and Focus

Based on this theoretical background, the main motivation of the present thesis is to respond to the identified challenges in the field through the development of population models capable of generating large-scale datasets with improved accuracy and spatial resolution. Here, we specifically focus on providing large-scale solutions for the first three limitations mentioned in the previous section, presenting a methodology based strictly on "top-down" methods of population disaggregation. For completeness, however, we also present a discussion on how the methods and data developed here could potentially contribute to address the fourth limitation, although the latter is out of the scope of the main research.

Concisely, this thesis focuses on developing and evaluating the effectiveness of top-down population models that rely solely on the novel World Settlement Footprint (WSF) suite—a set of global layers that describe the extent, location, characteristics and spatial distribution of human settlements at with unprecedented accuracies and spatial resolutions. The main hypothesis is that each layer that compose the WSF-suite, namely the WSF2015 (binary and density layers), the WSF2019 (binary and imperviousness layers) and the WSF3D, respectively, has the potential to address one or multiple limitations affecting large-scale top-down gridded population models today; thus, allowing generating population distribution maps with higher accuracy than existing models.

This hypothesis is built upon four premises: First, that compared with any other built-area dataset available today, the new WSF-suite presents great accuracy in terms of settlement identification, both in urban and rural areas. Second, that compared with any other built-area dataset available today, the WSF-suite is produced at unprecedented high spatial resolutions of 10m and 12m at the Equator. Third, that compared with any other built-area dataset available today, the WSF-suite is one of the first datasets to offer 3D-information on the built-up environment, with accuracies comparable to very high-resolution products (e.g. Light Detection and Ranging (LiDAR) data). And fourth, that compared with highly accurate building footprint datasets which are only available for a limited number of countries, the WSF-suite offers global coverage.

In this framework, to test the proposed hypothesis, the design, theoretical and practical approach of this PhD research follows a series of evolving and systematic analyses that are inheritably linked to the chronological release of each layer. Accordingly, different research questions and objectives are addressed, covered in the three stand-alone peer-review publications that form the cumulative part of this thesis. These are summarised as follows:

1.3 Research Objectives and Research Questions

- **Research Article 1:** Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., Sorichetta, A., Zeidler, J., Kuenzer, C., & Dech, S. (2019). New perspectives for mapping global population distribution using world settlement footprint products. Sustainability, 11(21). https://doi.org/10.3390/su11216056

The first release within the WSF-suite included the WSF2015 and the WSF2015-Density layers. Both layers describe the extent, location and distribution of human settlements at a global scale for the year 2015; the first as a binary layer, and the second as a continuous layer depicting the PIS within built-up pixels. Following the premise that binary built-area datasets are by far the most informative proxy layers used for predicting population densities and distributions, and extending on previous local-scale research that have demonstrated that impervious surfaces are highly correlated to population counts, the first objective of this PhD research is to demonstrate if quality and accuracy improvements in population disaggregation can be achieved with the WSF2015-Density layer compared to the already established binary-dasymetric approach employed by other population dataset and the baseline built-area layers. The main research question is formulated as follows:

How effective is the WSF2015-Density layer in improving the accuracy of large-scale population models compared to the WSF2015 layer?

From this main research question, the following sub-questions are kept in focus:

1. What are the reported accuracies of population distribution maps produced on the basis of the WSF2015-Density compared to those produced on the basis of the WSF2015 layer?
2. How does changes in the spatial granularity of the available administrative units affect the accuracy of population models produced on the basis of the WSF2015-Density and the WSF2015 layers?
3. What is the quantitative relationship among the number of identified settlement pixels, the amount of population that needs to be distributed and the accuracy of the final population models produced with the WSF2015 layers?
4. What are the characteristics of the built-environment in which population models produced on the basis of the WSF2015-Density layer outperform those produced on the basis of the WSF2015 layer?
5. What advantages does the WSF2015-Density layer offer in support of large-scale population modelling in comparison to existing population models?
6. What are some of the remaining limitations affecting the qualitative and quantitative accuracy of population models produced on the basis of the WSF2015-Density layer?

To answer these questions the WSF2015 and WSF2015-Density layer are incorporated with sub-national census/estimate-based population data, to produce and validate high-resolution population datasets for nine low, middle and highly urbanised countries located across the four macro-regions of the world.

- **Research Article 2:** Palacios-Lopez, D., Bachofer, F., Esch, T., Marconcini, M., MacManus, K., Sorichetta, A., Zeidler, J., Dech, S., Tatem, A. J., & Reinartz, P. (2021).

High-Resolution Gridded Population Datasets: Exploring the Capabilities of the World Settlement Footprint 2019 Imperviousness Layer for the African Continent. *Remote Sensing*, *13*(6), 1142. https://doi.org/10.3390/rs13061142

The second release within the WSF-suite included the WSF2019 and the WSF2019-Imperviousness layer (WSF2019-Imp, previously referred to as "Density"). These layers represent an improvement over the 2015 products in relation to their production framework, which has led to a more accurate detection of settlement pixels and PIS calculation. Hence, building upon the limitations the WSF2015 layers and following the premise that the WSF2015-Density outperformed the qualitative and quantitative performance of population models produced with the binary WSF2015, the second main objectives of this PhD research are to demonstrate if the WSF2019-Imp is capable of producing systematically comparable population estimates under extremely varying spatial resolutions of the input population data, and to determine whether or not the layer can be used as single proxy for large-scale population models, reducing the complexities of multi-layer models (e.g. WorldPop and LandScan). The main research question is formulated as follows:

How effective is the WSF2019-Imp layer as a single proxy for population modelling in a continental-scale framework?

From this main research question, the following sub-questions are kept in focus:

1. Is the performance of the WSF2019-Imp as a single proxy layer for population modelling consistent within and across countries at a continental scale?
2. What are the spatial patterns of accuracy and how are these linked to the characteristics of the built-up environment and the population density?
3. What improvements can be reported over population maps produced on the basis of the WSF2015-Density layer?
4. What advantages does the WSF2019-Imp layer offer in support of large-scale population modelling in comparison to existing population models?
5. What current limitations are persistent in the final population models produced on the basis of the WSF2019-Imp layer?

To answer these questions the WSF2019-Imp is incorporated with an open archive of sub-national census/estimated-based population data to produce and validate high-resolution population maps for the entire African continent.

- **Research Article 3:** Palacios-Lopez, D., Esch, T., MacManus, K., Marconcini, M., Sorichetta, A., Yetman, G., Zeidler, J., Dech, S., Tatem, A. J., & Reinartz, P. (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. https://doi.org/doi.org/10.3390/rs14020325

The final release within WSF-suite pertaining to this PhD research included the WSF3D dataset. This dataset is integrated by a set of layers depicting the area, height, volume and fraction of the built-up environment at a global scale. Following the premise that volume and land-use information improve the qualitative and quantitative accuracy of population models, the third and fourth main objectives of this PhD research are to develop a novel and fully automatic framework for settlement-use classification solely

based on the WSF3D layer, and to demonstrate how the inclusion of our derived settlement-use layers and volume information improve the accuracy of population models compared with -so far- 2D population models (e.g. all existing large-scale gridded population datasets). The main research questions are formulated as follows:

How effective is the WSF3D for the classification of industrial and non-industrial settlements in the framework of population modelling?

How accurate are population maps that incorporate settlement use and volume information derived solely from the WSF3D?

From these main research questions, the following sub-questions are kept in focus:

1. Can the spatial metrics derived solely from the WSF3D layers be used to classify the built-up environment into industrial and non-industrial areas using a Random Forest (RF) algorithm?
2. What are the classification accuracies delivered by the WSF3D dataset compared to existing methods that rely on Very High Resolution (VHR) data?
3. Are the RF-models produced on the basis of the WSF3D dataset transferable across space?
4. What are the reported accuracies of population maps that integrate volume and settlement use information and how the final accuracy correlates with the quality of the classified maps?
5. How does the WSF3D dataset allow addressing limitations reported in population maps produced on the basis of the WSF2019-Imp, and the WSF2015-Density layer?
6. What advantages does the WSF3D layer offer in support of large-scale population modelling in comparison to existing population models?
7. What current limitations are persistent in the final population models produced on the basis of the WSF3D layer?

To address these questions the WSF3D dataset is incorporated with census/estimate-based national population data to produce high-resolution binary classification maps and population maps at the Pan-European scale.

1.4 Thesis outline

This is a *cumulative* dissertation which is organised as follows:

- **Chapter 1** provides a brief introduction into the scientific relevance of human population data, explaining how large-scale gridded population datasets have emerged as a solution to the increasing demands of spatially explicit, high-resolution and comparable population datasets. Moreover, it presents a short summary of the current limitations of state-of-the-art large-scale gridded population models, describing the key challenges that should be addressed in future research (sub-chapter 1.1). Subsequently, it outlines the main research motivation and goal (sub-chapter 1.2), and finalises with a summary of the research objectives and questions of this PhD thesis (sub-chapter 1.3).

- **Chapter 2** describes important theoretical background regarding the main purpose of gridding population data (sub-chapter 2.1). This is followed by a review of the most employed top-down population disaggregation techniques used in the context of

large-scale population modelling (sub-chapter 2.2), and a brief description of the state-of-art large-scale gridded population datasets (sub-chapter 2.3). It also focuses on the revision of current studies that have employed large-scale gridded population datasets with the aim of identifying the main research gaps and challenges that need to be addressed in the field (sub-chapter 2.4). Findings from this theoretical background were used as the basis for the selection of methods and to outline the objectives of this PhD thesis.

- **Chapter 3** provides an introductory review of the fundamental characteristics of the WSF suite employed in this PhD thesis. It examines the methods, data and validation results reported for each dataset, namely the WSF2015 (binary and density), WSF2019 (binary and imperviousness) and WSF3D.
- **Chapter 4** to **Chapter 6** comprise the core research of the cumulative thesis presented in terms of three stand-alone manuscripts that have been published in international, peer-review journals.
- **Chapter 7** presents a summary of the technical and practical achievements of this thesis and its contributions to the field of large-scale population modelling. Furthermore, it provides an outlook into the remaining limitations and opportunities for the short and long-term future of large-scale population modelling.

Chapter 2

2. Theoretical Background

The following chapter is an extended version of the theoretical literature review introduced in the three peer-review publications in chapters four, five and six. It focuses specifically on introducing the concept of "gridding" population data (sub-chapter 2.1), presenting specifically the main top-down population disaggregation techniques currently employed the field of large-scale top-down gridded population modelling (sub-chapter 2.2), describing the main characteristics of state-of-the-art top-down large-scale gridded population datasets (sub-chapter 2.3) and summarising the remaining limitations and challenges found the field (sub-chapter 2.4). In this context, this chapter does not address the long history of gridded population modelling which dates back to ca. 1936 (Wright, 1936), nor introduces/describes "bottom-up" methods of population distribution (see sub-section 7.2 for more details). For a thorough literature review on these topics, publications by Leyk et al. (2019) , Gregory (2002) and Wardrop et al. (2018) are suggested.

2.1 The concept of "gridding" population data

As outlined in Chapter 1.1, for most countries the traditional form of collecting population data is through national population or housing censuses, where information on the number of people and their main characteristics are collected approximately every 10 years. According to the UN, to this day censuses are the most accurate and rich source of population data, gathering information from the lowest geographical divisions (e.g. household or building level), covering small areas up to the national and international scales (UN-Statistics Division, 2022). However, even when census-based population data provide countries with the most complete demographic information, from an analytical point of view there are a series of limitations that affect their usability and effectiveness for a large number of applications. For example, census-based population datasets are hardly ever released with the same spatio-temporal detail as they were collected. In many countries, before being released to the public, population counts collected at the household or building level are aggregated and linked to larger administrative boundaries (e.g. enumeration areas, blocks, municipalities, districts) to protect the privacy of citizens. This aggregation process on its own, comes with a number of disadvantages. First, the recency, quality, size and number of administrative units used to report population counts vary substantially across and within countries. For example, Figure 2-1 illustrates the number of years since the last reported census used (compared to the year 2022) and the number of administrative units that were used/collected in the production and modelling of the GPWv4.11 (see sub-chapter 2.3 for more details)[1]. As observed, across countries the

[1] The GPWv4.11 is used here simply to exemplify the variability that might exist in terms of the census-year and number of reported administrative units across countries. The data presented corresponds strictly to that collected during the production of the GPWv4.11, which means that for some countries the last conducted census and number of administrative might differ from other official sources.

number of administrative units can vary from as little as 1 administrative unit up to 10,535,212, whereas the last reported census year could be as old as 25 years. A lack of standardisation of this type affects mainly the implementation of robust comparative analyses at local, national and global scales needed for a large number of applications. Second, population data aggregated to large administrative units assume a uniform distribution of the population across space. Here, the information on the true spatial patterns of the population distribution are lost, hiding the real heterogeneity and possible disparities that may exist among areas that report the same population characteristics, which consequently, affect the accuracy of subsequent analyses. Finally, as large administrative boundaries do not normally correlate with other geographical factors (e.g. natural hazards), census-based population datasets cannot be easily integrated with other datasets, limiting their overall usability for a large variety of studies.

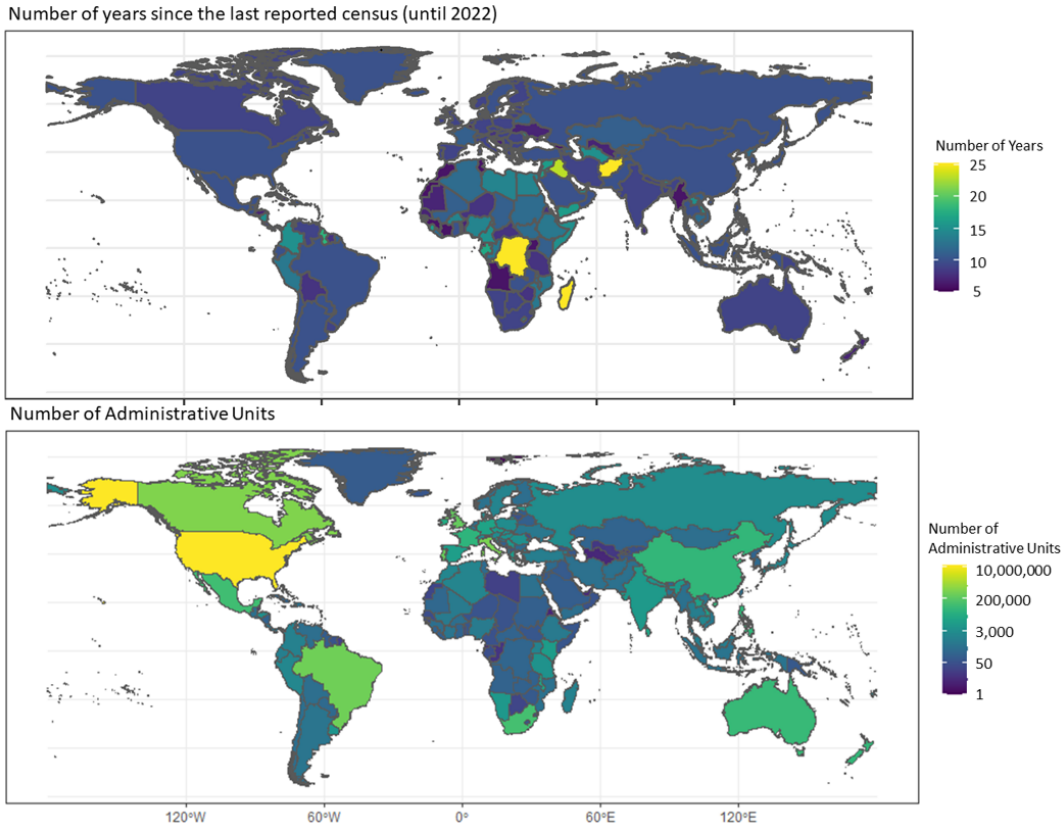

Figure 2-1. The following maps depict the years since the last reported census and available number of administrative units used in the production of the GPWv4.11.

To overcome the limitations of such aggregated and inconsistent data and to better characterise the distribution of populations, much research has been done around the concept of "gridding" population data. Here, the main objective is to produce alternative representations of the population distribution as continuous surfaces, where population counts from the administrative units are transferred to grids of a given spatial resolution (e.g. pixels) using different techniques (Langford, 1991). This gridding process has several advantages: First, in their raster format, gridded population counts can be more easily integrated with other gridded data such as environmental geospatial datasets, allowing a deeper understanding of human-environment interrelationships. Second, gridded population counts can be aggregated to arbitrary spatial units, including hazard zones,

climate zones, risk areas, etc., facilitating spatial and statistical analyses. And third, gridded population datasets provide consistent and comparable data across space, allowing implementing cross-comparison analyses within and across regions.

2.2 Top-down population disaggregation approaches

Before the development of the first global gridded population dataset in the mid-1990s (Tobler et al., 1997) many top-down methods to disaggregate population counts from spatially-coarse source units (e.g. polygons of irregular shape) into spatially-fine target units (e.g. a regular grid), had already been proposed in the literature, focusing on the production of local-level population grids (Goodchild et al., 1993; Langford, 1991; Wright, 1936). In the field of large-scale gridded population modelling, however, only two main approaches or techniques have largely dominated the field, namely areal-weighting and dasymetric modelling. These are presented in Figure 2-2, and described as follows:

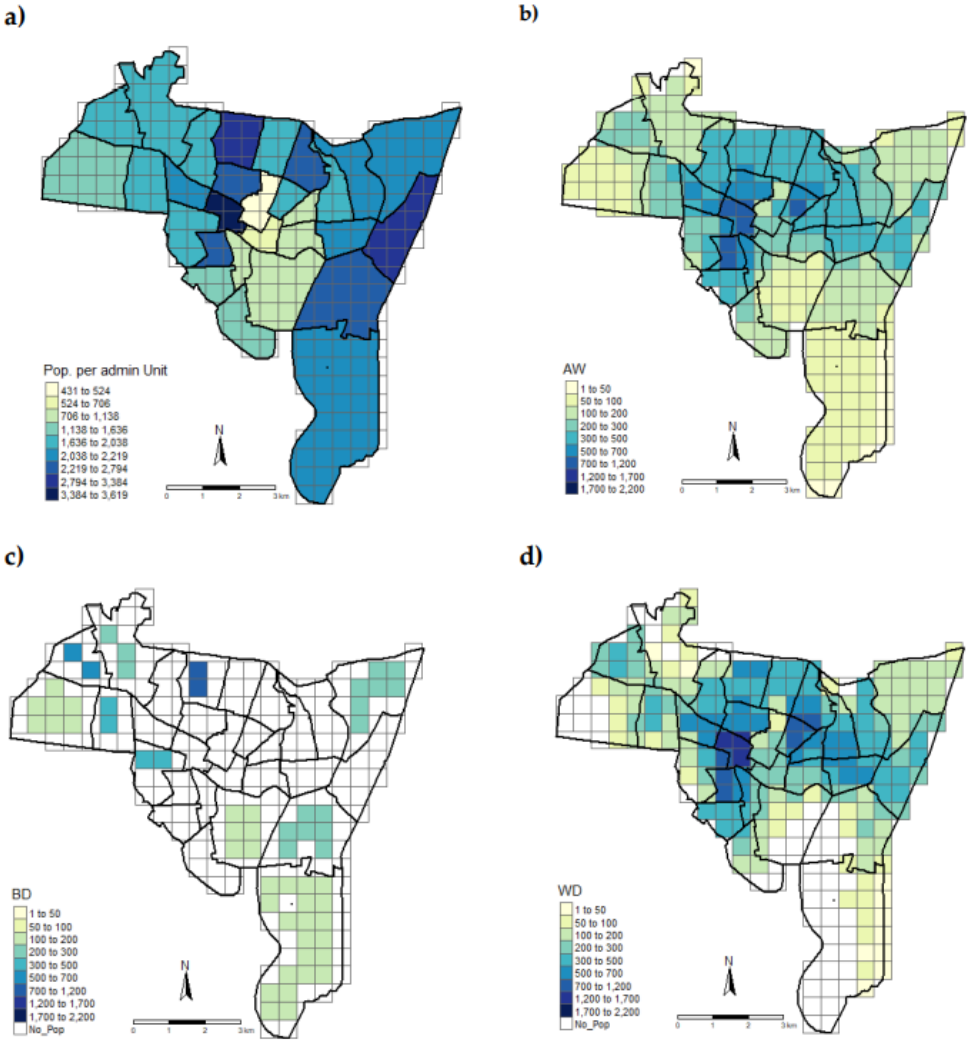

Figure 2-2. Schematic representation of the different top-down population disaggregation techniques: **a)** example of input census-based population data for the area of New Haven, USA with a 500m grid overlay, **b)** AW: areal-weighting output, **c)** BD: binary dasymetric output restricted by e.g. built-up areas and **d)** WD:

weighted dasymetric output considering e.g. land-use densities. Figures produced using the open code published by Comber and Zeng (2019).

2.2.1 *Areal-weighting*

Areal weighting is one of the simplest modelling techniques in which population counts from the source units (e.g. administrative units) are spatially reallocated into target units (e.g. pixels) using as weight the proportion of the area that overlaps between the two units (Goodchild & Lam, 1980), as illustrated in Figure 2-2(b). Population models that employ areal-weighting techniques produce a volume-preserving, homogeneous or proportional reallocation of population counts. This means that all the target units within a source unit allocate the same number of people, and the sum of population counts from all target units adds-up to the original population total of the source unit.

According to multiple researches, this disaggregation method reports two main advantages. On the one hand, it is easy to operate and has high calculation efficiency. On the other hand, as it does not rely on any additional data (e.g. other geospatial datasets), the output population datasets do not suffer from endogeneity problems. This means that in terms of applicability restrictions, the final population datasets can be integrated with any other geospatial datasets, without limitations or complex uncertainties (Balk et al., 2006). At the same time, as this method is only based on the geometrical properties of the source and target units, the accuracy of the output population datasets produced through this method is only linked to the accuracy and spatial resolution of the input population data (Doxsey-Whitfield et al., 2015; Hallisey et al., 2017; Sadahiro, 2000; Thomson et al., 2021a).

However, even when areal weighting is a straightforward method its main limitation is the implicit assumption of a homogenous population distribution within each source unit, which is rarely true in the real world. The lack of spatial patterns together with the strong discontinuities between administrative boundaries, has also shown to affect both the qualitative and quantitative accuracy of the final population datasets, limiting their usability for subsequent analyses (Fisher & Langford, 1996). Here, efforts to smooth transitions between administrative boundaries using a *pycnophylatic interpolation* algorithms (Tobler, 1979), for example, have been employed as a post-processing solution. However, these type of methods have not been largely adopted, as they do not draw on information about real population distribution (Kim & Yao, 2010).

2.2.2 *Dasymetric modelling*

To respond to the limitations of the areal-weighting technique, dasymetric modelling is a disaggregation technique that refines population distributions by employing a "restrictive" and/or "probability" layer (hereinafter referred to as weighting layer), that defines the amount of population counts that need to be allocated in each grid cell within a source unit. This weighting layer is derived from single or multiple ancillary datasets (often referred to as geospatial covariates or proxy layers), that are presumably related to population presence and densities (Goodchild et al., 1993; Langford, 1991; Mennis, 2003). Depending on the method used to derive this weighting layer, dasymetric modelling techniques can range from binary-dasymetric, to more complex weighted-dasymetric techniques.

Binary dasymetric techniques employ one or more proxy layers that simply restrict the redistribution of population to a limited set of areas inside the source units. The most commonly employed proxies include mainly binary datasets representing built-areas, building footprints, restricted or protected areas, and water bodies datasets (Langford & Unwin, 1994). As exemplified in Figure 2-2c, population models that rely on binary dasymetric techniques produce a volume-preserving, homogenous or proportional reallocation of population counts, in which all the target units within a source unit allocate the same number of people.

Compared with the areal-weighting techniques, binary-dasymetric techniques are also simple to implement but suffer from the quantitative and qualitative restrictions of delivering a proportional allocation of population counts. For example, in any given population model that relies only on binary built-area layers (e.g. built-up vs non-built-up) to redistribute populations, it is assumed that urban areas and rural areas allocate the same proportion of the population, which is generally inaccurate. This leads to either great errors of over— and underestimation in the final population grids, which overall affect their usability for subsequent analyses. At the same time, the reallocation accuracy of these population models is highly dependent on the spatial resolution and quality of the employed built-area layers. Here, the proper identification and classification of "populated" areas, in particular, is of paramount importance, as any misclassification resulting in predicting no population in a particular area may be quite undesirable for many applications (e.g. emergency responses) (Stevens et al., 2020).

Weighted dasymetric techniques rely on one or more proxy layers to produce a probability scheme that determines the amount of population to be allocated in each grid cell within a source unit (Su et al., 2010). These weights represent a measurement of the presumed relationships that might exist between the amount of population that needs to be allocated and the geographical factors represented by each proxy. Depending on the simplicity of the relationships, the weights needed to redistribute populations can be directly derived from the geospatial covariates (e.g. using the percent of built-up density), or can be derived through other more complex methods like empirical sampling (Mennis, 2003), regression analyses (Mennis & Hultgren, 2006), machine learning or deep learning approaches (Stevens et al., 2015b). In case of the latter, weighted-dasymetric techniques are commonly referred to as *intelligent dasymetric techniques*, in which the most commonly employed geospatial proxies include a combination of built-up density layers, urban/rural extents, topographic layers, climatic factors, environmental datasets, land-use and land-cover datasets, infrastructure data (e.g. roads, points of interest, transportation network) and night-time lights imagery.

As seen from Figure 2-2d, population models that use a weighted dasymetric technique produce a volume-preserving, heterogenous reallocation of population counts, in which the target units allocate different amounts of population. From a comparative point of view, this can be considered as one of their main advantages, as not only the spatial distribution of the population adheres more to the reality, but the accuracy of population estimates reported in the final population grids has also shown to be more accurate that those produced by areal-weighting and binary dasymetric techniques (Mennis & Hultgren, 2006; Palacios-Lopez et al., 2019; Palacios-Lopez et al., 2021; Stevens et al., 2015b; Su et al., 2010). In this context, however, it is important to mention that one

of the main limitations of weighted dasymetric models is that it is frequently challenging to acquire geospatial covariates with the same quality in terms of spatial resolution, and temporal—and spatial coverage. Furthermore, it has been found that the use of multiple proxy layers can reduce the applicability of the final population grids, due to the increased probability of endogeneity issues (Nagle et al., 2014).

2.3 State-of-art top-down large-scale gridded population datasets

Today, there are six state-of-the-art, top-down large-scale gridded population datasets that are commonly used in academic research, and which have been produced to support governments, organisations and institutions around the world (Allen et al., 2021; Freire et al., 2018). These datasets include[2]:

1. The Gridded Population of the World, version 4 (GPWv4.11) produced by CIESIN, Columbia University (Doxsey-Whitfield et al., 2015).
2. The Global Rural-Urban Mapping Project (GRUPM) produced by CIESIN (CIESIN, 2011).
3. The Global Human Settlement Population layer (GHS-POP; R2015A and R2019A) produced by the European Commission Joint Research Centre (EC-JRC) in collaboration with CIESIN (Freire et al., 2016).
4. The High-Resolution Settlement Layer (HRSL) produced by Facebook Connectivity Lab, in collaboration with CIESIN (Tiecke et al., 2017).
5. The LandScan dataset produced by the Oak Ridge National Laboratory (ORNL) (Bhaduri et al., 2007; Dobson et al., 2000) .
6. The WorldPop datasets produced by the WorldPop project; University of Southampton (Stevens et al., 2015b).

The primary characteristics of the latest versions of these datasets, including the population model employed to produce them (input data + disaggregation technique), their spatio-temporal resolution and population concept are presented in Table 2-1. These can be briefly summarised as follows:

GPW4.11

The GPW4.11 population datasets are the only grids that are produced using an area-weighting technique, relying simply on a water mask to ensure that population counts are only assigned to land pixels. The final datasets represent the residential population (e.g. people counted at their place of living) either as population counts (people per pixel) or population density (people per km^2) for the years 2000,2005,2010,2015 and 2020, respectively. The datasets are available at a global scale and are produced at a spatial resolution of 30 arc-seconds, which corresponds to ~1km at the Equator. The datasets are published and made available in WGS1984 geographic coordinate system in ASCII, GeoTiff and NetCDF formats.

[2] The large-scale top-down gridded population datasets presented here are constantly evolving and different versions have existed through time. The thesis describes those which were available at the time of writing.

GRUMP

The GRUMP datasets are based on population data collected from the GPW, version 3. They are produced using a binary dasymetric technique to allocate population according to rural or urban gradients, derived—in part—on night-light imagery. The final datasets represent the residential population, either as population counts or population densities for the years 1990, 1995 and 2000, respectively. The datasets are available at a global scale and are produced at a spatial resolution of 30 arc-seconds, which corresponds to ~1km at the Equator. The datasets are published and made available in WGS1984 geographic coordinate system in ASCII, BIL and GRID formats.

GHS-POP (R2015A, R2019A)

The GHS-POP datasets are produced using a density weighted dasymetric technique that relies on the distribution of population counts from administrative units into settlement pixels describing built-up density, as defined by the GHS- BUILT datasets (R2015B, R2018B) (Pesaresi et al., 2016; Pesaresi et al., 2013) (see Table 2-2). The population data are reallocated in one of three ways: if the administrative area is large enough to generate 250 m grids and contains built-up areas, then the population for that administrative area is assigned in proportion to the density of the built-up areas. If the administrative area is large enough to generate 250 m grids but does not contain any built-up areas, then the population is allocated using an area-weighting technique. If a cell is located on the border of an administrative area, it is assigned to the administrative area its centroid falls in. And finally, if the administrative area is smaller than a 250-m grid cell, then a centroid is generated for the area and the population of all centroids found within a cell is added (Archila Bustos et al., 2020). The final datasets represent the residential population, either as population counts or population densities for the years 1975, 1990, 2000 and 2015, respectively. The datasets are available at a global scale and are produced at a spatial resolution of 250m and 1km at the Equator. The datasets are published and made available in a World Mollweide projection in GeoTiff format.

HRSL

The HRSL is produced using a binary dasymetric technique that redistributes population from administrative units to built-up areas as defined by proprietary settlement layer produced using high resolution (0.5m at the Equator) satellite imagery from Digital Globe (Tiecke et al., 2017). The final datasets represent the residential population, people per pixel, for 2015. The datasets are currently available for 140 countries and are produced at a spatial resolution of 1 arc-second, which corresponds to ~30 at the Equator. The datasets are published and made available in WGS1984 geographic coordinate system in GeoTiff format.

LandScan

The LandScan population datasets are produced using a weighted intelligent dasymetric technique that consists of dynamically adaptable algorithms used to generate a weighting layer based on statistically-derived relationships among multiple proxy layers. As seen from Table 2-2, some of the most commonly employed geodatasets are available from open and free sources, however, it is known that other commercial and local data are also employed, especially to manually fine-tune the accuracy of the grids.

So far, the methods employed to produce the LandScan datasets are not publicly available, and the population grids are only free for researchers and students. The final datasets represent the ambient population (e.g. population at the place of work) as population people per pixel, for the years 2000-2018. The datasets are available at a global scale and are produced at a spatial resolution of 30 arc-seconds, which corresponds to ~1km at the Equator. The datasets are made available in WGS1984 geographic coordinate system, in GRID and binary raster formats.

WorldPop

The WorldPop population datasets are produced using a weighted intelligent dasymetric technique that consists of locally modelled RF algorithms used to generate a weighting layer based on the relationships between population densities and multiple proxy layers (see Table 2-2). Here, redistributions are done in two ways, first where population counts are redistributed to all grid cells or pixels (unconstrained), and second, where population counts are redistributed only within areas identified as settlements (constrained). In the case of the latter, different built-area layers are employed depending on the location. For Africa, for example, satellite-derived building footprint data from Maxar/Ecopia (Maxar Technologies, 2020) are used, whereas for the rest of the countries a novel built settlement growth model is employed (Nieves et al., 2020a; Nieves et al., 2020b), derived from other built-area layers such as the GUF (Esch et al., 2018a; Esch et al., 2017), the GHSL and the European Space Agency (ESA) CCI land cover 300m (ESA, 2015) (see Table 2-2). The final datasets represent the residential population as people per pixel, for the years 2000-2020. The datasets are available at a global scale and are produced at a spatial resolution of 3 arc-seconds, which corresponds to ~100m at the Equator. The datasets are made available in WGS1984 geographic coordinate system in GeoTiff format.

Input data: population data and geospatial proxies

According to the information presented in Table 2-2, each gridded population dataset is produced using different sources of input population data and proxy layers. For example, most datasets employ as input population data, population totals adjusted to the UN-Population Division (UNPD) estimates and projections produced by CIESIN. The only exception are the Land-Scan datasets, which are produced using United States (USA) Census global population estimates, respectively.

Concisely, CIESIN collected census data at the highest spatial detail available from the results of the 2010 round of Population and Housing Censuses, which occurred between 2005 and 2014. CIESIN data include two types of population estimates: census-based and UN-adjusted, both estimated for the years 2000, 2005, 2010, 2015 and 2020. Initial population estimates were derived for each administrative unit by means of an exponential model fitted on at least two census counts for each country (Doxsey-Whitfield et al., 2015). However, to allow for global comparisons, CIESIN adjusted the census counts to the target year of 2010, which were then then interpolated and extrapolated to produce the UN-adjusted estimates with the objective to correct for over- or under estimations (CIESIN, 2018b; Doxsey-Whitfield et al., 2015). For the vector data or boundaries, the global administrative areas version 2 (GAMv2) was used to ensure consistent alignment between countries. For more details in the production of the CIESIN database, the following literature is suggested: Doxsey-Whitfield et al. (2015); (Freire et al., 2018) .

Finally, in terms of the most commonly employed openly-available proxies, each dataset uses either single proprietary layers (e.g. GHSL by GHS-POP or Digital Global-based settlement layer by HRSL), or a combination of multiple remotely-sensed datasets, including land-cover/land-use layers, topographic data, night-time-imagery and OSM data. Table 2-2 presents a list of datasets that are available at large-scales, however, other proxy layers which are available at country-by-country basis are also employed, especially in the production of the WorldPop and LandScan datasets. As explained in the previous paragraphs, each population model used to produce each one the gridded population dataset will process these geospatial proxies in different ways, resulting in varied outputs, especially at the local level (Archila Bustos et al., 2020; Chen et al., 2020). For a visual assessment, Figure 2-3 compares all of datasets for a small area near Puerto Vallarta, Mexico.

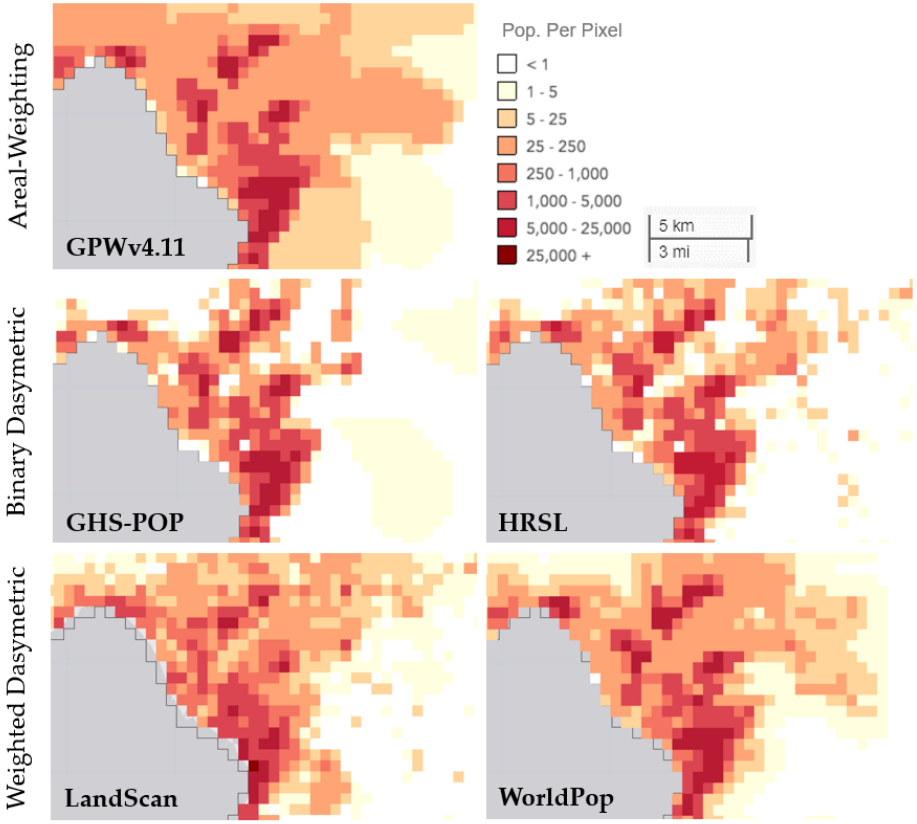

Figure 2-3. Visual comparison of five large-scale gridded population datasets for an area close to Puerto Vallarta, Mexico. Each dataset has been resampled to a 1km by 1km grid. Values represent population per pixel for the year 2015. Images have been produced using the PopGRID Viewer available at: https://sedac.ciesin.columbia.edu/mapping/popgrid/comparison-view/, from which the GRUMP dataset is not available for visualisation.

Table 2-1. Main characteristics of state-of-the-art gridded population datasets. BD: Binary Dasymetric. WD: Weighted Dasymetric, WID: Weighted Intelligent Dasymetric.

Gridded Population Dataset	Producer	Population Model										Spatial Resolution	Temporal Resolution	Coverage	Population Concept
		Main Data Inputs									Technique				
		Population Data	Roads	Land cover	Built-areas	Urban/rural extents	Nighttime lights	Environmental Data	Protected Areas	Water Bodies					
GPWv4.11	CIESIN	x							x	x	AW	1 km	2000;2005;2010; 2015;2020	Global	Residential
GRUMP	CIESIN	x				x			x	x	WD	1 km	1990;1995;2000	Global	Residential
GHS-POP	EC-JRC and CIESIN	x			x						WD	250m 1km	1975;1990;2000;2015	Global	Residential
HRSL	Facebook Connectivity Lab; CIESIN	x			x						BD	30m	2015;2018	140 countries	Residential
LandScan	ORNL	x	x	x	x	x	x	x	x	x	WID based on statistical modelling	1 km	annual releases 2000-2020	Global	Ambient
WorldPop (constrained & unconstrained)	WorldPop; University of Southampton	x	x	x	x	x	x	x	x	x	WID based on RF	100 m	2000-2020	Global	Residential

Table 2-2: Most commonly employed proxy layers used in the production of state-of-the-art large-scale gridded population datasets, available at large-scales.

Proxy layers	Used by	Proxy layers	Used by
Population Data		**Nighttime lights**	
CIESIN-UN Adjusted	GPWv4.11; GRUMP; GHS-POP; HRSL; WorldPop	Defence Meteorological Program-Operational Line-Scan System (DMSP-OLS)	WorldPop
Country's best available	LandScan	Visible Infra-red Imaging Radiometer Suite (VIIRS)	WorldPop
Built-area layers or Building Footprints		**Land-Cover**	
Global Human Settlement Layer (GHSL R2015B, R2018A)	GHS-POP WorldPop	National Land Cover Database (NLCD)	LandScan
Digital Global-based settlement layer	HRSL	Digital Chart of the World (DCW)-Landcover	LandScan
Global l Urban Footprint (GUF)	WorldPop	500m Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover	LandScan
Built-Settlement Growth Model	WorldPop-2020 Constrained	Global Land Cover Characterisation (GLCC)	LandScan
Ecopia-Maxar Building Footprints	WorldPop-2020 Africa	ESA-Climate Change Initiative (CCI-300m) Land Cover	WorldPop
Elevation		**OpenStreetMap (OSM) Data**	
Viewfinder Panoramas - Shuttle Radar Topography Mission (SRTM)	WorldPop	Infrastructure, POI's, Transportation Network	WorldPop; LandScan

2.4 Current limitations in the field of top-down large-scale population modelling

Over the last decade, the field of large-scale gridded population modelling has seen several advances that have allowed producing population datasets with increased spatial resolution and improved qualitative and quantitative accuracy. According to the information presented in the review of Leyk et al. (2019), the most important advances influencing the field include:

1. The increased availability of more accurate, updated, and spatially refined census-based population data for many countries.
2. The increased availability of spatially refined remotely sensed satellite imagery needed to derive proxy layers for disaggregation.

3. The increased computing power which has allowed scaling the production of proxies, and thus, population datasets from local to global scales.
4. The development of more sophisticated methods to a) extract and classify populated land (e.g. development of accurate built-area and building footprints datasets), and b) estimate population distributions which, nowadays, include machine– and deep learning algorithms.

However, even when the latest versions of existing state-of-the-art products have leveraged these recent developments and do report drastic qualitative and quantitative improvements , most datasets still report extreme limitations and estimation errors that affect the accuracy and precision of subsequent analyses. These limitations are mainly still derived from issues in the quality of the input census-based population data (e.g. recency, completeness, reliability), but also from the shortcomings of the input proxy layers used for disaggregation and the employed modelling methods. In this context, focusing only on the limitations that are strictly derived from the employed modelling frameworks (e.g. areal-weighting, binary or weighted-dasymetric techniques), as well as from the qualitative and quantitative characteristics of the proxy layers used for disaggregation (e.g. scale, accuracy, thematic representation)[3], some of the most noticeable challenges documented in contemporary studies are presented in the following paragraphs.

In a general assessment done to evaluate the accuracy of different global built-area layers for large-scale population modelling (e.g. GUF, GHSL, 500m MODIS Land Cover and ESA CCI land cover 300m), Stevens et al. (2020) and Reed et al. (2018) demonstrated that population models that only employ built-area layers for disaggregation (e.g. based on binary techniques) or not at all (e.g. based on areal-weighting), are less accurate than those that combine these proxies with other geospatial layers (e.g. based on weighted techniques) . The authors argue, that while the integration of these proxy layers is crucial to produce accurate population distributions, the use of single binary layers leads to qualitatively less detailed population distributions on the one hand, and less quantitatively accurate estimates in rural areas, on the other, as most built-area layers used today fail to identify small settlements.

Contrastingly, in the studies presented by Schug et al. (2021) and Balk et al. (2006), the authors argue that weighted-approaches that rely on multiple proxy layers (e.g. WorldPop and LandScan models) suffer for quality biases that are introduced by inconsistencies on the input data and modelling frameworks. Schug et al. (2021) and Nagle et al. (2014) add to these conclusions, expressing that the physical relationships between population and multiple ancillary datasets are hard to quantify when multiple layers are employed, and that weighted layers derived from "intelligent techniques" (e.g. RF-based) can be regionally specific leading to differences across space. Balk et al. (2006) additionally argue that the use of multiple layers for population modelling can lead endogeneity issues, and that overall, the collection of multiple proxy layers with large-extent coverage and

[3] Limitation or errors in large-scale gridded population dataset are also linked to the quality of the input population data, however, in the framework of their validation the input population data is normally considered "accurate" as independent population data to validate their accuracy is normally not available. A broader discussion of this is presented in the subsequent sections.

spatio-temporal agreement is an exhaustive task that makes producing frequent updates infeasible.

While the aforementioned studies focused more on comparing the underlying population models employed to produce large-scale gridded population datasets, the conclusions reached by these authors have been reinforced in other more practical or applied studies. For example, in the research presented by Doll and Pachauri (2010) , the authors show that the equal or proportional distribution of population produced by the GPW4.11 and GRUMP datasets is sub-optimal to quantify rural population without access to electricity. The author showed that due to the equal distribution (e.g. proportional allocation), the density of population that should be found in highly luminous areas was greatly underestimated by the datasets.

Accordingly, in an assessment carried out to evaluate the accuracy of population counts in slums and deprived areas in Kenya (Nairobi) and Nigeria (Lagos and Port Harcourt), the authors of Thomson et al. (2021a) reported that different large-scale gridded population datasets, including the GPWv4.11, the GHS-POP, the HRSL, the WorldPop (constrained & unconstrained) and the LandScan datasets, respectively, vastly underestimate the total populations, with the most severe errors reported in the most populous and densest slums. As explained by the authors, reasons for this underperformance is attributed to the use of binary proxy layers which limits the highest population value that can be assigned to a cell (e.g. the homogeneous distributions of GPWv4.11, HRSL and GHS-POP), and the poor detection of slum areas (e.g. omission of settlements), which coupled with the lack of information on settlement use, building heights and building densities, produces underestimations in these highly dense areas.

De Mattos et al. (2020) add on the same topic, where the authors report that the coarse spatial resolution of the LandScan datasets (e.g. 1 km) affected the extraction of accurate populations living in slums in a selected area of Brazil. Here, the WorldPop datasets with their 100m spatial resolution reported less drastic underestimations; however, they still produced some critical errors due to the omission of built structures in slums.

Comparably, in the research presented by Smith et al. (2019), it was demonstrated that the homogeneous distribution of the GHS-POP and HRSL datasets was insufficient for analyses aimed at extracting populations at risk of flood-hazards. The authors argue that the restriction of allocating the same number of people per grid does not only affect the quantitative estimations of people at risk, but also produces unrealistic distributions of the populations around different river basins. In the same study, the authors also argue that the coarse spatial resolution of the WorldPop and LandScan datasets (e.g. 100m and 1km, respectively) restricts the integration of the population data with other high-resolution datasets such as flood hazard data available at 90m at the Equator. The authors show how coarsening the resolution of the hazard data to match that of the population data can lead to critically overestimated population counts, and suggest not to pre-process the data but rather to found solutions to improve the spatial resolution of existing gridded population datasets.

The same conclusions were reached in the studies presented by Calka and Bielecka (2019) and Calka and Bielecka (2020), where the authors demonstrated that the coarse

spatial resolution of the GHS-POP and LandScan datasets (e.g. 250m and 1 km) produces erroneous population counts, that become larger as the size of spatial unit used for analysis becomes smaller. Overall, the authors reported that for densely populated regions the GHS-POP underestimates the population numbers, while for thinly populated regions it overestimates.

Finally, moving from the limitations derived from the spatial resolution, accuracy of the proxy layers (e.g. identification of settlements) and the employed modelling framework , another major limitation affecting all existing large-scale population models is that none of them integrate proxy layers that provide information on the built-up environment in terms of use (e.g. residential or non-residential) and heights of buildings (Schug et al., 2021). Inclusion of this type of information has shown to increase the accuracy of population models at local-scales, but the methods that are employed at this scale are still not transferable to large-scale models. Hence, in current large-scale population datasets, the lack of settlement use information has led to large overestimation errors in industrial and commercial centres on the one hand, while the lack of height information has led to large underestimation errors in high-rise building areas, on the other (Huang et al., 2021b; Thomson et al., 2021a; Thomson et al., 2021b).

To consolidate the aforementioned information, Table 2-3 presents a summary of the current limitations affecting large-scale gridded population models. Specific focus is placed on models based on dasymetric modelling techniques, to reflect on the limitations derived both from the technique as well as from the shortcomings of the employed proxy layers.

Table 2-3. Summary of the limitations affecting large-scale gridded population models based on dasymetric modelling techniques.

Limitation	Affected large-scale population models	
	Binary-Dasymetric	Weighted-Dasymetric
Low spatial resolution	X	X
Omission and commission of settlement areas	X	X*
Homogeneous representations of population distributions	X	
Endogeneity issues		X
Difficult transferability and replicability		X
Quality inconsistency across-space		X
Exclusion of building use and building height information	X	X
Bad quality of the input-population data (e.g. recency)	X	X

*For multi-layer models that use built-area layers to constrain population distributions.

Chapter 3

3. The World Settlement Footprint suite

Improving the quality and accuracy of geospatial datasets describing the built-up environment at global scales has important scientific applications. One such application is estimating where and in what density humans live across the world, information –that on its own– is of great value for many research fields (Stevens et al., 2020). As such, over the last two decades, mapping the built-up environment with unprecedented spatial detail and accuracy has been facilitated by the increasing availability of free and open, high-resolution remotely sensed imagery and the continuous development of image processing methods. Here, the advantages made through time have led to the proliferation of many global (to near-global) built-up area datasets which have evolved from low resolution (1 km -500m at the Equator) to medium resolution (100m at the Equator) to high resolution (30m to 10m at the Equator); and from which the most representative ones have been used to refine gridded population datasets at near-global extents. As described in Table 2-2, some of the most widely employed built-area datasets, used in field of large-scale population modelling include the GUF (Esch et al., 2018a; Esch et al., 2017), the GHSL (Pesaresi et al., 2016; Pesaresi et al., 2013), the Digital Global-based settlement layer (Tiecke et al., 2017), the WorldPop growth built-up models (Nieves et al., 2020b) (Nieves et al., 2020a) and the Ecopia/Maxar (Maxar Technologies, 2020).

Concisely, the particular focus placed on built-area datasets for population modelling arises from the fact that this type of datasets have frequently proven to be stronger predictors of population inhabitation in comparison with other geospatial layers such as land-cover, elevation, slope and nightlight imagery (Linard et al., 2011; Nieves et al., 2017; Reed et al., 2018; Stevens et al., 2020; Tatem et al., 2007). Different research has demonstrated that when built-area datasets are used to model/restrict the distribution of population, the final products deliver better qualitative and quantitative results in comparison to those models where the datasets were not included (Reed et al., 2018; Rubinyi et al., 2021; WorldPop, 2020). More recently, it has been shown that when a given built-area dataset is accurate and coherent enough with population densities, it has the potential to be used as a single proxy for population modelling, overcoming some of the limitations of simple areal-weighting or multivariate techniques (Stevens et al., 2020)

However, despite the emergence of more accurate and detailed built-area layers, existing state-of-the-art large-scale gridded population models still suffer from qualitative and quantitative limitations derived in part from the inaccuracy of the geospatial datasets used to distribute population across space (see sub-chapter 2.4). The most prominent limitations related to the currently employed built-area layers (in either binary- or weighted-dasymetric models) include their inability to map fine-scale population distributions from their coarse resolution (e.g. 500m MODIS Land Cover and ESA CCI land cover 300m), the poor identification of human settlements in rural settings (or

misclassification errors in general), and the lack of integration of settlement use and settlement 3D information, respectively (Archila Bustos et al., 2020; Schug et al., 2021) .

On the one hand, as stated by Marconcini et al. (2020), the consistent low classification accuracy reported in existing built-area products is linked to their employed processing frameworks which, –so far–, still rely on the exclusive use of either optical (e.g. GHLS and HRSL) or radar imagery (e.g. GUF). This has led to errors of commission and omission, first because the spectral information in optical images tend to confuse built-areas with other bare land classes (e.g. in arid and semi-arid regions, second, because extreme topologies produce high backscattering comparable with built-areas, and third, because the resolution of currently employed optical imagery, is not enough to identify small settlement located in cliffs, valleys or complex topographies (e.g. Landsat-8 30m at the Equator).

On the other hand, if information on building use/type has not yet been integrated into modern large-scale population models because appropriate data to derive this information at/for large extents (e.g., national, continental, global) does not exist. Currently, contemporary research that focuses on extracting use and volumetric semantic information of built-up structures employ a combination of regionalized building footprints, cadastral data, LiDAR data, social media data, aerial imagery and/or commercial (and frequently expensive) very high-resolution imagery (e.g. < 5m optical data or orthoimage) (Du et al., 2015; Jochem et al., 2021; Lloyd et al., 2020; Ma et al., 2015; Stéphane et al., 2020; Zhang et al., 2017a), which restricts the implementation of the developed methods to the specific areas where these data are available, reliable, replicable and - more importantly - complete. Similarly, efforts to derive building heights at national or continental scales are either limited to specific regions (e.g. mainly Europe, North America or Asia), or their spatial resolutions are still quite coarse (>100m at the Equator) (Falcone, 2016; Frantz et al., 2021; Li et al., 2020).

To overcome these limitations, the German Aerospace Centre (DLR) in collaboration with the ESA and the Google Earth Engine (GEE) team has been working on the development and open-release of the WSF suite, which includes a set of high-resolution datasets describing the extent, location, PIS and 3D characteristics of the built-up environment at global scales. This product represents a follow-on development to DLRs' previous global built-area dataset –the GUF– developed by Esch et al. (2013). Currently, the WSF suite is composed of three main layers: the WSF2015, the WSF2019 and the WSF3D dataset, respectively. Each of these layers has been developed using novel and robust methodologies that jointly exploit, – for the first time ever–, open and free multi-temporal optical and radar data.

The following sub-chapters present a brief summary of the characteristics of each WSF layer used in this PhD research. The author of this thesis actively contributed to the qualitative and quantitative validation of the following layers, as well as to the preparation of their respective peer-review papers and delivered reports.

3.1 The WSF2015 and WSF2015-Density Layer

- **Main Reference:** Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., & Kakarla, A. (2020). Outlining where humans live--The World Settlement Footprint 2015. *Sci Data*, *7*(242). https://doi.org/10.1038/s41597-020-00580-5

The WSF2015 is a binary mask that describes the extent and location of human settlements at global scale at an unprecedented spatial resolution of 10m at the Equator for the year 2015. It was processed using a novel and robust methodology that relied on multi-temporal statistics extracted from ~107,000 and ~217,000, 2014-2015 Sentinel-1 (S1) and Landsat-8 scenes, to produce a binary classification of settlement and non-settlements based on an advanced machine learning approach.

The main rationale followed in the production of the WSF2015 is rooted in the assumption that the temporal dynamics of human settlements, compared with other non-settlement classes, remain constant over time. This means that over time, the spectral and backscatter characteristics that differentiate built-up structures from other features are consistent over time, allowing a proper classification of these structures on the ground. Following this premise, the processing framework of the WSF2015 can be divided in four main steps:

First, for a selected target region of interest, S1 and Landsat-8 images were acquired for a period of ~1 year, from which key temporal statistics were extracted, forming two separate feature stacks, respectively. From the S1 images, five temporal statistics were extracted, including the minimum, maximum, mean, standard deviation and mean slope of the backscattering values. The coefficient of variation (COV) of the temporal mean backscattering and the total and the number of available scenes per pixel were also calculated, resulting in a 7-feature stack. From the Landsat-8 imagery a set of indices were extracted including the Normalised Difference Build-Up Index (NDBI), the Modified Normalised Difference Water Index (MNDWI), the Normalised Difference Vegetation Index (NDVI), the Normalised Difference Middle Infrared (NDMIR), the Normalised Difference Red Blue (NDRB) and the Normalised Difference Green Blue (NDGB). Accordingly, for each of the 6 indices, the same 5 temporal statistics used for the S1-imagery were also extracted, including the COV of each of the derived 6 temporal mean indices. This led to a final 37-feature stack.

Second, training and label data for the settlement and non-settlement classes were generated by setting thresholds to 3/44 features based on an extensive empirical analysis against Google Earth VHR imagery, carried over more than 450 tiles of 1x1 degree. Concisely, a given point sample x, would be classified as either settlement or non-settlement if it satisfied a number of conditions within these thresholds. To compensate for the variations derived from different climate zones, the thresholds for each class were fine-tuned in relation to the 30 climate types of the Köppen Geiger scheme (Peel et al., 2007).

Third, using the training data, a binary classification based on Support Vector Machine (SVM) with Radial Basis Function (RBF) Gaussian Kernel was separately applied to the optical and radar-based feature stacks to classify the remaining scenes. Considering

that a large number of training points exist, a total of 1000 random points are used per every 1x1 degree tile. This operation was repeated 20 times in an ensemble of SVM classifiers, from which the majority vote is used to define the final class.

Fourth, as optical and radar stacks were processed separately, a post-processing phase was then applied to properly combine the Landsat 8– and S1-based classification maps, and automatically identifying and deleting false alarms.

The validation of the WSF2015 was performed using a thorough protocol that consisted in collecting reference data for ~900,000 sample points using Google Earth VHR satellite/areal imagery. This was done through visual assessment and an established crowdsourcing collaboration between Google and DLR. The results of this validation campaign indicated that in comparison with the GUF, the GHSL (R2015) and the GLC30 layers, respectively, the WSF2015 exhibited the best percentage average accuracy; reaching a value of 86.37%, which represented a mean increase of +6.24%, +15.28% and +18.58% over the rest of the layers. Alongside, it resulted in an average Kappa coefficient of 0.68, which represented an increase of +0.07, 0.23 and 0.29 over the GUF, GHLS and GLC30, respectively.

The WSF2015-Density is one of the first experimental developments of the WSF suite and service portfolio, aiming at enhancing the semantic and thematic scope of the WSF2015; in particular, the layer describes the PIS within areas categorised as settlements in the WSF2015. Effectively mapping the PIS is of high importance to assess—among others—the risk of urban floods, the urban heat island phenomenon as well as the reduction of ecological productivity. Furthermore, it is generally considered as an effective proxy for the housing density, thus making it particularly suitable for supporting spatial population distribution (Azar et al., 2010; Li & Weng, 2005; Lu et al., 2006). The current processing methodology follows the approach originally described by Marconcini et al. (2015) and is based on the assumption that a strong inverse relation exists between vegetation and impervious surfaces (i.e., the higher the presence of vegetation is, the lower the corresponding imperviousness is). Accordingly, the core idea is to compute and analyse for each pixel the temporal maximum of the Normalised Difference Vegetation Index (NDVI), which depicts the status at the peak of the phenological cycle. To this purpose, the NDVI available from the TimeScan dataset (Esch et al., 2018a; Esch et al., 2018b) has been used, which has been derived globally from Landsat-8 scenes acquired during 2014–2015. Figure 3-1 shows different subsets of the WSF2015 binary and Density layers for the cities of Hai Phong, Vietnam, Abidjan, in Côte d'Ivoire and Berlin, Germany. For the WSF2015 layer values are either 0 or 1 for settlement and non-settlement areas, for the WSF2015-Density layer values range between 0 and 100, with red and green tones highlighting high and low PIS, respectively.

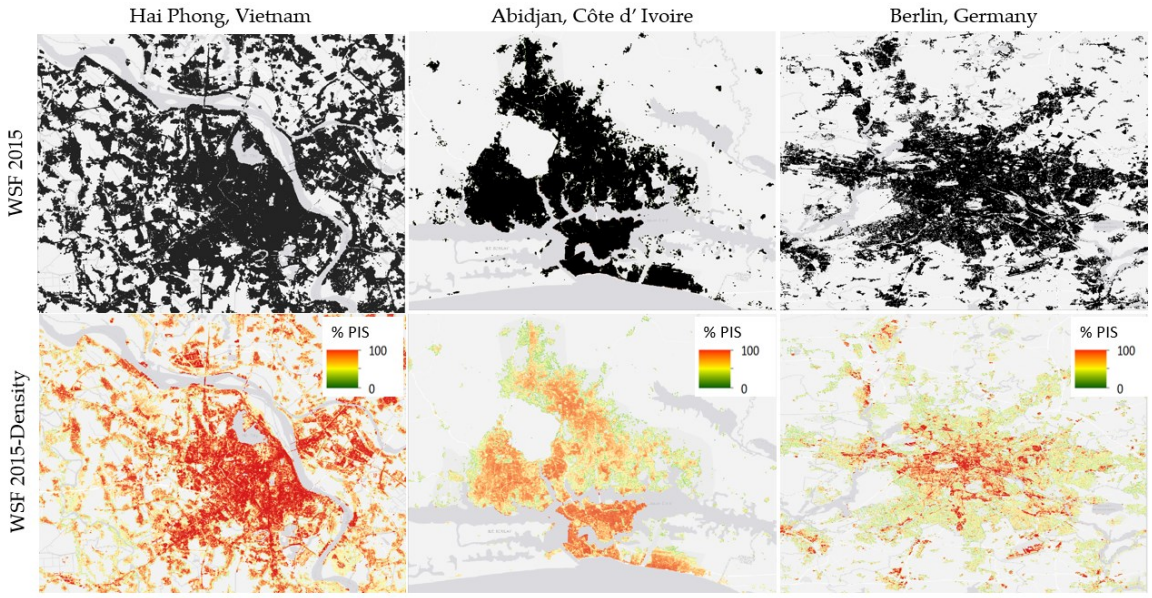

Figure 3-1. Subsets of the WSF2015 and WSF2015-Density layers for the cities of Hai Phong, Vietnam; Abidjan, Côte d' Ivoire and Berlin, Germany produced with a spatial resolution of 10m at the Equator. PIS values are city dependent.

3.2 The WSF2019 and the WSF2019-Imperviousness

- **Main Reference:** Marconcini, M., Metz-Marconcini, A., Esch, T., & Gorelick, N. (2021). Understanding Current Trends in Global Urbanisation-The World Settlement Footprint Suite.*GI-Forum 2021 (1)* https://doi.org/ 10.1553/giscience2021_01_s33

The WSF2019 and the WSF2019-Imp layers represent follow-on products to the WSF2015 and the WSF2015-Density datasets; however, they are produced using different input data and a slightly modified processing framework. The WSF2019 settlement layer, unlike the WSF2015 layer, is produced by means of a novel methodology that jointly exploits multi-temporal ~286,000 S1-radar imagery and ~2,000,000 Sentinel-2 optical (S2) images. The processing method is based on the same rationale employed in the production of the WSF2015, following a set of processing steps briefly described as follows:

First, key temporal statistics were extracted from the S1 and S2 dataset, using a selected target region during a time period where no significant changes could be expected to the settlement environment (e.g. a 1x1 degree tile, in a 1-year period). From the S2-imagery a total of 455 features were extracted, obtained from calculating approximately 55 normalised indexes (e.g. water, vegetations, soils/desert, snow, etc), and their corresponding mean, media, standard deviation and 5th and 95th percentiles, together with the original 10 band that integrate the dataset. From the S1 images, temporal statistics such as the mean, median, standard deviation, and the 5th and 95th percentiles were extracted from the backscattering, together with additional values corresponding to the sum of the total backscattering of all polarisation channels and the intensity difference. This led a total of 21 from the sentinel datasets.

Second, for the 456-stack of features, a feature selection process was carried out by means of a RF algorithm using pre-selected labelled data collected from OpenStreetMap (OSM) building footprints. From this process, a total of 31 features (6-S1 and 25-S2) were chosen as candidates for the classification task.

Third, once the 31 temporal features were selected, training samples for the settlement and non-settlement classes were extracted by means of an automated process that included removing complex topographies, applying a set of thresholds to 16/31 features and by including previously classified areas in the WSF2015 layer. Similar to the process followed by the WSF2015, the thresholding was fine-tuned according to each Köppen-Geiger climate type (Peel et al., 2007).

Fourth, unlike with the WSF2015, the classification process of the WSF2019 layer consisted of a binary RF classifier. The selection of the RF classifier over the SVM relied on its robustness to deal with unbalanced training data and its easier-to-implement method with similar high generalisation characteristics. In the 1x1 degree study area a total of 10,000 training samples were used, 5,000 for each class, within a RF with 50 trees.

Fifth, once the RF model was trained this was applied to a collection of temporal statistics over ~286K and ~2M S1 and S2 imagery. From here, a post-classification process was employed to a) remove misclassified pixels corresponding to roads, railways and rivers and b) incorporate potentially missing areas. The first process was done using a combination of dedicated ancillary datasets that included OSM roads and railway datasets, the most recently released road dataset from Facebook (Basu et al., 2019) and the JRC Global Surface Water Mapping Layers (Pekel et al., 2016). The second process was done by evaluating OSM and Google Open Building Footprints, in missing areas, and merging only potential pixels that share characteristics with the labelled training data.

The validation of the WSF2019 layers was done through a quantitative exercise that used ~700,000 reference labels collected from photointerpretation of 2019 VHR Google and Maxar Imagery. This was done through a crowdsourcing collaboration with a team of expert operators from Google and the World Bank. Concisely, a set of 220/220 settlement/non-settlement points were collected from the WSF2019 layer in 200 randomly selected areas. These points were used then extended to a 3x3 (440*9*200 =792,000), in which each cell was then labelled by the aforementioned photo-interpretation process. Accordingly, the same cells were extracted from the following datasets to perform comparative analyses: the GHSL2018 layer (Corbane et al., 2021), the Global Annual Impervious Area (GAIA) (Gong et al., 2020), the Global Impervious Surface Area (GISA) (Huang et al., 2021a), the ESA WorldCover2020 maps (Buchhorn et al., 2020) and the global land use/land cover (LU/LC) map released from ESRI (Karra et al., 2021).

As with the WSF2015, the accuracy of the layers was compared through common statistical metrics derived from a confusion matrix. Here, the results suggest that, —systematically—, the accuracy of the WSF2019 was superior than those products which still rely on Landsat data for their production, including the GAIA and the GISA. The average difference, in terms of the kappa coefficient, for example was close to 0.26 and 0.23 points against these two layers, respectively. Compared to the rest of the layers, the WSF2019 also reported higher accuracies.

In the production of the WSF2019-Imp layer, the calculation of the PIS value,—which was previously derived through a multi-temporal analysis of the maximum Normalised Difference Vegetation Index (maxNDVI) extracted from the TimeScan dataset (Esch et al., 2018a; Esch et al., 2018b)—, is now derived from the multi-temporal analysis of S2 data. Here, just as before, the employment of higher resolution optical imagery has resulted in remarkable improvements to the thematic accuracy of the layer, delivering a more consistent product compared to the WSF2015-Density layer. The current processing is based on the same assumption that was used to produce the WSF2015-Density layer (Marconcini et al., 2015). To create the layer, the first step is to compute the maximum temporal NDVI (maxNDVI) from all S2 scenes acquired in 2019, considering only Level 2A bottom of the atmosphere reflectance imagery available globally from December 2017. From there, for each of the Köppen–Geiger climate zones, areas associated with impervious surfaces are extracted from OSM where these are available, and then rasterized and aggregated at S2 ~10m spatial resolution. An ensemble of support vector regression (SVR) modules is then employed for properly correlating the resulting training information with the maxNDVI to finally derive the PIS of the pixel marked as settlements in the WSF2019 layer.

For a visual comparative analysis Figure 3-2 shows subsets WSF2015, WSF2019 and WSF2019-Imp layers for the cities of Dar Es Salaam in Tanzania; Dallas in the United States; Lima in Peru; and Kolkata in India.

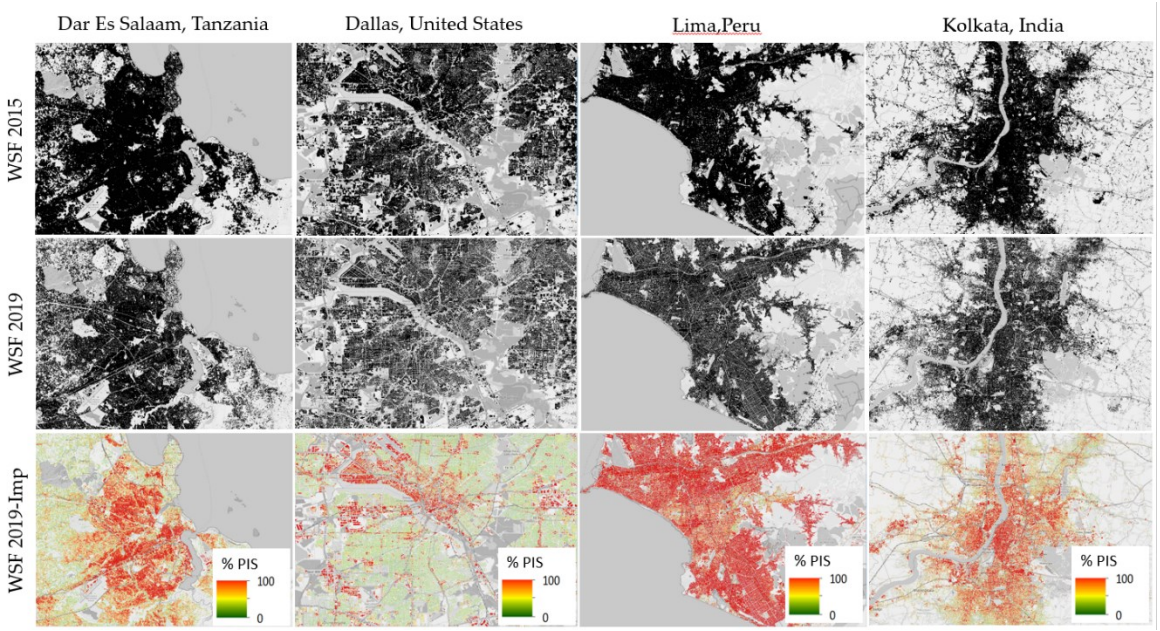

Figure 3-2. Subsets of the WSF2015, WSF2019 and WSF2019-Imp layers for the cities of Dar Es Salaam, Tanzania; Dallas, USA; Lima, Peru and Kolkata, India, produced with a spatial resolution of 10m at the Equator. PIS values are city dependent.

3.3 The WSF3D dataset

- **Main Reference 1:** Esch, T., Zeidler, J., Palacios-Lopez, D., Marconcini, M., Roth, A., Mönks, M., Leutner, B., Brzoska, E., Metz-Marconcini, A., & Bachofer, F. (2020). 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*(15), 2391. https://doi.org/10.3390/rs12152391
- **Main Reference 2:** Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., & Zeidler, J. (2022). World Settlement Footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sensing of environment*, *270*, 112877. https://doi.org/https://doi.org/10.1016/j.rse.2021.112877

The WSF3D is the first global dataset that provides detailed quantification of the fraction, total area, average height and total volume of buildings within the built-area at an unprecedented spatial resolution of 90m at the Equator. Its original processing framework is based on the preliminary methods presented in Esch et al. (2020), which were later modified to produce the final version of the WSF3D dataset presented here. In its specification, one global coverage of the WSF3D dataset consists of 18,634 files considering all of its layers. It is presented as 1x1 degree tiles with at least one settlement pixel.

The approach to produce the WSF3D dataset consist of three main modules that integrate the information of the 10m WSF2019-Imp layer, and the 12m TanDEM-X Digital Elevation Model (TDX-DEM) datasets, including its underlying 3m Synthetic Aperture Radar (SAR) amplitude images (TDX-AMP). The technical steps performed in each module are briefly summarised as follows:

First, within the settlement areas identified by the WSF2019-Imp layer, the first module measures the height variations in the 12m TDX-DEM, assumed to represent building edges (BE). The height differences are then spatially aggregated at a 90m grid delivering the average Building Height (BH) layer measured in meters [m], shown in Figure 3-3a.

Second, using the WSF2019-Imp, the TDX-AMP and the 12m BE produced in module one, the second module delivers a binary building coverage (BC) layer, which is later aggregated to the same 90m grid to calculate the Building Fraction (BF) layer. The BF, presented in Figure 3-3b, is measured in percentage [%], and is later used to define the total Building Area (BA) layer, which is measured in square meters [m2], and presented in Figure 3-3c.

Third, the final module, module number three, combines the information of the BH and the BA to calculate the Building Volume (BV) layer. This layer is measured in cubic meters [m3] and presented in Figure 3-3d.

The validation of the WSF3D datasets was done by comparing each one of its layers against 19 regions located in different part of the words, were VHR data in the form of building models (e.g. spatial resolution of <50 cm, level of detail 1) was available. Here,

the building models were rasterised to the 90m resolution of the WSF3D to perform fair comparisons. Results reveal that in terms of the BF, there is tendency towards overestimation. The mean error among all sites was of ~3%, with errors as large as 12% and as low as 0.22% across 13/19 sites. In terms of the BH, the tendency was mainly of underestimation, with an average error of -2.30m. Here, the main outliers were reported in cities with considerable high buildings, where it has been reported that errors of underestimation are quite considerable for building >30m. Comparably, in terms of the BV, the general tendency was also of underestimation, following the same trends reported for the BH. Across all cities, the average underestimation reached values of -2,080 m^3.

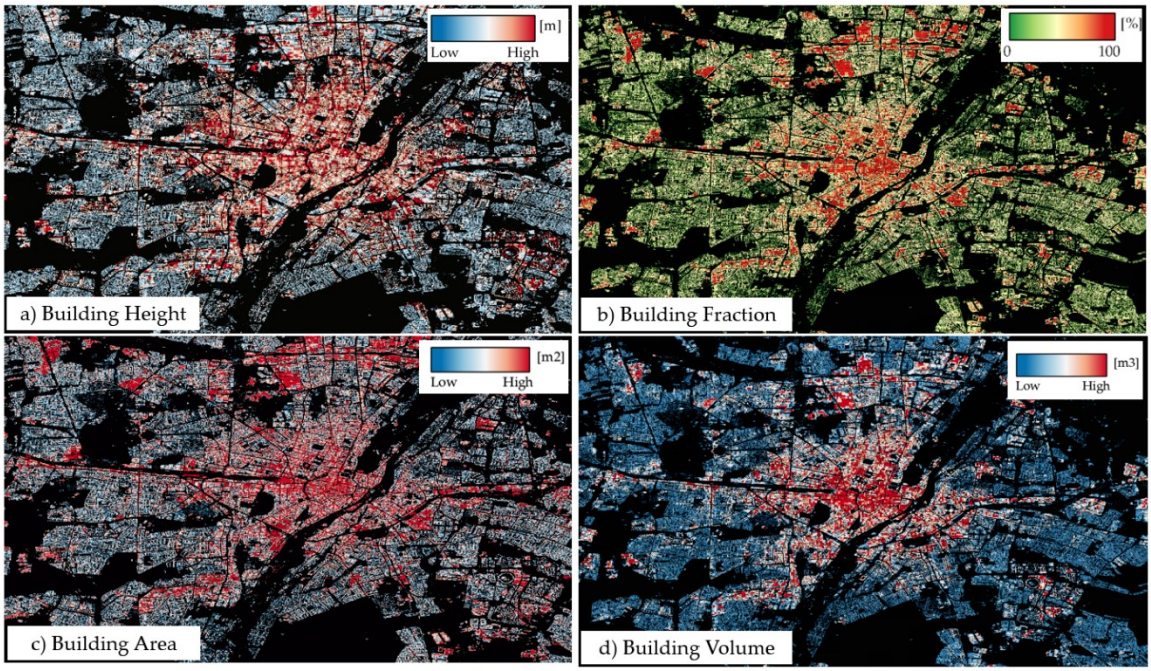

Figure 3-3. Subset of the WSF3D for the city of Munich, Germany. **a)** Building Height layers in [m], **b)** Building Fraction in [%], **c)** Building Area in [m^2] and **d)** Building Volume in [m^3]. Each layer is produced at a spatial resolution of 90m at the Equator.

Chapter 4

4. New Perspectives for Mapping Global Population Distribution Using World Settlement Footprint products

The following section represents a modification of the first peer-reviewed research article of this cumulative thesis. Some chapters have been adapted to include only relevant information that has not already been presented in the previous chapters (e.g. introduction and materials). Where needed, the reader will be referred to the corresponding sub-chapters.

- Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., Sorichetta, A., Zeidler, J., Kuenzer, C., & Dech, S. (2019). New perspectives for mapping global population distribution using world settlement footprint products. *Sustainability*, *11*(21). https://doi.org/10.3390/su11216056.

Abstract

In the production of gridded population maps, remotely sensed, human settlement datasets rank among the most important geographical factors to estimate population densities and distributions at regional and global scales. Within this context, the DLR has developed a new suite of global layers, which accurately describe the built-up environment and its characteristics at high spatial resolution: (i) the WSF2015 layer (WSF2015), a binary settlement mask; and (ii) the experimental WSF Density 2015 layer (WSF2015-Density), representing the percentage of impervious surface. This research systematically compares the effectiveness of both layers for producing population distribution maps through a dasymetric mapping approach in nine low-, middle-, and highly urbanised countries. Results indicate that the WSF2015-Density layer can produce population distribution maps with higher qualitative and quantitative accuracies in comparison to the already established binary approach, especially in those countries where a good percentage of building structures have been identified within the rural areas. Moreover, our results suggest that population distribution accuracies could substantially improve through the dynamic preselection of the input layers and the correct parameterisation of the Settlement Size Complexity (SSC) index.

4.1 Introduction: Problem Statement

In the field large-scale population modelling, the most commonly employed geospatial proxies used to redistribute population counts from administrative units to pixels of a given spatial resolution include: land cover and land use types, intensity of nightlights, climatic factors, human settlements, urban/rural extents, water features, road networks and topographic elevation and slope. However, not all these proxies are equally important for the process of disaggregation, as some of them present stronger correlations with population densities than others.

According to the research presented by Nieves et al. (2017), geographical data pertaining to the built-up environment and urban extents are the two most important proxies for predicting population densities and are significantly more important than other proxies at both regional and global scales. In this respect, for example, the GHS-POP and HRSL population grids are processed using a binary-dasymetric mapping technique (see Table 2-1), restricting the distribution of population only to those grid cells identified as human settlements. The GHS-POP uses the GHSL built-up grids (GHSL-BUILT) (Pesaresi et al., 2016), while the HRSL uses a binary mask of areas identified as human-made buildings extracted from very high-resolution satellite imagery (see Table 2-2).

The PIS, on the other hand, has proven to have an even higher correlation to population counts than simple binary-weighting, making it a good predictor of population distribution. Some studies, like the ones presented in (Azar et al., 2010; Lu et al., 2006), showed some preliminary results of this premise; nevertheless, they only focused on limited areas, thus leading to results and methodologies that are not globally transferable.

In this framework, the main objective of this research is to examine the suitability of the WSF2015 and the—thus far experimental—WSF2015-Density layers as input covariates for the development of a new global population distribution dataset. Population distribution maps were produced using a dasymetric mapping approach in combination with the finest population census/estimate-based data available at global scale at the time of writing. Here, we specifically focus on the systematic cross-comparison between the performance of the binary and the impervious layer, to investigate if quality and accuracy improvements in population disaggregation can be achieved with the WSF2015-Density layer, compared to the already established binary approach that has been employed by other population datasets and their baseline settlement layers.

Through a comprehensive quantitative assessment, we evaluated the mapping performance of each covariate layer, addressing the influence of: (i) the spatial resolution of the input census/estimate-based data; (ii) the quality of the input covariate layers; and (iii) the spatial distribution of the built-up environment on the final results. The corresponding analyses were conducted for nine representative countries of different size and different levels of urbanisation and population aggregation

4.2 Material and Methods

4.2.1 *Input Geospatial Covariates: WSF2015 and WSF2015-Density Layers*

The WSF2015 and WSF2015-Density layers used for this research have been previously described in sub-chapter 3.1.

4.2.2 *Input Census Data*

For this research, population census/estimate-based data for nine low-, middle- and highly urbanised countries (Ritchie & Roser, 2019) located in four different macro-regions of the world were collected to analyse how the differences in the level of spatial granularity of the available administrative boundaries and the variability in the morphology of built-up landscapes influence the accuracy of each covariate layer. To achieve these objectives, countries were selected on basis of the availability of population census/estimate-based data at different spatial aggregation levels. In other words,

countries were selected only if the census/estimate-based data allowed for the spatial aggregation of the administrative boundaries up to four administrative levels.

CIESIN provided geographic administrative boundaries and corresponding population counts for Cambodia, Côte d'Ivoire, England, France, Germany, Malawi, Mexico, and Vietnam. CIESIN population data were selected for this research, as it has been used in the production of other population dataset such as GPWv4, GHS-POP, WorldPop and the HRSL. The 2015 UN-adjusted estimates were used in this research. The collection and standardisation of the CIESIN data has been previously described in Chapter 2.3.

For Myanmar, population data were collected from the Ministry of Immigration and Population in reference to the Population and Housing Census of 2014 (Taw, 2015) and was joined with publicly available geographic administrative boundaries (GeoNode, 2019). The population data were released on May 2015 and the original population counts were used in this research.

For each country, administrative boundaries and population counts were aggregated at four levels of spatial resolution using attribute information stored within the data. Table 4-1 shows the total population for 2015 for each country as well as the official administrative unit nomenclature at each spatial aggregation level, the number of administrative units, the average area and the average spatial resolution (ASR). The ASR is calculated as the square root of each country total area divided by the number of administrative units, representing the effective resolution units within each country (Balk et al., 2006).

4.2.3 *Population Distribution: Dasymetric Mapping Approach*

Population distribution maps for 2015 were generated for each country at each administrative unit level using a dasymetric mapping approach, where population census/estimate-based data from administrative boundaries (source zones) are disaggregated into smaller areal units of fixed spatial resolution (target zones). The size of the target zones is normally defined by the pixel resolution of the different ancillary datasets employed to restrict and refine the distribution of the population within each administrative unit (Li & Lu, 2016). The estimated population per grid cell is defined in Equation 4-1:

$$\mathrm{Pop_t} = \mathrm{Pop_s}\frac{\mathrm{A_t} * \mathrm{W_p}}{\sum_{t \in s}(\mathrm{A_t W_p})} \qquad \text{Eq. 4-1}$$

where $\mathrm{Pop_t}$ is the population of the target zone, $\mathrm{Pop_s}$ is the population of the source zone, $\mathrm{A_t}$ represents the area of the target zone and $\mathrm{W_p}$ is the weight of a grid cell within the target zone. With this modelling approach, population counts are maintained (volume-preserving property) at each original input source zone.

In this research, two types of dasymetric mapping techniques were used. The first method is the traditional binary approach, which relies on the WSF2015 layer to assign a weighting factor of 1 to built-up pixels and a 0 for non-built-up pixels. The second method uses the WSF2015-Density layer to assign a weighting factor that ranges from 0 to 100, estimating the PIS for the pixels classified as settlement in the WSF2015.

Table 4-1.Input census/estimate-based data characteristics.

Country (ISO)/*Census Year*	Total Population 2015	Official Admin. Unit Nomenclature	No. of Units	Average Area of Units (km^2)	ASR (km)
CIV Côte d'Ivoire *2014*	22,701,552	Sub-Prefectures (Adm 3)	517	621.85	24.99
		Departments (Adm 2)	110	2907.6	54.17
		Region (Adm 1)	35	9220.92	96.03
		National (Adm 0)	1	322,744.29	568.11
DEU Germany *2014*	80,688,539	Enumeration Area (EA Level)	11,292	31.26	5.59
		Districts (NUTS3)	402	878.25	29.64
		States (NUTS1)	16	22,066.28	148.55
		National (NUTS 0)	1	353,060.51	594.19
ENG England *2014*	54,376,281	Enumeration Area (EA Level)	6791	19.2	4.38
		District (Adm 2)	326	400.16	20.00
		Region (Adm 1)	9	14,494.94	120.39
		National (Adm 0)	1	130,454.54	361.18
FRA France *2009*	64,395,348	Enumeration Area (EA Level)	36,562	15.09	3.89
		Departments (NUTS3)	96	5749.86	75.83
		Regions (NUTS2)	22	25093.51	158.41
		National (NUTS 0)	1	552,057.38	743.01
KHM Cambodia *2008*	15,394,276	Commune (Adm 3)	1633	109.66	10.47
		District (Adm 2)	197	909.06	30.15
		Province (Adm 1)	25	7163.40	84.64
		National (Adm 0)	1	179,084.95	423.18
MEX Mexico *2010*	129,731,190	Enumeration Area (EA Level)	65,477	27.7	4.91
		Municipality (Adm 2)	2456	804.65	25.36
		States (Adm 1)	32	59,898.45	222.15
		National (Adm 0)	1	1,579,248.33	1256.68
MMR Myanmar *2014*	50,279,900	Township (Adm 3)	330	2032.66	45.09
		District (Adm 2)	74	9064.6	95.21
		Regions (Adm 1)	15	44,718.7	211.47
		National (Adm 0)	1	670,780.63	819.01
MWI Malawi *2010*	17,215,235	Enumeration Area (EA Level)	12,550	7.19	2.68
		Trad: Authority (Adm 3)	357	252.92	15.90
		District (Adm 2)	32	2821.69	53.12
		National (Adm 0)	1	90,294.35	300.49
VNM Vietnam *2009*	93,447,596	District (Adm 3)	688	477.52	21.85
		Municipality-Province (Adm 2)	63	5214.87	72.21
		Region (Adm 1)	6	54,756.19	234.00
		National (Adm 0)	1	328,537.15	573.18

4.2.4 *Quantitative Accuracy Assessment*

As stated by Bai et al. (Bai et al., 2018) "quantifying the accuracy of population distribution maps has been recognized as a critical and challenging task". Determining the spatial and quantitative uncertainties of population distribution products is fundamental yet very difficult due to the lack of independent and compatible reference data (Freire et al., 2016). Nevertheless, through well-established accuracy methods, it is possible to assess the effectiveness of new models (disaggregation methods and/or covariate layers) and investigate if higher population distribution accuracies can be reached in comparison to previous approaches. For this research, the accuracy of the two covariate layers was

assessed by computing the difference between the estimated population counts extracted from maps produced using coarser administrative units (input units) and the actual population counts of the finest administrative units (validation units). This accuracy method has been widely employed in previous research (Merkens & Vafeidis, 2018; Reed et al., 2018; Stevens et al., 2015b; Tatem et al., 2007; Tiecke et al., 2017); however, it still presents some limitations, as high-resolution boundaries and population data (e.g., enumeration area level) are not publicly available for all countries.

For this reason, to gain a more comprehensive and detailed understanding of the mapping capabilities of each covariate layer, the final population distribution maps were evaluated following a series of thorough quantitative analyses performed at the validation unit level and the input level of the administrative units. The analysis at the validation unit level was divided in two parts. In the first part, an overall accuracy assessment was carried out to examine the influence of the spatial resolution of the input census/estimate-based data on the results. Here, population distribution maps were produced using three spatial aggregation levels of the administrative boundaries as input units (Analyses I–III in Table 4-2).

For each analysis, four main descriptive statistics were calculated to measure the overall accuracies of each layer. These metrics are briefly described in Table 4-3 and include: the Mean Absolute Error (MAE), the normalised Mean Absolute Error (%MAE), the Root Mean Square Error (RMSE) and the coefficient of determination (R^2).

Table 4-2. Spatial aggregation levels of the administrative boundaries used as input units and validation units for each analysis (finest to coarser spatial detail) (EA, Enumeration Area).

Country (ISO)	Analysis	Level of Administrative Input Units	Level of Administrative Validation Units
KHM	I	Adm 2	Adm 3
CIV	II	Adm 1	
MMR	III	Adm 0	
VNM			
ENG	I	Adm 2	EA
	II	Adm 1	
	III	Adm 0	
FRA	I	NUTS 3	EA
	II	NUTS 2	
	III	NUTS 0	
DEU	I	NUTS 3	EA
	II	NUTS 1	
	III	NUTS 0	
MWI	I	Adm 3	EA
	II	Adm 2	
	III	Adm 0	
MEX	I	Adm 2	EA
	II	Adm 1	
	III	Adm 0	

Table 4-3. Descriptive statistics for overall accuracy assessment at the validation unit level for Analyses I-III.

Metric		Description
$\mathrm{MAE_i} = \frac{\sum_{VU=1}^{n} \lvert \mathrm{PE_{VU}} - \mathrm{P_{VU}} \rvert}{n}$	Eq. 4-2	MAE is the mean absolute error at each level of analysis (i), calculated as the average of the sum of the absolute differences between the estimated population (PE_{vu}) and the actual population (P_{VU}) at each validation unit.
$\%\mathrm{MAE_i} = \frac{\mathrm{MAE_i}}{\mathrm{Av.\,Pop}} x100\%$	Eq. 4-3	%MAE is the mean absolute percentage error at each level of analysis (i), calculated as the MAE_i divided by the average population of each country.
$\mathrm{RMSE_i} = \sqrt{\frac{\sum_{VU=1}^{n} (\mathrm{P_{VU}} - \mathrm{PE_{VU}})^2}{n}}$	Eq. 4-4	RMSE is the root mean square error at each level of analysis (i), calculated as the square root of the mean of the sum of squares of the differences between the estimated population at (PE_{vu}) and the actual population (P_{VU}) at each validation unit.
R^2		Defined as the coefficient of determination at each level of analysis, derived from classical linear least square modelling with constant intercept at 0. It is also defined as the square of the Pearson correlation coefficient, to measure the variation between the estimated population and the actual population of all validation units. Readers can refer to (Anderson-Sprecher, 1994) for detailed calculations.

The second part of the analysis was carried out only for the population maps produced using the finest input units (Analysis I in Table 4-2). Here, similar to the

methodology and classification presented by Bai et al. (Bai et al., 2018), the Relative Estimation Error (REE) metric was used to identify the amount and distribution of error produced by each covariate layer. The REE for each validation unit was calculated as:

$$\mathrm{REE}_{\mathrm{VU}} = \left((\mathrm{PE}_{\mathrm{VU}} - \mathrm{P}_{\mathrm{VU}})/\mathrm{P}_{\mathrm{VU}}\right) * 100\% \qquad \text{Eq. 4-5}$$

where $\mathrm{PE}_{\mathrm{VU}}$ is the estimated population of the validation unit and P_{VU} is the actual population of the validation unit. Using the $\mathrm{REE}_{\mathrm{VU}}$, validation units were grouped and classified into different REE ranges **(**Table 4-4**).**

Table 4-4. REE classification (Bai et al., 2018).

REE Ranges	Description
[−100%, −50%)	Greatly underestimated
[−50%, −25%)	Underestimated
[−25%, 25%]	Accurately estimated
(25%, 50%]	Overestimated
(50%, ≥100%]	Greatly overestimated

From this classification, two sub-analyses were conducted for each country. First, for a better understanding of the error distribution associated with each covariate layer, we calculated the percentage of each country's total population that fell within each error range. Second, for each country, we calculated the average actual population and average number of settlements pixels for the validation units that fell within each error range. This last analysis was done to identify if there is any relationship between the amount of population that needs to be distributed (P_{VU}) and the number of available settlement pixels, and if the ratio between these two parameters can explain the REE values reported in the validation units.

Finally, as the reported accuracy at the validation unit level is only a reflection of the capability of each input covariate layer to correctly allocate population counts at the input unit level, a series of analyses were carried out at the input unit level, focusing only on Analysis I (Table 4-2). First, we used the RMSE metric as a summary of the error within each original input unit, following the methodology presented by Mennis and Hultgren (2006). RMSE was calculated as the square root of the mean of the sum of squares of the difference between the actual population counts and estimated population counts of all validation units within an input unit:

$$\mathrm{RMSE}_{\mathrm{IU}} = \sqrt{\frac{\sum_{\mathrm{VU}\in\mathrm{IU}}(\mathrm{P}_{\mathrm{VU}} - \mathrm{PE}_{\mathrm{VU}})^2}{\mathrm{n}}} \qquad \text{Eq. 4-6}$$

where P_{VU} is the actual population at validation unit, $\mathrm{PE}_{\mathrm{VU}}$ is the estimated population at validation unit and n is the number of validation units within an input unit. To compare the effectiveness of the covariate layers, input units were grouped according to the layer that produced the lowest RMSE values and for each group the percentage of each country's total population was calculated.

Second, on basis of these results, we undertook a series of analyses to identify and describe the regions where one layer outperformed the other. For this analysis, we derived the SSC index, which classifies each input unit according to (i) the number of small,

medium- and large-settlement objects that can be found within each unit; and (ii) the proportion of each input unit's total area that is covered by these settlement objects. To calculate the SSC index, settlement objects were created, where each object is composed of connected settlement pixels via at least one pixel edge or corner (8-neighbourhood), as described by Esch et al. (2014). The SSC index within each given input unit was derived as:

$$SSC_{IU} = \left(\frac{\#\text{settlement pixels}}{\#\text{ settl. objects}}\right) * \left(\frac{\text{Sum of the area settl. objects}}{\text{Total area of input unit}}\right) * \left(\frac{\text{Area of largest settl. object}}{\text{Mean area of settl. objects}}\right) \quad \text{Eq. 4-7}$$

where high SSC_{IU} values indicate dense built-up environments and low SSC_{IU} values indicate sparse built-up environments. To allow country cross-comparisons, we normalised the SSC index values from 0 to 10 and divided it into three classes, as shown in Table 4-5.Thresholds were visually derived and evaluated against all available countries. For each SSC class, we calculated the average RMSE produced by each layer.

Table 4-5. SSC Index classifications scheme.

SSC Index Class	Description
Low (>0–1)	Small size settlements and low coverage of the total area of the input units
Medium [1–1.8)	Mix of small and medium size settlements and medium coverage of the total area of the input units
High [1.8–10)	Mix of medium and large size settlements with high coverage of the total area of the input units

4.3 Results

4.3.1 *Visual Assessment of the Population Distribution Maps*

The WSF2015-Density and the WSF2015 layers were used to produce population distribution maps for each country at each spatial aggregation level of the administrative units, representing the estimated residential population (population counted at place of domicile) as the number of people per grid cell for the year 2015. The final spatial resolution of the population distribution maps equals the spatial resolution of the input covariates (~10m at the equator).

Because the volume of results (72 population distribution maps) is too large to present here in full, we focused on one representative country to visually inspect the thematic differences between the maps produced using the WSF2015-Density and the WSF2015 layers before turning to the quantitative analyses of all the maps. Figure 4-1 shows the final population distribution maps produced using the finest administrative units for Germany (enumeration areas), depicting the local metropolitan areas of Berlin and Munich. Note that, for the finest administrative units, these two areas have been modelled using a single administrative unit were local differences between the binary and the weighted disaggregation approaches are rather clear.

In this context, population disaggregation based on the WSF2015 layer produces homogeneous population counts within each administrative unit in comparison to the

WSF2015-Density layer, which offers more spatial heterogeneity. As a result of the proportional allocation produced by the binary layer, it is possible to observe abrupt changes from high to low population counts between neighbouring administrative units. The transitions are considerably smoother when using the WSF2015-Density layer, due to the weight given by the PIS, which rarely changes abruptly at the boundaries of the administrative units.

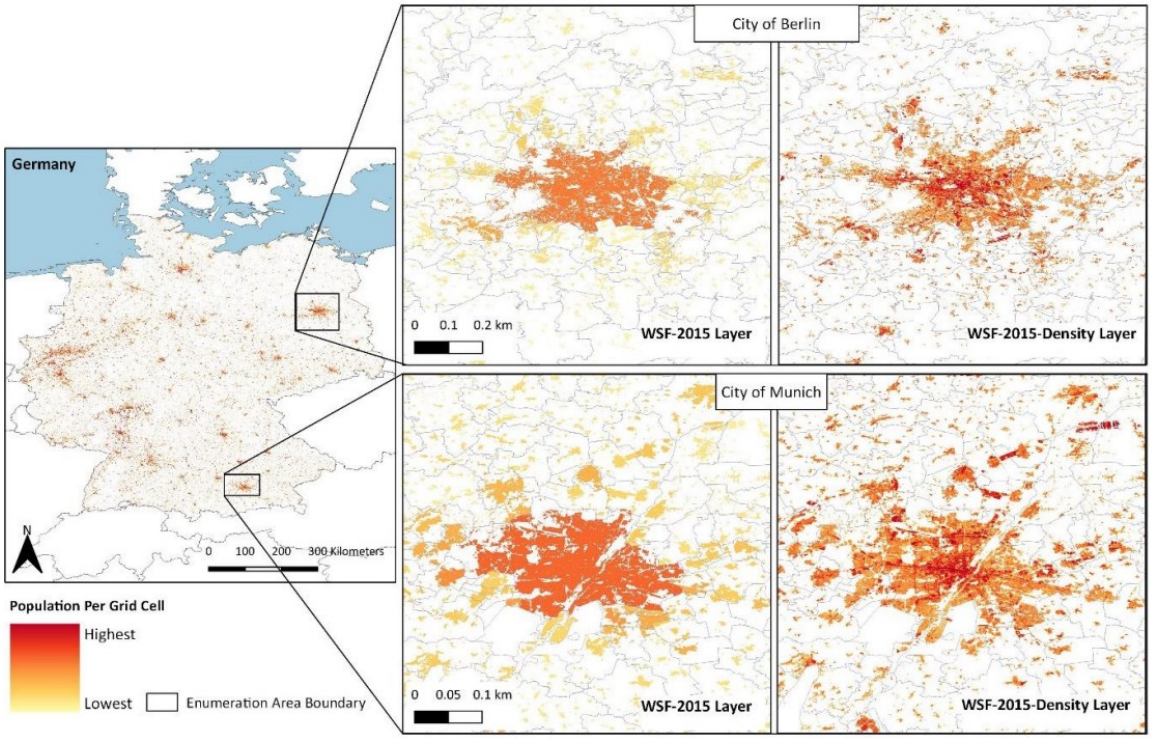

Figure 4-1. In this example: Estimated population as the number of people per grid cell for Germany in 2015 produced at the finest aggregation level of the input data (enumeration areas). The population distribution is displayed as the result of dasymetric approach using the WSF2015 layer and the WSF2015-Density layer. Detailed examples show the metropolitan areas of Berlin and Munich.

4.3.2 *Accuracy Assessment*

4.3.2.1 Analyses at the Validation Unit Level

A summary of the accuracy assessment results using the WSF2015-Density and WSF2015 layers is presented in Table 4-6. Results show that, for each layer and each country, the highest R^2, the lowest MAE, the lowest %MAE and the lowest RMSE values are reached using the finest administrative input units (Analysis I, Table 4-2). Furthermore, from one level of spatial aggregation to the next, the values for the R^2 decrease, while the MAE, %MAE and RMSE values increase.

From the RMSE and MAE metrics, it can be seen that, for Analysis I, for most countries, errors remain below the size of the average population using any of the two covariate layers. While for all countries the MAE values remain below the average population size for Analyses I-III, RMSE exceeds this threshold in Analysis II in Germany and France and in Analysis III in Mexico and Myanmar. Additionally, the difference between the RMSE and the MAE values tends to increase as the spatial detail of the input

units decreases, with significant higher differences in countries such as Côte d'Ivoire, France, Myanmar and Vietnam. In case of the three latter, the large differences can be explained by the large variances between the errors of the validation units within each country.

Comparing the results between the WSF2015-Density and the WSF2015 layers, it can be seen that, for Cambodia and Malawi, the best overall accuracies are reported using the WSF2015 layer at all levels of aggregation. For the rest of the countries, the WSF2015-Density layer performs better at all levels of aggregation, except for Mexico and Myanmar where there is a transition between layers in Analysis III.

Focusing only on the population distribution maps produced using the finest input units (Analysis I, Table 4-2), further analyses were performed at the validation unit level. First, classifying the REE values in different error ranges (Table 4-4), we calculated the percentage of each country's total population that fell within each REE range for each covariate layer, as shown in Figure 4-2 and Table 4-7.

The percentage bar charts in Figure 4-2 show that for each country both covariate layers distribute approximately the same amount of population with comparable accuracies. From here, it can be seen that for all countries, the largest percentage of the population was "accurately estimated" with estimation errors ranging from −25% to 25% for both covariate layers. For Côte d'Ivoire, Germany, England and Myanmar, this represents more than 50% of the total population; for France, Cambodia and Vietnam, between 40% and 50% of the total population, and for Malawi and Mexico between 30% and 40% of the total population. Moreover, for the majority of the countries the second largest percentage of the population was either "underestimated" or "overestimated" (from ±25 to ±50%). For all countries, less than 15% of the total population was overestimated, while for most countries, except Germany and Myanmar, from 15% to 25% of the total population was underestimated. Finally, the smallest percentage of the population for all countries was "greatly underestimated" or "greatly overestimated" (≥50% or ≤−50%), with Malawi reporting an average of ~30% of the total population within these ranges, followed by Mexico with ~25%, and France and Vietnam with ~17%.

Table 4-6. Accuracy assessment results using the WSF2015 and the WSF2015-Densinty covariate layers. Values of MAE and RMSE represent number of people.

					WSF2015-Density				WSF2015			
Country ISO	**Average Population**	**Analysis**	**No. of Input Unit**	**No. of Validation Units**	**MAE**	**%MAE**	**RMSE**	**R^2**	**MAE**	**%MAE**	**RMSE**	**R^2**
CIV	43,910.16	I	110	517	10,029.04	22.84%	40,198.00	0.7803	10,375.16	23.63%	44,814.26	0.7224
		II	35		11,851.45	26.99%	41,343.98	0.7725	11,862.96	27.02%	45,593.19	0.7130
		III	1		15,016.82	34.20%	47,045.80	0.5684	15,118.44	34.43%	50,124.64	0.3891
DEU	7145.64	I	402	11,291	828.86	11.60%	2261.67	0.9975	984.10	13.77%	2824.88	0.9961
		II	16		1897.35	26.55%	12,580.46	0.9316	2281.22	31.92%	14,409.45	0.9094
		III	1		2481.30	34.72%	23,280.14	0.9170	2999.64	41.98%	26,407.33	0.9010
ENG	8007.11	I	326	6791	2218.00	27.70%	3309.71	0.1744	2347.93	29.32%	3401.02	0.1415
		II	9		2776.75	34.68%	4310.88	0.1000	3208.51	40.07%	4619.30	0.0474
		III	1		3098.81	38.70%	4666.95	0.0634	3642.90	45.50%	5017.18	0.0167
FRA	1761.26	I	96	36,562	589.00	33.44%	4605.53	0.8777	685.47	38.92%	5242.17	0.8352
		II	22		702.31	39.88%	9543.18	0.7698	817.24	46.40%	10,950.33	0.6333
		III	1		821.41	46.64%	11435.96	0.5279	954.06	54.17%	12495.90	0.3390
KHM	9426.99	I	197	1633	3425.38	36.34%	4898.26	0.6174	3241.26	34.38%	4694.16	0.6204
		II	25		4325.54	45.88%	6680.15	0.5244	4078.17	43.26%	6027.73	0.5371
		III	1		4738.49	50.27%	8363.82	0.5333	4343.88	46.08%	6270.24	0.5662
MEX	2915.00	I	2456	65,477	954.40	32.74%	2424.57	0.3841	1031.51	35.39%	2599.99	0.3672
		II	32		1080.44	37.06%	2440.33	0.3176	1194.89	40.99%	2611.97	0.3162
		III	1		1719,04	58.97%	30507.37	0.2326	1702,60	58.41%	3464.93	0.2604
MMR	76,263.92	I	75	330	32,257.60	42.30%	47,374.91	0.8214	34,301.82	44.98%	49,602.98	0.7986
		II	15		41,755.91	54.75%	58,807.41	0.7611	44,506.83	58.36%	64,708.38	0.7071
		III	1		83,960.45	110.09%	111,546.15	0.5243	66,606.76	87.34%	88,449.93	0.4051
MWI	1371.73	I	357	12,550	712.08	51.91%	1038.03	0.3231	687.40	50.11%	1001.41	0.3290
		II	32		795.36	57.98%	1219.17	0.1732	766.46	55.88%	1177.45	0.2050
		III	1		836.53	60.98%	1310.94	0.1924	792.69	57.79%	1182.53	0.2423
VNM	135,824.99	I	63	688	46,646.67	34.34%	76,804.15	0.6018	47,837.20	35.22%	87,481.13	0.5218
		II	6		57,187.23	42.10%	94,536.29	0.4317	61,288.84	45.12%	99,151.92	0.3578
		III	1		61,323.29	45.15%	95,472.76	0.3617	63,825.03	46.99%	100,829.93	0.2636

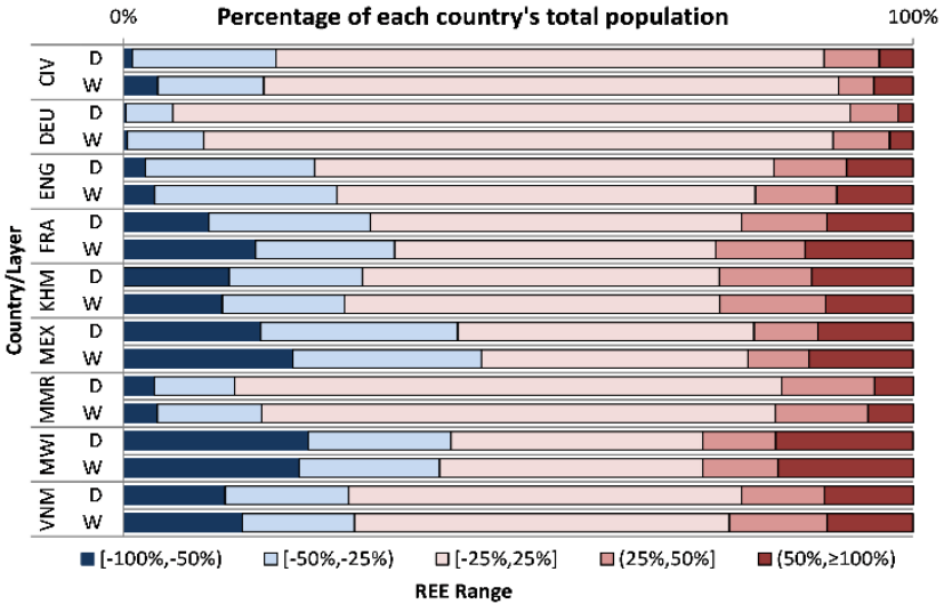

Figure 4-2. Percentage of each country's total population that fell within each REE range. D, using the WSF2015-Density layer; W, using the WSF2015 layer.

Table 4-7. Summary of the percentage of each country's total population that fell within each REE range.

	REE Range	[−100%, −50%)		[−50%, −25%)		[−25%, 25%]		(25%, 50%]		(50%, ≥100%]	
		D	W	D	W	D	W	D	W	D	W
CIV	%Population	1.11%	4.30%	18.22%	13.44%	69.40%	72.84%	6.98%	4.47%	4.29%	4.95%
DEU	%Population	0.34%	0.51%	5.90%	9.63%	85.78%	79.69%	6.05%	7.19%	1.92%	2.97%
ENG	%Population	2.78%	3.94%	21.40%	23.09%	58.22%	53.00%	9.21%	10.30%	8.39%	9.67%
FRA	%Population	10.82%	16.73%	20.42%	17.57%	47.06%	40.70%	10.79%	11.34%	10.92%	13.66%
KHM	%Population	13.35%	12.50%	16.87%	15.48%	45.23%	47.55%	11.73%	13.40%	12.82%	11.07%
MEX	%Population	17.37%	21.42%	24.97%	23.90%	37.50%	33.73%	8.09%	7.76%	12.07%	13.19%
MMR	%Population	3.92%	4.27%	10.14%	13.22%	69.30%	65.06%	11.74%	11.73%	4.92%	5.73%
MWI	%Population	23.44%	22.23%	18.03%	17.80%	31.87%	33.33%	9.25%	9.54%	17.41%	17.11%
VNM	%Population	12.84%	15.04%	15.66%	14.20%	49.78%	47.47%	10.50%	12.43%	11.23%	10.86%

To identify if there is any significant relationship between the actual population to distribute in a particular validation unit and the number of available settlement pixels, we calculated the average actual population and the average number of settlement pixels for the validation units that fell within each REE range. Figure 4-3a shows the ratio between these two parameters for each REE range, where the general tendency indicates that, for most countries, errors of underestimation are mainly reported in validation units where a relatively low number of settlement pixels were identified in comparison to the average actual population reported for those validation units. In other words, errors of underestimation tend to increase as the ratio between the population and the number of settlement pixels increases. On the other hand, for most countries, errors of overestimation tend to increase as the ratio between the average actual population and the number of settlement pixels decreases, indicating that a large number of settlement pixels have been detected in relation to the average actual population reported on those validation units.

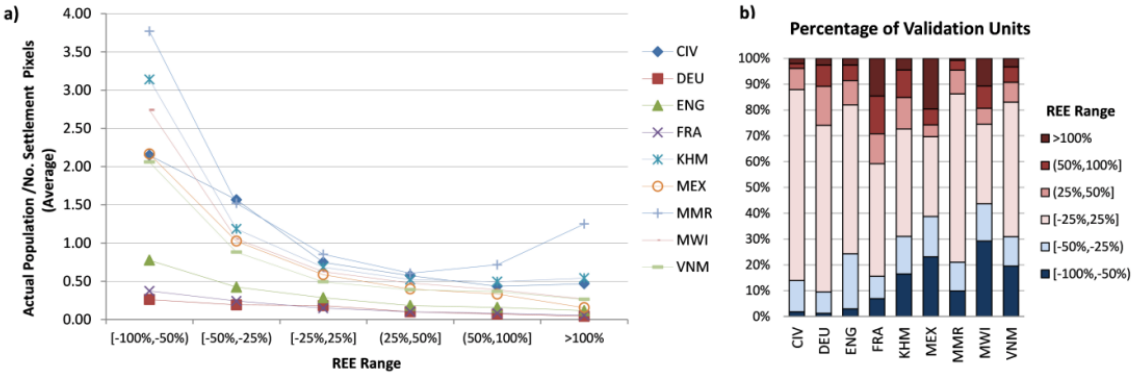

Figure 4-3. REE distribution: (**a**) ratio between the average population and the average number of settlement pixels for the validation units that fell within each REE range; and (**b**) percentage of validation units that fell within each REE range.

For a better understanding of the error distribution, the percentage of validation units that reported similar ratios and fell within each REE was quantified for each country. From the percentage bar charts in Figure 4-3b, it is possible to observe that, for countries such as Cambodia, Mexico, Malawi and Vietnam, more than 30% of the validation units reported errors of underestimation (from −100% to −25%), with Mexico, Malawi and Vietnam reporting ~20% of the validation units "greatly underestimated" (from −100% to −50%). In the same way, France reported the largest percentage of the validation units (~41%) with errors of overestimation (from 25% to ≥100%), followed by Mexico (~30%), Malawi and Germany (~25%). Here, Mexico reported the largest percentage of validation units "greatly overestimated", with ~20% of the validation units with REE larger than 100%.

4.3.2.2 Analyses at the Input Unit Level

To evaluate the actual performance of each covariate layer, results at the validation unit level were used to calculate the $RMSE_{IU}$ metric of the original input units used for population disaggregation according to Eq. 4-6. Input units were grouped according to the input covariate layer that produced the lowest $RMSE_{IU}$ values and for each group the percentage of each country's total population was calculated.

Figure 4-4 illustrates the percentage bar charts for each country. As one can notice, for Germany, France and Mexico, the predominance of the WSF2015-Density is clear, distributing more than 75% of each country's total population with overall lower RMSE values in comparison to the WSF2015 layer. On the other hand, for Cambodia and Malawi, the WSF2015 layer performs better, distributing more than 75% of the population more accurately compared to WSF2015-Density layer. In the rest of the countries (i.e., Côte d' Ivoire, England, Myanmar and Vietnam), both layers perform equally, with the WSF2015-Density layer distributing a slightly larger amount of the population better than the WSF2015 layer.

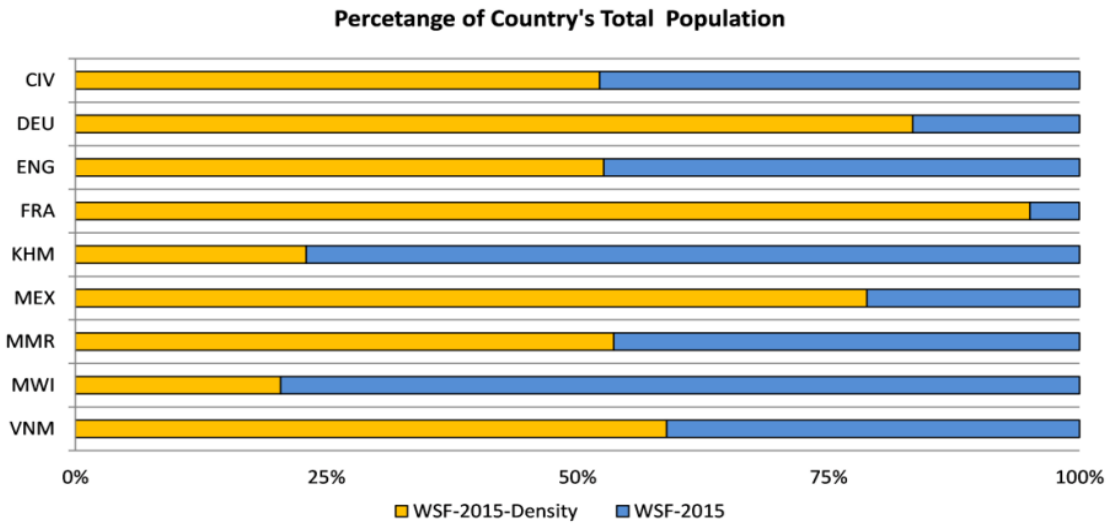

Figure 4-4. Percentage bar-charts of each country's total population distributed with higher accuracy by each covariate layer. Orange bars, WSF2015-Density layer; Blue bars, WSF2015 layer.

To identify the regions where each covariate layer produced higher accuracies, the input units of each country were classified according to the SSC index (Eq. 4-7,Table 4-5).The map in Figure 4-5 illustrates the results of this classification for Côte d' Ivoire. Here, most of the input units fell within the "low" SSC class, which is characterised by small size settlement objects that cover a low percentage of each input unit's total area. A few input units fell within the "medium" SSC class, characterised by a mix of medium and small size settlements objects, and only two input units fell within the "high" SSC class, characterised by large size settlement objects that cover a large extent of each input unit's total area. For Côte d' Ivoire, some of the most populated cities are located within the "high" and "medium" input units, such as Abidjan, Bouake, Korhogo and Divo.

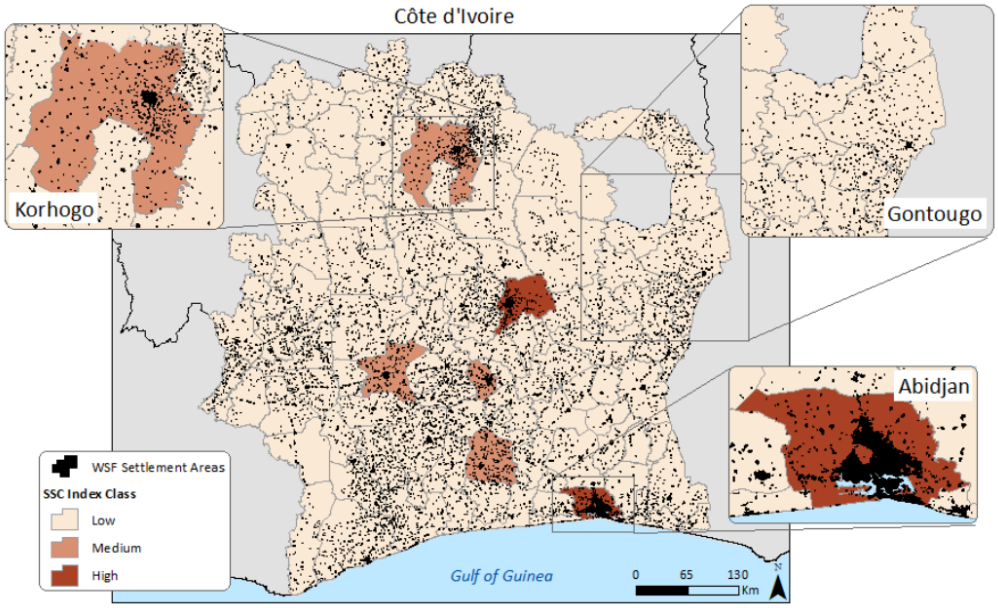

Figure 4-5. Input units classified according to the SSC-Index for Côte d'Ivoire.

Following this classification, the same analysis was carried out for each country. Figure 4-6 shows the percentage of each country's total area (pie charts) and corresponding population (boxes) derived from the input units according to the SSC index.

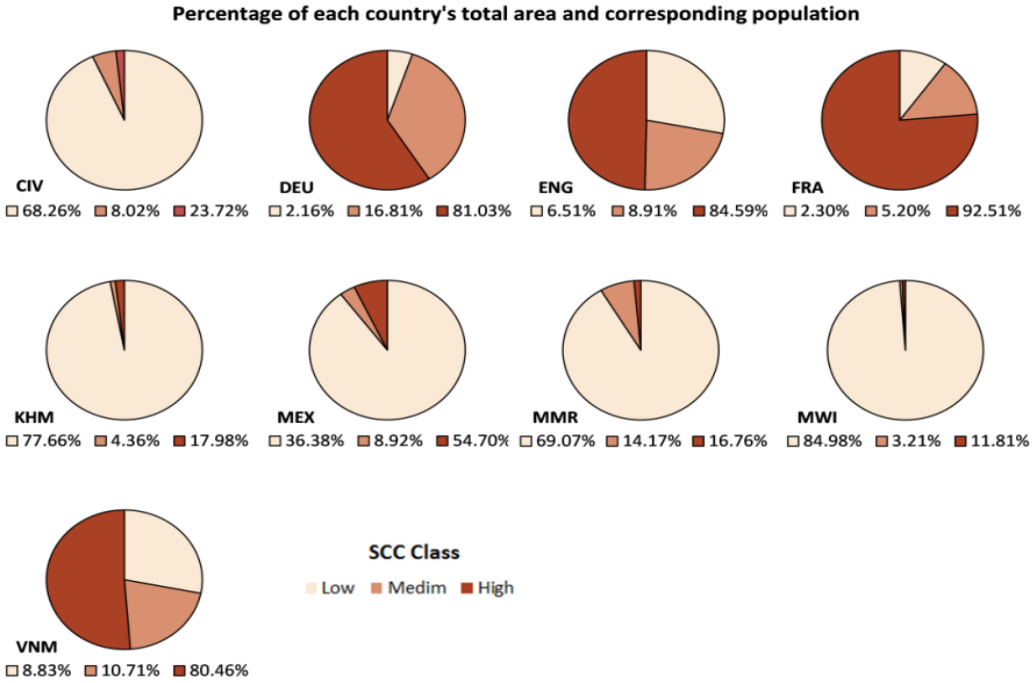

Figure 4-6. Percentage of each country's total area (pie charts) and corresponding population (boxes), classified according to the SSC index.

For Côte d' Ivoire, Cambodia, Mexico, Myanmar and Malawi, the largest percentage of the total area fell within the "low" SSC class. For all these countries, more than 70% of the population is located within these areas, except for Mexico, where the majority of the population (54.79%) is located within areas belonging to the "high" SSC class. For Germany, England, France and Vietnam, the largest percentage of the total area fell within the "high" SSC class, where more than 80% of the population is located. For most countries, the second largest percentage of the area fell within the "medium" SSC class. In these areas, the second largest percentage of the population is located, which does not exceed more than 17% of the total population.

For each SSC class, we computed the average RMSE error produced by each covariate layer and the percentage difference between the two layers (Table 4-8). The results indicate that, for all countries, the WSF2015-Density layer performed better in regions that fell within the "high" SSC class, with improvements ranging from 1.12% to 31.20% over the WSF2015 layer. For regions within the "low" or "medium" SSC classes, the behaviour of the covariate layers is more variable among the countries. For Germany, England, France and Myanmar, the WSF2015-Density layer performed better for regions within the "low" SSC class, with improvements ranging from 4.36% to 22.40%, while. for Côte d' Ivoire, Cambodia, Malawi and Vietnam, the WSF2015 layer performed better with improvements ranging from 2.12% to 9.82%. For the regions within the "medium" SSC class, the WSF2015-Density layer performed better in Germany, England, France, Malawi and Vietnam, with improvements ranging from 6.62% to 21.03%, as opposed to Côte d'

Ivoire, Cambodia, Mexico and Myanmar, where the WSF2015 layer performed better with improvements ranging from 6.69% to 30%.

Table 4-8. RMSE (number of people) and percentage difference reported for each covariate layer at each SSC index class. D, results of the WSF2015-Density layer; W, results of the WSF2015 layer; positive bold values, countries where the WSF2015-Density performed better; negative values, countries where the WSF2015 performed better.

	Low SSC Class			Medium SSC Class			High SSC Class		
	RMSE (D)	RMSE (W)	%Diff.	RMSE (D)	RMSE (W)	%Diff.	RMSE (D)	RMSE (W)	%Diff.
CIV	6195.88	5824.85	−6.17%	13,385.74	9893.41	−30.00%	121,500.76	138,430.55	**+13.03%**
DEU	598.65	701.99	**+15.89%**	1169.94	1422.94	**+19.51%**	1715.78	2100.37	**+20.16%**
ENG	2449.15	2879.85	**+16.16%**	2580.89	3013.39	**+15.46%**	2908.04	2980.60	**+2.46%**
FRA	517.12	647.56	**+22.40%**	975.40	1207.01	**+21.03%**	4391.66	5124.74	**+15.41%**
KHM	4041.02	3785.02	−6.54%	3536.39	3084.83	−13.64%	6372.06	6443.97	**+1.12%**
MEX	892.80	874.05	−2.12%	2107.69	2253.53	−6.69%	2376.74	2626.13	**+9.97%**
MMR	33,452.74	34,943.76	**+4.36%**	39,432.66	32,580.79	−19.03%	43,682.69	59,832.04	**+31.20%**
MWI	819.79	768.93	−6.40%	778.43	831.73	**+6.62%**	1150.90	12,20.03	**+5.83%**
VNM	47,476.56	43,030.73	−9.82%	32,471.05	27,000.96	**+18.40%**	63,679.29	65,272.30	**+2.47%**

4.4 Discussion

In the above sections, we present a set of comprehensive analyses to compare the relative accuracies of population distribution maps produced using the WSF2015 and the experimental WSF2015-Density layers. The first analysis consisted of an overall accuracy assessment carried at the validation unit level, where metrics such as MAE, %MAE, RMSE and R^2 (Table 4-3) were used to evaluate maps produced using three spatial aggregation levels of the administrative units (Table 4-2). The results presented in Table 4-6 show that, for all countries and both covariate layers, the highest accuracy values were reported for population maps produced using the finest input units (Analysis I, Table 4-2), with accuracies decreasing from one level of spatial aggregation to the next. These results are directly in line with previous findings (Hay et al., 2005a; Tatem et al., 2007; Tiecke et al., 2017), and confirm the premise that higher accuracies in population mapping can be achieved with improvements in the resolution of the input census/estimate-based data. In the same way, from a comparative point of view, the overall accuracy results showed that, for the majority of the countries, except Cambodia and Malawi, the WSF2015-Density layer performed better than the WSF2015.

When interpreting and comparing the overall accuracy results between countries and between covariate layers, there are, however, a set of considerations that need to be considered. First, it is important to understand, that regardless of the input covariate layer used for population disaggregation, high accuracies can be reached, when the number and ASR of the of the administrative units used for validation are similar to those of the administrative units used as input data (Table 4-2). This can be seen, for example, by examining the results of Analysis I for Côte d'Ivoire, Myanmar and Vietnam (Table 4-6). The fact that these countries reported relatively good accuracy results is more likely to be due to the small difference between the number of administrative units used as input and validation units (407, 225 and 625, respectively) and the small ratio between their ASR (2.16, 2.11 and 3.30, respectively). These results are linked to the scale effect of the

modifiable areal unit problem (MAUP), where the correlation between variables increases as the areal unit size becomes similar (Qi & Wu, 1996).

A second consideration to keep in mind is to avoid the use of the R^2 metric as a unique statistical indicator to report the accuracy of population distribution models. Previous research has demonstrated that the lack of variability in the data influences the coefficient of determination (Goodwin & Leech, 2006). For example, for England, where significantly low R^2 values were obtained in comparison to the MAE, %MAE and RMSE metrics, these can be related to the fact that the original census/estimate-based data reports similar population counts for a large number of the administrative units used for validation. This can be seen in the boxplots of Figure 4-7, where the reported actual population counts of the validation units of England are constrained within a small range of values. This small variability in the data, according to Goodwin et al. (Goodwin & Leech, 2006), results in a poor correlation between the estimated population counts and actual population counts as exemplified in the scatter plots of Figure 4-8. Here, it is possible to observe an amorphous or non-structured appearance of the data points for England in comparison to France, which results in a poor correlation, signalised by the almost horizontal trend-line.

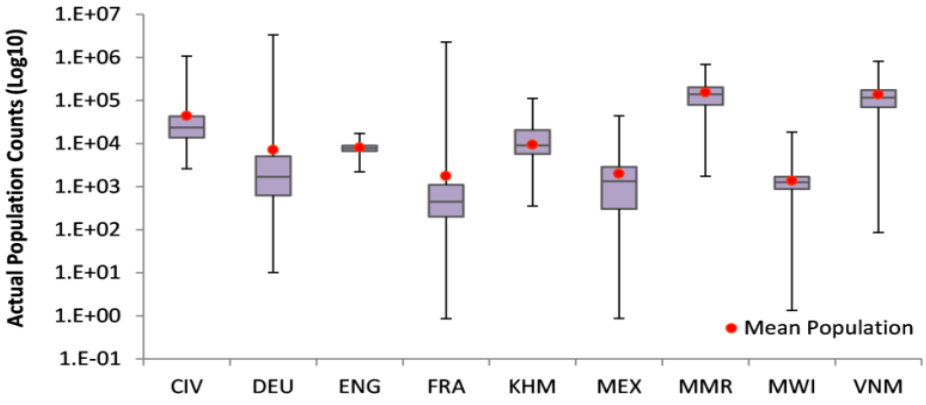

Figure 4-7. Boxplots of the distribution of the actual population counts of the validation units for each country with the inter-quartile range demarcated by the purple box.

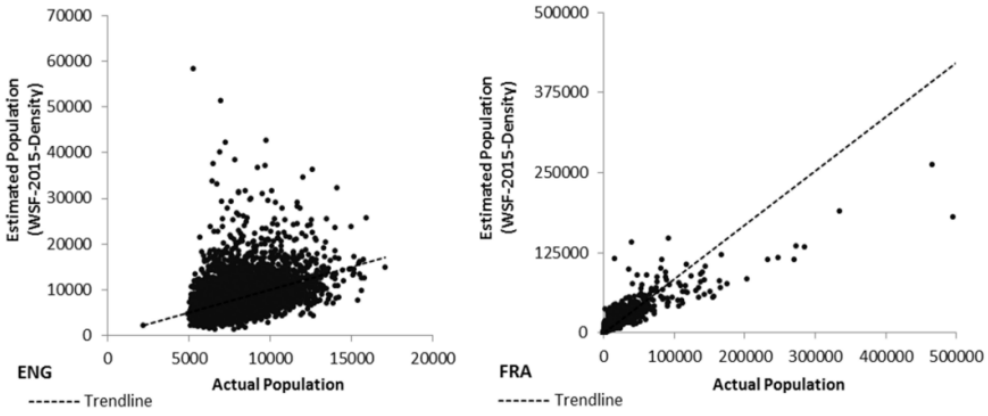

Figure 4-8. Scatter plot of estimated population and actual population for England and France at the validation unit level. Data show the results of population estimates using the WSF2015-Density layer.

The aforementioned findings indicate that the use of single statistics metrics can be misleading and that population distribution maps can report high accuracy results independently of the quality of the underlying covariate layers used for population disaggregation. Therefore, it is important to emphasise, not only that full dissemination of the data used for modelling and validation is essential when reporting accuracy results (Bai et al., 2018; Balk et al., 2006; Doxsey-Whitfield et al., 2015; Hay et al., 2005a), but also that, to evaluate the real effectiveness of the covariate layers, it is necessary to undertake more in-depth analyses using complementary metrics.

In this research, with the use of the REE statistical metric (Eq. 4-5) it was possible to evaluate the amount and distribution of error generated by each covariate layer (Figure 4-2 and Table 4-6), and identify the areas where large errors of underestimation and large errors overestimation can be expected (Figure 4-3). Our results show that both layers perform similarly, distributing approximately the same percentage of each country's total population with the same REE values. For all countries, the largest percentage of the population has been estimated with errors ranging from −25% to 25%, which in previous research has been considered as "accurately estimated" (Bai et al., 2018). Nevertheless, only in Côte d' Ivoire, Germany, England and Myanmar this represent more than 50% of the total population, which indicates that, for the rest of the countries, a significant percentage of the total population was distributed with larger errors of underestimation and errors overestimation.

We attribute these errors to the quality (completeness) of the covariate layers and to the fact that they do not consider information on the land or building use. On the one hand, our findings indicate that errors of underestimation are reported in validation units where not enough settlement pixels have been found for population disaggregation. These errors increase as the ratio between the actual population and the number of settlement pixels increases (Figure 4-3a). This means, for example, that in countries where a large percentage of the population and validation units were "greatly underestimated" (Table 4-6 and Figure 4-3b) such as France, Cambodia, Mexico and Malawi, this can be explained by the large amount of validation units where zero or very few settlement pixels have been identified (Figure 4-9). Therefore, despite the fact that the thematic accuracy of the WSF2015 layer clearly outperforms any of the currently existing global human settlements masks (Marconcini et al., 2020), it is clear the data still show limitations with respect to a complete detection of all building structures. This can be explained by the spatial resolution of the Sentinel-1 and Landsat imagery used as input data, which restricts the identification of building structures, especially in regions where the settlement pattern is characterised by wide-spread single houses or very small hamlets.

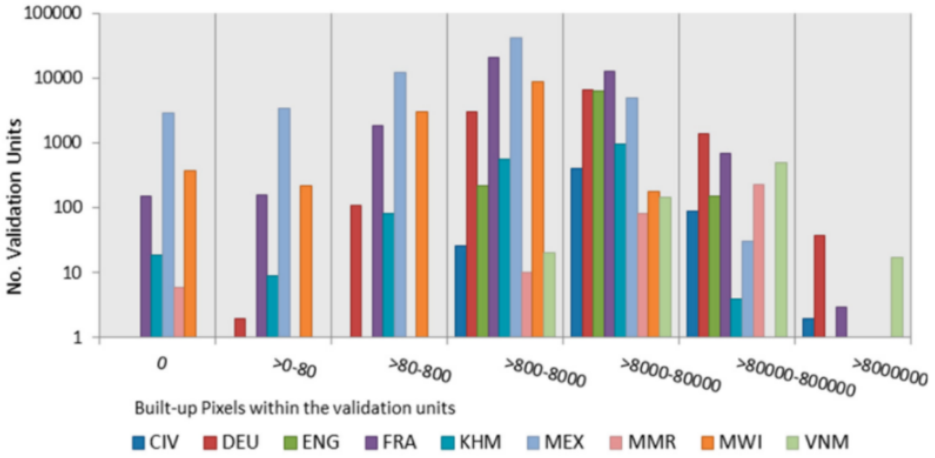

Figure 4-9. Number of settlement pixels identified within the validation units.

On the other hand, errors of overestimation are reported in validation units where a large number of settlement pixels have been reported in comparison to the amount of actual population, and that they increase as the ration between these two parameters decreases (Figure 4-3a). After a visual analysis of VHR satellite imagery, we found that large errors of overestimation are mainly reported in validation units where seaports and industrial complexes exist. Figure 4-10 shows an example of the population distribution results for an input unit in England with this particular built-up environment. The red line represents the geographical boundary of the input unit used for population disaggregation and the blue lines represent the geographical boundaries of the validation units. Here, it is possible to observe industrial areas in the southern parts of the input unit. These areas capture many of the population counts comparable to high-density residential areas, reporting large errors of overestimation in the validation units. In the selected validation unit, for example the WSF2015-Density layer reported a higher REE (186.56%) in comparison to the WSF2015 (154.49%). This does not mean, however, that in every validation unit where this built-up environment exists the binary layer will perform better than the impervious layer. Depending on the extent and geographical boundaries of the input units, industrial or port areas can be mixed with residential areas, influencing the performance of each layer. More detailed information on and discussion of this aspect is provided at the end of this section in the context of the SSC index.

Similar accuracy limitations have been reported in the production of the GHS-POP and the HRSL population datasets (Freire et al., 2016; Tiecke et al., 2017). Even when several local studies have demonstrated that information on the building use has the potential to improve population distribution results (Biljecki et al., 2016; Goerlich, 2016), this remains a major source of limitation in the production of global population datasets, as it is not possible to derive detailed semantic information on the building use through RS methodologies. Population datasets such as LandScan and WorldPop integrate land use and land cover covariates to improve their results; however, as mentioned above, this introduces global transferability limitations and applicability restrictions.

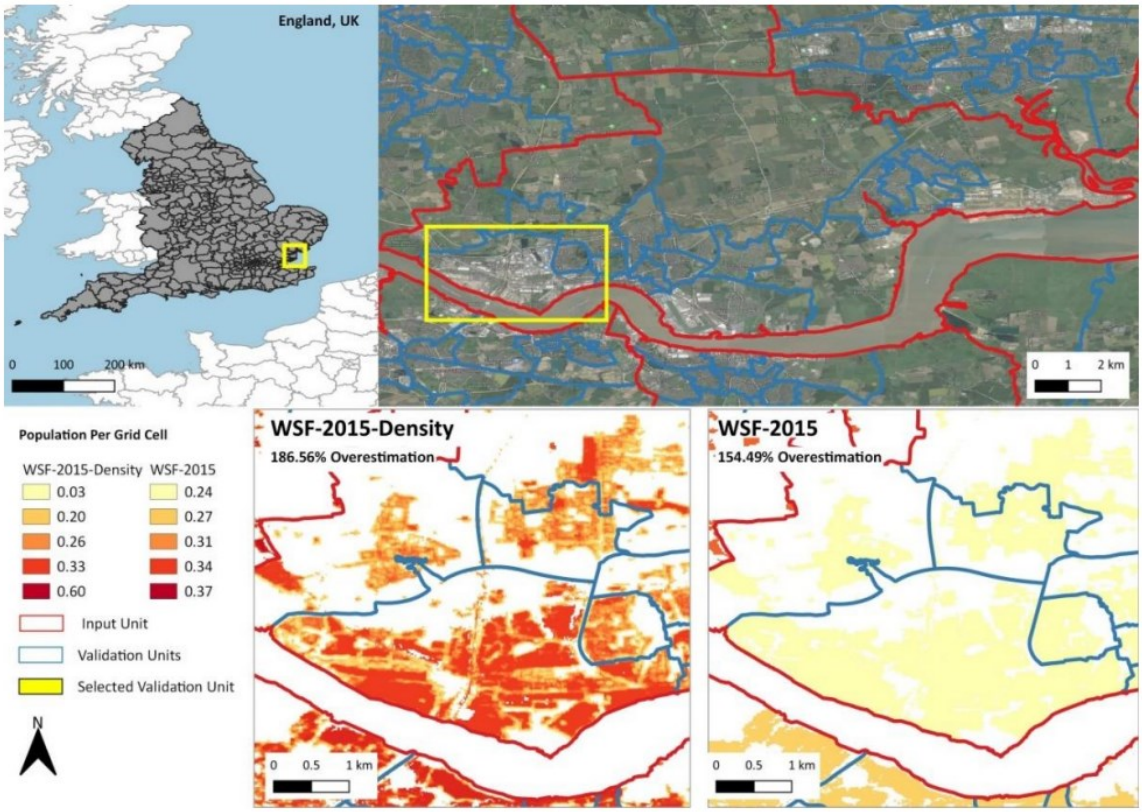

Figure 4-10. Influence of the building use in the population distribution results. Industrial areas capture large population counts resulting in large errors of overestimation within the validation units.

For this reason, in this study, we began to analyse the relationship between the inherent characteristics of the underlying built-up environment and the performance of each covariate layer, as an alternative approach that could be used to minimise the errors introduced by the quality and lack of functional characterisation of the input covariate layers. Here, we introduced the SSC index as a globally transferable metric to categorise the input units in terms of the size and coverage of the underlying settlement objects. Our results clearly indicate that WSF2015-Density layer distributes population with higher accuracies in regions with high SSC index values, reaching improvements up to ~30% over the WSF2015 layer (Table 4-8). For regions with low and medium SSC index values, the performance of each covariate layer varies from country to country. Figure 4-11 shows the distribution of the SSC index values and the mean SSC index value for the "low" and "medium" SSC classes for each country.

Focusing on the distribution of the "low" SSC class, countries where the WSF2015 reported in average less RMSE values are also the countries where more than half of the input units reported SSC index values lower than 0.40. In other words, the SSC index values fell below the mean of the "low" SSC class that ranges from >0 to 1. For the "medium" SSC class, the distribution of the SSC index values among countries is relatively similar. The mixture of medium to highly populated cities and rural areas within these input units represent challenging modelling regions where further analyses are required to identify the particular circumstances where one layer outperforms the other.

Nevertheless, it is important to notice that the WSF2015-Density layer performed better in all three classes of the SSC index for countries such as Germany, France and

England, hence suggesting that the overall performance is largely driven by an accurate identification of building structures within the rural areas of each country. In this context, it is expected that limitations derived from the current underestimation of smaller settlements and isolated buildings can be overcome by the future integration of Sentinel-2 data in the production of future WSF datasets, due to its increased spatial resolution (Esch et al., 2018a).

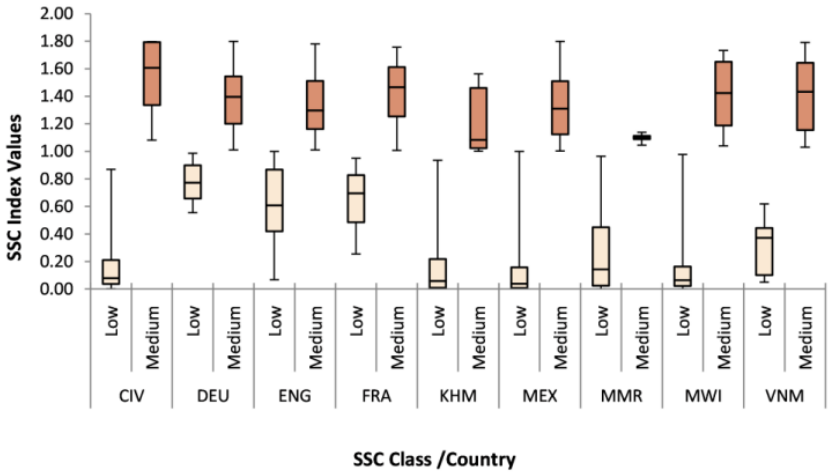

Figure 4-11 Boxplots of the distribution of the SSC index values for the "low" (yellow boxplots) and "medium" (green boxplots) classes for each country.

As a final note, it is important to mention that, even when the population distribution maps presented in this research have been produced using the most frequently employed population census/estimate-based data, the difficulties in the acquisition of the finest census/estimate-based data, the challenges in integrating census data with spatial boundaries and the uncertainties of population estimates based on statistical projections, are additional sources of errors and uncertainty limiting the accuracy of the population distribution models. Therefore, as stated by (Doxsey-Whitfield et al., 2015) acquiring up-to-date global population census data at the highest spatial detail possible should remain a priority for improving global population mapping.

4.5 Summary

The presented study focused on the cross-comparison of population distribution maps produced using the WSF2015 and the experimental WSF2015-Density layers. The main objective was to investigate if higher accuracies in population distribution mapping can be achieved using additional information on the build-up environment, such as the percentage of impervious surface, in comparison to the already established binary approach employed by other population datasets and their baseline settlement layers.

The results of the quantitative assessment showed that the overall accuracies between both covariate layers are comparably similar, with the best accuracy results reported for population distribution maps produced using the finest input census/estimate-based data. Our results indicate that, while both layers distribute the largest percentage of each country's total population with estimation errors ranging from −25% to 25%, remaining limitations derived from: (i) the incomplete identification of settlement pixels; and (ii) the lack of information on the building use, still introduce large

errors of underestimation and errors overestimation in a considerable percentage of the population.

Notwithstanding these limitations, from a comparative point of view, our results have shown that population distribution maps produced on basis of the WSF2015-Density layer provide a more realistic representation of the spatial distribution of the population, as the heterogeneous allocation of population counts prevents the appearance of artificial patterns between neighbouring administrative units. Furthermore, it has been demonstrated that the WSF2015-Density layer produces higher accuracies in high-density built-up environments and is capable to improve the estimation accuracies of the WSF2015 layer up to ~30%, especially in those countries where a good percentage of building structures have been identified within the rural areas. The fact that the WSF2015-Density layer is derived from RS approaches that do not require a priori knowledge of the land cover makes it a strong suitable proxy capable to improve global population distribution methodologies, and, as it is not based on local relationships, it has no applicability restrictions in comparison to other existing products. Moreover, it provides global coverage and can be straightforwardly updated allowing time agreement with census population data, enabling the production of a consistent global population distribution dataset with higher accuracy and spatial resolution than those currently available.

One of the strengths of our study is the implementation of the SSC index, used to investigate the correlation between the built-up environment and the performance of each covariate layer. Our results suggest that higher accuracies in population disaggregation could be achieved with the correct preselection of the input covariate at the input unit level; however, to implement this preselection, additional research is still necessary, as the SSC index cannot provide a complete distinction between the covariate layers in areas with middle SSC index values.

However, in the light of these highly promising results, future research will focus on the validation and open release of the WSF2015-Density layer, expanding the accuracy assessment of population mapping to other regions of the world, with special focus on arid and semi-arid areas, and comparing the results against other existing global population distribution datasets. Within this outlook, deeper research on the SSC index will also be included, to develop a methodology that can help minimise the inherent distribution errors derived from the quality and functional characterisation of the input covariates, as well as in the production of a new global population dataset.

Chapter 5

5. High-Resolution Gridded Population Datasets: Exploring the Capabilities of the World Settlement Footprint 2019 Imperviousness Layer for the African Continent

The following section represents a modification of the second peer-reviewed research article of this cumulative thesis. Some chapters have been adapted to include only relevant information that has not already been presented in the previous chapters (e.g. introduction and materials). Where needed, the reader will be referred to the corresponding sub-chapters.

- Palacios-Lopez, D., Bachofer, F., Esch, T., Marconcini, M., MacManus, K., Sorichetta, A., Zeidler, J., Dech, S., Tatem, A. J., & Reinartz, P. (2021). High-Resolution Gridded Population Datasets: Exploring the Capabilities of the World Settlement Footprint 2019 Imperviousness Layer for the African Continent. *Remote Sensing*, 13(6), 1142. https://doi.org/10.3390/rs13061142.

Abstract

The field of human population mapping is constantly evolving, leveraging the increasing availability of high-resolution satellite imagery and the advancements in the field of machine learning. In recent years, the emergence of global built-area datasets that accurately describe the extent, location, and characteristics of human settlements has facilitated the production of new population grids, with improved quality, accuracy, and spatial resolution. In this research, we explore the capabilities of the novel WSF2019-Imp layer, as a single proxy in the production of a new high-resolution population distribution dataset for all of Africa—the WSF2019-Population dataset (WSF2019-Pop). Results of a comprehensive qualitative and quantitative assessment indicate that the WSF2019-Imp layer has the potential to overcome the complexities and limitations of top-down binary and multi-layer approaches of large-scale population mapping, by delivering a weighting framework which is spatially consistent and free of applicability restrictions. The increased thematic detail and spatial resolution (~10m at the Equator) of the WSF2019-Imp layer improve the spatial distribution of populations at local scales, where fully built-up settlement pixels are clearly differentiated from settlement pixels that share a proportion of their area with green spaces, such as parks or gardens. Overall, eighty percent of the African countries reported estimation accuracies with percentage mean absolute errors between ~15% and ~32%, and 50% of the validation units in more than half of the countries reported relative errors below 20%. Here, the remaining lack of information on the vertical dimension and the functional characterisation of the built-up environment are still remaining limitations affecting the quality and accuracy of the final population datasets.

5.1 Introduction: Problem Statement

The aim of this research was to explore the capabilities of the novel WSF2019-Imp layer in the production of a new high-resolution large-scale gridded population distribution dataset—the WSF2019-Population (WSF2019-Pop). Using a simple and semi-automatic weighted-dasymetric modelling approach, we incorporated the imperviousness layer with an open archive of subnational census/estimate-based estimates to produce high-resolution population distribution datasets for the African continent. Employing a well-established validation method (Leyk et al., 2019) and leveraging the variably in quality and spatial granularity of the input population data, the main focus of our research was to systematically investigate how accurate and stable the WSF2019-Imp layer is as a single proxy for population modelling. Here, we specifically explore if the WSF2019-Imp layer delivers consistent patterns of accuracy/uncertainty within and among countries, and address the main advantages and limitations of the WSF2019-Imp layer and WSF2019-Pop datasets in support of large-scale population modelling and future research applications, respectively.

5.2 Materials and Methods

Figure 5-1 outlines the general process used for the modelling and validation of the WSF2019-Pop dataset for Africa.

Steps concerning this research include the production of the end-user WSF2019-Pop dataset (Step 1) and the accuracy assessment of the population datasets of each country (Step 2). Input data, namely, the WSF2019-Imp layer for Africa and the 2019 subnational population data were either made available or downloaded ready-to-use. A detailed description of the main elements (grey labels) of each step are described in more detail in the following sections.

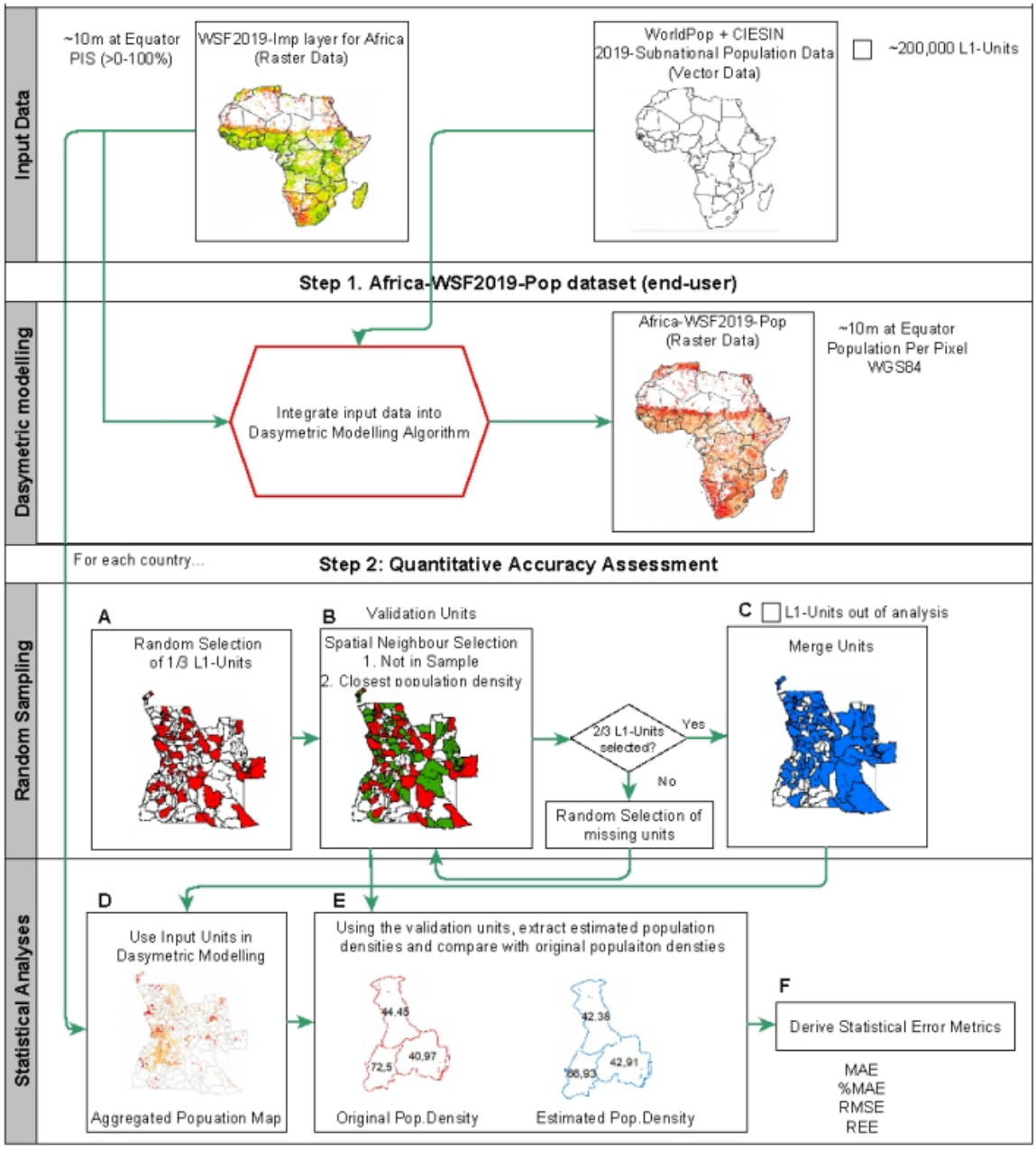

Figure 5-1. General workflow for the modelling and validation of the WSF2019-Pop dataset for Africa.

5.2.1 *WSF2019-Imperviousness layer*

The processing framework to produce the WSF2019-Imp layer was described in detailed in sub-chapter 3.2. In this research a previous version of the dataset was employed, in which the masking of the roads was still not included. For the specific case of this study, Figure 5-2 provides five different examples of the WSF2019-Imp layer. The first three images (top–bottom) refer to the city of Niamey (Niger), characterized by a hot semi-arid climate; Cairo (Egypt) characterized by a hot desert climate and the city of Antananarivo (Madagascar) characterized by a subtropical highland climate according to the Köppen Climate classification system, respectively. The last two examples show suburban areas and rural areas in South Africa and Nigeria, and are used to exemplify the

local spatial details of the layers in different vegetation cover and urbanised settings. For each of these test sites, additional subsets are compared against VHR satellite imagery.

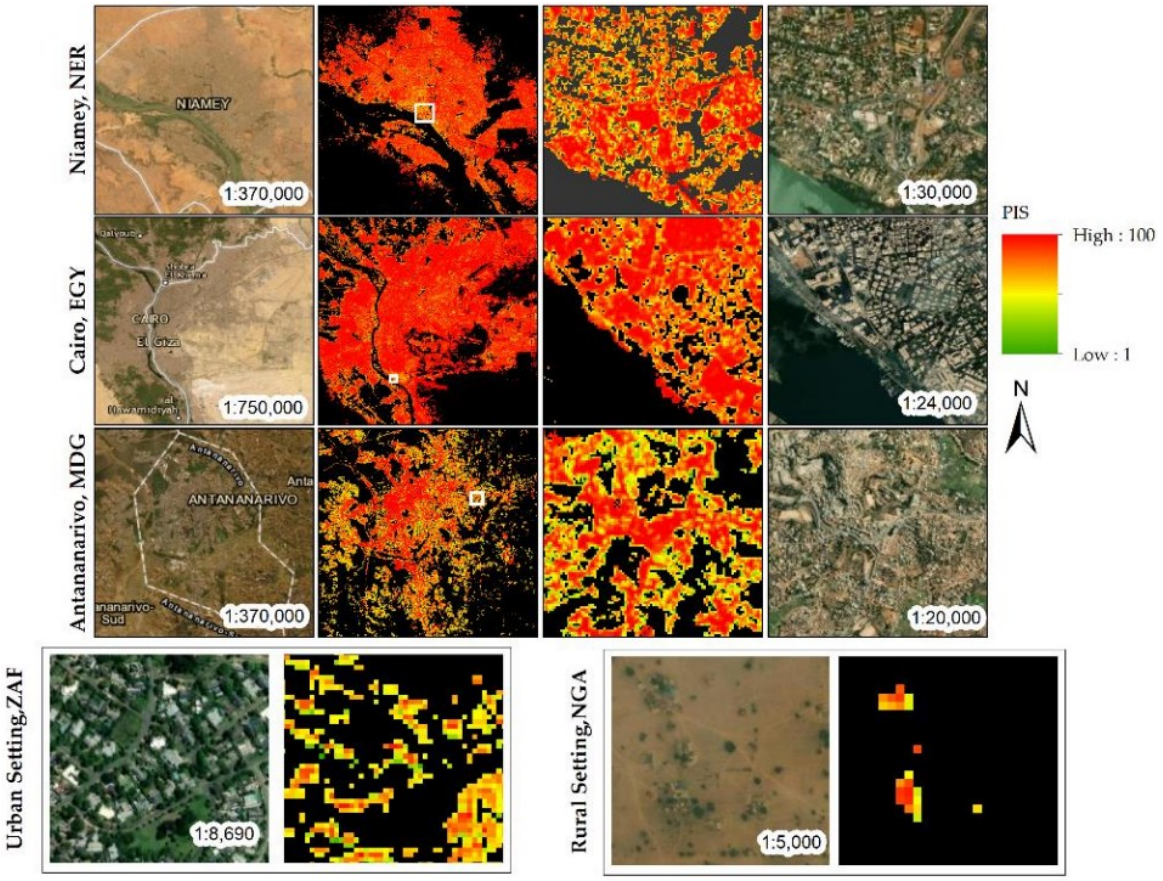

Figure 5-2. WSF2019-Imp. Top to bottom: areas of Niamey (Niger), Cairo (Egypt), Antananarivo (Madagascar), and urban (left) and rural (right) areas in South Africa and Nigeria. PIS legend from >0% to 100% with country-specific minimum and maximum values. Additional subsets (white boxes) compared against VHR imagery. Black areas: pixels outside the WSF2019 settlement mask.

5.2.2 *Subnational 2019 Population Data*

The population estimates for the year 2019 and corresponding subnational administrative unit boundaries (vector data) for all African countries employed in this research were prepared by the CIESIN. The data was produced in the context of a cross-organizational collaboration with WorldPop and accounts for the period of 2000 to 2020 (Lloyd et al., 2019). For most countries (except Kenya and Malawi), the data were directly downloaded from the open archive of the WorldPop Global Project available at https://doi.org/10.5258/SOTON/WP00650. The population data for Kenya and Malawi were provided by CIESIN.

All of the population datasets employed here were standardised by CIESIN based upon the methodology described in (CIESIN, 2018a). The subnational administrative unit boundaries and population counts follow the cartography and official estimates collected in the 2010 round of Population and Housing Censuses, which occurred between 2005 and 2014 (and data from the 2020 round for Kenya and Malawi). From these data, annual exponential growth rates were calculated using two census dates (between circa 2000 and 2010 for most countries) to interpolate and forecast population counts for each subnational

administrative unit for the period 2000 to 2020 (Lloyd et al., 2019). The exception is for Kenya, where the cartography (Kenya National Bureau of Statistics, 2020) and official estimates are from the 2019 census (Kenya National Bureau of Statistics, 2019), and for Malawi, where the cartography (Humanitarian Data Exchange, 2018) and official estimates are from the 2018 census (Malawi National Statistical Office, 2018), both of which are part of the 2020 round of Population and Housing Census. This was necessary due to restrictive licenses and significant administrative realignments between the 2010 and 2020 rounds in those countries. For each subnational administrative unit, two types of population estimates are available—census/estimate-based and UN-adjusted—with the latter employed for this research following the criterion of existing population datasets, which use UN-adjusted counts as a method of harmonisation (Freire et al., 2018). The subnational administrative unit boundaries, referred hereinafter as "L1-units"—according to their original description (Lloyd et al., 2019)—represent the highest available administrative unit level specific to each country, and are not comparable within and among countries, in terms of size and administrative level.

Table 5-1 shows a summary of the input population data. These include the three letter International Organisation for Standardization (ISO) identification code, total population for 2019 adjusted to the UN estimates, the base year of either the census or derived estimation, the number of subnational administrative units and the average ASR of the administrative units for each country. The data are presented divided in the five subregions according to the UN geoscheme for Africa (UN-Statistics Division, 2020).

In this research, the countries of Seychelles and Cape Verde were not included, as consistent S2 data for the selected period were not available when the employed version of the WSF2019-Imp layer was produced.

Table 5-1. Summary of 2019 UN-adjusted subnational population census/estimate-based data (2019-UNPop) for each African country: 3 letter ISO code, census or estimation year, number of L1-units (L1-U), and the average spatial resolution (ASR). ASR represents the effective resolution of the L1-units in km, calculated as the square root of each country's total area divided by the number of units.

Eastern Africa

ISO	Year	2019-UNPop	L1-U	ASR	ISO	Year	2019-UNPop	L1-U	ASR
BDI	2008	11,530,577	66	13	**MWI**	2019	18,628,749	73	14
COM	2013	850,891	93	21	**RWA**	2012	12,626,938	67	7
DJI	2009	973,557	77	52	**SOM**	2005	15,442,906	68	78
ERI	2012	3,497,117	82	127	**SSD**	2008	11,062,114	69	83
ETH	2007	112,078,727	67	35	**TZA**	2012	58,005,461	67	14
KEN	2019	52,573,967	68	36	**UGA**	2014	44,269,587	70	11
MDG	1993	26,969,306	69	19	**ZMB**	2010	17,861,034	69	67
MOZ	2007	30,366,043	65	40	**ZWE**	2012	14,645,473	80	63
MUS	2011	1,269,670	55	3					

Central Africa

ISO	Year	2019-UNPop	L1-U	ASR	ISO	Year	2019-UNPop	L1-U	ASR
AGO	2014	31,825,299	161	87	**GAB**	2003	2,172,578	48	73
CAF	2012	4,745,179	174	58	**GNQ**	2014	1,920,917	39	29
CMR	2005	25,876,387	58	89	**STP**	2012	215,048	7	12
COD	1984	86,790,568	188	106	**TCD**	2009	15,946,882	62	142
COG	2007	5,380,504	12	166					

Northern Africa

ISO	Year	2019-UNPop	L1-U	ASR
DZA	2008	43,053,054	1540	41
EGY	2006	100,388,076	385	49
ESH	2014	582,455	27	103
LBY	2006	6,777,453	22	280
MAR	2014	36,471,766	1657	17
SDN	2008	42,813,237	130	114
TUN	2014	11,694,721	270	26

Southern Africa

ISO	Year	2019-UNPop	L1-U	ASR
BWA	2011	2,303,703	29	141
LSO	2006	2,125,267	80	20
NAM	2011	2,494,524	5473	12
SWZ	2007	1,148,133	55	17
ZAF	2011	58,558,267	86814	4

Western Africa

ISO	Year	2019-UNPop	L1-U	ASR	ISO	Year	2019-UNPop	L1-U	ASR
BEN	2013	11,801,151	77	39	**MLI**	2009	19,658,023	765	38
BFA	2006	20,321,383	351	28	**MRT**	2013	4,525,698	218	71
CIV	2014	25,716,554	519	25	**NER**	2012	23,310,719	66	127
GHA	2010	30,417,858	170	37	**NGA**	2006	200,963,603	774	34
GIN	2014	12,771,246	340	27	**SEN**	2013	16,296,362	45	66
GMB	2010	2,347,696	40	16	**SLE**	2004	7,813,207	160	21
GNB	2009	1,920,917	39	29	**TGO**	2010	8,082,359	40	38
LBR	2008	4,937,374	136	27					

5.2.3 *Dasymetric Modelling Approach*

Gridded population distribution maps for each African country were modelled using a weighted dasymetric mapping approach, where the 2019 UN-adjusted population counts from the input L1-units were redistributed into pixels classified as settlements in the WSF2019-Imp layer (Figure 5-1, Step 1). For each pixel within an L1-unit, the estimated population count is defined as follows:

$$Pop_{(p \in IU)} = Pop_{IU} \frac{PIS_p}{\sum_{(p \in IU)}(PIS_p)} \qquad \text{Eq. 5-1}$$

According to Eq. 5-1, each pixel within a given input unit $Pop_{(p \in IU)}$ is given a proportion of the input unit's total population Pop_{IU}, relative to their percent of impervious value PIS_p. This means, for example, that within a single input unit, the population count of a pixel with a 50% PIS value is twice as high as in a pixel with a 25% PIS value. This modelling technique preserves population input totals, where the sum of population counts of all pixels within an input unit matches the input unit's original total population.

5.2.4 *Quantitative Accuracy Assessment*

In this research, we applied as quasi-similar validation method that the one presented in sub-chapter 4.2.4 to systematically investigate the relative accuracy and mapping capabilities of the WSF2019-Imp layer. The quantitative accuracy assessment presented here comprised two main steps, described as follows.

5.2.4.1 Random Sampling

To produce the population distribution maps needed for validation, we first generated the aggregated version of the L1-units, following a sampling and merging methodology similar to that employed by Stevens et al. (2020). For each country, we started by randomly selecting one third of the L1-units. For each L1-unit in the sample we then selected a spatial neighbour unit that (1) was not already in the random sample, and (2) had the closest value in population density (Figure 5-1, Step2-B). This process was performed iteratively until approximately two thirds of the original L1-units were selected. From here, the one third random sample units and the one third selected spatial neighbour units were merged, and their population counts summed to produce coarser units for population modelling (Figure 5-1,, Step 2-C). These coarser units were then used as input units to produce population distribution maps (Eq. 5-1) (Figure 5-1, Step 2-D), while the two thirds of sampled L1-units were used for validation (Figure 5-1, Step2-E). All the remaining unsampled/unmerged L1-units were excluded from the analyses, as their reported differences would have been zero.

The implementation of this aggregation method was deemed necessary, because in each country, the original L1-units represent a mixture of administrative levels, where no attribute is available to identify their administrative levels. Hence, aggregating the L1-units into a common official level, comparable across all countries, was not possible to implement. Consequently, due to the fact that some countries have very large L1-units (Table 5-1), we selected a merging criterion based on the similarity of population densities, in order to reduce the effect that the size of the input units used for modelling have on the estimation error. Here, research has shown that larger input units tend to present larger estimation errors simply due to their size (Palacios-Lopez et al., 2019; Sinha et al., 2019). Finally, we also excluded all the L1-units that reported zero population counts from the sampling process. These units would have generated errors of overestimation of 100%, derived solely from the quality of the input population data, and unrelated to the capabilities of the modelling framework.

The aforementioned sampling method was applied to all African countries, except Comoros. Comoros' input population data consisted of only three geographically separated polygons representing each of the islands: Grande Comore (Ngazidja), Mohéli

(Mwali), and Anjouan (Ndzuani). For the validation of Comoros, the two randomly selected L1-units were merged into a "multi-part" polygon, and their populations were summed. The two L1-units were further used for validation.

5.2.4.2 Statistical Analyses

From the gridded population distribution maps produced using the coarser input units, population density estimates were extracted for all the sampled L1-units (also referred to as validation units from here on) using the Zonal Statistic tool of ArcGIS (Figure 5-1, Step 2-E). For each country, the reported differences between the actual population densities and the estimated population densities of the sampled L1-units were then used to derive aggregated error metrics, such as MAE (Eq. 5-3), the normalised MAE (nMAE or %MAE) (Eq. 5-4) and RMSE (Eq. 5-5) and individual error metrics, such as the REE (Eq. 5-7) and the SSC-Index (Eq. 5-7) (Figure 5-1, Step2-F).

For this research, total population densities were used instead of total population counts to more easily perform comparisons within and among countries with varying population sizes, and with varying numbers and ASR of the sampled L1-units. Statistical analyses were carried out in two ways. First, to perform direct comparisons among countries, the aggregated error metrics were calculated taking into consideration the size/area (km²) of all sampled L1-units that make up each country. This weighting factor removes the bias caused by the differences in size and number of the sampled L1-units among countries, allowing the evaluation of the relative accuracy and modelling stability of the WSF2019-Imp layer at a continental scale. Here, the average population density of each country D_i is then calculated as the conventional population density as follows (Ottensmann, 2018) (Eq. 5-2):

$$D_i = \frac{P_i}{A_i} = \frac{\sum_{j \in i} p_j}{\sum_{j \in i,} a_j} = \frac{1}{A_i} \sum_{j \in i} a_j d_j \qquad \text{Eq. 5-2}$$

where p_j, a_j and d_j represent the population, area and density of each individual sampled L1-unit within a country j, respectively. Consequently, the MAE is the average of the sum of absolute differences between the estimated $\hat{d}_j$ and actual d_j weighted population densities divided by the total area, and the %MAE is the MAE divided by the total population density. Dividing the MAE by the average population density of each country D_i additionally removes the bias caused by the differences in population sizes (Vandeput). The %MAE was chosen over the %RMSE metric, due to the fact that the RMSE is likely to report higher values influenced solely by a larger sample size (Chai & Draxler, 2014). Both error metrics measure the average of the absolute errors in the sampled L1-units; however, while MAE weights each error equally, the RMSE gives more weight to larger differences, skewing the errors towards the odd outliers (Chai & Draxler, 2014). This quality is useful to check, for example, whether the MAE reported for each country originates from extreme errors or not.

$$MAE_i = \frac{1}{A_i} \sum_{j \in i} (|\hat{d}_j - d_i| * a_i) \qquad \text{Eq. 5-3}$$

$$\%MAE_i = \frac{MAE_i}{D_i} \quad \text{Eq. 5-4}$$

$$RMSE_i = \sqrt{\frac{\sum_{i \in i}((\hat{d}_j - d_j)^2 * a_j)}{\sum_{j \in i} a_j}} \quad \text{Eq. 5-5}$$

In a broad sense, the area-weighted aggregated metrics assume a proportional distribution of error within each country, allowing us to derive meaningful comparisons among countries. However, as the population density of the individual sampled L1-units varies from unit to unit, so do errors, which are unevenly distributed across space. Therefore, to properly investigate the error distribution within each country, for the second part of the statistical analyses, we calculated the percent REE and the SSC-Index for each sample L1-unit as follows:

$$REE_j = \frac{|\hat{d}_j - d_j|}{d_j} * 100\% \quad \text{Eq. 5-6}$$

$$SSC - Index_j = \left(\frac{\#settlement\ pixels}{\#\ settl.objects} * \frac{\sum\ area\ settl.objects}{Total\ area\ of\ \text{j}} * \frac{RSdD\ settl.objects}{Av.area\ sett.objects}\right) \quad \text{Eq. 5-7}$$

The REE is derived by calculating the absolute error between the actual and estimated population density, divided by the actual population density of each unit. Using this metric, each validation was categorised into REE ranges of 20%, following the thresholding criterion employed by (Da Costa et al., 2017). The Settlement SSC-Index is a metric that was first introduced by Palacios-Lopez et al. (Palacios-Lopez et al., 2019) to categorise the built-up environment within any given area (polygon boundary) in terms of the size, number, distribution (compacted/spread) and coverage of built-up objects derived from the WSF2015 layer. On the one hand, high SSC-Index values indicate dense built-up environments, where the total area derived from the settlement pixels is almost proportional to the total area of the sample L1-units. Low SSC-Index values, on the other hand, indicate the presence of small and sparse built-up environments, where the coverage of the built-up settlement is proportionally low compared to the total area of the input units. For this research, built-up objects are constructed from the WSF2019-Imp layer, where every object is composed of an 8-neighbourhood connected settlement pixel.

Using a 2D density analysis, we integrated the REE, the population density and the SSC-Index value of each unit to investigate if the REE of a given range was found in validation units with similar characteristics. The 2D density analysis uses contour plots that replace the scatter plot distribution, allowing for better visualisations of clustered data. Contour lines connect the points (validation units) that have the same response value (REE) with regard to two predictors (population density and SSC-Index values) (Minitab LLC, 2021).

5.3 Results

5.3.1 *Africa – WSF2019-Pop Dataset*

The end-user WSF2019-Pop dataset for the African continent depicts the residential population for the year 2019 adjusted to the UN national total estimates. The final dataset

has a spatial resolution of 0.3 arc-sec (~10m at the Equator), a WGS84 Geographic Coordinate System projection, and represents the number of people per pixel. Figure 5-3 shows the WSF2019-Pop dataset that Africa produced on basis of the L1-units of each country. It depicts the areas within the five regions of the continent, using the country boundaries for better visualization. As illustrated, the use of the WSF2019-Imp layer as proxy for population modelling delivers a heterogenous distribution of population guided by the underlying PIS value. The colour scales are country specific.

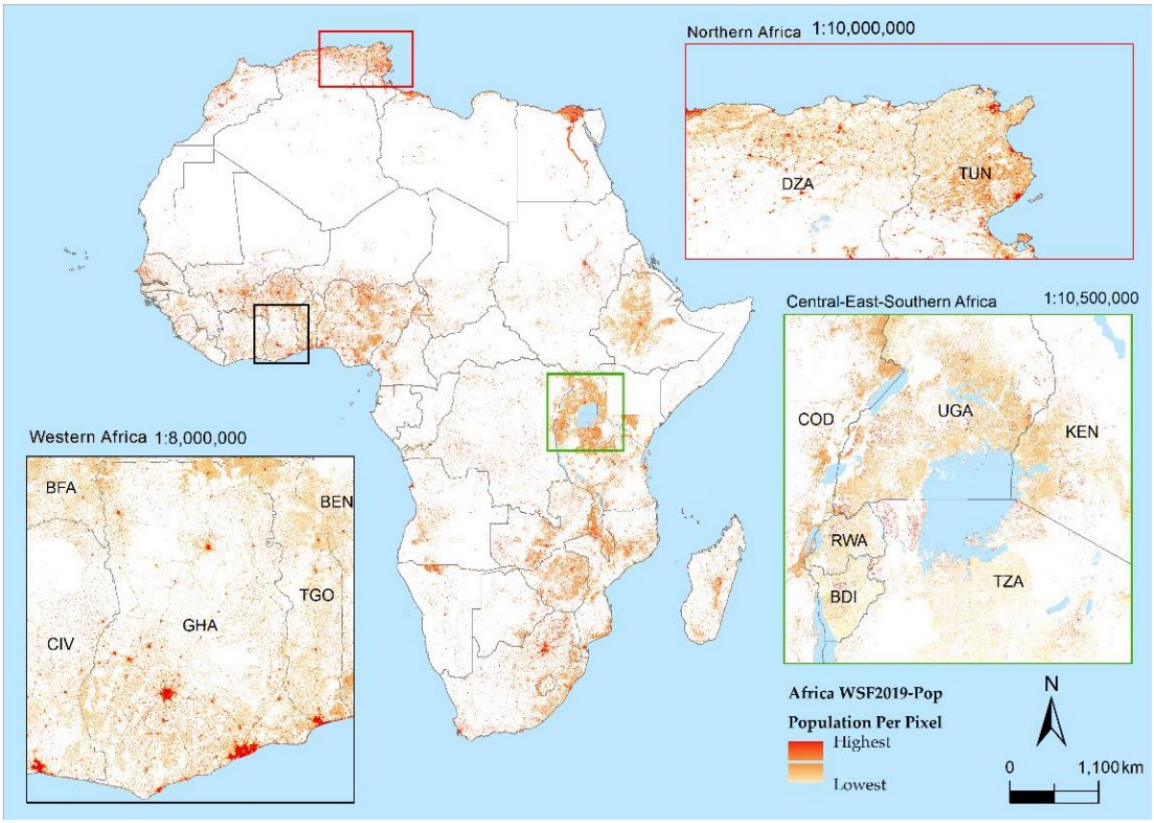

Figure 5-3 WSF2019 Population dataset for Africa (WSF2019-Pop). Colour ranges and values are country specific and represent the estimated population per every ~10m pixel.

5.3.2 *Quantitative Accuracy Assessment*

5.3.2.1 Random Sampling – Validation Unit description

For each country, the results of the sampling process described in sub-chapter 5.2.4.1 are presented in Table 5-2. From an inspection across all African countries, it is possible to observe that the final sample size *(n)* varies greatly among countries, with values ranging between two sampled L1-units for Comoros (COM), and up to 56,478 sampled L1-units for South Africa (ZAF). Independently of the sample size, results show that for most countries, more that 50% of the total population was covered by the sample, with the exceptions of Congo (COG, 25.62 %), Sao Tome and Prince (STP,48.08 %) and Liberia (LBR, 46.35%). Similarly, for most countries, more than 50% of the total area was covered by the sampled L1-units, with the exceptions of Djibouti (DJI, 20.14%) and Egypt (EGY, 14.12%). Overall, ~70% of Africa's total population and total area was covered by the random sample.

For a better visual comparison of each country's random sample, the distribution of the population density (ppl/km^2) and the size (km^2) of the sampled L1-units are displayed in the form of violin plots in Figure 5-4 and Figure 5-5. The shape of the violin plots describes the probability density or frequency of the sampled L1-units within each value range, and the black dots represent the mean value of each metric. From these plots, it is possible to observe, on the one hand, that a large proportion of the sampled L1-units in countries such as Burundi (BDI), Mauritius (MUS), Rwanda (RWA), Uganda (UGA), Egypt (EGY) and South Africa (ZAF) report population densities higher than 100 ppl/km^2. A total of 16 countries reported sampled L1-units with population densities higher than 10,000 ppl/km^2, with Egypt (EGY) and South Africa (ZAF) among the most representative.

Table 5-2. Summary of the sampled L1-units for each country grouped by region.

Eastern Africa							
ISO	n	%Pop	%Area	ISO	n	%Pop	%Area
BDI	86	67.43	64.84	**MWI**	283	67.46	67.66
COM	2	92.51	74.32	**RWA**	277	66.39	61.97
DJI	3	70.30	20.14	**SOM**	50	57.57	72.94
ERI	4	82.12	81.48	**SSD**	51	66.51	71.01
ETH	490	68.42	76.15	**TZA**	2428	64.29	70.85
KEN	229	63.50	71.34	**UGA**	918	68.88	74.57
MDG	828	68.70	67.05	**ZMB**	99	62.65	66.01
MOZ	275	62.54	71.95	**ZWE**	59	72.38	82.04
MUS	105	68.07	64.78				

Central Africa							
ISO	n	%Pop	%Area	ISO	n	%Pop	%Area
AGO	108	55.79	72.15	**GAB**	31	76.64	72.99
CAF	115	71.90	72.34	**GNQ**	3	57.15	79.77
CMR	37	56.16	71.30	**STP**	4	48.01	60.36
COD	120	60.24	67.35	**TCD**	41	64.89	69.39
COG	7	25.62	66.64				

Northern Africa			
ISO	n	%Pop	%Area
DZA	1026	60.40	81.71
EGY	225	70.70	13.14
ESH	16	73.94	62.85
LBY	13	65.01	58.99
MAR	1072	64.82	74.57
SDN	85	68.41	62.40
TUN	176	67.13	76.69

Southern Africa			
ISO	n	%Pop	%Area
BWA	17	77.81	61.66
LSO	53	65.61	76.83
NAM	3645	67.27	72.71
SWZ	35	63.95	67.39
ZAF	56478	68.16	76.45

Western Africa							
ISO	n	%Pop	%Area	ISO	n	%Pop	%Area
BEN	51	70.29	86.78	**MLI**	507	60.98	77.50
BFA	233	55.40	68.76	**MRT**	143	62.12	86.23
CIV	344	65.74	70.21	**NER**	44	70.28	55.12
GHA	113	60.45	75.22	**NGA**	515	65.33	68.31
GIN	226	67.46	68.79	**SEN**	29	58.33	78.74
GMB	25	80.82	70.63	**SLE**	106	66.34	72.30
GNB	26	75.88	76.50	**TGO**	25	70.76	76.84
LBR	89	46.35	74.22				

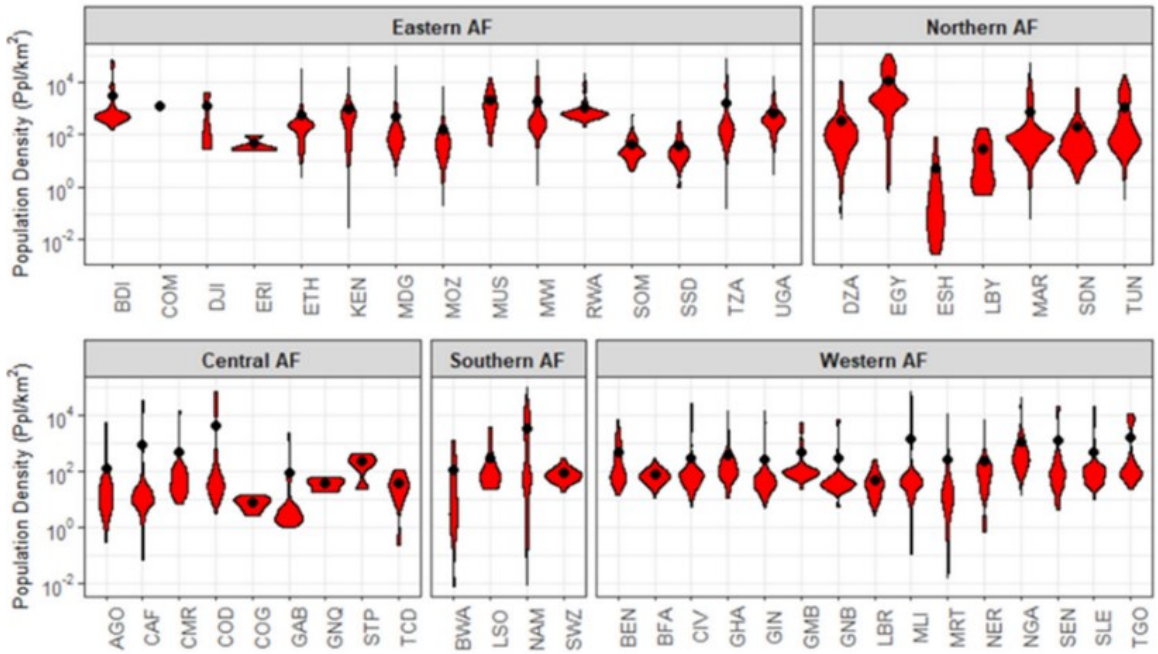

Figure 5-4 Violin plots illustrating the distribution of the population density (ppl/km^2) of the sample L1-units within each country. Black dot: mean value of the distribution, not to be confused with the average population density of the country.

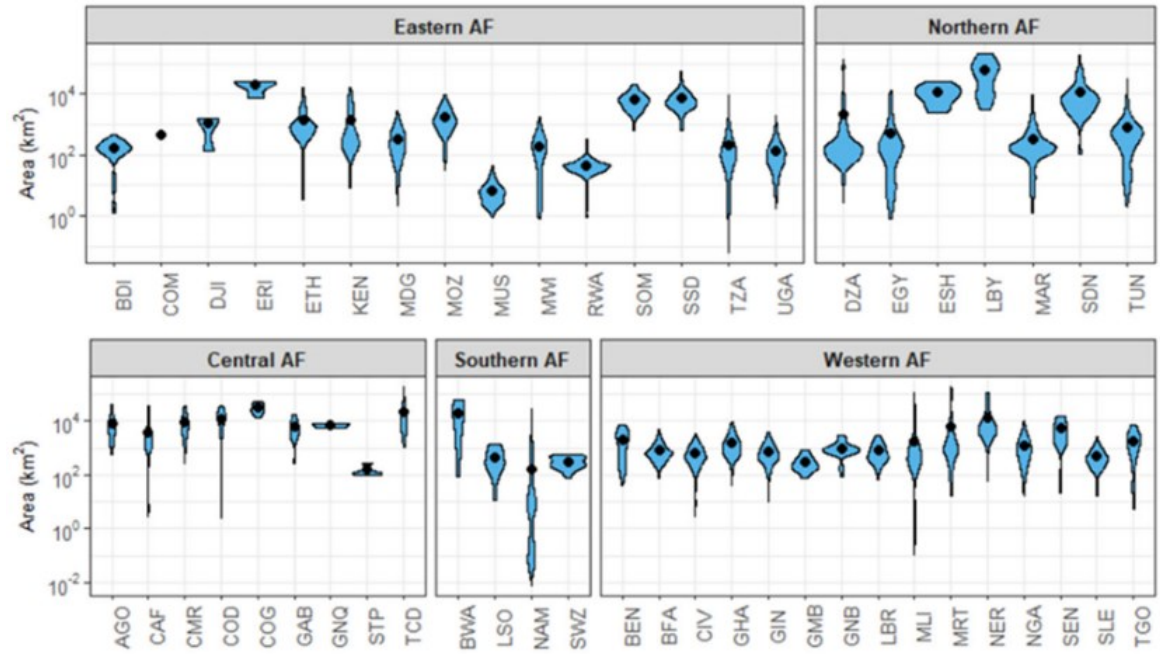

Figure 5-5. Violin plots illustrating the distribution of the sampled L1-units within each country in terms of their actual size (km^2). Black dot: mean value of distribution.

5.3.3 *Statistical Analyses*

Table 5-3 summarises the results of the first part of the statistical analyses displaying the average population density (Eq. 5-2) the MAE (Eq. 5-3), the %MAE (Eq. 5-4) and the RMSE (Eq. 5-5) for each country. A look at the results in terms of the %MAE indicates that the performance of the WSF2019-Imp layer has some minor variabilities across countries. For 80% of the countries located in the upper 10% and lower 90% percentiles (41 countries),

the %MAE values ranged from 13.95% to 32.10% with a standard deviation of ±5.32%. Twenty-one of the 41 countries reported %MAE values below or equal to ~20%, ten between ~20% and ~25%, and the last ten between ~25% and ~32%. The lower 10% of the countries reported %MAE values between 6.64% and 12.16%, and the upper 10% reported %MAE values between 35.13% and 72.22%. Within each main region, the lowest and highest %MAE values were reported for Mauritius (MUS,15.51%) and Comoros (COM, 72.22%) in Eastern Africa, Sao Tome and Prince (STP, 12.17%) and Gabon (GAB, 46.57%) in Central Africa, Western Saharan (ESH, 6.64%) and Morocco (MAR, 31.07%) in Northern Africa, South Africa (ZAF,16.72%) and Botswana (BWA, 38.24%) in Southern Africa, and Senegal (SEN, 7.82%) and Mauritania (MRT, 31.66%) in Western Africa, respectively. In terms of the MAE and the RMSE metrics, for all countries, the MAE remained below the average population density value. This behaviour was not the same for the RMSE metric, where for 24 countries, this value exceeded the average population density. According to the distribution of these metrics shown in Figure 5-6, the difference or ratio between the two metrics is relatively large for countries such as Algeria (DZA), Mauritania (MRT), Mali (MLI), Namibia (NAM), and Angola (AGO). These differences indicate that a large variability exists between the errors of the sampled L1-units within each country.

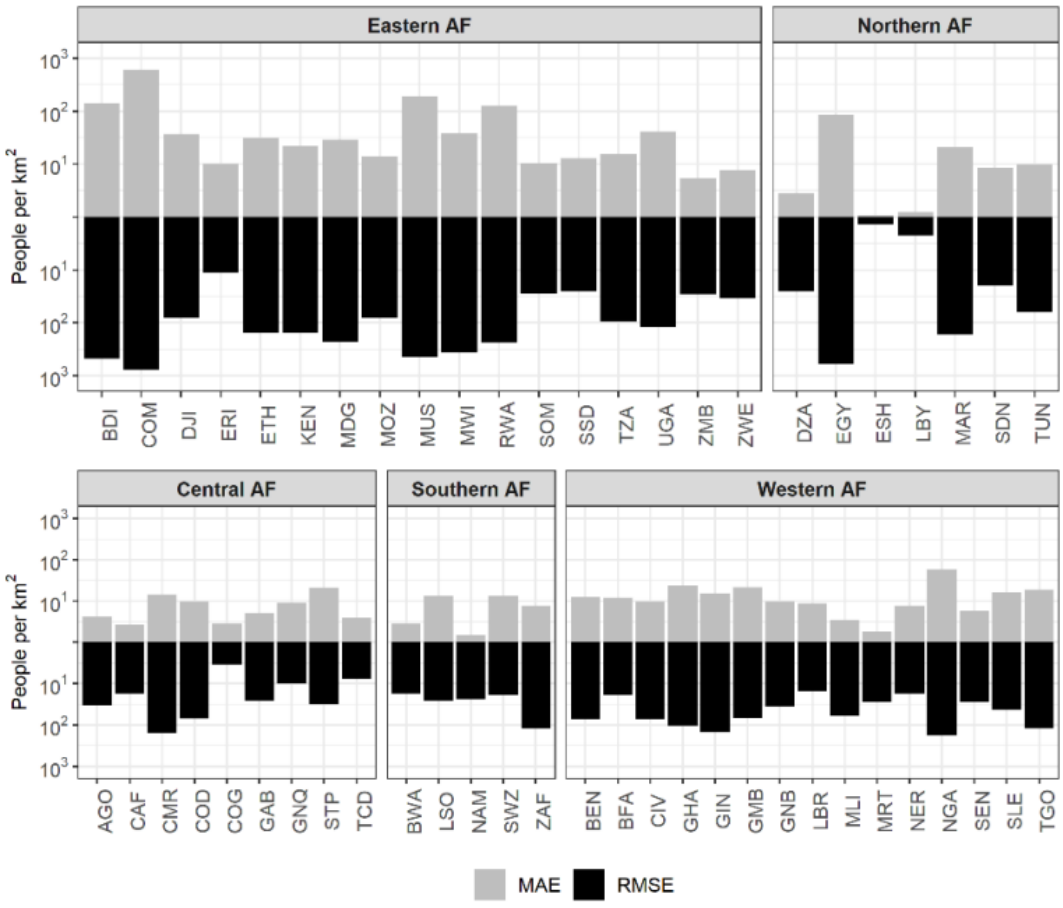

Figure 5-6. Bar plots of the distribution of the mean absolute error (MAE) (grey) and Root Mean Square Error (RMSE) (black) of the population density for each country (ISO) within each region.

Table 5-3. Statistical metrics for population density.

Eastern Africa											
ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE	ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE
BDI	86	549.84	24.95	137.1	480.5	**MWI**	283	230.4	15.86	36.54	358.5
COM	2	837.55	72.23	604.9	788.3	**RWA**	277	672.7	18.47	124.2	237.2
DJI	3	208.99	17.36	36.28	77.82	**SOM**	50	27.31	35.14	9.60	27.68
ERI	4	36.18	25.50	9.23	10.26	**SSD**	51	19.05	62.82	11.97	24.79
ETH	490	114.34	26.12	29.87	155.4	**TZA**	2428	71.92	20.55	14.78	95.83
KEN	229	101.59	20.98	21.32	153.7	**UGA**	918	234.3	17.07	40.01	120.3
MD	828	64.28	43.49	27.96	226.8	**ZMB**	99	25.13	17.98	4.52	28.23
MOZ	275	39.80	32.10	12.78	80.92	**ZWE**	59	36.30	18.88	6.85	33.05
MUS	105	1236.8	15.51	191.7	450.6						

Central Africa											
ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE	ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE
AGO	108	20.12	16.28	3.28	32.75	**GAB**	31	8.85	46.57	4.12	24.57
CAF	115	8.03	21.76	1.75	16.42	**GNQ**	3	36.39	22.97	8.36	8.96
CMR	37	44.54	31.03	13.82	154.3	**STP**	4	167.9	12.17	20.43	30.39
COD	120	36.72	24.14	8.86	68.36	**TCD**	41	11.89	26.19	3.11	6.96
COG	7	6.29	30.34	1.91	2.49						

Northern Africa					
ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE
DZA	102	12.13	15.75	1.91	24.44
EGY	225	593.57	13.96	82.86	602.4
ESH	16	2.38	6.64	0.16	0.49
LBY	13	2.00	16.49	0.33	1.35
MAR	107	65.41	31.07	20.32	166.3
SDN	85	27.96	27.40	7.66	19.13
TUN	176	55.43	16.00	8.87	63.06

Southern Africa					
ISO	n	$\overline{\text{Pop. D}}$	%MAE	MAE	RMSE
BWA	17	5.07	38.24	1.94	16.25
LSO	53	58.78	21.21	12.47	25.41
NA	3645	2.72	22.49	0.61	22.51
SWZ	35	66.11	18.89	12.49	18.45
ZAF	5647	41.41	16.72	6.92	119.9

Western Africa											
ISO	n	Pop.D	%MAE	MAE	RMSE	ISO	n	Pop.D	%MAE	MAE	RMSE
BEN	51	81.40	14.74	12.00	73.62	**MLI**	507	13.95	18.45	2.57	59.47
BFA	233	58.05	19.19	11.14	17.85	**MRT**	143	2.97	31.66	0.94	26.61
CIV	344	74.58	11.67	8.70	72.79	**NER**	44	27.93	24.08	6.73	16.25
GHA	113	103.44	21.61	22.36	103.1	**NGA**	515	210.9	26.74	56.42	182.4
GIN	226	51.09	28.53	14.57	150.3	**SEN**	29	61.85	7.82	4.84	26.93
GMB	25	255.63	8.16	20.86	66.46	**SLE**	106	99.47	15.23	15.15	40.94
GNB	26	57.22	15.37	8.80	33.80	**TGO**	25	128.7	13.88	17.86	117.4
LBR	89	32.24	24.90	8.03	14.05						

For the second part of the analyses, we first compared the actual and estimated population density of the validation units of each county. Figure 5-7 shows these distributions as scatterplots and marginal histograms, depicting the concentration of underestimated (grey) and overestimated (red) validation units. Each plot aggregates the information of all countries within one main African region, so that countries with a small number of units can also be represented. As observed in the tails of the histograms and the scatter of the validation units, there is a tendency of overestimating values below 10 ppl/km^2 and underestimating values >10000 ppl/km^2. Within the ranges where a larger number of validation units are concentrated, there seems to be a larger tendency towards underestimations; however, the distribution between underestimations and overestimations is somehow proportional across the different population density ranges.

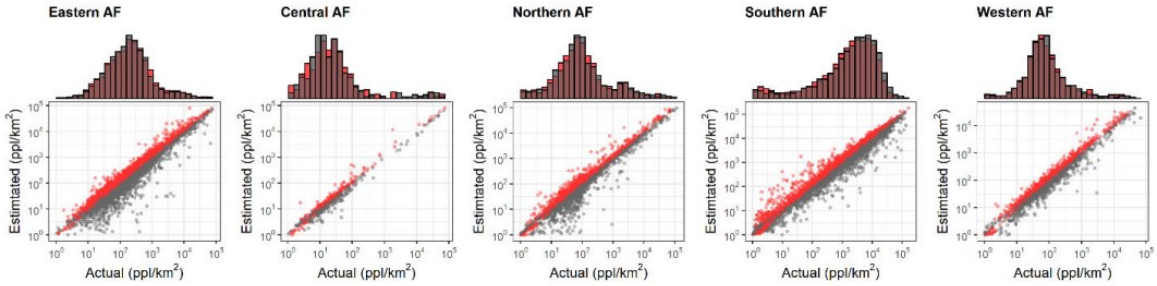

Figure 5-7. Scatter plots of estimated population density and actual population density for the validation units of each country within each region. Marginal histograms depict the concentration of underestimations (grey) and overestimations (red). Each panel shows the log population density.

To investigate the general patterns of error distribution within the validation units of each country, Figure 5-8 shows the percentage of validation units that fall within REE ranges of 20%. From here, it is possible to observe that all countries have at least 20% of their validation units within the >0%-20% REE range. For 32 of the 53 countries, this proportion increases to at least 50%, and up to 60% for 16 countries. Sao Tome and Principe (STP), Côte d'Ivoire (CIV), Senegal (SEN), and Togo (TGO) all have at least 75% of the validation within this range, followed by Gambia (GMB) with 100%. For most countries, the second largest proportion of validation units fall within the >20%-40% REE range, where at least ~10% but not more than ~30% of the validation units fall within this range. Some exceptions are Zimbabwe (ZWE), Libya (LBY), and Eritrea (ERI), where ~40%, ~50% and ~75% of the validation units fall in this range, respectively. Similarly, the proportion of validation units within the >40%-60% range is of at least ~1% for all countries, but no more than ~16%. Here, only Gabon (GAB), Eritrea (ERI), Congo (COG), Djibouti (DJI), and Equatorial Guinea (GNQ) report that ~20% up to ~30% of the validation units fall within this range. From here, 42 of the 53 countries report validation units within REE >60%-80%, with 29 of them reporting a proportion of less than 10% of the validation units, from 10% to 20% for 11 countries and 50% for Comoros (COM). Similarly, 35 of the 53 countries report validation units within REE >80%-100%, with 30 of them reporting a proportion of less than 10% of the validation units, from 10% to 20% for four countries, and 50% for Comoros (COM). Finally, 38 of the 53 countries report validation units with REE >100%, where 30 of them report a proportion of less than 5%; six from 5% to 7%; and ~10% to ~18% for Botswana (BWA) and Western Sahara (ESH), respectively.

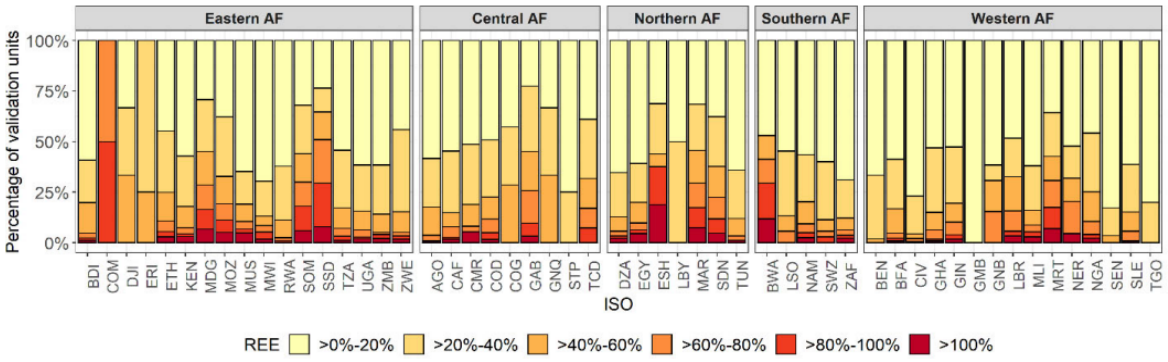

Figure 5-8. Stacked bar plots showing the percentage of validation units within each 20% Relative Estimation Error (REE) range.

To explore whether general trends of error distribution are delivered by the WSF2019-Imp layer, for the last part of the statistical analyses, we investigated the relationships among the REE, the population density and the SSC-Index of the validation units. Figure 5-9 shows the 2D-density plots for the validation units grouped according the different REE ranges. Here, we only present the results for a set of countries where validation units fell within each error range, and where the amount of validation units within each range was enough to produce the contour lines. For comparison purposes, the population density and the SSC-Index values were log-transformed.

From these plots, it is possible to observe that the distribution of the different ranges of REE can be found in the validation units with similar degrees of population density and SSC-Index. There are, however, some general tendencies that can be seen within each error range across most countries, which potentially explain the transitions from one REE range to another. These trends are summarised as follows:

1. For all countries, the majority of the validation units with REE between >0% and 40% are located in units with moderately high population densities and moderately high SSC-Index values (top-right quadrant);
2. Errors tend to increase as the population density increases and the SSC-Index decreases (shift towards the bottom-right quadrant);
3. Large errors (>100%) tend to be located in validation units with extremely high population density and extremely high SSC-Index values;
4. Most of the validation units with low population densities and low SSC-Index generally fall within error ranges of REE >60%.

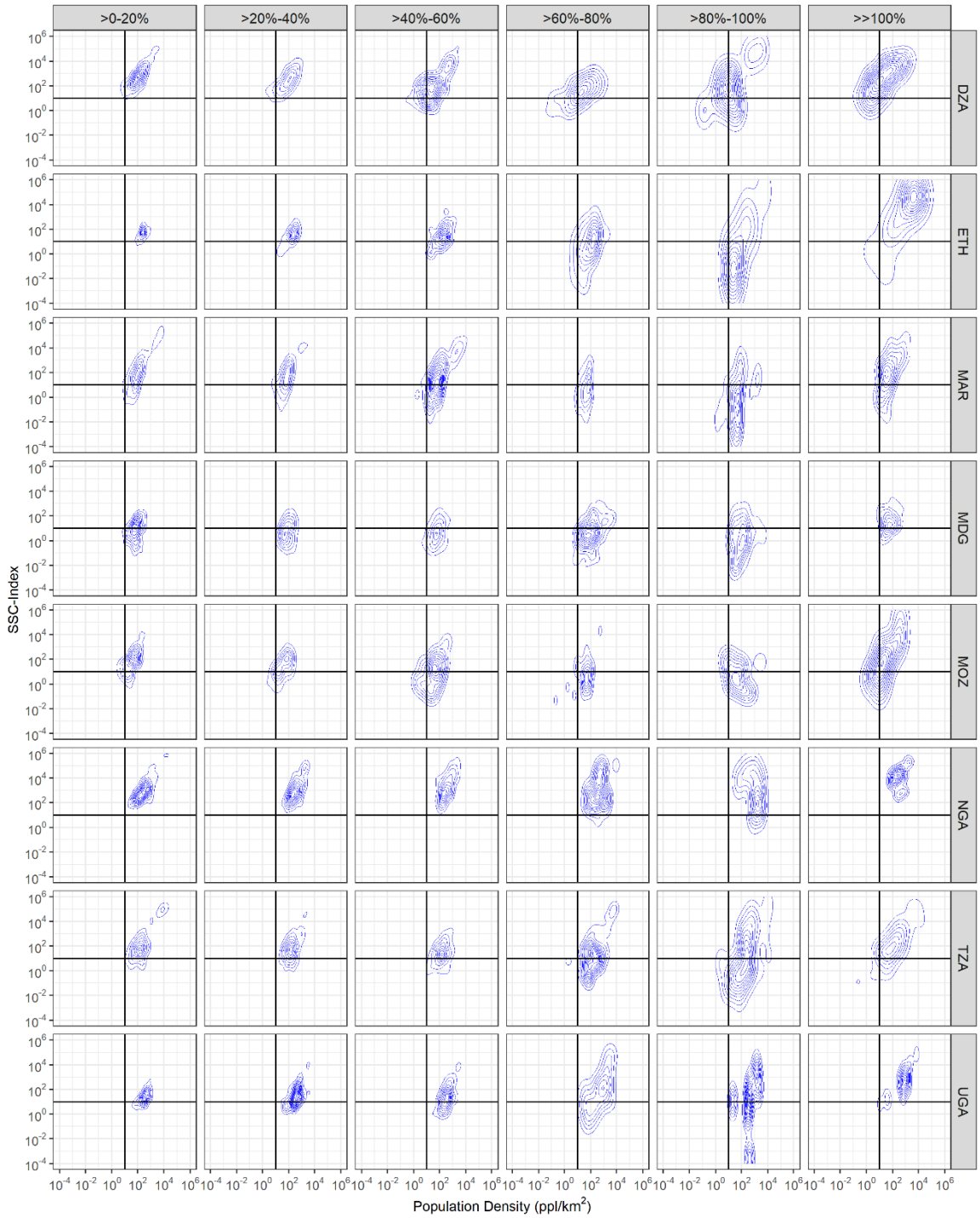

Figure 5-9.Two-dimensional-density plots showing the relationship among the population density, the SSC-Index, and the REE at different ranges of the REE.

5.4 Discussion

5.4.1 *WSF2019-Pop Dataset: Qualitative Assessment*

In this research, we presented the production of a new large-scale high-resolution gridded population distribution dataset for the African continent produced on the basis of the WSF2019-Imp layer and openly available subnational census/estimate-based population data. From Figure 5-1, it is possible to observe that the WSF2019-Imp layer

depicts a high likelihood between the estimated PIS values and the underlying built-up environment. High, medium, and low PIS values are proportionally assigned to every 10 x 10m pixel depending on the density of built-up and green spaces (e.g., parks and gardens) found within them. Here, the specific climate zone of the given region of interest does not seem to generate significant discrepancies in the final calculation of the PIS values, which indicates that the layer is potentially robust, consistent, and comparable across space.

From a practical point of view, the WSF2019-Imp layer provides a weighting framework that is calculated independently of other geospatial layers. This independence provides the final WSF2019-Pop dataset with several advantages over existing binary- and multi-layer products in the following ways. First, as seen from Figure 5-3, when employed as proxy in a dasymetric modelling approach, the WSF2019-Imp layer produces a heterogenous allocation of population counts that adheres to the variations of PIS values within the L1-units. From a strictly qualitative point of view, this asymmetric distribution of population has shown improvement over the homogeneous/uniform distribution delivered by the traditional binary dasymetric approach, revealing more detailed spatial distribution patterns. Previous comparisons presented in Stevens et al. (2020), Reed et al. (2018), and Palacios-Lopez et al. (2019) demonstrated, for example, that binary dasymetric modelling techniques tend to produce visible abrupt changes between census administrative units, whereas weighted approaches (including multi-layer and intelligent dasymetric) smooth these transitions. Second, compared to multi-layer products, another main advantage of the WSF2019-Imp layer is that it allows for the final WSF2019-Pop dataset to be more easily updated and replicated in other areas, without the extensive work that is needed for acquiring multiple geospatial layers of equal quality, extent, spatial resolution, and spatio-temporal coverage (Lloyd et al., 2019). Modelled with a single layer, the final population datasets are potentially more consistent across space in comparison to multi-layer products, in which the quality varies from location to location depending on the number and quality of geospatial datasets available for a given area (Dobson et al., 2000). In addition to this, as there are no other geospatial datasets involved in the production of the final WSF2019-Pop dataset, the dataset does not suffer from applicability restrictions derived from endogeneity issues (Leyk et al., 2019). For example, when land-cover data are used to model population datasets, these consequently should not be used for applications focused on understanding correlations between population and land-cover changes.

Notwithstanding these qualitative and practical advantages, as with any other global and regional population distribution dataset, the quality of the final WSF2019-Pop dataset is unavoidably affected by errors and anomalies derived from (1) the completeness and lack of functional characterization of the WSF2019-Imp layer, and (2) the quality of the input population data. Errors derived from the WSF2019-Imp layer include, first of all, a mismatch in the total population counts resulting from the absence of settlement pixels in some populated units. This type of error was identified in three countries: Mauritius (MUS), Morocco (MAR), and South Africa (ZAF). Within each country 8, 49, and 57 populated L1-units reported zero settlement pixels, with a total population sum of 43,931(3.4%), 337,647 (0.9%), and 230,829 (0.03%), respectively. Through a visual assessment of these countries, we were able to confirm the presence of built-up structures

within the reported L1-units. For the most part, the structures were very small and sparse, and were located in environments such as deserted areas or deep valleys. While this underestimation of built-up settlements was also reported for the population distribution datasets produced using the previous WSF2015-Density layer, the amount of validation units with no settlement pixels reported here is considerably less in comparison to the results presented in Palacios-Lopez et al. (2019). For example, in that research, where the African countries of Malawi and Côte d' Ivoire were also analysed, it was found that ~500 units were missing building structures. With the current WSF2019-Imp layer, these two countries reported full coverage, which indicates that the identification of settlement pixels has improved considerably as a result of the integration of S1 and S2 data into the underlying classification framework of the WSF2019 layer.

In the same context, an additional type of error derived from the WSF2019-Imp layer is the allocation of population counts to settlement pixels which are of non-residential use, such as industries, ports, and stadiums. The lack of functional characterization of existing built-up structures is still a persistent limitation that also affects other large-scale gridded population distribution products, such as the HRSL and the GHS-POP datasets. This qualitative limitation has additional quantitative implications, as non-residential, highly impervious surfaces will capture large proportions of the population counts, leading to underestimation in the surrounding settlement pixels. To solve this issue, machine learning methodologies, which are able to classify the residential status of urban buildings from LiDAR data at local scales (Lu et al., 2014; Xie & Zhou, 2017), are now applied to large territorial extents using satellite images (Lloyd et al., 2020; Sturrock et al., 2018). For example, in the recent work presented by Lloyd et al. (2020), the authors combine satellite image-derived building footprint and OSM-label data to classify buildings as residential and non-residential in Democratic Republic of Congo and Nigeria. Their results show that the method classifies buildings with accuracies from 85% to 93% across both countries. Overall, the potential for the large-extent applicability and transferability of this new method will more likely influence the field of large-scale population modelling in the near future.

From the qualitative errors derived from the input population data, the first kind of error is related to the presence of unpopulated units within the population data, where a considerable number of settlement pixels were detected, and where actual populated areas exist. Freire et al. (Freire et al., 2018) recently addressed this issue, explaining that while the CIESIN census database is the most detailed, complete and coherent database available at global scales, it still presents some anomalies which are derived from the source population statistics (e.g., National Statistic Offices). In this research, ~ 2099 L1-units were reported as unpopulated, and while some of these units are actually non-enumerated units, some of them still cover large built-up areas according to Freire et al. (Freire et al., 2018). In terms of the mapping outcomes, for these L1-units,"NoData" values were assigned to the final settlement pixels resulting in visual inconsistencies in the final population distribution maps. While de-facto no quantitative errors exist in the final population maps in relation to the total input population, the missing counts of these areas can have relevant impacts on further analyses, highlighting the importance of full disclosure on the uncertainties present in the final datasets. To the best of our knowledge,

other top-down large-scale gridded population datasets that are based on the CIESIN data currently present the same anomalies.

Finally, the age and spatial detail of the input population data are other factors that without a doubt affect the quality of the final population distribution maps. As seen from Table 5-1, for many African countries, the last official population data are from more than 10 years ago, resulting in potentially inaccurate estimates, a low number of administrative units, and outdated administrative boundaries. To be sure, significant improvements have been made in the frequency of population data collection in Africa. Countries such as Burkina Faso, Kenya, Madagascar and Malawi, for example, carried out their last population census between 2018 and 2019, while approximately 80% of the African countries conducted their last census between 2005 and 2015. However, limited financing and poor budgeting strategies for data collection are concurrent issues in many African countries, which result in incomplete or outdated demographic statistics (Tuholske et al., 2019). Under any context, from policy making to scientific research, acquiring up-to-date population data at the highest available resolution should remain the main priority (Balk et al., 2006).

5.4.2 *WSF2019-Pop Dataset: Quantitative Assessment*

To evaluate the relative accuracy, effectiveness, and stability of the WSF2019-Imp layer, for each country, statistical analyses were carried out in two ways: (1) at the country level, where aggregated metrics were computed to allow for cross-country comparisons; (2) at the validation unit level, where individual metrics were computed to establish correlations between the error distribution and the built-up environment. Together, the results presented in Table 5-3, Figure 5-8 and Figure 5-9 show that WSF2019-Imp produces a systematic distribution of error, where estimation accuracies remain relatively consistent among and within countries. At the country level, the population distribution maps of 80% of the countries reported %MAE values between ~15% and ~32%, with a standard deviation of ±~5%. At the validation unit level, for 32 out of 53 countries, at least half of the validation units reported REE values between 0% and 20%, followed by errors of >20%-40% and >40%-60%. In terms of the error distribution, REE values between >0% and 40% were concentrated in validation units with medium ranges of population density and medium ranges of SSC-Index values, with errors increasing as the SSC-Index decreased and the population density increased. Large estimation errors (>100%) were found in validation units with extremely high population densities and extremely high SSC-Index values.

On that note, whether the presented accuracies can be considered low or high is still a debatable topic (Calka & Bielecka, 2019). Only a few studies have classified the accuracy results into levels or degrees, but a single threshold of reliability has not yet been established. For example, in the uncertainty quantification of the GRUMP dataset for Poland, Da Costa et al. (2017) established that units deviating < 20% from the actual population can be considered as "reliable data" and >20% considered as having "medium reliability". In the accuracy assessment of the GRUMP, GPW, and WorldPop datasets for China presented by Bai et al. (2018), the authors established that REE errors < ±25% can be considered as "accurately estimated", between ±25% and ±50% as "under or overestimated", and from ±50% to >±100% as "greatly under- or overestimated".

Following these criteria, in this research, 25 to 36 countries would be considered as "reliable" or "accurately estimated", 15 would have "medium reliability", and two would be found to be poorly reliable. Consequently, within each country and for most countries, the largest proportion of validation units would be "reliable" or "accurately estimated", while the second largest would have "medium reliability".

In general, the analyses presented showed that the accuracy of the WSF2019-Imp layer follows the premise established by Stevens et al. (2020), who stated that high accuracies in population modelling can be expected when built-up area datasets are proportionally coherent with the population density. The lowest estimation errors in all countries were, for the most part, located in those validation units where the SSC-Index showed a linear correlation with the population density. Notably, as soon as these two factors started to decorrelate, the REE (mainly errors of overestimation) started to increase. Exceptions to this rule applied only to extremely populated units with extremely dense built-up environments, where the largest REE >100% (mainly errors of underestimation) corresponded to units delineating small cities within the countries.

Overall, the general trends found here are derived from limitations that are consistent across all existing top-down large-scale gridded population datasets. The distribution of error can be explained by four main factors summarised as follows: (1) errors of omission in the identification of built-up settlements in rural settings, which causes the allocation of large population counts into only a few settlement pixels; (2) the potential overestimation of population totals in units with a low number of settlement pixels derived directly from the outdated input population data (Tiecke et al., 2017); (3) the lack of characterisation of the built-up environment (residential/non-residential), which causes the underestimation of population counts in surrounding settlement pixels; and (4) the lack of height and volume (3D) information on the building structures, which causes underestimations, especially in areas with a mix of low- and high-rise buildings.

Nevertheless, there are, however, additional factors that affect the estimation accuracies which are unrelated to the WSF2019-Imp layer. These uncertainties are mainly derived from a) the nature of the input population data and b) the sampling process. First, for the majority of countries, there were not enough L1-units to produce significant sample sizes (Table 5-2). To be able to meet the requirements of a random sampling process that, in parallel, was capable of selecting 2/3 spatially united L1-units as validation units, it was necessary (and unavoidable) to produce sample sizes below 100 units for almost half the countries. Therefore, countries with an already low number of large sampled L1-units, such as Western Sahara (ESH), Senegal (SEN), Gambia (GMB), and Sao Tome and Principe (STP), reported some of the lowest %MAE values, simply due to the small differences in the sizes between the coarser input units used for modelling and the fine units used for validation. This is known as MAUP (Duque et al., 2018), which in the context of this research was difficult to avoid without compromising the random sampling process. Second, it goes without saying that different samples for each country will produce different results. This particular limitation was pointed out by Stevens et al. (2020) and Sinha et al. (2019), who demonstrated that the RMSE and MAE metrics are sensitive to the generated sample in terms of their size and the spatial autocorrelation of the sampled units. Moreover, additional research has also shown that when the sample sizes are very small (4-10 samples), aggregated metrics, such as the RMSE and the MAE, cannot produce

robust results (Chai & Draxler, 2014), highlighting the importance of using individual metrics, such as the REE employed here.

In this context, it is important to understand that the accuracies reported here are constrained to the employed validation method. The final usability and effectiveness of the WSF2019-Pop dataset will also be determined by the accuracy of population estimates extracted in the context of different application scopes. As an example, Figure 5-10 shows the differences that could be obtained from extracting population counts at very local scales from the WSF2019-Pop dataset and mock-datasets produced using the WSF2019-binary layer. Coastal areas in Morocco and Tanzania illustrate the final population distribution maps produced by each WSF2019 product in medium-to-high urbanised environments. The yellow polygons represent arbitrary areas where population counts were extracted.

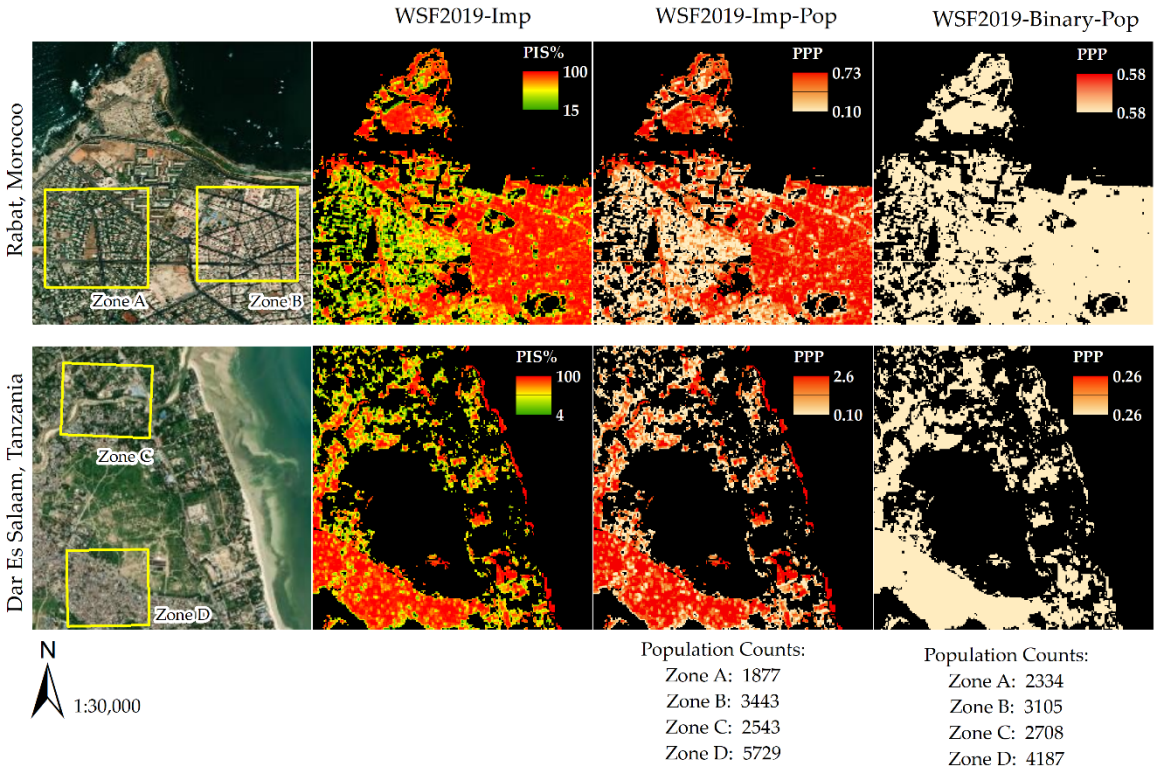

Figure 5-10. Final population distribution maps produced using the WSF2019-Imp layer and WSF2019 layer for the coastal areas of Rabat, Morocco, and Dar Es Salaam, Tanzania. From each map, population counts were extracted for Zones A, B, C, and D, respectively. Colour ramps depict values in the current extent.

As seen from Figure 5-10, extracted population estimates can vary greatly from one dataset to the other. Low impervious areas, such as Zone A and Zone C, allocate less population counts in the WSF2019-Imp layer in comparison to the binary approach. The opposite applies for highly impervious areas, such as Zone B and Zone D, where the binary approach allocates less people per pixel in comparison to the WSF2019-Imp layer. Differences between population datasets range from ~~150 to ~1500 people. Depending on the application field where the datasets are employed, the magnitude of these differences can have significant implications, especially in studies where accurate population counts are necessary, such as emergency response or risk assessments.

The results presented here are simply used to provide complementary qualitative and quantitative insights into the capabilities of the WSF2019-Imp layer. A complete validation of the results would require real application cases and the availability of reference data. Nonetheless, considering the very local nature of many socio-environmental phenomena (Smith et al., 2019), it could be expected for the WSF2019-Pop dataset to potentially produce more accurate population estimates compared to currently available binary products (e.g., HRLS and GHS-POP datasets) and coarse spatial resolution products (e.g., WorldPop and LandScan datasets).

On that note, in this research, we did not include quantitative accuracy comparisons against other available large-scale population grids, as many of the current products do not have datasets representing the year 2019. The closest datasets from the GPWv4, HRSL, and GHS-POP products, for example, represent population distributions for the years 2015 or 2020. Under these conditions, the temporal disagreement among the different datasets would have introduced a certain level of uncertainty too complex to account for, especially when independent validation data do not exist to verify the results. Here, the lack of validation data is also the reason restricting comparisons with other 2019 population grids, namely, the 2019-WorlPop and 2019-LandScan datasets. Accordingly, comparisons to other built-area datasets (e.g., the 2019-WorldPop building-patterns (Dooley et al., 2020), the 2015-HRSL settlement mask (Tiecke et al., 2017), or the 2020 GHSL layer (Pesaresi et al., 2016)) and modelling methods (e.g., areal-weighting, binary dasymetric, or multi-layer dasymetric) were not included for two main reasons. For the first case, with the validation of the WSF2019-Imp layer in terms of settlement identification still pending, the differences in population estimations between built-area datasets derived from the omission or commission of settlement pixels would not have been possible to address. This means that to properly interpret the differences between the outputs of each built-area dataset, first, we need to know which dataset is more accurate and complete in its own framework. For the second case, comparisons to methods such as areal-weighting and binary-dasymetric were not included, as previous research has already shown that weighted dasymetric mapping is by far more accurate than these two methods (Palacios-Lopez et al., 2019; Sorichetta et al., 2015; Stevens et al., 2015a). For the case of multi-layer approaches, comparisons were not included, as the overall objective focuses on exploring the particular advantages or limitations of employing the layer on its own.

5.5 Summary

The present study focused on systematically evaluating how accurate and effective the novel WSF2019-Imp layer is in the production of a new large-scale gridded population dataset—the WSF2019-Population dataset (WSF2019-Pop). Employed as a single proxy in a dasymetric mapping approach, the WSF2019-Imp layer was used in combination with an open archive of census/estimate-based population data to construct population datasets for each African country.

Results of our qualitative and quantitative assessment indicate that the main advantages of the WSF2019-Imp layer as a proxy for large-scale population modelling, are derived from its robustness, spatial consistency, independent weighting framework, and improved spatial resolution. These characteristics allow the layer to produce spatially detailed population datasets that could potentially be more accurate than binary-derived

products, on the one hand, and that could potentially overcome the local qualitative variations, applicability restrictions, and production complexities of multi-layer-derived products, on the other. The results of our statistical analyses additionally confirm that the WSF2019-Imp layer is capable of producing a systematic distribution of error that remains stable independently of the quality and spatial granularity of the input population data. Overall, the WSF2019-Imp layer reported %MAE values between ~15% and ~32% for close to 80% and REE below 20% for up to 50% of the validation units of most countries. Following the pre-established classification criterion, these error ranges indicate that the WSF2019-Imp layer produces, for the most part, "accurately estimated" population datasets. Notwithstanding these promising results, there are, however, some limitations that still need to be addressed, as high errors of underestimation and overestimation are still present in the final WSF2019-Pop dataset. In particular, the omission of settlement pixels in rural settings and the lack of information on the use and height of the building structures are factors that currently affect the quality and accuracy of the final population datasets. In this context, it is expected that with the upcoming validation of the WSF2019 products, these remaining uncertainties can be assessed, allowing a focus on further technical improvements to the WSF2019-Pop dataset. Considering this, future research will also include quantitative comparisons with other built-area datasets and population grids, and the integration of other geospatial layers into the modelling framework, such as the newly developed GUF3D dataset (Esch et al., 2020). Furthermore, as the semi-automatic methods presented here are completely transferable, future research will also focus on expanding the accuracy assessment of the WSF2019-Pop dataset to other countries. Within this outlook, the WSF2019-Pop dataset will also be evaluated in the framework of different application fields, especially those related to risk assessment and emergency response. Here, additional comparisons with other population grids will be performed to assess their accuracy, usability, and limitations.

To conclude, the WSF2019-Population dataset developed in this research represents an important contribution to the field of large-scale gridded population mapping, helping to improve and enhance the spatial granularity and local detail of census population data needed for a wide range of research and governmental applications. In the context of risk assessment, the WSF2019-Pop dataset is currently used by the World Bank to identify all localities on the African continent with an estimated population of >10,000 inhabitants. Additionally, the population at risk with respect to urban hazard zones, such as seismic, landslides, flooding, and storm surge, is determined based on a combination of the WSF2019-Pop layer and risk data, such as those provided by the Think Hazard! datasets (GFDRR, 2007). Open and free provision of the WSF2019-Pop dataset is foreseen through the Urban Thematic Exploitation Platform (https://urban-tep.eu) and the EO Center Geoservice (https://geoservice.dlr.de).

Chapter 6

6. Towards an Improved Large-Scale Gridded Population Dataset: A Pan-European Study on the Integration of 3D Settlement Data into Population Modelling

The following section represents an adapted version of the third peer-reviewed research article of this cumulative thesis. Some chapters have been modified to include only relevant information that has not already been presented in the previous chapters (e.g. introduction and materials). Where needed, the reader will be referred to the corresponding sub-chapters.

- Palacios-Lopez, D., Esch, T., MacManus, K., Marconcini, M., Sorichetta, A., Yetman, G., Zeidler, J., Dech, S., Tatem, A. J., & Reinartz, P. (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. https://doi.org/doi.org/10.3390/rs14020325.

Abstract

Large-scale gridded population datasets available at the global or continental scale have become an important source of information in applications related to sustainable development. In recent years, the emergence of new population models has leveraged the inclusion of more accurate and spatially detailed proxy layers describing the built-up environment (e.g., built-area and building footprint datasets), enhancing the quality, accuracy and spatial resolution of existing products. However, due to the consistent lack of vertical and functional information on the built-up environment, large-scale gridded population datasets that rely on existing built-up land proxies still report large errors of under- and overestimation, especially in areas with predominantly high-rise buildings or industrial/commercial areas, respectively. This research investigates, for the first time, the potential contributions of the new WSF3D dataset in the field of large-scale population modelling. First, we combined a RF classifier with spatial metrics derived from the WSF3D to predict the industrial versus non-industrial use of settlement pixels at the Pan-European scale. We then examined the effects of including volume and settlement use information into frameworks of dasymetric population modelling. We found that the proposed classification method can predict industrial and non-industrial areas with overall accuracies and a kappa-coefficient of ~84% and 0.68, respectively. Additionally, we found that both, integrating volume and settlement use information considerably increased the accuracy of population estimates between 10% and 30% over commonly employed models (e.g., based on a binary settlement mask as input), mainly by eliminating systematic large overestimations in industrial/commercial areas. While the proposed method shows strong promise for overcoming some of the main limitations in large-scale population modelling, future research should focus on improving the quality of the WFS3D dataset and the

classification method alike, to avoid the false detection of built-up settlements and to reduce misclassification errors of industrial and high-rise buildings

6.1 Introduction: Problem Statement

While gridded population datasets produced on the basis of the binary and imperviousness WSF layers (2015 and 2019) have already shown some qualitative and quantitative advantages over other existing products (Palacios-Lopez et al., 2019; Palacios-Lopez et al., 2021), to the best of our knowledge, no assessment that reports on the suitability of the WSF3D dataset in the framework of large-scale top-down population modelling has been undertaken. Therefore, in this research, we examine the utility of the WSF3D dataset as a single proxy for top-down, large-scale population modelling. This examination was carried out following a two-step approach briefly described as follows:

1. In the first step, we investigate if the WSF3D dataset can be used to effectively identify and eliminate large industrial/commercial areas from the built-up environment, which in the past have been reported as major sources of under/overestimation errors in population modelling. To this end, we present a methodology that combines a RF algorithm with a set of spatial metrics derived solely from the WSF3D dataset to predict the "Industrial" versus "Non-Industrial" class of built-up settlements across 38 countries located in Europe. Reference datasets to collect training data and validate our classification results are produced using the Urban Atlas 2018 dataset. Overall, the main objectives of this part of the research are to build an automatic classification model for each country, and to produce binary classification maps that can be used to refine population distribution datasets.
2. In the second step, we evaluate the accuracy of population distribution maps produced on the basis of the new WSF3D data and the integration of information on industrial/non-industrial land use from step 1. For this assessment, we specifically employ the information of the WSF3D building fraction (BF) and building volume (BV) layers, downscaled to 12 m (see sub-chapter 6.2.2), to generate population distribution maps using a weighted dasymetric mapping approach together with 2020 census-derived population data. We then compared the outcomes to the results achieved with a binary settlement mask as input. Overall, the main objectives of this part of the research are to investigate if improvements can be gained from the inclusion of settlement information related to the use and/or volume of building structures, and to assess under which circumstances these improvements are more significative, and how they correlate with the quality of our classification maps.

6.2 Material and Methods

6.2.1 *Study Area*

The study area of our analyses covered the 38 countries in the European Union Area (EEA), including the member state countries of the European Union (EU), the countries of the European Free Trade Association (EFTA), the West Balkans countries, Turkey and the United Kingdom, as illustrated in Figure 6-1.The selection of this study area was primarily guided by the parallel availability of standardized land-use data from the Urban Atlas

dataset (see sub-chapter 6.2.3 for more details) and contemporary high-resolution population data needed for model training, population modelling and validation. In addition, the unique characteristics of each country in terms of the 3D morphology and functional use of the built-up environment, provided with an excellent set up in which to test whether the contributions of the WSF3D dataset in the field of large-scale population modelling were systematically consistent across variable landscapes.

On the one hand, binary classification maps differentiating industrial and non-industrial built-up settlements (see sub-chapter 6.2.6) were produced for the EEA38 countries using training, tests and validation datasets collected from a number of Functional Urban Areas (FUAs) (see sub-chapter 6.2.3) spread across all countries (red points). Tasks related to population modelling and comparative analyses in population estimates (see sub-chapter 6.2.6), on the other hand, were carried out only in 30 countries, excluding Austria, Cyprus, Hungary, Latvia, Malta, Netherlands, Portugal and Romania (crossed-out polygons) where no open population data were available (see sub-chapter 6.2.4).

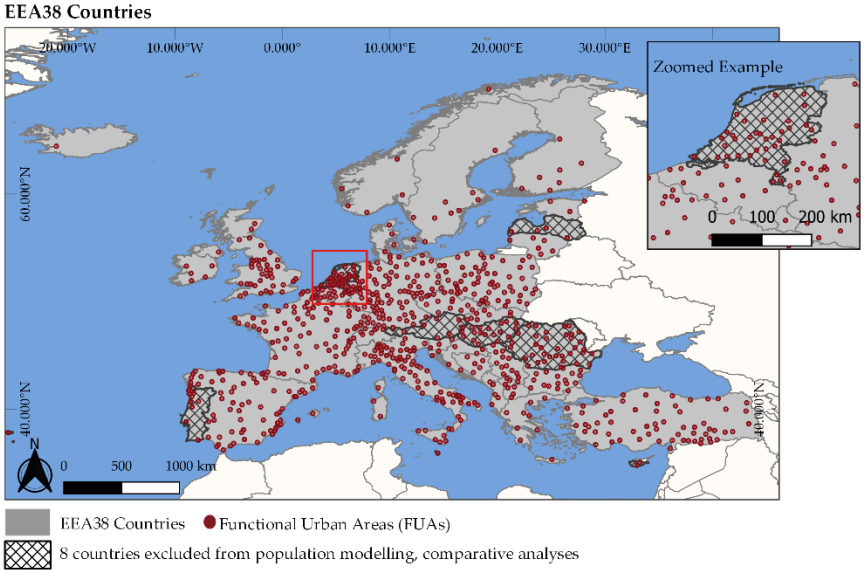

Figure 6-1. Study area covering the EEA38 countries (grey-polygons), with Functional Urban Areas (red points) and countries excluded from population modelling (8-crossed-out polygons).

6.2.2 *World Settlement Footprint 3D Dataset*

For each of the EEA38 countries, the layers employed for this research were provided ready-to-use by DLR. Here, we include a short description of the production process of each layer, focusing specifically on the 12 m versions displayed previously presented in Figure 3-3:

1. Building Height (BH): The 12 m BH layer represents a spatial disaggregation of the standard 90 m WSF3D BH layer, which was derived by measuring the height variations of vertical edges related to building edges (BE) in the 12 m TDX-DEM within the settlement areas defined by the WSF-Imp layer. The height is reported in meters (m) in the final product.
2. Building Fraction (BF): This layer was produced by quantifying the built-up coverage at 12 m derived from the joint analysis of the WSF-Imp, TDX-AMP

and BE. The values in the final product range from 0–100, measured in percentage.

3. Building Area (BA): This layer was derived by multiplying the BF times the area of each ~12 m grid cell (~144 m^2 at the equator). The area is reported in square meters (m^2) in the final product.
4. Building Volume (BV): This layer was derived by multiplying the BH with the area of the 12 m pixels. The total volume is expressed in cubic meters (m^3) in the final product.
5. Building Mask (BM): This layer represents the binary version of the BF layer, where all pixels PIS > 0 have been converted to values of 1.

6.2.3 *European Urban Atlas Dataset*

The European Urban Atlas is a dataset produced and supported by the ESA and the European Environment Agency (EEA). It provides standardized vector LU/LC data covering more than 700 Functional Urban Areas (FUAs) and their immediate rural vicinity, with more than 50,000 inhabitants across the EEA38 countries. The cartography of the Urban Atlas polygons is based on image interpretation of very high-resolution satellite data (2 m or 4 m spatial resolution). The LU/LC nomenclature is composed of 27 classes distributed in five major groups (Level 1) as described in Figure 6-2. Within each FUA, LU/LC polygons have a minimum mapping size of 0.25 ha for classes with class code 1, and 1 ha for classes with class code 2 to 5, which are spatially distributed in heterogenous patterns (Batista e Silva et al., 2013).

For each country in our study area, the 2018 versions of 13 datasets were downloaded from the Copernicus land monitoring services website (Copernicus Programme, 2022). Accordingly, for each country, Table 6-1 summarizes the number of FUAs employed in this research.

Table 6-1. Number of available FUAs per country.

Country Name	ISO	No. FUAs	Country Name	ISO	No. FUAs	Country Name	ISO	No. FUAs
Albania	ALB	4	Un. King	GBR	40	Norway	NOR	6
Austria	AUT	6	Greece	GRC	9	Poland	POL	58
Belgium	BEL	11	Croatia	HRV	7	Portugal	PRT	11
Bulgaria	BGR	17	Hungry	HUN	19	Romania	ROM	35
Bos. and Her	BIH	5	Ireland	IRL	5	Serbia	SRB	13
Switzerland	CHE	10	Island	ISL	1	Slovakia	SVK	8
Cyprus	CYP	2	Italy	ITA	81	Slovenia	SVN	2
Czechia	CZE	15	Lithuania	LTU	6	Sweden	SWE	9
Germany	DEU	96	Luxembourg	LUX	1	Turkey	TUR	62
Denmark	DNK	4	Latvia	LVA	4	Kosovo	UNK	3
Spain	ESP	69	Macedonia	MKD	4		**Total**	753
Estonia	EST	3	Malta	MLT	1			
France	FRA	83	Montenegro	MNE	1			
Finland	FIN	7	Netherlands	NLD	35			

Level 1	Level 2	Level 3	Level 4
1 Artificial Surfaces	1.1 Urban Fabric	111 Continuous Urban Fabric (S.L >80%)	11100 Continuous urban fabric (S.L >80%)
		112 Discontinuous Urban Fabric (S.L 10%-80%	11210 Discontinuous dense urban fabric (S.L. 50%-80%) 12220 Discontinuous medium density urban facric (S.L. 30%-50%) 12230 Discontinuous low density urban fabric (S.L. 30%-50%) 12240 Discontinuous very low density urban fabric (S.L. <10%)
		113 Isolated Surfaces	11300 Isolated Surfaces
	1.2 Industrial, commerical,public, military, private and transport units	121 Industrial, commerical,public, military, private units	12100 Industrial, commercial, public, military and private units
		122 Road and rail network and associated land	12210 Fast transit roads and associated land 12220 Other roads and associated land 12230 Railways and associated land 12300 Port areas 12400 Airports
	1.3 Mine,dump and construction sites	131 Mineral extraction and dump sites	13100 Mineral extraction and dump sites 13300 Construction sites 13400 Land without current use
	1.4 Artificial, non-agricutural vegetated areas	141 Green Urban Areas	14100 Green Urban Areas
		142 Sports and leisure facilities	14200 Sports and leisure facilities
2 Agricultural Areas			
3 Natural and (semi-) natural areas			
4 Wetlands			
5 Water			

Figure 6-2. Urban Atlas dataset nomenclature.

6.2.4 *Population Data and Administrative Boundaries for 2020*

CIESIN provided upon request the subnational administrative boundaries and the corresponding 2020 census/estimate-based population data for 30 of the EEA38 countries in our study area. For each subnational boundary two types of population data estimates were provided: (i) census/estimated-based numbers calculated using annual exponential growth rates, and (ii) United Nation-adjusted estimates (Freire et al., 2018), which were used in this research. The collection and standardisation of the CIESIN data has been previously described in Chapter 2.3.

Table 6-2. Summary of 2020 UN-adjusted census-based population data for each country, including 3-letter ISO-Code, census or estimation year, total population and highest administrative level plus number of units (admin. level/count).

ISO Code	Census Year	UN2020 Estimation	L1-unit /Count	ISO Code	Census Year	UN2020 Estimation	L1-Unit /Count
ALB	2011	2,935,145	3/365	IRL	2011	4,874,291	4/18,488
BEL	2014	11,634,330	4/589	ISL	2010	342,140	2/73
BGR	2011	6,884,343	2/263	ITA	2011	59,741,323	3/317
BIH	2013	3,758,147	3/141	LTU	2011	2,794,897	2/60
CHE	2010	8,654,270	3/2514	LUX	2011	605,110	4/139
CZE	2011	10,573,292	3/6249	MKD	2010	2,088,374	2/78
DEU	2011	80,392,210	3/11,185	MNE	2011	625,837	1/21
DNK	2010	5,775,633	3/2135	NOR	2011	5,490,394	2/429
ESP	2011	43,931,099	3/7931	POL	2011	38,407,264	4/2500
EST	2011	1,295,158	3/4587	SRB	2011	6,641,618	5/4616
FIN	2011	5,554,886	2/320	SVN	2010	2,075,010	3/5969
FRA	2009	65,720,028	5/36,602	SWE	2010	10,120,395	3/14,605
GRB	2011	66,700,124	6/232,296	TUR	2010	82,255,778	2/957
GRC	2011	10,825,409	5/6121	UNK	2011	2,031,895	1/37
HRV	2011	4,162,498	2/556				

6.2.5 *Industrial and Non-Industrial Classification of Built-up Settlements Using Random Forest*

In the field of land-use mapping, research has shown that spatial metrics derived from remotely sensed data combined with a RF algorithm can be used to effectively to identify different land-use/land-cover classes on the ground (Du et al., 2015; Grippa et al., 2018; Ruiz Hernandez & Shi, 2018; Zhang et al., 2017b). On the one hand, spatial metrics quantitatively describe the configuration of the landscape in terms of the structure (e.g., shape, size, number, density) and the arrangement of elements (e.g., buildings) across space (Herold et al., 2005). At a specific scale and resolution, differences in these metrics are normally an indicator of different land-use classes, thus allowing the production of LU/LC classification maps. In the framework of this research, for example, previous studies have shown that spatial metrics such as the average, median and standard deviation of the density, height and volume of building structures can be used to discriminate industrial (and large commercial) buildings from residential and other non-industrial buildings (Ma et al., 2015). Overall, industrial buildings are generally larger, higher and denser in comparison with residential buildings, allowing their identification through different RS techniques.

The RF classifier, on the other hand, is a robust ensemble machine algorithm that has proven to be a powerful tool capable to perform accurate supervised classification tasks (Rodriguez-Galiano et al., 2012). Essentially, the RF classifier builds multiple decision trees, each one constructed using a random subset of the training data. Each individual tree delivers a class prediction, and the class with most votes becomes the model's prediction. Compared to other classification algorithms which are also known to produce robust classifications in remote sensing problems (e.g., support vector machines SVM), the RF performs equally, with the advantage that is easier to implement as it requires less parametrization (Pal, 2005).

Following these premises, in this research we combined an RF classifier with a set of spatial metrics derived solely from the WSF3D layers to predict the "industrial" versus "non-industrial" class of the built-up settlement pixels in each country of our study area. The whole workflow for the production and validation of the final binary classification maps is shown in Figure 4, followed by a detailed description of the main steps in the following sub-chapters. Unless indicated otherwise, all processing steps were carried out using GDAL-commands in a Linux environment and Python programming language and libraries.

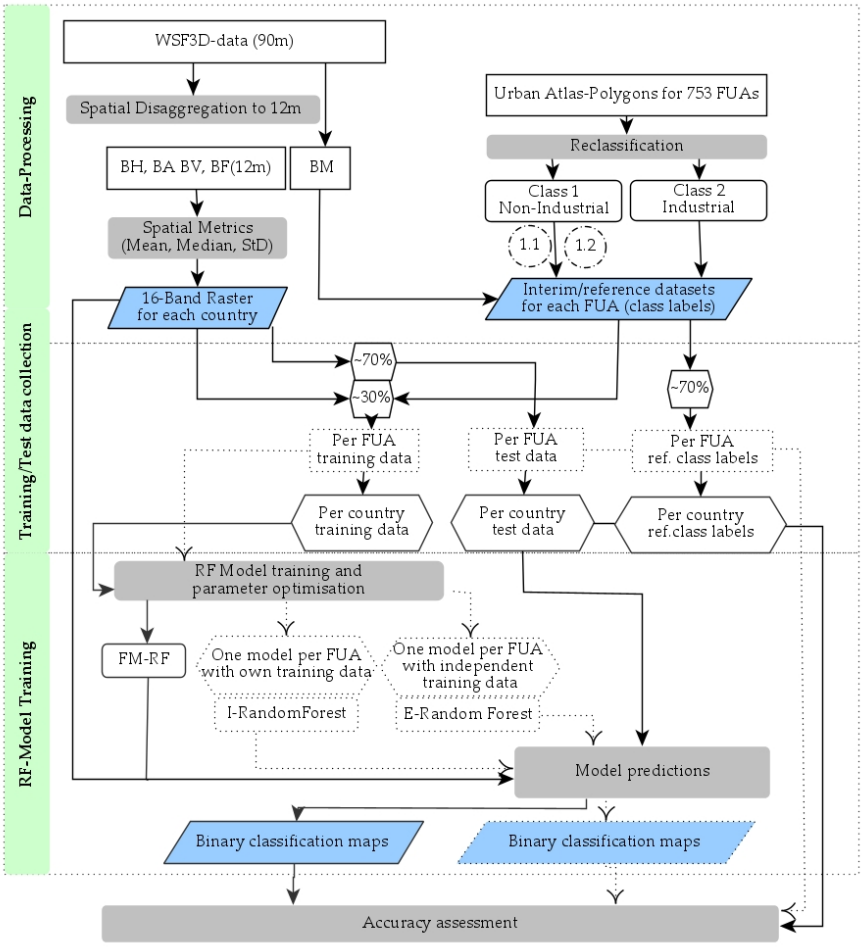

Figure 6-3. Workflow for model training, classification and validation of industrial and non-industrial binary classification maps using RF.

6.2.5.1 Derivation of Spatial Metrics

In this study, a total of 16 spatial metrics derived from the WSF3D dataset were used as variables to train the RF models for each country. These included the four basic components of the WSF3D dataset: BA, BH, BF and BV, and 12 additional metrics based on distributional statistics calculated over a 25 × 25 window size (300 × 300 m): mean, median, and the standard deviation. The window size was chosen to ensure that the surroundings of the potentially smallest "non-industrial" areas were evaluated, using as reference the minimum size employed in the cartography of the Urban Atlas datasets (1 ha, 100 × 100 m) (ESA, 2016). For each country, a 16-band raster composite was generated, which included the total of all parameters derived.

6.2.5.2 Interim Reference Datasets

Following methodologies similar to the ones presented in (Khryashchev et al., 2018; Leinenkugel et al., 2019), interim and reference datasets needed to automatically collect training and test data, and to validate the final classified maps, were produced using the Urban Atlas datasets. For each country the interim and reference datasets were produced by classifying the built-up pixels of the WSF3D building mask (BM) layer according to the class of the Urban Atlas polygons their centroid fell into. As seen in Figure 6-4, the final interim/reference datasets covered a little more than 50% of the total built-up area for six countries, between 40 and 50% for 24 countries, and between 25% and 35% for eight countries.

Using the reclassification scheme presented in Table 6-3 for each FUA, built-up pixels were classified as "non-industrial" (Class 1) if their centroids were within those UA polygons with Level 2 codes: 111, 112, 113, 121, 122, 131, 141 and Level 1 codes: 2, 3, 4 and 5, respectively (see Figure 6-2). On the one hand, according to the Urban Atlas-Mapping Guide (ESA, 2016), polygons within the classes 111, 112 and 113 encompass built-up structures that have a predominant residential component, with the occasional presence of mix-use buildings. Polygons within the remaining classes, on the other hand, encompass built-up structures that have industrial, commercial, public and military use, or small built-up structures with non-residential use located in the proximity of roads and train stations, or within construction sites, gardens, zoos, parks or marinas.

As such, in the particular case of class 121, to exclude polygons representing large industrial and commercial units (e.g., energy plants, production sites, retail parks) only polygons with areas below 10 km^2 were considered as "non-industrial". This threshold was selected after a visual assessment of more than 50% of the Urban Atlas polygons across all FUAs, using VHR optical imagery. Consequently, all built-up pixels whose centroids were within Urban Atlas polygons with class code 121 and areas >10 km^2 were classified as "industrial" (Class 2), including those located within class code 122 polygons, corresponding to ports and airports.

Table 6-3.Reclassification scheme using for reference and interim datasets.

Class	Major Classes	Urban Atlas Codes	Sub-Classes
1	Non-industrial	111: 11100	1.1 High-dense residential
		112: 11210, 11220, 11230, 11240 113: 11300 121: 12100 (area < 10 km^2) 122: 12210, 12220, 122230 131: All 141: All Level 1: 2, 3, 4, 5	1.2 Low-dense residential + Small non-residential
2	Industrial	121: 12100 (area >= 10 km^2) 122: 12300, 12400	

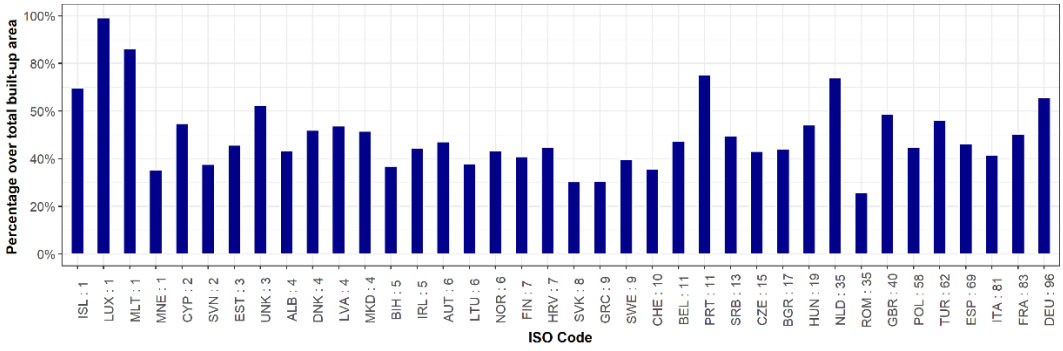

Figure 6-4. Bar plots depicting the percentage of built-up area covered by the interim/reference datasets per country. Countries are ordered according to the total number of available FUAs.

At the same time, for the purpose of training data collection (Section 2.5.3), the built-up pixels within the "non-industrial" class were further differentiated into two sub-classes, namely "High-dense residential" (Class 1.1) and "Low-dense residential + Small non-residential" (Class 1.2) as noted in Table 6-3 and Figure 6-5a–c. This sub-categorization was simply carried out to ensure that enough samples were collected within areas where built-up structures could potentially present similar metrics to the "industrial" class, such as the case of high-rise buildings within the 111 class.

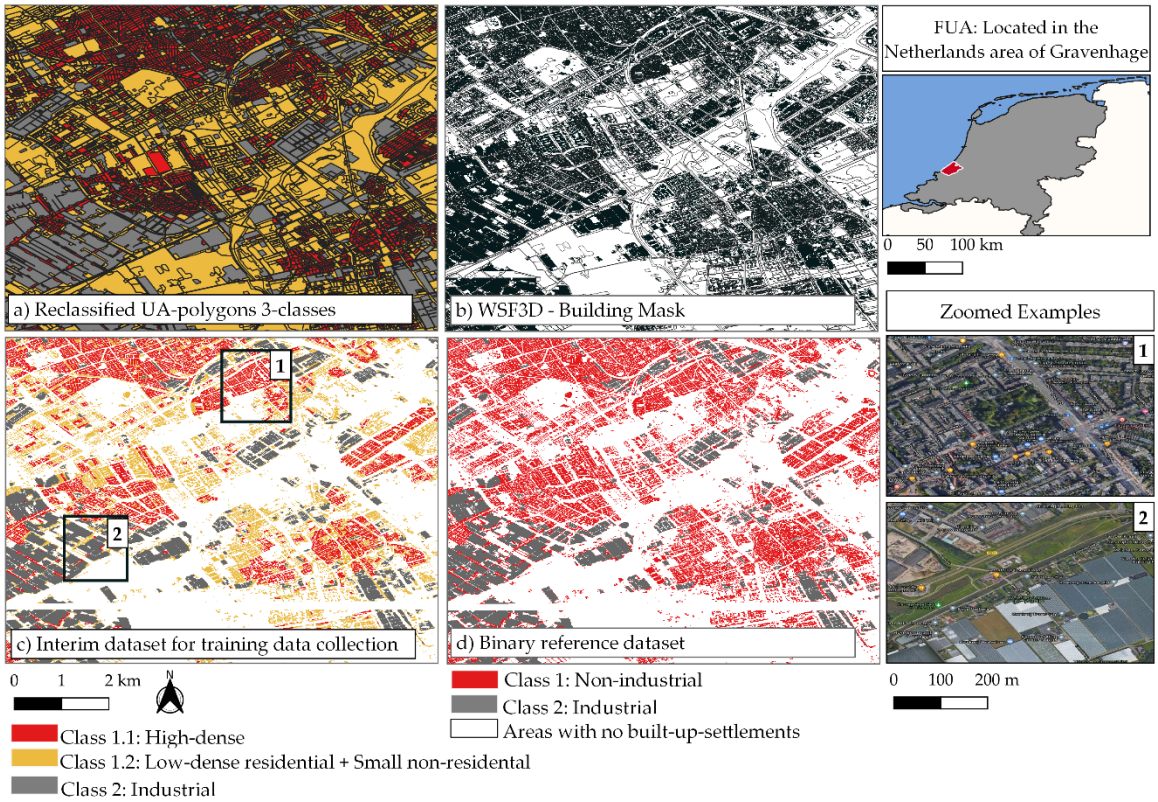

Figure 6-5. Example of reference dataset for a FUA located in Netherlands, with (a) reclassified Urban Atlas (UA) polygons according to Table 3 using three sub-classes, (b) WSF3D building mask overlapping Urban Atlas (UA) polygons, (c) interim dataset for training data collection and (d) binary reference dataset.

6.2.5.3 Automatic Training Data Collection

Once the interim reference datasets for each FUA within a country were produced, these were used to automatically collect point training data by means of a proportionally

allocated stratified random design. From each class (Figure 6-5c), we collected 1000 samples (or class labels), which resulted in a total of 3000-point samples per FUA per country. The location of these samples was then used to extract the 16 spatial metrics from each country's 16-band composite, to finally construct the input training datasets for model training (see sub-chapter 6.2.5.4). In this research, the selected sample size represented the maximum size in which the training data for the less represented class, in the less represented FUA, was less or equal to 30% of total available pixels. This means that for each class in each FUA, 70% or more pixels were left as independent test data for model prediction and reference class labels for validation purposes (see sub-chapter 6.2.5.5). These ratios are inline to the ones employed in Zhang et al. (2017a), for a similar assessment.

6.2.5.4 Model Training

To produce the final binary classification maps of "industrial" versus "non-industrial" classes, for each country, a single RF model (hereinafter referred to as Full Model (FM-RF) was built using Python's scikit-learn libraries (*Scikit-learn. Machine Learning in Python*). As described in Figure 4, for each country a single FM-RF was trained using the entire set of training data collected from all the FUAs belonging to a particular country. Here, it is important to clarify that all of the 16 spatial features derived from the WSF3D were used for model training, without the implementation of feature selection, as internal results (not included here) showed that removing features did not improved model predictions. Accordingly, in order to produce the most robust predictions, during the training process, model hyperparameters were independently optimized for each country, and the final model was used to predict over (1) the entire test data and (2) entire country. Hyperparameter selection was carried out using scikit-learn's "GridSearchCV" functionalities, to select the number of trees, the maximum depth of a tree, the minimum number of samples required to split an internal node or the minimum number of samples required at a leaf node.

6.2.5.5 Quantitative Accuracy Assessment

To evaluate the classification accuracy of the FM-RF models, for each country, the predictions made over the test data were compared against the reference data by means of a confusion matrix. For this assessment, built-up pixels predicted as classes 1.1. and 1.2 were first merged into a single class, representing the final "Non-industrial" class to match the final binary reference datasets (Figure 6-5d). From here, for a balanced accuracy assessment, an equal number of pixels were randomly selected for each class, equal to the size of the least represented class (excluding training data points). This sample was then used to derive common statistical accuracy metrics including the Overall Accuracy (OA) and Cohen's kappa coefficient (k), and the Producer's and User's Accuracy (PA, UA) for each class, respectively.

However, considering that a proper accuracy assessment can only be performed in the areas where reference data is available (see Figure 6-4), to provide a general overview of the relative accuracy that can be expected in the country-wise classification maps, we produced two alternative models to (1) analyse the spatial transferability of our RF-models and (2) compare their performance against "optimal" scenarios (Figure 6-3, dashed process). First, following recent methodological guidelines (Jin et al., 2018; Orynbaikyzy

et al.; Ploton et al., 2020), for each country, *k* external models (E-RF, *k* = no. of FUAs) were trained and optimized by excluding the training data of one FUA at the time. In each iteration, the FUA that was left out was used as a spatially independent test area, and the accuracy of its classified map compared against the reference data. Accordingly, for each FUA an internal model (I-RF) was trained and optimized using only each FUAs' training data. This model was then applied to the test data of same FUA and the accuracy of the classified map compared against the reference data. The results of the I-RF and E-RM were then aggregated at the country level and compared against the accuracy of the FM-RF.

6.2.6 *Population Modelling and Comparative Analyses*

Figure 6-6 illustrates the workflow followed for the assessment of population models built on the basis of the original building mask (BM), building fraction (BF) and building volume (BV) layers of the WSF3D, and a combination of building volume and industrial settlement use information (exclusion of industrial settlements BV-IS).Accordingly, the main steps included the production of gridded population distribution datasets using top-down dasymetric modelling techniques, followed by a well-established quantitative accuracy assessment. These steps are described in more detailed in the following sub-chapters.

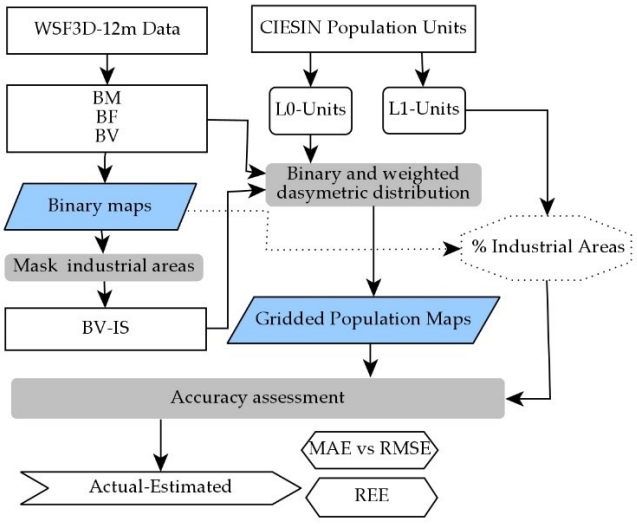

Figure 6-6. Workflow for the population datasets and comparative analyses.

6.2.6.1 Top-Down Dasymetric Modelling

A total of four gridded population maps for each country were modelled using a dasymetric binary technique or weighted technique, where the 2020 UN-adjusted population counts from L0-units were redistributed into the built-up settlement pixels of the WSF3D datasets. First, gridded population datasets were produced on the basis of the building binary mask (BM) as proxy layer using Eq. 6-1. In this technique, each built-up pixel within a given L0-unit $Pop_{(p \in L0)}$ has a weight of one $W_p = 1$, resulting in each pixel being allocated an equal number of people. This approach is similar to the one employed by the HRSL and the GHS-POP datasets and produces a homogenous distribution of the population within each L0 unit, preserving the original population counts of the input unit.

$$Pop_{(p \in L0)} = Pop_{L0} \frac{W_p}{\sum_{p=1}^{n}(W_p)} \begin{cases} W_p = 1, & BM \\ 0 < W_p \leq 100, & BF, \\ 0 < W_p \leq \max(p_v), & BV, BV - IS \end{cases} \quad \text{Eq. 6-1}$$

Second, gridded population datasets were produced on the basis of the different continuous layers, including the building fraction (*BF*), building volume (*BV*) and building volume minus industrial settlements (*BV-IS*). Here, unlike in the binary technique, each built-up pixel is allocated a proportion of the input unit's total population Pop_{L0}, relative to their density ($0 < W_p \leq 100$) or volume ($0 < W_p \leq \max(p_v)$) pixel values p_v. This approach produces heterogenous population distributions, comparable to the ones provided by the WSF-Pop datasets, preserving the original population counts of the input unit.

6.2.6.2 Quantitative Accuracy Assessment

As shown by previous studies ((Bai et al., 2018; Freire et al., 2016; Palacios-Lopez et al., 2021), in the field of large-scale population modelling, a "true-validation" of gridded population datasets remains a very challenging task due to the lack of high-resolution ground-truth data (e.g., population counts at the pixel level) needed for an independent quantitative assessment. Therefore, in order to test the accuracy of population distribution datasets, the research community has developed an empirical validation method that measures the internal accuracy of population distribution maps in terms of "how well and plausibly populations were distributed" (Leyk et al., 2019).

Overall, in this method a series of statistical analyses are performed using the differences between population counts extracted from maps modelled using a coarser level of the administrative units (here, L0 units or input units), and the population counts of the finest administrative units (here, L1 units or validation units). Here, the main assumption is that input population data is accurate, and as such, the resulting empirical analyses only measure the relative accuracy, effectiveness and stability of the employed disaggregation method and/or proxy layers.

For this research we applied the same validation method to systematically compare the quantitative accuracies of the four different population datasets described in the previous section. First, as explained in sub-chapter 6.2.6.1, we have chosen to model the final gridded population datasets using the national level administrative units for all countries (L0 units). This was carried out to reduce the bias that is normally introduced when the input units and validation units have a similar size (Hay et al., 2005b; Palacios-Lopez et al., 2019), on the one hand, and to be able to evaluate each country with the largest number of validation units possible (L1 units), on the other (Palacios-Lopez et al., 2021). Second, using the L1 units (validation units), from each population dataset we extracted the estimated population counts using the Zonal Statistic tool of ArcGIS. For each country and each gridded population dataset, the reported differences between the actual and estimated values were then used to derive the following error metrics:

$$MAE_c = \frac{\sum_{i \in L1=1}^{n} |pop_a - pop_e|}{n} \quad \text{Eq. 6-2}$$

$$\%MAE_c = \frac{MAE_c}{\overline{pop_c}} \quad \text{Eq. 6-3}$$

$$RMSE_c = \sqrt{\frac{\sum_{i\in L1=1}^{n}(|pop_a - pop_e|)^2}{n}} \quad \text{Eq. 6-4}$$

$$REE_{i\in L1}^{n} = \frac{pop_a - pop_e}{pop_a} * 100 \quad \text{Eq. 6-5}$$

On the one hand, for a given country, the Mean Absolute Error (Eq. 6-2) and Root Mean Square Error (Eq. 6-4) ($MAE_c, RMSE_c$) both measure the average of the absolute differences between the actual (pop_a) and estimated population (pop_e) counts of the L1 units. However, unlike the *RMSE*, which penalizes larger errors by squaring the differences, the *MAE* weights each error equally, allowing the identification of outliers in the data. On the other hand, the percentage *MAE* (Eq. 6-3), which is the *MAE* divided by the average population of each country, allows the comparisons across countries by removing the bias of different population totals and number of L1 units. This metric can be used to determine if the errors/improvements generated by the different proxy layers are similar and systematic, or if different behaviours are observable across countries.

The Relative Estimation Error (REE, Eq. 6-5) measures the error in each L1 unit in proportion to their actual population counts. By reducing the bias caused by differences in population counts across L1 units, this metric is useful in comparing the distribution of errors within countries and across countries produced by each covariate layer. In this research, the REE was used in two ways:

(i) Firstly, for each country we calculated the proportion of industrial areas found within the L1 units according to the final binary classification maps. For all L1 units with the same amount of industrial presence, we then calculated the average REE produced by each gridded population map.
(ii) Secondly, similar to (Schug et al., 2021), we grouped the L1 units into REE ranges of 25% according to the results produced by each gridded population map. For each country, we then calculated the percentage of each countries' total population found in these units.

6.3 Results

6.3.1 *Industrial and Non-Industrial Binary Classification Maps*

To evaluate the overall performance of our automated FM-RF models for the classification of industrial versus non-industrial built-up settlements, Figure 6-7 together with Figure 6-8 show the results of comparing the percentage of total area covered by each class according to the reference (R) and the predicted (P) classified maps. First, at the country scale, as seen from the distribution of the per-class percentage share presented in Figure 6-7, the proportion of built-up settlements pixels predicted as industrial (grey bar) and non-industrial (red bars) types by the FM-RF models were fairly comparable to those reported by the reference maps. As observed, for most countries, there are slight overestimations in the predicted industrial share. Overall, according to the Pearson's correlation (r) values, the agreement between the reference and predicted maps at the

country level ranged from 0.40 (MLT) to 0.65 (LTU), with an average value of 0.54, a median of 0.55 and a standard deviation of ±0.05, reported at the Pan-European scale.

At the FUA level, a closer look at the distribution of the absolute differences in class proportions presented in Figure 6-8, reveals that for most countries (24/38) at least 75% of the FUAs' predicted maps (IQR range box) showed differences in class proportions below 10% compared to their respective reference maps. For the remaining 25% of the maps, and for 9 of the 14 left countries, differences did not exceed 15%. As such, differences in class proportions between the predicted and reference maps equal or larger than 20% (but lower than 40%) were only found in a small number of outlier-FUAs in BEL, NDL, TUR, ITA and FRA.

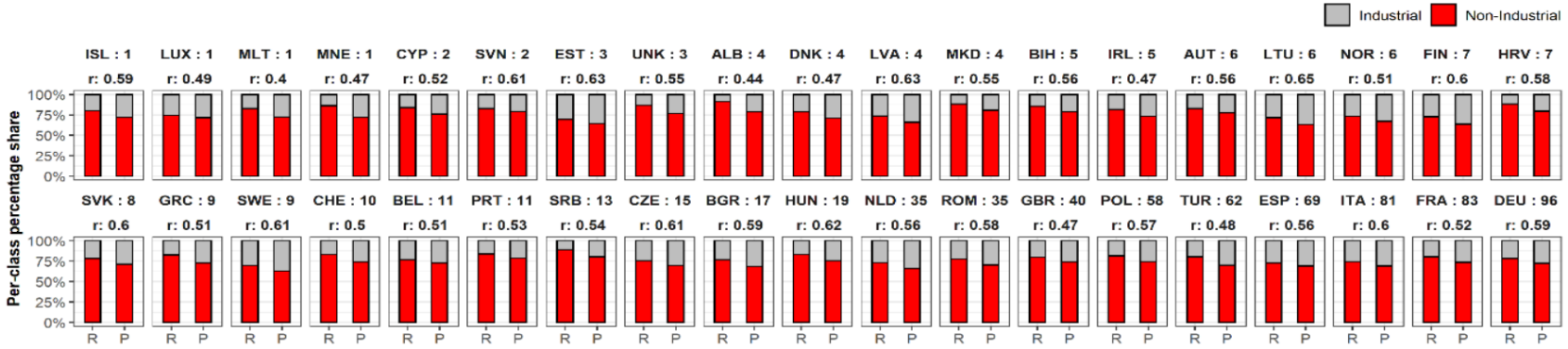

Figure 6-7. Stacked bar plots showing the Persons' correlation (r), and percentage shared of each class (grey: industrial, red: non-industrial) within the reference (R) and predicted (P) classification maps at the country level. Countries order according to the no. of available FUAs.

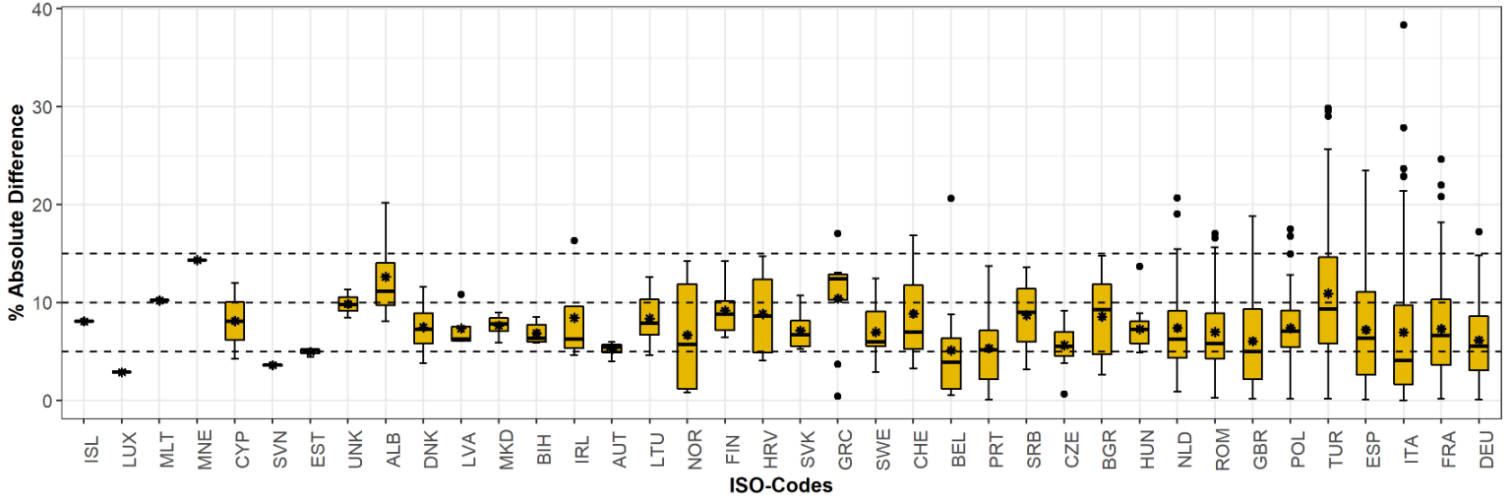

Figure 6-8. Box plots of the distribution of absolute difference in class proportions for all FUAs within a country. Middle line of each boxplot showing the position of the median difference, asterisk (*) showing the position of the average and yellow boxes showing the 75% inter-quantile range (IQR). Countries ordered according to no. of available FUAs

To complement the aforementioned results, Figure 6-9 shows the accuracy metrics reported by the confusion matrix analysis, for the FM-RF (black), the I-RF (red) and the E-RF (blue) models, respectively. Focusing first on the results produced by the FM-RF, results show that in terms of the overall accuracy (OA), for most countries, the accuracies were higher than 85%. The highest value of 90% was reported in LTU, whereas the lowest value of 75% was reported in MLT. Overall, at the Pan-European scale, the OA reported by the FM-RF reached an average of 84.32%.

Accordingly, for most countries (22/38), the Cohen's kappa coefficient (k) values remained higher than 0.7, with the highest value of ~0.82 reported in LTU. Fifteen out of the remaining sixteen countries reported k-values higher or close to 0.6, with the lowest value of ~0.52 reported in MLT. At the Pan-European scale, the k value reported by the FM-RF reached an average of 0.68.

In terms of the Producer's Accuracy (PA) for class 1: non-industrial and class 2: industrial, results reveal that in 37 of the 38 countries, the PA of class 1 was close or higher than 90%, with the lowest metrics of 85% reported in MLT and the highest value of ~95% reported in ISL, SVN, NOR, HRV, GRC and HUN, respectively. Conversely, the PA of class 2 was higher or equal to 80% for 23 countries and between 70% and 80% for the rest of the countries. The highest value of ~87% was reported in LTU and the lowest value of ~72% was reported in MLT and GBR, respectively. Overall, at the Pan-European scale the PA of the non-industrial and industrial classes, reached an average of 92% and 79%, respectively. Accordingly, results reveal than in 21 of the 38 countries, the User's Accuracy (UA) of class 1 was higher or equal to 80%, between 70% and 80% for nine countries, and below 70% for eight countries. The highest value of ~87% was reported in LTU, and the lowest value of ~65% was reported in MLT and GBR, respectively. Conversely, the UA of class 2 was higher, equal or close to 90% for 36 of the 38 countries. The highest value of 95% was reported in HRV, and SVK, and the lowest value of ~82% was reported in MLT, respectively. Overall, at the Pan-European scale, the UA of the non-industrial and industrial classes, reached averages of 76% and 91%, respectively.

Finally, from a comparative point of view, the results produced by the different RF-models reveal that the FM-RF models performed fairly comparable to the I-RF models, while marginally improved over the E-RF models across all countries. As observed, in terms of the OA, the difference between the FM-RF and the I-RF was only 2%, and between the FM-RF and the E-RF only 3% at the Pan-European scale. Accordingly, the kappa coefficient dropped 0.04 between the FM-RF and the I-RF, while it improved 0.02 points between the FM-RF and the E-RF. In terms of the PA and UA the trends are similar, where the largest difference of 4% can be seen between the PA of the I-RF and the FM-RF models.

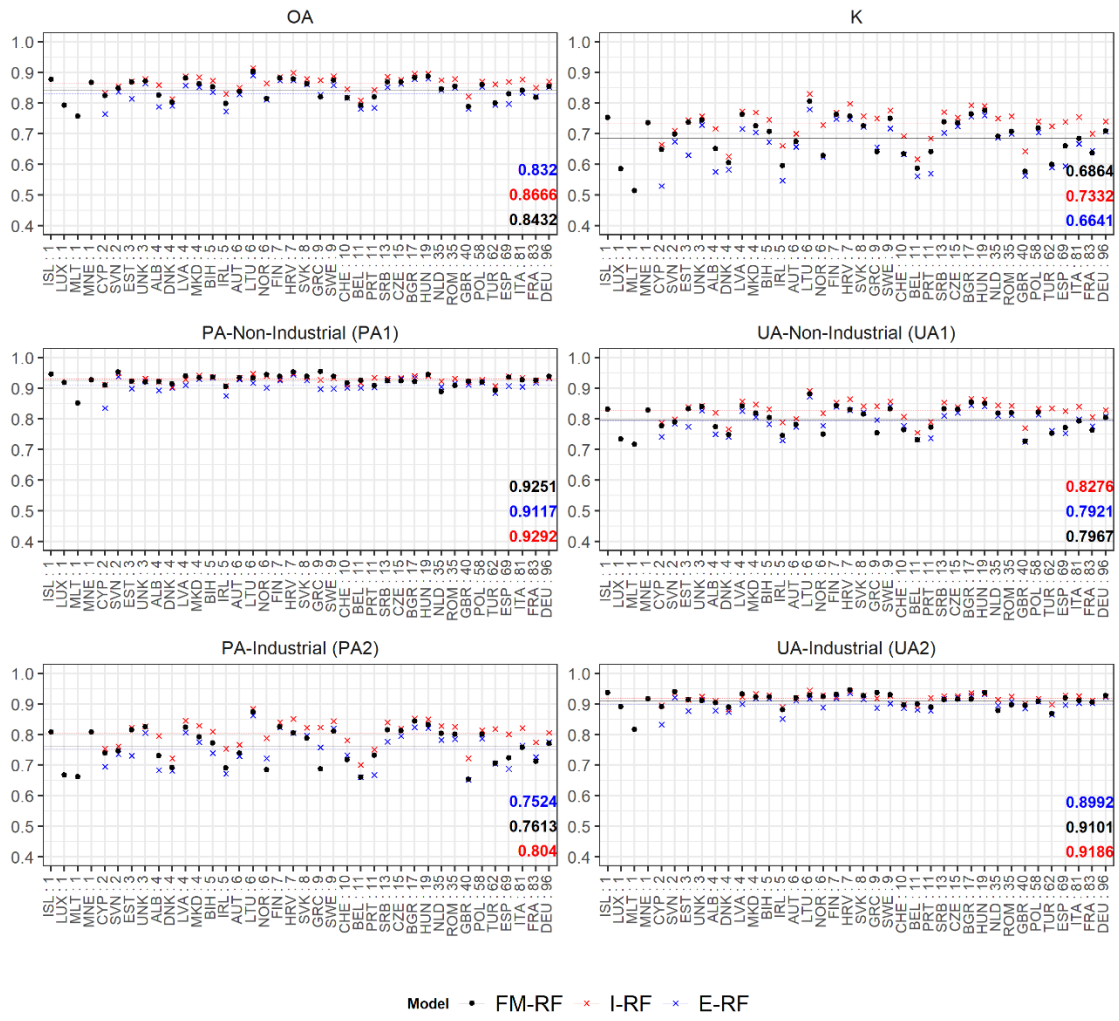

Figure 6-9. Confusion matrix average accuracy metrics reported in each country on the basis of the FM-RF (black points), I-RF (blue crosses) and E-RF (red crosses) models. First row: overall accuracy (OA) and kappa coefficient (K). Following rows: producers "accuracy (PA) and Users" accuracy (UA) for class 1: Non-Industrial (left), class 2: Industrial (right). Pan-European results represented by the average lines and bold numbers. Countries ordered according to no. of available FUAs.

6.3.2 *Population Modelling: Output Gridded Population Maps*

Figure 6-10 shows several extracts of the output gridded population maps produced using the four proxy layers (BM, BF, BV and BV-IS) and the national level (L0 units) administrative units. Each map was produced at a spatial resolution of 12 m, representing the estimated amount of people per pixel for the year 2020. To visually inspect the thematic differences between maps, we focused on representative areas with a mix of non-industrial and industrial areas, including ports and commercial centres. As observed, gridded population maps produced on the basis of the BM proxy layer delivers homogenous distributions of the population, where each pixel holds the same amount a people. Gridded population maps produced on the basis of the BF, BV and BV-IS, on the other hand, offer more spatial heterogeneity, adhering to the relative changes in the density and volume values, respectively. Without the inclusion of settlement use information, it is possible to observe that maps produced with the BF and BV layers allocate a large proportion of the population in areas identified as Industrial in the BV-IS

maps. The BV proxy layer, however, seems to minimize this effect, by allocating a higher proportion of the population in the dense non-industrial areas, as opposed to the BF, where the allocation of people in dense non-industrial areas and industrial areas appears balanced

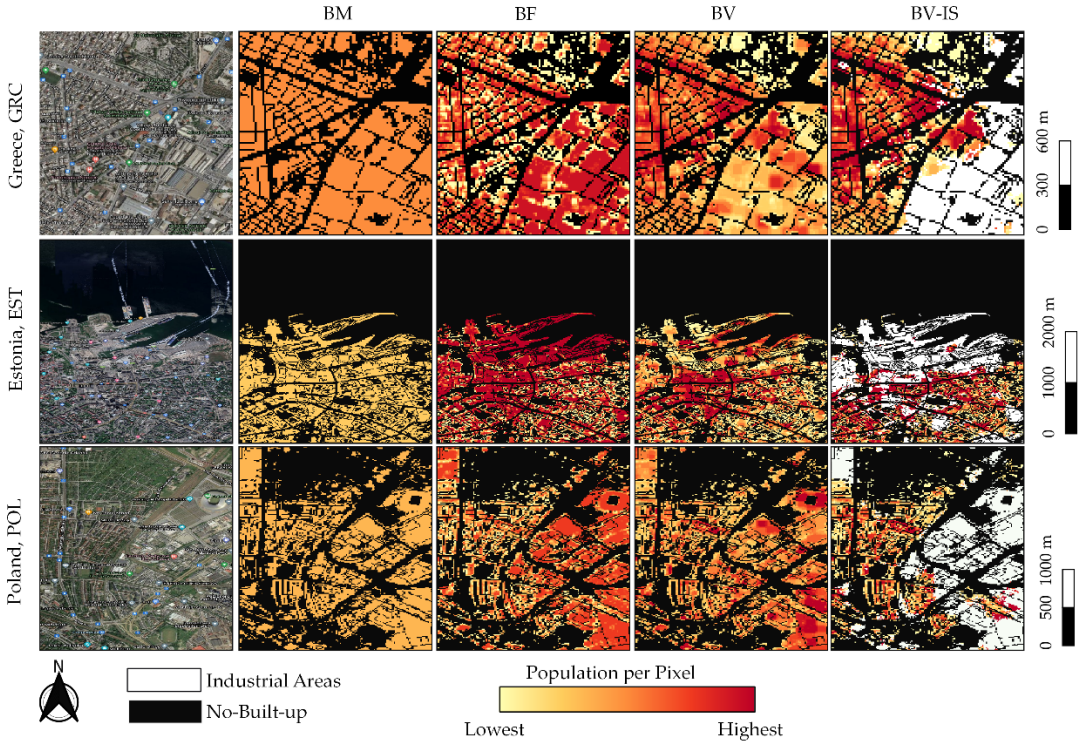

Figure 6-10. Local-area examples of the output population distribution maps produced on the basis of the BM, BF, BV and BV-IS layers and the national administrative units. Each map represents the UN-2020 population per pixel at a spatial resolution of ~12 m at the equator. Population per pixel is country/area dependent.

6.3.3 *Population Modelling: Quantitative Comparative Analyses*

The results of the accuracy assessment for each of the gridded population dataset in terms of the %MAE are presented in Table 6-4. Overall, results indicate that the integration of volume and industrial settlement use information (BV-IS) produced the lowest %MAE errors in the majority of the countries (bold numbers), whereas the BM produced the highest %MAE (italic numbers), respectively. At the Pan-European scale, 75% of the countries reported %MAE equal or below 47.26%, 46.06%, 42.93% and 37.72%, using the BM, BF, BV and BV-IS proxy layers, respectively. Here, the lowest %MAE value reported by each layer was 16.15%, 14.88%, 11.66% and 8.47%, respectively; while the highest %MAE value was 68.37%, 67.25%, 80.13% and 56.92%, respectively.

To evaluate the correlation between industrial coverage and the %MAE reported by each proxy layer at the country scale, we categorized the countries into three industrial levels, namely "Low" (0–10%), "Medium" (10–20%) and "High" (>20%) according to the share of industrial areas found according to the classified maps. With this, we then evaluated %MAE that was reported by each country's population map transitioning from one proxy layer to the next (BM to BF, BF to BV and BV to BV-IS) as described in Figure 6-11.

As observed, results indicate that only two out the 30 evaluated countries fall within the "Low" category of industrialized level, whereas 15 and 13 countries fall within the "Medium" and "High" category (see Table 6-4), respectively. For countries with "Low" industrial coverage, the %MAE remained some-what stable from one proxy layer to the next, where improvements of 10% are only reported in SVN by the BV-IS proxy layer. For countries with "Medium" and "High" coverage this behaviour is a more variable. First, independently of the error range, for most countries (22/28) the %MAE produced by the BM proxy layer remained within the same range as %MAE produced by the BF proxy layer. Here, five countries reported improvements of 10% (4 "Medium",1 "High"), while one country (GBR) reported 10% worsening. Consequently, for 12/28 countries errors remained within the same range between the BF and BV proxy layers. Here, 12/28 reported improvements (1 "Low", 6 "Medium", 5 "High"), elven countries of 10% and one country (GRC) of 20%. Three countries reported worsening of 10% and one country of 20% (GBR). When transitioning from the BV to the BV-SI, for 11/28 the %MAE remained within the same error range: Here, 14/28 countries reported improvements (1 "Low",6 "Medium", 7 "High"), twelve of 10%, one of 20% (IRL) and one of 30% (GBR). Three countries reported 10% worsening. Finally, a general evaluation transitioning from the BM to the BV-IS shows that 9/30 countries remained within the same %MAE errors range, from which one had "Low" industrial coverage, seven "Medium" industrial coverage and 1 "High" industrial coverage. Therefore, 21/30 countries show improvements, 15 of 10% (1 "Low", 4 "Medium", an" 10 "High"), 5 of 20% (3 "Medium" and 2 "High") and one country of 30% ("Medium").

Table 6-4. Accuracy assessment results.

		%MAE						%MAE					
ISO	**Av. Pop**	**BM**	**BF**	**BV**	**BV-IS**	**Ind. Level**	**ISO**	**Av. Pop**	**BM**	**BF**	**BV**	**BV-IS**	**Ind. Level**
ALB	7869.02	*60.70*	55.33	46.05	**41.69**	Medium	IRL	263.65	60.11	67.03	*78.43*	**56.92**	Medium
BEL	19,752.68	*36.12*	33.74	32.54	**26.22**	Medium	ISL	4623.53	*37.57*	26.31	20.54	**15.52**	Medium
BGR	26,077.06	*45.39*	41.73	**31.26**	37.49	High	ITA	189,654.99	*16.15*	14.88	11.66	**8.47**	High
BIH	26,465.83	28.11	*29.17*	**25.18**	25.73	Medium	LTU	46,581.63	*36.93*	28.26	**19.54**	26.15	High
CHE	3441.06	40.27	*41.39*	34.75	**27.43**	High	LUX	4353.32	*33.76*	30.88	31.64	**27.73**	High
CZE	1691.46	*40.51*	37.58	30.38	**28.89**	Medium	MKD	26,774.03	*34.19*	33.11	**28.21**	30.07	Medium
DEU	7119.40	*37.36*	36.23	31.83	**28.37**	High	MNE	29,801.81	*25*.**08**	26.65	28.77	**29.48**	High
DNK	2701.42	48.01	*48.48*	44.02	**30.86**	High	NOR	12,768.36	33.59	**35.39**	30.85	**26.99**	High
ESP	5541.96	43.52	*44.25*	34.08	**28.03**	High	POL	15,362.91	*46.61*	42.87	**34.47**	38.88	Medium
EST	282.35	*57.59*	56.7	54.8	**54.00**	Low	SRB	1438.83	*47.41*	44.94	**38.23**	40.78	Low
FIN	17,359.02	*39.55*	36.38	28.03	**20.06**	High	SVK	1857.59	*37.6*	34.04	**31.38**	33.00	Medium
FRA	1795.43	*46.83*	45.49	39.69	**30.55**	High	SVN	347.63	35.14	36.1	**36.22**	**28.86**	Low
GBR	287.13	56.55	67.25	*80.13*	**51.63**	Medium	SWE	6931.78	43.24	46.25	**47.66**	**37.80**	High
GRC	1768.57	*68.37*	64.44	47.77	**35.11**	Medium	TUR	7869.02	*60.7*	55.33	46.05	**41.69**	Medium
HRV	7486.51	*39.3*	39.28	32.76	**30.90**	Medium	UNK	54,918.53	*17.45*	18.78	**22.08**	19.55	Medium

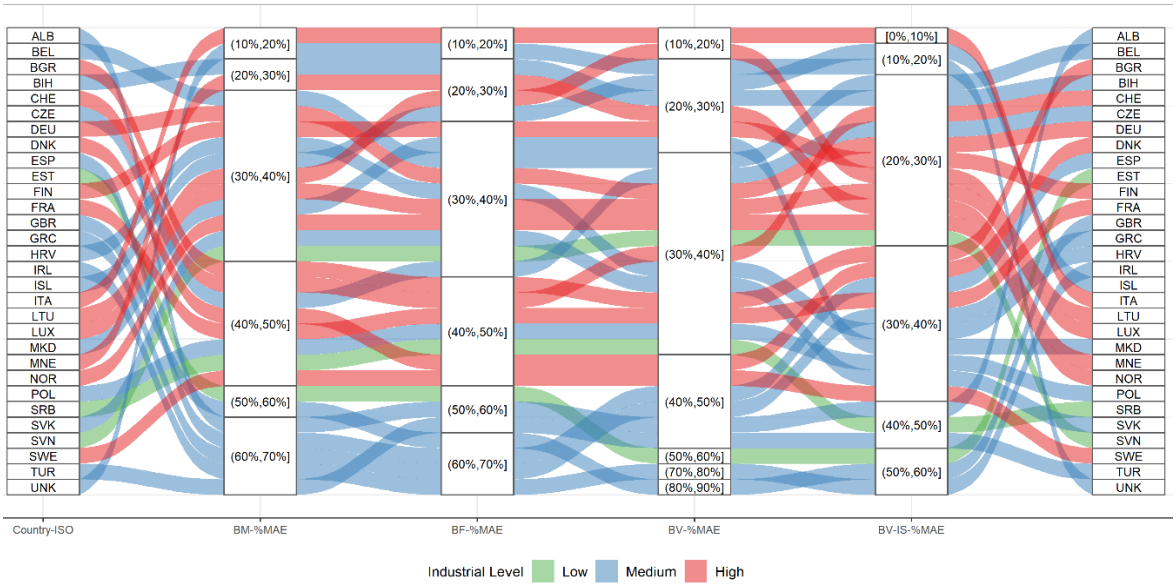

Figure 6-11. Alluvial plot showing the transitions of the %MAE across each proxy layer for all countries. Colours represent the industrial level of each country; *x*-axis elements represent the %MAE aggregated in 10% intervals.

In terms of the MAE and the RMSE, the results in Figure 6-12 show that for most countries and proxy layers, the MAE value remained at least twice as low as the average population at the country level. This behaviour was not similar for the RMSE, where for most countries this value was higher than the average population with the BM, BF, BV and BV-IS proxy layers, respectively. Accordingly, within most countries the distance between these two metrics is relatively shorter for models produced with the BV and BV-IS layers, respectively. This means, that in models produced with the BM and the BF layer, a larger variability exists between errors, which also suggest the presence of one or multiple outliers.

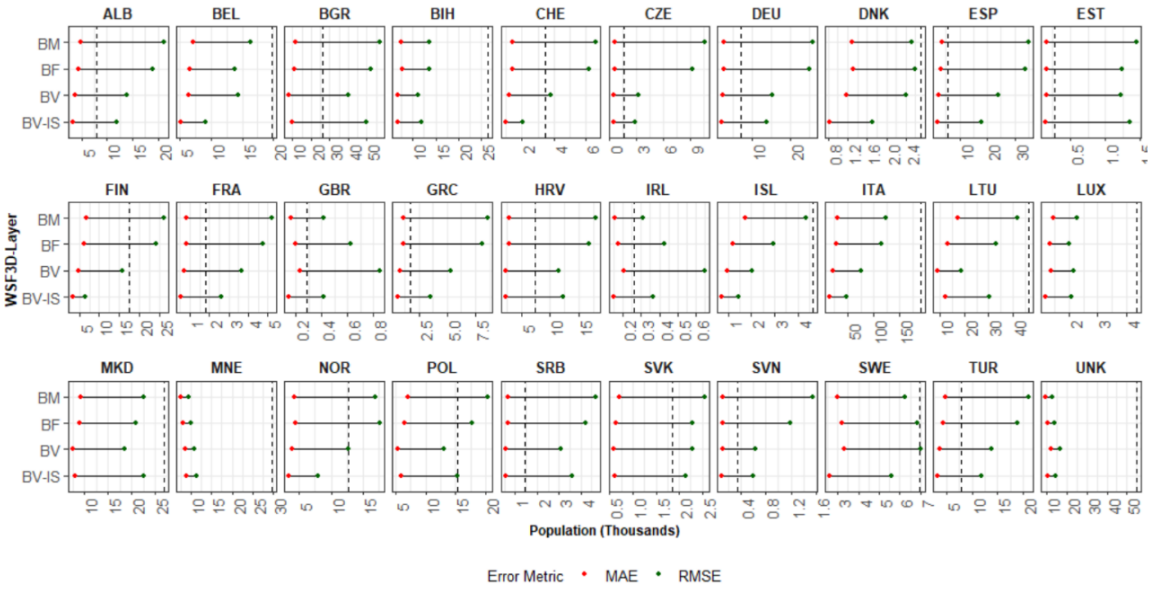

Figure 6-12. Lollipop plot showing the distribution of the MAE (red-dot) with respect to the RMSE (green-dot) for each country, and the average population (dashed line).

As described in sub-chapter 6.2.6.2 to compare the general trends of error distribution delivered by each proxy layer, we investigated the relationship among the Relative Estimation Error (REE), the total population and the share of industrial areas found within the validation L1-units of each country. Figure 6-13 presents the results of this assessment, where we have included only those countries where the majority of industrial ranges were present, with the rest of countries showing similar trends.

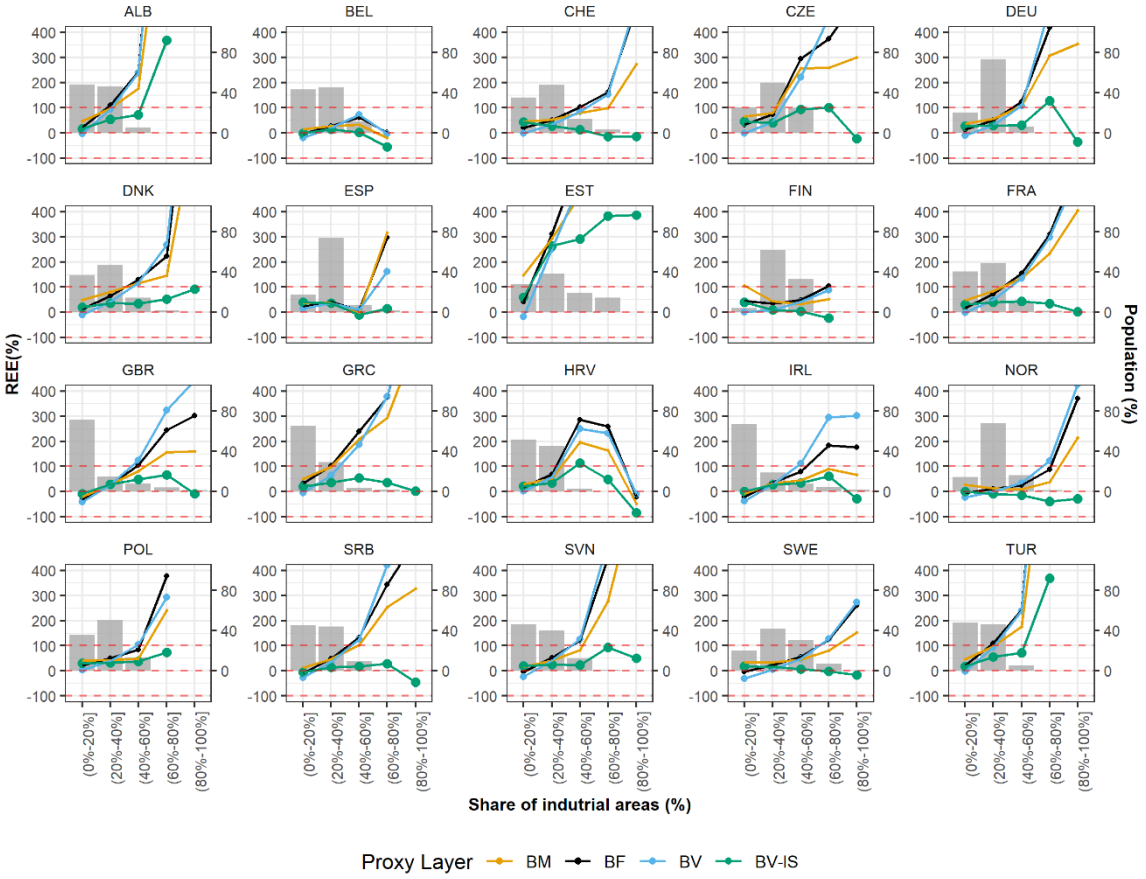

Figure 6-13. Point-line plots: average REE (left-y axis) produced by each proxy layer in relation to the share of industrial share found within the validation units. Bar plots: average percent population (right-y axis) found in validation units grouped by share of industrial areas.

First, as seen from these plots, the largest proportion of each countries' the population (bar plots, y-right-axis), is mainly found in units where the calculated industrial presence is below 40%. As the share of industrial areas increases, the average population decreases, reaching values equal or below 20% for most countries. Second, by analysing the distribution of the REE (line/point y-left-axis), one of the general trends we can observe, is that the majority of the proxy layers produced errors of overestimation across all ranges of industrial share (points above the "0" horizontal line). Errors of underestimations, produced mainly by the BV (light blue) and BV-IS (green) proxy layers, can be seen in some countries, especially in validation units with an industrial share lower than 20%, and in some few cases in ranges higher than 80%. Accordingly, for the majority of the countries, the BV-IS proxy layer produced overall the lowest REE. While for most countries, this tendency started from validation units with industrial share larger than 20%, improvements over the BM (yellow), BF (black) and the BV (light blue) proxy layers

became more pronounced in units with industrial share >40%. For the BM, BF and BV proxy layer, the largest (visible) overestimations are present in units with more than 60% of industrial share. In these units, the BV-IS proxy layer reduces the overestimations by as much as 700%, reaching either overestimations or underestimation in the range of 25–50%.

This behaviour, however, is different in units with industrial shares lower than 20%. For most of the countries, the BV proxy layer produced the smallest errors (underestimation) in the range of 10–15%, followed by BF and BM proxy layers, respectively. Finally, while the REE increased with increasing values of the industrial share for the BM, BF and BV layer, the errors reported by the BV-IS were consistently more stable, remaining systematically between −50% and 100% error ranges, in comparison with the other proxy layers, were errors reached overestimation higher than 400%.

Finally, to evaluate the distribution of error across all countries, Figure 6-14 shows the share of total population that fell within different REE ranges according to each proxy layer summarized at the Pan-European scale. At this level of evaluation, it is possible to observe that the BV-IS proxy layer estimates close to half of the population with errors ranging from −25% to 25%, with most errors being of underestimation. Comparably, within the same ranges, the BM, BF and BV proxy estimate 37%, 40% and 47% of the population, respectively, also with a tendency to underestimate.

Accordingly, with the BV-IS the second largest proportion of the population (~30%) was estimated with errors ranging from ± (25% to 50%), ~11% was estimated with errors ranging from ± (50% to 75%), ~3% was estimated with errors ranging from ± (75% to 100%), and ~3% with errors >100%. For the BM proxy layer, ~32% of the population was estimated with errors ranging from ± (25% to 50%), ~20% was estimated with errors ranging from ± (50% to 75%), ~7% was estimated with errors ranging from ± (75% to 100%), and ~5% with errors >100%. For the BF proxy layer, ~30% of the population was estimated with errors ranging from ± (25% to 50%), ~20% was estimated with errors ranging from ± (50% to 75%), ~6% was estimated with errors ranging from ± (75% to 100%), and ~5% with errors >100%. Finally, for the BV proxy layer, ~32% of the population was estimated with errors ranging from ± (25% to 50%), ~13% was estimated with errors ranging from ± (50% to 75%), ~5% was estimated with errors ranging from ± (75% to 100%), and ~4% with errors >100%.

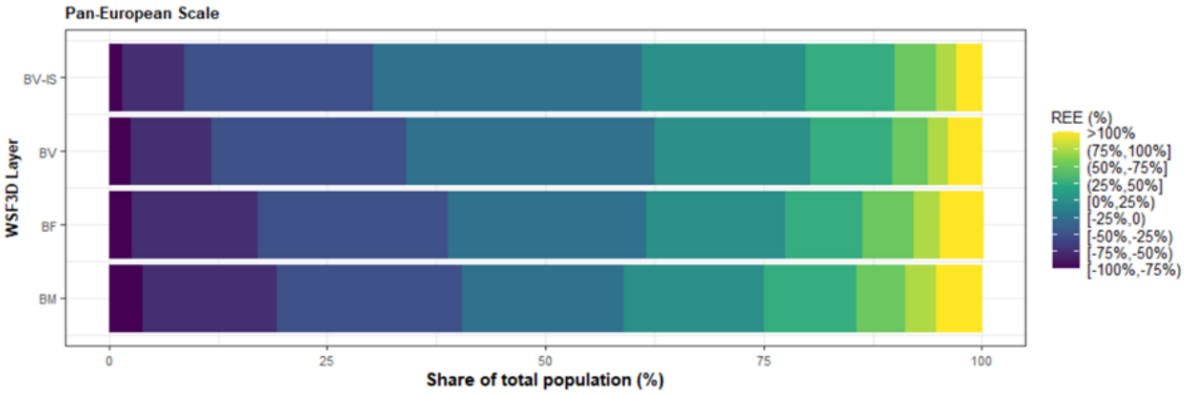

Figure 6-14. Percentage of total population aggregated over the 30 countries that fell within each error range.

6.4 Discussion

6.4.1 *Industrial and Non-Industrial Classification of Built-Up Settlements Using Random Forest*

The results of the classification tasks reveal that spatial metrics derived solely from the WSF3D dataset in combination with an RF classifier can be used to effectively identify and discriminate industrial versus non-industrial settlement use over large territorial extents. First, according to the results presented in Figure 6-7 and Figure 6-8, for most countries, the binary maps produced on the basis of the FM-RM models showed good agreement with the reference datasets in terms of the share of built-up settlements belonging to one or the other class. According to the standard interpretation of the Pearson's' correlation metric in the context of intraclass correlation (Cicchetti, 1994), the agreement between the refence and predicted maps was- between "fair" (0.4–0.59) and "good" (0.6–0.74) for all countries, highlighting the overall robustness of the presented approach.

Furthermore, as seen from Figure 6-9, the FM-RF models for most countries delivered average Overall Accuracy (OA) and Kappa coefficient (k) metrics that, at the Pan-European scale, remained above 84% and 0.68, together with Producer's Accuracy (PA) and User's Accuracy (UA) metrics that remained above 92–79% and 76–91% for each class, respectively. As observed, these metrics were not only fairly comparable to those reported by the I-RF models which, in the framework of our analyses, can be as "the best case scenario"(Orynbaikyzy et al.), but they also showed high correlation with the metrics of the E-RF models demonstrating that (1) the spatial metrics that characterized the training data for each class were heterogenous for most FUAs, and (2) that that these spatial metrics were similar across FUAs, allowing for the spatial transferability of our approach (Meyer & Pebesma, 2021).

In this context, from a comparative point of view, it is also worth noting that the results presented here are in line with those reported in other fine-scale studies that have employed an RF classifier in combination with more accurate and VHR remotely sensed data. For example, for an assessment of the classification accuracy of residential and industrial areas in Yangtze River Delta, China, the authors of (Ma et al., 2015) reported an OA of 87% and k of 0.74 from RF-models trained using spatial lacunarity metrics derived from VHR-LiDAR data. In the same way, using feature spatial metrics derived from VHR-LiDAR data, building footprints, VHR ortho-imagery (HRO) and Google Street View (GSV) images (GSV), the authors of (Zhang et al., 2017a) obtained and OA of 51.4% for commercial and industrial buildings, focusing on two small districts in Brooklyn, New York. Comparably, using spatial metrics derived from VHR ortho-imagery (0.5 m) and OSM parcel data, the authors of (Grippa et al., 2018) reported average OA of ~81.5% for the non-residential class (incl. administrative and commercial services), in Ouagadougou, Burkina Faso and Dakar, Senegal. Here, it is important to note that the WSF3D-based approach presented in our study is globally applicable—in contrast to methods requiring VHR ortho-imagery. Thus, it can be assumed that the approach developed here can easily be applied worldwide and, at the same time, accuracies can be achieved that are in the range of results obtained on the basis of commercial, high-resolution satellite images and building models.

With that being said, while the results presented here illustrate the high potential of the WSF3D dataset for identifying non-industrial versus industrial areas, there were a set of basic components that without a doubt influenced the accuracy of the final binary maps. These can be summarized as follows:

- *WSF3D*: The quality and accuracy of the WSF3D in terms of settlement detection (building mask-BM) and the final derived spatial metrics, played a fundamental role in the final accuracies reported in this research. A thorough inspection of the classified maps revealed that in areas identified as "industrial" by the reference datasets, many pixels representing actual green areas or parking lots were detected as built-up in the BM layer. Considering their low spatial metrics, the FM-RF then predicted these pixels as "non-industrial" leading to errors of omission in the industrial class, and errors of commission in the non-industrial class as summarized in Figure 6-9. In this context, from the average 25% errors of omission reported at the Pan-European level for the industrial class (100–75%, PA2 = 25%), it was found that approximately 15% of the errors came from confusing class 2 for class 1.1, and 10% for class 1.2 during the prediction process. While the classification of these pixels as "non-industrial" could be in reality "thematically correct", for the purpose of population modelling the presence of these pixels are detrimental, as population counts are allocated within these areas. Therefore, to potentially reduce the misclassification caused by the false detection of settlement pixels, future research should explore improving the final BM layer by integrating thresholds in the BF layer. Similarly, the integration of additional post-classification steps should also be considered, such as employing a broader number of window sizes for the extraction of spatial metrics as carried out in (Jochem et al., 2021), or by reclassifying the pixels according to their RF-class probability as carried out in (Grippa et al., 2018).

- *Automatic training data collection:* Unlike some local-scale research where a manual collection of training samples allows for a visual qualitative assessment of the training data (Schug et al., 2021), in this research we relied on an automatic procedure that did not include performing any sort of quality control over the training datasets. In correlation with our previous point, in a few cases, this lack of assessment resulted in poorly heterogenous training samples among classes, which without a doubt lead to some misclassification errors. For example, by evaluating the training data of a FUA in Ireland that shows large differences between classes (see Figure 6-8), it was possible to observe that the pixels values for class 1.1 "High-dense residential" and class 2: "Industrial" were similar across the bands corresponding to the BA, BF, BH and BV as represented in Figure 6-15. Within the selected FUA, this homogeneity led to errors of omission in the "non-industrial" class close to 20%, which meant that many pixels were erroneously classified as "industrial". In this framework, while the errors of omission in the non-industrial class at the Pan-European scale are considered low (see Figure 6-9), these misclassification errors had repercussions on the final population datasets, as seen by the results of Figure 6-13. Therefore, while a manual collection of training samples within the extent of our study area would have translated into a time-consuming task, further research should focus on the implementation of automatic intra-class separability analyses like the one presented in (Wicaksono & Aryaguna, 2020), with the objective to produce more significative training datasets.

- *Equal number of training samples per FUA, per class*: With the aim of producing robust comparative analyses within a country, in this research, an equal sample size was kept among all FUAs so that the I-RF and E-RF models were trained under similar conditions. In this framework, while the less represented class for some FUAs would reach close to 30% of the total available built-up pixels, in many cases less than 5% of the available pixels per class were used for training. This under-representation, coupled with the limitation mentioned on our previous point, affected the classification accuracy, especially in areas with a high inter-class diversity. In light of this, to improve the classification accuracy, future research should consider the inclusion of additional re-sampling steps. Here, post-classification approaches such as the one presented in (Schug et al., 2021) could become beneficial, where resampling is carried out using the RF class probabilities to concentrate in areas with high model-uncertainty.

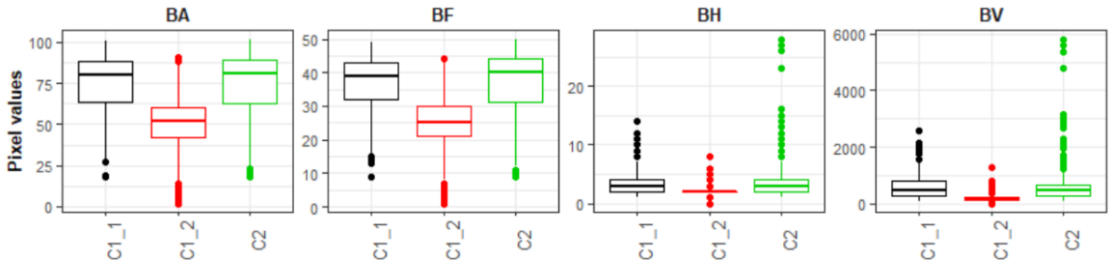

Figure 6-15. Distribution of pixel values in the four basic bands of the 16-band WSF3D composite used from training. Sample collected from a FUA in Ireland.

While the aforementioned points refer to the components affecting the accuracy of the final binary maps, there are also a couple of limitations and restrictions that need to be pointed out. On the one hand, while our results suggest that the presented approach has good spatial transferability within countries, it should not be assumed that models trained in one country can successfully be applied to another country. As demonstrated in Figure 6-9, the E-RF models did not perform better that the FM-RF model, even when the differences were minimal, this indicates that local training data is still preferable to achieve good classification results. In this context, future research should focus on evaluating the spatial transferability across countries, with analyses carried out outside Europe, to include a larger variety of built-up environments.

With that being said, it is then important to recognise that a common limitation of the presented approach, is the impossibility of discriminating small non-residential built-up types such as schools, hospitals, churches, etc. from the WSF3D alone, that in the end affect the final population distributions. This type of function is purely social, and as such, impossible to retrieve from the spatial metrics employed here. In this framework, even when the overall objective of this research was not to generate a land-use maps, we suggest that future work should focus on exploring the synergies between the WSF3D dataset with other remotely sensed datasets, where the inclusion of nightlight-imagery, building footprints and OSM tags/point of interest (Lloyd et al., 2020), for example, could result beneficial in the identification and refinement of a more extensive set of land-use classes.

Finally, considering that for this research only an RF classifier was employed, to fully evaluate the suitability of the WSF3D and derived spatial metrics for settlement classification tasks, future research should also explore the implementation of other

commonly employed pixel-based classifiers including, but not limited, to K-nearest neighbour and SVM, and compare if higher classification results can be achieved, holding the same degree of automation and spatial transferability.

6.4.2 *Population Modelling*

In this research, we produced gridded population datasets across 30 countries located in the EEA area to quantify the improvements in population estimates gained from the inclusion of volume (BV) and settlement use information (BV-IS) derived from the WSF3D dataset. For our assessment, we performed comparisons against other proxy layers, to simulate the thematic characteristics of covariate layers that are currently employed in the production of large-scale gridded population datasets, including the binary approach employed for the GHS-POP and the HRSL datasets (BM), and the density approach employed for the WSF-Pop datasets (BF), respectively.

From many perspectives, the results and conclusions obtained in this research are in line with the results found in other research. The main points can be summarizing as follows:

- *Weighted approaches perform better than binary approaches:* As already demonstrated in many other studies (Palacios-Lopez et al., 2019; Palacios-Lopez et al., 2021; Reed et al., 2018; Stevens et al., 2020), weighted dasymetric approaches produce higher accuracies than binary dasymetric approaches. First of all, as observed in Figure 6-10 from a qualitative point of view, the output population maps produced with the "density" layers (BF, BV and BV-IS) show a higher spatial correlation with the underlying rural-urban gradient in comparison to those produced with the binary layer (BM). The results of the quantitative assessment, further confirm that the spatial representations of the population distribution are not only more "realistic", but also more accurate, as all "density" layers consistently reported better aggregated statistics (%MAE, MAE and RMSE values) compared with the binary layer. On this note, however, it is worth noticing, that at level of validation units (REE), the BM proxy layer is capable to outperform the results of the "density" layers, especially in areas with a large share of industrial areas (see Figure 6-13). This makes sense if one considers that the "density" layers (BF and BV) amplify the errors of overestimation in these areas, by erroneously allocating more population due to their weighting value.

- *Building volume and settlement use information improve population estimates:* Comparable to the conclusions reached in local- and national-scale studies (Biljecki et al., 2016; Rubinyi et al., 2021; Schug et al., 2021), the inclusion of volume and settlement use information derived from the WSF3D dataset produced by far the best estimation accuracies across the majority of the countries. First, according to the results presented in Table 6-4 and Figure 6-11, the BV-IS proxy layer produced improvements over the BM, BF and BV proxy layers that reached %MAE values up to 30%. These were more frequently present in countries with "High" industrial coverage, where large errors were remarkably reduced as observed from Figure 6-13. Second, as observed in Figure 6-12, the BV-IS proxy layer remarkably reduced the differences between the MAE and RMSE metrics. This was, once again, correlated to the fact that large errors of overestimations were drastically reduced by the proxy layer, especially in areas with

a high share of industrial land cover (Figure 6-13). In this context, the inclusion of settlement use information played a major role, as it allowed the BV-IS proxy layer to produce systematically more stable results across all countries, and across all validation units, while REE errors remained between −50% and 100% with the BV-IS layer for most countries, the BM, BF and BV proxy layers produced variable results that reached overestimations in the rage of 500% or even higher (close to 4500% for EST). Accordingly, at the Pan-European scale, the BV-IS proxy layer estimated a larger proportion of the total population with errors in ±25% in comparison to BM add BF proxy layer, which according to pre-established rankings of accuracy (Bai et al., 2018), can be considered as "accurately" estimated.

- *The input and validation units influence accuracy results:* It is important to note that the maps evaluated in this research have been produced with the coarser administrative population units for each country (national scale). In some countries, where the BV-IS did not report the same systematic improvements compared to the other proxy layer, this characteristic might have influenced the accuracy results. For example, in the case of ALB, EST and TUR (Figure 6-13), many validation units that reported high coverage in industrial areas still reported large errors of overestimation with the BV-IS layer. This is, because even when industrial areas were successfully identified, the very low population counts of the validation units (sometimes less than 100) were difficult to match from a national-scale disaggregation. In this context, it can be expected for maps produced with the highest level of administrative units to be more accurate than the maps presented here. This assumption is supported by the large amount of research that has already proven that population maps produced with the finest administrative units produces the most accurate population maps (Hay et al., 2005a; Palacios-Lopez et al., 2019; Schug et al., 2021). However, considering that this only affected a few countries, also indicates that BV-IS proxy layer is capable to produce more accurate population maps than the rest of the proxy layers, when high-resolution input population data is not available.

- *The relative effectiveness of the BV-IS proxy layer is heavily dependent on the quality of the classified maps:* While the BV-IS produced consistently great improvements over the BM and BF proxy layers, in a few cases the performance of the proxy layer was improved by the BM and BV proxy layers, respectively. For example, as observed from Figure 6-13, in the majority countries the BV proxy layer produced better results than the BV-IS layer in validation units with industrial share below 20% (or 80% non-industrial). In these units, errors were ~50% higher (mostly overestimations) with the BV-IS layer, which suggest that a number of non-industrial built-up settlements pixels were erroneously removed, causing an allocation of a larger population in the remaining pixels. In this context, as expressed in the previous section, these results correlate to the difficulties of accurately classifying built-up settlements pixels in complex urbans settings, where high-rise buildings are mixed with industrial areas. Here, improvements in the classification process should then reflect in improvements in the population distribution.

With that being said, when evaluating the aforementioned points, a consideration to keep in mind is to recognize that the results presented here are strictly constrained to the employed validation method. The quantitative assessment was conducted following two

main assumptions (1) that the population data used for disaggregation were accurate and (2) that WSF3D dataset and its derived layers (BM, BF, BH and BV) were also complete and accurate. Therefore, while discussing the quality of these two main inputs is out of the scope of this research, uncertainties derived from the inherent shortcomings of each input dataset will by default affect the overall accuracy of the final population dataset.

For example, even when the CIESIN census dataset is the most detailed and complete database available at the global scale (Freire et al., 2018), we can observe that for many countries the last official population data are from more than 10 years ago (Table 6-2). This means, that potentially, both the population projections and administrative boundaries are outdated, which translate into errors in the final population maps. In this framework, similar to the points presented in the previous section, future research should focus on extending the methods presented here to areas outside Europe. Testing the presented approach in countries where fine resolution population data is not available, such as many countries in Africa (Palacios-Lopez et al., 2021), would be of great interest, especially in relation to local-scale applications where current gridded population dataset have presented major limitations (Fries et al., 2021; Thomson et al., 2021a).

6.5 Summary

In this research, we explored the contributions that the new WSF3D dataset can bring into the field of large-scale top-down population modelling. We performed a series of quantitative analyses that investigated the potential of the dataset from two main perspectives: (1) its ability to discriminate large industrial areas which in the past have been reported as major sources of under- and overestimation in population estimates, and (2) its capabilities to improve population estimates by integrating volume and settlement use information into population modelling frameworks.

To this end, we first proposed a method that relied on spatial metrics derived solely from the WSF3D dataset in combination with a RF classifier to discriminate industrial and non-industrial areas. Here, our results revealed not only that the WSF3D dataset is capable of producing accurate binary classification maps, but that its performance is comparable to other, more spatially granular, VHR remotely sensed datasets that have been used for the same purpose. Foremost, the findings also indicated that the presented method has strong spatial transferability, which means that the dataset poses a viable solution to the existing gap between local- and large-scale analyses.

Accordingly, by integrating the resulting classified maps into frameworks of population modelling, the results were also promising. The results of our quantitative assessment indicate that inclusion of volume and settlement use information (industrial, non-industrial) derived from the WSF3D dataset produced, by far, the best population estimates in comparison to other commonly employed proxy layers. For the most part, the main advantages delivered from the layer include (i) a remarkably, systematic and consistent reduction in errors of overestimation in areas with a high share of industrial areas, (ii) an improved distribution of population estimates in high-density built-up settings and (iii) an increased ability to produce accurate population estimates in the presence of less detailed input census-based population data.

Notwithstanding these promising results, there is however, room for improvement. The results of the classification tasks, for example, can be further enhanced with the integration of post-classification methods and a more careful collection of training data. These improvements will be directly reflected in final population estimations, where misclassification errors proved to be detrimental in highly dense and highly populated built-up settings.

Overall, the results of this study provide a valuable contribution to the field of large-scale population modelling. The methods presented here show strong promise for helping to bridge the gap between fine- and large-scale efforts aimed at improving top-down population distribution models. As shown, the synergies between volume and settlement use information derived solely from the WSF3D dataset provide the basis to create more accurate global population distribution dataset or related updates for arbitrary regions or countries worldwide. In this context, future developments of this work will include the final production and open release of global gridded population dataset, with unprecedented accuracy and spatial resolution.

Chapter 7

7. Synthesis and Outlook

This final chapter presents a brief summary on the results of this dissertation. Sub-chapter 7.1 and sub-chapter 7.2 outline the main technical and practical achievements of this PhD research, presenting a discussion on how they make an important contribution to the field of top-down large-scale population modelling. Furthermore, sub-chapter 7.3 presents a brief summary of past–, ongoing–, and future projects and use-case scenarios that had, are and will leverage the outcomes of this PhD research. And finally, sub-chapter 7.4, presents the main conclusions and recommendations on future research.

7.1 Technical achievements and findings

Building on the individual discussions presented in each peer-review article, the most prominent contributions of this PhD research to the field of top-down large-scale population modelling can be summarised as follows.

1. **Enabling the integration of state-of-the-art WSF datasets into population modelling frameworks.**

First, through the thorough and comprehensive analyses presented in this dissertation, the first direct contribution of this PhD research is the demonstration of how the employment of the WSF-suite into top-down population modelling frameworks helped to overcome some of the existing limitations in the field. The results presented in this thesis provide a deep and transparent measure of the effectiveness of each layer for top-down population modelling, highlighting not only their advantages over other existing models, but the remaining limitations that will affect any future developments produced with the layers.

2. **First in-depth quality assessment of the WSF-dataset as proxy layers for top-down large-scale population modelling**.

In this context, building on the previous point, the second contribution of this PhD research is a first "in-depth" quality assessment of the effectiveness and suitability of each WSF-layer as a proxy for large-scale top-down population modelling. Emphasis in the word "in-depth" is made, as the results presented here extend on a series of spatial and statistical analyses that, –at the time of writing–, not many of the state-of-the-art datasets have been presented with, or at least not at same geographical extent. This statement is supported by the research presented by Bai et al. (2018), who reported in the middle of 2018 that "overall, past research inadequately addresses the accuracy and uncertainties of the current gridded population distribution datasets".

For example, to-date other than a series of comparative analyses done to understand the differences between population grids, (Archila Bustos et al., 2020; Bai et al., 2018; Calka & Bielecka, 2019, 2020; Da Costa et al., 2017) neither of the data producers of the

GPWv4.11, GHS-POP and HRSL internally validate the accuracy of the disaggregated estimates at a large-scale (Thomson et al., 2021a; Tuholske et al., 2021), mainly due to absence of independent or authoritative validation data (POPGRID, 2021). The only datasets that frequently (and openly) report validation metrics at the scale of the input population data are the WorldPop datasets, but even here, a thorough evaluation like the ones conducted in Chapter 4, Chapter 5 and Chapter 6 are not usually performed mainly due to the complexities introduced by the employed multi-layer approach. In other words, due to the use of a large number of proxy layers it is more complex to assess the uncertainty delivered by each layer; thus, leading to results that are not universal (Nagle et al., 2014). This is inclusively more difficult in the case of the LandScan datasets, where information on the input population data and ancillary datasets used for modelling are not completely documented (Archila Bustos et al., 2020).

On this basis, to ensure a thorough validation of the population models produced in this research, within each publication a series of spatial and statistical analyses were implemented, gathering the most representative and useful analyses that have been employed in other studies of top-down population modelling at smaller-scales. For example, to fulfil the objectives of the first research publication (Chapter 4), a multi-scale analysis was employed in nine representative countries, in which the size of the input units used for modelling were progressively increased from Admin 2 – 3 (Enumeration Areas or District level) to Admin 0 (National level). This type of analysis has been employed in Hay et al. (2005a) and Tatem et al. (2007), and is useful to assess the effects of the spatial resolution of the input data on the final accuracy results on the one hand, but most importantly, to reveal the possible bias in the accuracy results due to the MAUP. Similarly, for all of the three research publications, a more thorough set of spatial and statistical analyses collected from Bai et al. (2018) were applied to identify the amount of error produced by each proxy layer at the validation unit level. These analyses relied on the Relative Estimation Error (REE) metric whose utility and advantages lie on 1) allowing moving from aggregated results like the RMSE, MAE, %MAE, etc. which are normally reported the country scale, to individual results reported at the validation unit-scale, 2) investigating whether spatial patterns of error distribution within countries and across-countries are systematic, 3) allowing establishing correlations between the reported errors and the employed proxy layers, and 4) ranking the overall accuracy results. Furthermore, these analyses help overcome the limitations of certain metrics like the R^2, which is sensitive to the number of validation units or the sampling process; or the RMSE and MAE which can be greatly affected by outliers.

3. Novel use of Settlement Size Complexity index for evaluation of uncertainty.

Accordingly, a major third contribution made by this research has been the design and implementation of the SSC index as a robust and globally transferable metric to evaluate the uncertainty of population models based on built-up area layers. On the one hand, as proposed in Chapter 4, the implementation of the SSC-Index was developed in an effort to correlate and quantify the characteristics of the underlying built-up environment as derived from the WSF2015 layer at administrative input unit level to the error produced at the validation unit level. The main assumption was that the reported accuracy at the validation unit level was bound to the characteristics of the built-up environment at the input unit level, in terms of settlement size, distribution and

compactness. Here, for example, it was found that the WSF-2015-Density layer outperforms the modelling capabilities of the WSF-2015 in "high" SSC-class units, due to the overall omission of settlements in "low" and "medium" SSC-class units. On the other hand, considering the analyses presented in Chapter 5, the SSC-Index was used at the validation level to correlate the REE with the properties of the built-up environment including population density and settlement properties. Here it was revealed, for example, that major errors were delivered by the WSF2019-Imp layer in highly-dense and population units with the presence of very high-rise buildings.

As such, the SSC-Index has two main functions that can be applied in any future research. First, it can be used as a tool to pre-identify areas with high probability of error, allowing for a quick visual inspection that can help refine the underlying built-up area layers. In the validation of the WSF2015-Density (Evolution) layer presented in Blersch (2020), for example, the SSC-Index was used to implement a transparent sampling design to evaluate the accuracy of the PIS values reported by the layer and propose targeted refinements. Second, the SSC-Index can be used to compare how different built-up area layers characterize the build-up environment within admirative census units' level. Together with information on the population density, these areas can be then reassessed to improve weighting frameworks that allow a better distribution population.

4. Support for "fitness for use" evaluation.

In the context of specific applications, it is important for end users to know about the magnitude of deviations when using different large-scale top-down gridded population datasets. Each dataset delivers a unique population distribution model, guided by the methodological approach and input data layers employed for their creation. Leyk et al (2019) provided one of the first in-depth evaluations of several global data products (e.g. GPWv4.11, GHS-POP, HRSL, WorldPop, etc.) examining the critical elements that make each dataset unique and suited for a given purpose. Here, the authors suggest that in order to choose the most appropriate population grid, users need to consider aspects such as (1) the spatial and temporal resolution, (2) the described population (residential or ambient) and (3) the underlying ancillary data that was used for modelling (e.g. avoid endogenous results). Based on different criteria, one dataset might be more suitable than another with regard to the target application.

With that being said, at a more overarching scale, the comprehensive assessments presented in this PhD research also contributed to existing efforts aiming at evaluating the "fitness for use" of large-scale gridded population datasets. Reporting on the actual level of qualitative and quantitative accuracy of the derived models produced on the basis of the WSF-suite, as well as the full and transparent dissemination of the input data and methods that were used to produce the final datasets will allow users to decide if the WSF-population datasets are suitable for a given application compared to other datasets.

5. Innovative method to identify and classify industrial and non-industrial areas.

While population models created on the basis of the WSF2015-Density and the WSF2019-Imp layer focused on the improvement of dedicated limitations of large-scale population modelling at the given time, the interim development of the WSF3D provided the missing piece which then allowed to tackle all existing limitations reported in the field.

The layers that form the WSF3D dataset, namely the BF, the BH and the BV, provided the framework in which to develop a spatially transferable and semi-automatic method for the identification of industrial and non-industrial areas based on a binary-RF classification approach. The synergies of volume and settlement use information derived solely from the WSF3D then allowed creating large-scale population models with higher accuracy than ever before.

As stated in Chapter 6, the methods to classify and derive settlement use information presented in this PhD research represents one of the first efforts helping to bridge the gap between fine and large-scale top-down population distribution models. While the process itself relies on well-known methods, the novelty and major contribution of the approaches presented here, is that the final classification of settlement use relies only on spatial metrics derived from the WSF3D, without the need of other geospatial dataset; thus, it can be assumed that the approach developed here can easily be applied worldwide. Furthermore, the accuracies reported here are in the range of results obtained on the basis of commercial, high-resolution satellite images and building models, which demonstrate the robustness of the metrics derived from the WSF3D to recognise the desired classes.

7.2 Practical achievements and findings

In this thesis it has been hypothesised that the improved characteristics of the new WSF-layers could potentially allow addressing some of the main challenges that affect the accuracy and usability of state-of-the-art large-scale top-down gridded population models. For each dataset, the remaining question was how and in which degree they could be used to achieve this, which lead to the establishment of a series of overarching research objectives tailored to specifically evaluate the effectiveness and suitability of each WSF-layer, namely the WSF2015 (binary and density), the WSF2019 (imperviousness) and the WSF3D, respectively, for large-scale top-down population modelling.

For this purpose, the layers were evaluated within a framework that was inherently linked to their chronological release. This means that while each layer proved its capabilities to improve large-scale populations models, within each optimisation step specific challenges arose in return. Therefore, the three approaches or implementations presented here are to be understood in the sense of an "evolution over time", linked to the availability of different data baselines. With that being said, this section provides a summarized discussion of the most relevant practical achievements of the PhD research, which altogether illustrate how the WSF-layers addressed the limitations affecting the accuracy of either binary or (multi-layer) weighted approaches presented in Table 2-3.

Increased spatial resolution of large-scale population models: Overall, the increased spatial resolution of the WSF-layers has allowed translating coarse large-scale top-down population model simulations to finer spatial scales needed for a range of applications that rely on high-resolution datasets (Smith et al., 2019). The spatial resolution of the final population models produced during this research are 3x, 10x and 100x times more granular compared with global state-of-the-art layers such as the HRSL, GHS-Pop, WorldPop, LandScan, GRUMP, and GPWv4.11 datasets. While further application-based analyses are needed to fully determine the benefits and trade-offs of this increased spatial granularity for different applications, the refined spatial resolution of the initial WSF2015 and WSF2015-Density population maps have already shown to be beneficial, in particular,

within the context of floor-risk assessments. For example, in the framework of the e-Drift (Disaster Risk Financing and Transfer) project which started ~2018 (CIMA Foundation, 2019) , internal evaluations of the WSF2015-Population datasets for flood risk assessment proved that the population maps of Myanmar and Cambodia produced within this PhD research (see Chapter 4) delivered far more accurate populations estimates (e.g. population at risk) compared to maps produced by the GHS-POP and WorldPop datasets (250m and 1km at the Equator). Improvements were linked to the increased spatial resolution (10m at the Equator) of the models. This resulted in a better integration of the datasets with the high spatial resolution (e.g. < 90m at the Equator) of the EO-flood models produced within the Southeast Asia Disaster Risk Insurance Facility (SEADRIF) initiative, which allowed practitioners to better estimate the number of people affected by floods.

Unprecedented accuracy of large-scale population maps: As described in Chapter 3, in the framework of global built-area datasets, the WSF2015 and WSF2019 layers were and continue to be the most accurate global settlement extent layers, outperforming all other existing similar datasets. This means that errors of omission and commission that frequently affect the accuracy of large-scale top-down population models (Stevens et al., 2020) can be potentially minimized with the integration of the WSF layers into the modelling frameworks compared with other commonly employed built-area layers.

Outside of this PhD research, one of the first experimental results supporting this claim were given in the work of Reed et al. (2018). In this study the authors showed that a 100m resampled version of the WSF2015 layer was capable of outperforming the accuracy of binary-dasymetric models produced with the GHSL dataset, while producing accuracies similar to those of the only regionally available HRSL layer derived from commercial VHR satellite imagery. Overall, the population models produced with the WSF2015 layer produced improvements over the GHSL-derived maps of ~14% (reported in terms of the MAE) for the six countries where it was tested. Furthermore, when the WSF2015 was integrated into "constrained" intelligent dasymetric models, comparable to those employed to produced WorldPop datasets, the final datasets showed higher accuracy than those produced with the GHSL layer, respectively.

However, according to the results presented in Chapter 4, errors of omission of small settlements or isolated houses in rural and sub-urban areas still affected the accuracy of the final population maps produced with the WSF2015-layers (binary and density), which were easily identified by the large number of administrative that reported no settlement areas (see Figure 4-9) These errors, nevertheless were later-on drastically minimised with the integration of the WSF2019 as proxy layer, where the improvements in settlement classification accuracy are intrinsically attributed to the integration of S2 optical imagery available at 10m spatial resolution, in the processing framework of the layer. For example, making reference to the discussion presented in sub-chapter 5.4.1 , in population models produced on the basis of the WSF2019-Imp layer all input administrative units reported settlement areas in the countries of Malawi and Côte d' Ivoire, as opposed to >500 units that were reported as "empty" by the models produced on the basis of the WSF2015 layers. These improvements in settlement classification directly lead to increased overall accuracies in the derived population models, as errors of underestimation did not longer reach the 100% mark. Accordingly, in view of the correlation existing between the

WSF2019 and the WSF3D, the same quality can be expected from the population maps produced on the basis of the WSF3D layer.

Increased spatial representations, applicability and replicability: All of the best performing WSF-layers (e.g. WSF2015-Density, WSF2019-Imp and BV-IS derived from the WSF3D layer, respectively) provide a "non-binary" weighting framework for population disaggregation that 1) are derived independently from other geospatial layers, and 2) have direct relationships with population densities without the need of additional statistical modelling. This has shown to have several qualitative and quantitative advantages, which overall tackle the limitations related to homogeneous spatial representations of the population distribution, poor transferability and replicability, and endogeneity issues reported by binary– and intelligent dasymetric models.

First, compared with binary-weighted dasymetric models, the non-binary weighted frameworks provided by the PIS (>0-100%) and the WSF3D (BV-SI) (>0-∞ m^3) showed considerably higher spatial correlation with the underlying rural-urban gradient, producing population maps that depicted the spatial distribution of the population in a more realistic manner. Extreme distribution artefacts between administrative units with different population counts were also minimized by the presented models, producing more accurate and visually appealing population maps as exemplified in the comparative visual assessment presented in Figure 4-1, Figure 5-10 and Figure 6-10.

Second, from a quantitative point of view, the PIS & BV-IS population models also outperformed the quantitative results over well-established binary-models. For example, in their initial version, the WSF2015-Density layer was able to consistently deliver more accurate population models than the binary WSF2015 layer, even when models were produced at the coarsest spatial resolution of the input population data Table 4-6. Accordingly, the results presented in Table 4-8, confirm that the improvements made over the WSF2015-models (based on RMSE metrics) ranged from as little as 1.12% to as high as 31.20%, with average increases of 14.70%, 16.20% and 11.29%, respectively in low, medium and highly dense built-up environments. Moving from the 2015 to the 2019 products, the results presented in Figure 5-10, show that at local-scales the WSF2019-Imp layer delivered population models that were tentatively more accurate than binary models (e.g. WSF2019 binary), especially in areas with marked transitions between high–and sparsely dense built-up environments. So far, with the WSF2015-Density and the WSF2019-Imp layers, the only areas where binary-weighted models are potentially more accurate than the PIS-weighted models, include those where industrial or large commercial centres were present (see Figure 4-10). This makes sense, if one considers that the high PIS-values amplify the errors of overestimations in these areas. This limitation, however, is resolved with models produced on the basis the WSF3D (BV-IS) layer, which allows on the one hand, to remove large un-populated areas (e.g. industrial and commercial areas), and increase the variability of weights in highly-dense built-up areas, on the other. As shown by the results presented in sub-chapter 6.3.3 , improvements in accuracy are not only made over well-established binary models, but also over the previously PIS-based WSF-models, with values as large as ~30% based in terms of the MAE, and as large as >4500% terms of the REE metrics, respectively

Third, from a technical point of view and compared to (multi-layer) intelligent-dasymetric models, the proposed models have been produced with a single proxy layer,

which overall reduces the processing times and the complexities that come with collecting and harmonising different geospatial covariates (Archila Bustos et al., 2020; Lloyd et al., 2017). Furthermore, due to their global coverage, using the WSF-layers as single also enables the easy and fast replicability and transferability of the derived population models on the one hand; while at the same time, completely eliminates the biases that are introduced by using multiple proxies that might report different quality, accuracy and temporal coverage, on the other. Comparably, the direct relationship that exists between population densities and the weighted-values (e.g. more volume, more people) reduces the ambiguities that might arise at the time of interpreting the final accuracy results. In other words, for each processed model, a single proxy allows a more straightforward/better/more simple understanding of the produced population maps.

Fourth, compared to intelligent-dasymetric models the WSF-layers additionally provide weighting frameworks that are processed independently of other thematic geospatial layers (e.g. the PIS values are calculated without the need of land-cover or land-use layers), from an application point of view, this increases the applicability of the final population datasets for many applications fields, as there are no endogeneity issues that arise, for example, when population distributions explain land-cover changes and land-cover explains population distributions.

Consistent and comparable data across multiple spatial scales: Another important advantage of the proposed models presented in this PhD research is that final datasets produced on the basis of all the WSF layers proved to deliver qualitative and quantitate accuracies that are consistent across space. For example, even when population models were produced and validated at different spatial scales and under varying qualities of the input census-based population data across different countries and continents, they all produced consistent and systematic patterns of error distribution. Here, the largest percentage of the population of each study area was always modelled with estimation accuracies ranked as "accurately estimated". Before the integration of settlement use and volume information derived from the WSF3D, for example, the WSF2015-Density and the WSF2019-Imp layers consistently and systematically produced population datasets where the highest errors of under—and overestimation (>±25%) were reported in areas (or validation units) where a poor correlation existed between the amount of population that and amount of available settlement pixels (e.g. as described by the SSC-Index) (see Figure 4-3 and Figure 5-9). First, it was determined that the errors of omission presented in the WSF2015- Density layer largely decreased the accuracy of population estimates in areas with large populations counts. Results generated with the WSF2019-Imp then confirmed this observation by demonstrating that not only the errors of omission on the settlement classification were related to these underestimation errors, but that a lack of information on height of the building structures negatively affected the quality of the population modelling in densely-populated areas. Similarly, for both layers, the largest errors of overestimation (>100%) were reported in validation units with high coverage of industrial and commercial areas.

In terms of population models produced with the WSF3D (e.g. models produced with the BV-SI), errors were consistently and systematically kept within the -50% to 100% range, with the largest improvements being reported in validation units with high industrial/commercial coverage. Accordingly, for all the studies sites, the vast majority of

underestimations observed were related to units with low industrial coverage (< 20%) and high-population density. This effect was resulted from the potential masking of residential buildings in the context of the classification of the type of settlement use (residential, industrial/commercial).

In this framework, it is important to understand that this consistent and systematic error distribution allows not only to deliver more robust interpretations of the final accuracies (e.g. deeper understanding of the correlation between the final population datasets and the quality of the input data), but also allows implementing comparisons across space without the biases that arise, for example, from intelligent-dasymetric models that rely on datasets that are just available locally (e.g. LandScan datasets). From a user's perspective this property is quite valuable because the uncertainties are more consistent.

First large-scale integration of information on building use and building volume: Finally, all of the aforementioned advantages have to be coupled with the unprecedented gains in accuracy that resulted from integrating not only building volume information, but also building use information into models of top-down population distribution. As mentioned in the summary section of Chapter 6, the synergies of ML algorithms with spatial metrics derived solely from the WSF3D represent one of the first efforts that have allowed improving large-scale top-down population models over large territorial extents. This has been done by integrating building use and volume information that is as accurate as that derived from VHR imagery or detailed building footprints at very local scales. Thereafter, by exploiting the capabilities of the WSF3D from large-scale population modelling, three of the four major challenges in the field of top-down large-scale gridded population modelling detailed in the introduction, have been successfully addressed.

New perspectives for "bottom-up" approaches of population modelling: Within the scope of the PhD research, addressing the current limitations in the field of large-scale gridded population modelling focused specifically in exploring solutions to improve the input methods and ancillary spatial data used for modelling. However, as expressed by Wardrop et al. (2018), "the central challenge to the accuracy of top-down disaggregation methods is the reliability of the input population data". In Chapter 5, sub-chapter 5.4.1 the implications related to the quality of the input population data in the final population models were addressed in detail. From this assessment we know that while the CIESIN census database can be considered as the most detailed, complete and comprehensive global database available today (Freire et al., 2018); the population counts (incl. estimations/projection), as well as the administrative boundaries, can be largely outdated and/or incomplete (see Figure 2-1). This is especially true for regions where economic, social or political limitations have not allowed to produce updated national censuses, shortcomings that affect the final accuracy of top-down population models (El-Khoury & Jaulin, 2012; UNFPA, 2012, 2014).

In this regard, as gridded population datasets produced using top-down approaches are only as good as the input population data on which they are based, in recent years novel "bottom-up" methods have emerged to produce up-to-date (and more frequent) gridded estimates. In nature, bottom-up approaches rely on micro-census data (e.g. data collected for a few and small representative regions) and different statistical models to predict population numbers in unsampled locations (see Wardrop et al. (2018) for a list of potential methods). At the grid level, the accuracy of these predictions depends largely on

the accuracy, completeness and reliability of different spatial proxies that are used to establish allocation-relationships. In this framework, similar to the early stages of top-down gridded population methods, research on the implementation of bottom-up approaches has been mainly restricted to small urban settings or high-income countries where access to high-quality, reliable and comparable data exists (Biljecki et al., 2016; Harvey, 2002; Schug et al., 2021; Sutton et al., 2001) . With a few new exceptions, methods have been applied in low-income regions (Darin et al., 2021a; Dooley et al., 2021), but overall results have suggested that the lack of information on the use and 3D characteristics of the build-up environment produce persistent over- and underestimation errors.

With that being said, while the scope of this research was not focused on exploring bottom-up approaches of population modelling, it can be argued that the integration of the WSF-layers into these modelling frameworks could potentially improve the accuracy of their predictions. The same advantages that the WSF-layers provide within top-down methods can be leveraged for bottom-up methods, especially those related to consistency, and increased detailed information on settlement use, settlement height and settlement volume, both to run predictions or improve sampling processes (Dooley et al., 2021). Thereinafter, if in future research the capabilities of the WSF-suite were to produce favourable results in bottom-up methods of population disaggregation, significant areas of active research would also benefit from these developments.

7.3 Supported projects and use cases

Finally, the main overarching contribution of the PhD has been the production and delivery of a first-round of large-scale gridded population datasets to a series "champion-users" undertaking a variety of projects in different research areas. Some of the most presentative uses that have been given to the WSF-Population datasets produced during this PhD thesis are presented in Table 7-1.

Table 7-1. Summary of external projects and applications where the WSF-Population dataset and methods were/are/will be employed.

Project/Funding agency	WSF-Population datasets	Use case or scenario
eDrift (CIMA Foundation, 2019) ESA; SEADRIF; World Bank	Population maps for Myanmar and Cambodia produced on the basis of the WSF2015-Density layer	The population maps were used to estimate of population at risk of flood, for the calculation of financial aid.
Satellite Monitoring Service of Urbanization in Africa World Bank	Population maps for entire Africa, produced on the basis of the WSF2019-Imp layer	The population maps were used to identify all localities with an estimated population of >10,000 inhabitants. And to estimate populations at risk of different hazards, including earthquakes, landslides, flooding and storm surge.
TraK - Transport and Climate * (DLR-Institute of Transport Reseach, 2018) DLR	Population maps for 33 cities located in different parts of the world produced on the basis of the WSF2015-Density and WSF2019-Imp layers.	The population maps were used for accessibility analyses aimed at producing metrics for SDG 11, indicator 11.2.1. In particular, the datasets were used to calculate the percentage of population with convenient access to public transportation from a walkable distance (500m to 1km).
e-Shape (Gilardi et al., 2021; H2020, 2022) H2020	Population maps for Munich and Belin in Germany, and Milan, Italy, produced on the basis of the WSF2019-Imp layer	The population maps were used to perform an assessment of the increased health risks within urban areas in Europe due to exposure of PM2.5.
Global Development Assistance Programme (GDA)- Fragility, Conflict and Security * (ESA, 2022) ESA	Population maps for countries in conflict, including Afghanistan, Tajikistan, Mozambique, produced on the basis of the WSF2019-Imp layer and WSF3D.	Population maps will be used to support international financial institutions in better targeting their development assistance activities in countries and regions subject to settings of fragility, conflict and violence. Use cases include for example, understanding/projection refugee dynamics from political and spatial context.

Project/Funding agency	WSF-Population datasets	Use case or scenario
GDA-Disaster Risk Analytics * (ESA, 2022) ESA	Population maps for the entire African continent, produced with the WSF2019 and WSF3D layers.	Population maps will be used to perform vulnerability analyses and to compute impacts related to different natural disasters
United Nations Office for the Coordination of Humanitarian Affairs (OCHA).	Population maps produced for Ukraine on the basis of the WSF3D + the industrial/non-industrial built-up use mask. Integrated within the RF-modelling framework of the WorldPop datasets.	Population maps will be used to support the Ukraine-Russia conflict, providing update input data to inform policy makers.

* Foreseen projects for 2022, or projects where the new methods based on the WSF3D datasets are yet to be implemented.

7.4 Conclusions and Future Research

The timing of the start of this PhD research was ideal to coincide with the internal project of the German Aerospace Centre aiming at improving the accuracy, spatial resolution and thematic representation of their previous global built-up layer the GUF. The development of the WSF suite, and the envision of the different geospatial layers that would comprise this product over time, represented a new window of opportunity in which to explore how these new layers could improve the qualitative and quantitative accuracy of large-scale gridded population models.

In this framework, the analyses presented in this PhD research were performed to investigate the effectiveness of the WSF-layers from four interconnected perspectives i) their ability to improve population estimates compared to binary-dasymetric models (Chapter 4), ii) their ability to produce high and comparable accuracies across space acting as single proxies for population modelling (Chapter 5, Chapter 6), iii) their ability to discriminate large industrial areas that affect population models (Chapter 6) , and iv) their ability to improve population estimates through the integration of volume and settlement use information (Chapter 6).

As such, the results produced in this PhD research confirmed that —at a given time– – each WSF-layer was capable of addressing one or more limitations in the field of large-scale top-down population modelling. The WSF2015-Density layer and the underlying weighting framework produced by the PIS values overcame both the qualitative and quantitative limitations of binary-dasymetric models. The improved accuracy in terms of settlement classification of the WSF2019-Imp layer and the independent framework provided by the PIS values allowed producing not only spatially comparable population datasets across space, but datasets with easy replicability and improved applicability overcoming the limitations of multi-layer intelligent-dasymetric models (on top of the binary-dasymetric). And the WSF3D provided the ultimate framework in which to extract settlement use information, to finally integrate settlement volume and use into large-scale top-down population modelling approaches, overcoming some of the major limitations in the field. Overall, the WSF-layers did not only allow producing large-scale top-down population models with unprecedented accuracy and improved spatial resolution reaching the main goal of the thesis, but also enabled the development of a settlement use classification method with potential transferability and applicability at a global scale.

With that being said, however, as in any other research there is still room improvement. For example, future research should focus on improving the methods used to derive settlement use information from the WSF3D layers. It is suggested that more accurate labelled data needs to be collected for model training, coupled with the integration of post-processing steps that allow increasing the accuracy of the derived binary classification. Accordingly, other ML methods should be tested for settlement use classification. These can range from other supervised methods like SVM which might improve classification accuracies; to unsupervised methods like Gaussian Mixture Models which might improve or facilitate spatial transferability. Furthermore, to the extent that is possible, comparative analyses should be done against other state-of-the-art large-scale gridded population datasets. This could include integrating the final weighting layer (BV-SI) into the multi-layer models employed by the WorldPop datasets; or preferably, within

the framework of tailored applications where independent ground truth data is available for validation.

Furthermore, as mentioned before, the limitations addressed in this thesis focused solely on issues that do not take into consideration the effects that the quality and accuracy of the input population data can have in the final population models. Under optimal circumstances, top-down population models are produced with the most recent (and available) census-based population data (incl. estimations/projections), but for many countries' the reality is that censuses are considerably outdated (e.g. 20 years or more). In this framework, the CIESIN databased was used due to its availability and coverage, however, as any other product, the data is far from perfect showing qualitative and quantitative limitations both in the boundaries and population estimations. Thereof, as rising number of scientific applications require more accurate and updated information on population estimates, as part of the outlook of this thesis, the weighted layer that integrates volume and settlement use information should be tested in bottom-up population modelling approaches to evaluate its suitability in the production of more current population estimates/projections derived from micro-census data. This implementation, will therefore, contribute to the efforts of some of the most contemporary research on the field aiming at bridging this gap (Darin et al., 2021b; Dooley et al., 2021; Qader et al., 2021; Wardrop et al., 2018).

At the same time, in view of the of the future development and production of a multi-temporal WSF-Evolution—a settlement layer describing the extent, location and distribution of human settlements at a multi-temporal scale (1985-2015), the methods presented here should be integrated in the production of a multi-temporal WSF-Population dataset that would allow to evaluate patterns of population change in a similar matter than current existing multi-temporal layers. This integration, however, is expected to add its maximum value to population modelling, when the volume and settlement use information can also be added in a temporal scale.

To summarise, the results of this PhD research provide a valuable contribution to the field of large-scale population modelling. Here it was confirmed not only that the WSF-layers have the potential to address some of the major identified challenges in the field, but that the methods develop here open a new set of opportunities in which to create more accurate population models with improved accuracy, ensuring gains in scalability, cost and time -optimisation, adaptability and automation. As such, future development of this work, include the production and final released of a first-time WSF-based global population dataset with improved accuracy and unprecedented spatial resolution.

Bibliography

Aguirre, M. S. (2002). Sustainable development: why the focus on population? *International Journal of Social Economics*. https://doi.org/10.1108/03068290210447978

Ajisegiri, B., Andres, L. A., Bhatt, S., Dasgupta, B., Echenique, J. A., Gething, P. W., Zabludovsky, J. G., & Joseph, G. (2019). Geo-spatial modeling of access to water and sanitation in Nigeria. *J. Water Sanit. Hyg. Dev.*, *9*(2), 258-280. https://doi.org/10.2166/washdev.2019.089

Allen, C., Smith, M., Rabiee, M., & Dahmm, H. (2021). A review of scientific advancements in datasets derived from big data for monitoring the Sustainable Development Goals. *Sustainability Science*, 1-16. https://doi.org/10.1007/s11625-021-00982-3

Amoah, B., Giorgi, E., Heyes, D. J., van Burren, S., & Diggle, P. J. (2018). Geostatistical modelling of the association between malaria and child growth in Africa. *International journal of health geographics*, *17*(1), 7. https://doi.org/10.1186/s12942-018-0127-y

Anderson-Sprecher, R. (1994). Model comparisons and R2. *The American Statistician*, *48*(2), 113-117. https://doi.org/10.1080/00031305.1994.10476036

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science*, *20*(2), 77-96. https://doi.org/10.1080/10095020.2017.1333230

Archila Bustos, M. F., Hall, O., Niedomysl, T., & Ernstson, U. (2020). A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990–2015. *Population and Environment*, *42*(2), 255-277. https://doi.org/10.1007/s11111-020-00360-8

Aubrecht, C., Özceylan, D., Steinnocher, K., & Freire, S. (2013). Multi-level geospatial modeling of human exposure patterns and vulnerability indicators. *Natural Hazards*, *68*(1), 147-163. https://doi.org/10.1007/s11069-012-0389-9

Azar, D., Graesser, J., Engstrom, R., Comenetz, J., Leddy, R. M., Schechtman, N. G., & Andrews, T. (2010). Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti. *International Journal of Remote Sensing*, *31*(21), 5635-5655. https://doi.org/10.1080/01431161.2010.496799

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., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I., & Nelson, A. (2006). Determining Global Population Distribution: Methods, Applications and Data. *Adv. Parasitol.*, *62*, 119-156. https://doi.org/10.1016/s0065-308x(05)62004-0

Barbier, E. B., & Hochard, J. P. (2018). Land degradation and poverty. *Nature Sustainability*, *1*(11), 623-631. https://doi.org/10.1038/s41893-018-0155-4

Basu, S., Gill, J., Yang, D., Bonafilia, D., & Kirsanov, D. (2019). *Mapping roads through deep learning and weakly supervised training*. Meta AI. Retrieved 03 from https://ai.facebook.com/blog/mapping-roads-through-deep-learning-and-weakly-supervised-training

Batista e Silva, F., Lavalle, C., & Koomen, E. (2013). A procedure to obtain a refined European land use/cover map. *Journal of Land Use Science*, *8*(3), 255-283. https://doi.org/10.1080/1747423X.2012.667450

Bhaduri, B., Bright, E., Coleman, P., & Urban, M. L. (2007). LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, *69*(1-2), 103-117. https://doi.org/10.1007/s10708-007-9105-9

Biljecki, F., Ohori, K. A., Ledoux, H., Peters, R., & Stoter, J. (2016). Population estimation using a 3D city model: A multi-scale country-wide study in the Netherlands. *PLoS One*, *11*(6), e0156808. https://doi.org/10.1371/journal.pone.0156808

Blersch, M. (2020). *Validation of global urban density products with very high resolution imagery* [Masters Degree, Heidelberg University].

Brown, S., Nicholls, R. J., Goodwin, P., Haigh, I., Lincke, D., Vafeidis, A., & Hinkel, J. (2018). Quantifying land and people exposed to sea-level rise with no mitigation and 1.5 C and 2.0 C rise in global temperatures to year 2300. *Earth's Future*, *6*(3), 583-600. https://doi.org/10.1002/2017EF000738

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing*, *12*(6). https://doi.org/10.3390/rs12061044

Calka, B., & Bielecka, E. (2019). Reliability analysis of LandScan gridded population data. The case study of Poland. *ISPRS International Journal of Geo-Information*, *8*(5), 222. https://doi.org/10.3390/ijgi8050222

Calka, B., & Bielecka, E. (2020). GHS-POP Accuracy Assessment: Poland and Portugal Case Study. *Remote Sensing*, *12*(7), 1105. https://doi.org/10.3390/rs12071105

Calka, B., Nowak Da Costa, J., & Bielecka, E. (2017). Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk*, *8*(2), 1440-1455. https://doi.org/10.1080/19475705.2017.1345792

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

Chen, R., Yan, H., Liu, F., Du, W., & Yang, Y. (2020). Multiple Global Population Datasets: Differences and Spatial Distribution Characteristics. *ISPRS International Journal of Geo-Information*, *9*(11). https://doi.org/10.3390/ijgi9110637

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

CIESIN. (2011). *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid* NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4R20Z93

CIESIN. (2018a). *Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 [data set]* Version Version 4). https://doi.org/10.7927/H4BG2KXS

CIESIN. (2018b). *Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11* NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PN93PB

CIMA Foundation. (2019). *eDFRIT Disaster RIsk Financing Transfer*. https://www.cimafoundation.org/foundations/projects/edrift.html

Comber, A., & Zeng, W. (2019). Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. *Geography Compass*, *13*(10), e12465. https://doi.org/10.1111/gec3.12465

Copernicus Programme. (2022). *Copernicus Land Monitoring Service*. Retrieved 28 July from https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018

Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2021). Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications*, *33*(12), 6697-6720. https://doi.org/10.1007/s00521-020-05449-7

Da Costa, J. N., Bielecka, E., & Calka, B. (2017, April 27-28). *Uncertainty Quantification of the Global Rural-Urban Mapping Project over Polish Census Data*. Environmental Engineering. 10th International Conference, Lithuania.

Darin, E., Boo, G., & Tatem, A. J. (2021a). *A bottom-up population modelling approach to complement the population and housing census* International Population Conference 2021, Hyderabad, India.

Darin, E., Leasure, D. R., & Tatem, A. J. (2021b). *Statistical population modelling for census support*. UNFPA and WorldPop. https://wpgp.github.io/bottom-up-tutorial/

De Mattos, A. C. H., McArdle, G., & Bertolotto, M. (2020). Assessing the Quality of Gridded Population Data for Quantifying the Population Living in Deprived Communities. *arXiv preprint arXiv:2011.12923*. https://doi.org/10.48550/arXiv.2011.12923

Dhewantara, P. W., Mamun, A. A., Zhang, W.-Y., Yin, W.-W., Ding, F., Guo, D., Hu, W., & Magalhães, R. J. S. (2018). Geographical and temporal distribution of the residual clusters of human leptospirosis in China, 2005–2016. *Scientific Reports*, *8*(1), 16650. https://doi.org/10.1038/s41598-018-35074-3

DLR-Institute of Transport Reseach. (2018). *TraK: Transport and Climate*. https://www.dlr.de/vf/en/desktopdefault.aspx/tabid-2974/1445_read-54393/

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, *66*(7), 849-857. https://doi.org/doi.org/10.1007/s10708-007-9105-9

Doll, C. N. H., & Pachauri, S. (2010). Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy*, *38*(10), 5661-5670. https://doi.org/https://doi.org/10.1016/j.enpol.2010.05.014

Dooley, C., Chamberlain, H., Leasure, D., Membele, G., Lazar, A., & Tatem, A. (2021). Description of methods for the Zambia modelled population estimates from multiple routinely collected and geolocated survey data, version 1.0.

Dooley, C., Gianluca, B., Leasure, D., & Tatem, A. (2020). *Gridded Maps of building patterns through sub-Saharan Africa*. WorldPop. Retrieved 15 December from https://doi.org//10.5258/SOTON/WP00677

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). 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*(3), 226-234. https://doi.org/10.1080/23754931.2015.1014272

Du, S., Zhang, F., & Zhang, X. (2015). Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS journal of photogrammetry & remote sensing*, *105*, 107-119. https://doi.org/10.1016/j.isprsjprs.2015.03.011

Duque, J. C., Laniado, H., & Polo, A. (2018). S-maup: Statistical test to measure the sensitivity to the modifiable areal unit problem. *PLoS One*, *13*(11), e0207377. https://doi.org/10.1371/journal.pone.0207377

El-Khoury, M., & Jaulin, T. (2012). *Country Report: Lebanon*. http://hdl.handle.net/1814/39894

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *35*(8), 1652-1660. https://doi.org/doi.org/10.1016/j.cageo.2009.01.009

ESA. (2015). *Climate Change Initiative*. https://www.esa-landcover-cci.org/?q=node/164

ESA. (2016). *Mapping Guide for a European Urban Atlas*. https://land.copernicus.eu/user-corner/technical-library/urban-atlas-mapping-guide

ESA. (2022). *Global Development Assistance Programme*. https://www.esa.int/ESA_Multimedia/Images/2021/11/Global_Development_Assistance_Programme

Esch, T., Bachofer, F., Heldens, W., Hirner, A., Marconcini, M., Palacios-Lopez, D., Roth, A., Üreyen, S., Zeidler, J., & Dech, S. (2018a). Where we live—A summary of the achievements and planned evolution of the global urban footprint. *Remote Sensing*, *10*(6), 895. https://doi.org/10.3390/rs10060895

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., & Zeidler, J. (2022). World Settlement Footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, *270*, 112877. https://doi.org/https://doi.org/10.1016/j.rse.2021.112877

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., & Sensing, R. (2017). Breaking new ground in mapping human settlements from space–The Global Urban Footprint. *ISPRS Journal of Photogrammetry*, *134*, 30-42. https://doi.org/10.1016/j.isprsjprs.2017.10.012

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., & Dech, S. (2013). Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geoscience and Remote Sensing Letters*, *10*(6), 1617-1621. https://doi.org/10.1109/LGRS.2013.2272953.

Esch, T., Marconcini, M., Marmanis, D., Zeidler, J., Elsayed, S., Metz, A., Müller, A., & Dech, S. (2014). Dimensioning urbanization – An advanced procedure for characterizing human settlement properties and patterns using spatial network analysis. *Applied Geography*, *55*, 212-228. https://doi.org/10.1016/j.apgeog.2014.09.009

Esch, T., Üreyen, S., Zeidler, J., Metz–Marconcini, A., Hirner, A., Asamer, H., Tum, M., Böttcher, M., Kuchar, S., Svaton, V., & Marconcini, M. (2018b). Exploiting big earth data from space – first experiences with the timescan processing chain. *Big Earth Data*, *2*(1), 36-55. https://doi.org/10.1080/20964471.2018.1433790

Esch, T., Zeidler, J., Palacios-Lopez, D., Marconcini, M., Roth, A., Mönks, M., Leutner, B., Brzoska, E., Metz-Marconcini, A., & Bachofer, F. (2020). 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*(15), 2391. https://doi.org/10.3390/rs12152391

España, G., Grefenstette, J., Perkins, A., Torres, C., Campo Carey, A., Diaz, H., de la Hoz, F., Burke, D. S., & van Panhuis, W. G. (2018). Exploring scenarios of chikungunya mitigation with a data-driven agent-based model of the 2014–2016 outbreak in Colombia. *Scientific Reports*, *8*(1), 12201. https://doi.org/10.1038/s41598-018-30647-8

Falcone, J. A. (2016). US national categorical mapping of building heights by block group from Shuttle Radar Topography Mission data.

FAO, UNICEF, WFP, & WHO. (2021). *Asia and the Pacific Regional Overview of Food Security and Nutrition 2020: Maternal and child diets at the heart of improving nutrition*. FAO. https://www.fao.org/documents/card/en/c/cb2895en/

Fisher, P. F., & Langford, M. (1996). Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping. *The Professional Geographer*, *48*(3), 299-309. https://doi.org/10.1111/j.0033-0124.1996.00299.x

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., & Hostert, P. (2021). National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, *252*, 112128. https://doi.org/10.1016/j.rse.2020.112128

Freire, S., Doxsey-Whitfield, E., MacManus, K., Mills, J., & Pesaresi, M. (2016, 14-17 June). *Development of new open and free multi-temporal global population grids at 250 m resolution*. AGILE, Helsinki, Finland.

Freire, S., Schiavina, M., Florczyk, A. J., MacManus, K., Pesaresi, M., Corbane, C., Borkovska, O., Mills, J., Pistolesi, L., Squires, J., & Sliuzas, R. (2018). Enhanced data and methods for improving open and free global population grids: putting 'leaving no one behind' into practice. *International Journal of Digital Earth*, *13*(1), 61-77. https://doi.org/10.1080/17538947.2018.1548656

Fries, B., Guerra, C. A., García, G. A., Wu, S. L., Smith, J. M., Oyono, J. N. M., Donfack, O. T., Nfumu, J. O. O., Hay, S. I., & Smith, D. L. (2021). Measuring the accuracy of gridded human population density surfaces: A case study in Bioko Island, Equatorial Guinea. *PLoS One*, *16*(9), e0248646. https://doi.org/10.1371/journal.pone.0248646

Galway, L. P., Bell, N., Al Shatari, S. A., Hagopian, A., Burnham, G., Flaxman, A., Weiss, W. M., Rajaratnam, J., & Takaro, T. K. (2012). A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth TM imagery in a population-based mortality survey in Iraq. *International journal of health geographics*, *11*(1), 12. https://doi.org/10.1186/1476-072X-11-12

GeoNode. (2019). *MIMU-Myanmar Information Managment Unit*. Retrieved 10 February from http://geonode.themimu.info/

GFDRR. (2007). *ThinkHazard!* Retrieved February from https://thinkhazard.org/en/

Gilardi, L., Metz-Marconcini, A., Marconcini, M., & Erbertseder, T. (2021, September 13-18). *Urban air pollution exposure: an assessment exploiting world settlement footprint and land use data*. Remote Sensing Technologies and Applications in Urban Environments VI, Spain.

Goerlich, F. (2016). A volumetric approach to spatial population disaggregation using a raster build-up layer, land use/land cover databases (SIOSE) and LIDAR remote

sensing data. *Revista de Teledetección*(46), 147-163. https://doi.org/10.4995/raet.2016.4710

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., & Zhou, Y. (2020). Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment*, *236*, 111510. https://doi.org/10.1016/j.rse.2019.111510

Goodchild, M. F., Anselin, L., & Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and planning A*, *25*(3), 383-397. https://doi.org/10.1068/a250383

Goodchild, M. F., & Lam, N. S.-N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo-processing*, *1*(3), 297-312.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education*, *74*(3), 249-266. https://doi.org/10.3200/JEXE.74.3.249-266

Gregory, I. N. (2002). The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, *26*(4), 293-314. https://doi.org/10.1016/S0198-9715(01)00013-8

Grippa, T., Georganos, S., Zarougui, S., Bognounou, P., Diboulo, E., Forget, Y., Lennert, M., Vanhuysse, S., Mboga, N., & Wolff, E. (2018). Mapping urban land use at street block level using openstreetmap, remote sensing data, and spatial metrics. *ISPRS International Journal of Geo-Information*, *7*(7), 246. https://doi.org/10.3390/ijgi7070246

H2020. (2022). *e-Shape*. https://e-shape.eu/index.php/showcases/pilot2-3-eo-based-pollution-health-risks-profiling-in-the-urban-environment

Hallisey, E. A.-O., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., Graham, S., Flanagan, B., & Lunsford, N. B. (2017). Transforming geographic scale: a comparison of combined population and areal weighting to other interpolation methods. *Int. J Health Geogr*, *16*(29). https://doi.org/10.1186/s12942-017-0102-z

Harvey, J. (2002). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, *23*(10), 2071-2095.

Hay, S., Noor, A., Nelson, A., & Tatem, A. J. (2005a). The accuracy of human population maps for public health application. *Trop Med Int Health*, *10*(10), 1073-1086. https://doi.org/10.1111/j.1365-3156.2005.01487.x

Hay, S. I., Noor, A. M., Nelson, A., & Tatem, A. J. (2005b). The accuracy of human population maps for public health application. *Tropical Medicine and International Health*, *10*(10), 1073-1086. https://doi.org/10.1111/j.1365-3156.2005.01487.x

Heris, M., Foks, N., Bagstad, K., & Troy, A. (2020). A national dataset of rasterized building footprints for the US. *US Geological Survey*. https://doi.org/10.5066/P9J2Y1WG

Herold, M., Couclelis, H., & Clarke, K. C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment & Urban Systems*, *29*(4), 369-399. https://doi.org/10.1016/j.compenvurbsys.2003.12.001

Herrmann, M., Guzman, J. M., & Schensul, D. (2012). *Population Matters for Sustainable Development*. T. U. N. P. F. (UNFPA). https://www.unfpa.org/sites/default/files/pub-

pdf/UNFPA%20Population%20matters%20for%20sustainable%20development_1.pdf

Huang, X., Li, J., Yang, J., Zhang, Z., Li, D., & Liu, X. (2021a). 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019. *Science China Earth Sciences*, *64*(11), 1922-1933. https://doi.org/10.1007/s11430-020-9797-9

Huang, X., Wang, C., Li, Z., & Ning, H. (2021b). A 100 m population grid in the CONUS by disaggregating census data with open-source Microsoft building footprints. *Big Earth Data*, *5*(1), 112-133. https://doi.org/10.1080/20964471.2020.1776200

Humanitarian Data Exchange. (2018). *Malawi Traditional Authority*. Retrieved 15 November from https://data.humdata.org/dataset/2018_malawi_ta_dataset-updated-admin3

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., & Guo, Q. (2018). The transferability of Random Forest in canopy height estimation from multi-source remote sensing data. *Remote Sensing*, *10*(8), 1183. https://doi.org/10.3390/rs10081183

Jochem, W. C., Leasure, D. R., Pannell, O., Chamberlain, H. R., Jones, P., & Tatem, A. J. (2021). Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science*, *48*(5), 1161-1179. https://doi.org/10.1177/2399808320921208

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021, 11-16 July 2021). *Global land use / land cover with Sentinel 2 and deep learning*. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels,Belgium.

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., & Chuvieco, E. (2020). Towards delivering on the sustainable development goals using earth observations. *Remote Sens. Environ.*, *247*, 111930. https://doi.org/10.1016/j.rse.2020.111930

Kellenberger, B., Vargas-Muñoz, J. E., Tuia, D., Daudt, R. C., Schindler, K., Whelan, T. T., Ayo, B., Ofli, F., & Imran, M. (2021). Mapping Vulnerable Populations with AI. *arXiv preprint arXiv:14123*. https://doi.org/10.48550/arXiv.2107.14123

Kenya National Bureau of Statistics. (2019). *Population and Housing Census 2019: Table: Census Volume 1 Question 1 Population by County and Subcounty*. Retrieved 15 November from https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county.

Kenya National Bureau of Statistics. (2020). *Sub-County Boundaries*. Provided under licensing by the Office of General Director

Khryashchev, V. V., Pavlov, V. A., Priorov, A., & Ostrovskaya, A. A. (2018, September 14-17). *Deep learning for region detection in high-resolution aerial images*. 2018 IEEE East-West Design & Test Symposium (EWDTS), Kazan,Russia.

Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, *31*(21), 5657-5671. https://doi.org/10.1080/01431161.2010.496805

Kuffer, M., Owusu, M., Oliveira, L., Sliuzas, R., & van Rijn, F. (2022). The Missing Millions in Maps: Exploring Causes of Uncertainties in Global Gridded Population Datasets. *ISPRS International Journal of Geo-Information*, *11*(7), 403.

Kuffer, M., Thomson, D. R., Boo, G., Mahabir, R., Grippa, T., Vanhuysse, S., Engstrom, R., Ndugwa, R., Makau, J., & Darin, E. (2020). The Role of Earth Observation in an Integrated Deprived Area Mapping "System" for Low-to-Middle Income Countries. *Remote Sensing*, *12*(6), 982. https://doi.org/10.3390/rs12060982

Langford, M. (1991). The areal interpolation problem: estimating population using remote sensing in a GIS framework. *In Handling geographical information: methodology and potential applications*, 55-77.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, *31*(1), 21-26. https://doi.org/10.1179/000870494787073718

Leinenkugel, P., Deck, R., Huth, J., Ottinger, M., & Mack, B. (2019). The potential of open geodata for automated large-scale land use and land cover classification. *Remote Sensing*, *11*(19), 2249. https://doi.org/10.3390/rs11192249

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., & Frye, C. (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). https://doi.org/10.5194/essd-11-1385-2019

Li, G., & Weng, Q. (2005). Using Landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. *Photogrammetric Engineering*, *71*(8), 947-958. https://doi.org/10.14358/PERS.71.8.947

Li, L., & Lu, D. (2016). Mapping population density distribution at multiple scales in Zhejiang Province using Landsat Thematic Mapper and census data. *International Journal of Remote Sensing*, *37*(18), 4243-4260. https://doi.org/10.1080/01431161.2016.1212422

Li, M., Koks, E., Taubenböck, H., & van Vliet, J. (2020). Continental-scale mapping and analysis of 3D building structure. *Remote Sensing of Environment*, *245*, 111859. https://doi.org/10.1016/j.rse.2020.111859

Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., & Tatem, A. J. (2012). Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One*, *7*(2), e31743. https://doi.org/10.1371/journal.pone.0031743

Linard, C., Gilbert, M., & Tatem, A. J. (2011). Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*, *76*(5), 525-538. https://doi.org/10.1007/s10708-010-9364-8

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., Gaughan, A. E., Nieves, J. J., Hornby, G., & MacManus, K. (2019). Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 1-32. https://doi.org/10.1080/20964471.2019.1625151

Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Sci Data*, *4*, 170001. https://doi.org/10.1038/sdata.2017.1

Lloyd, C. T., Sturrock, H. J., Leasure, D. R., Jochem, W. C., Lázár, A. N., & Tatem, A. J. (2020). Using GIS and machine learning to classify residential status of urban buildings in low and middle income settings. *Remote Sensing*, *12*(23), 3847. https://doi.org/10.3390/rs12233847

Lu, D., Weng, Q., & Li, G. (2006). Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, *27*(16), 3553-3570. https://doi.org/10.1080/01431160600617202

Lu, Z., Im, J., Rhee, J., & Hodgson, M. (2014). Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. *Landscape Urban Planning*, *130*, 134-148. https://doi.org/10.1016/j.landurbplan.2014.07.005

Ma, L., Wu, D., Deng, J., Wang, K., Li, J., Gu, Q., & Dai, Y. (2015). Discrimination of residential and industrial buildings using LiDAR data and an effective spatial-neighbor algorithm in a typical urban industrial park. *European Journal of Remote Sensing*, *48*(1), 1-15. https://doi.org/10.5721/EuJRS20154801

Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., & Dow, P. A. (2019, 19-22 May). *Facebook Disaster Maps: Aggregate Insights for Crisis Response and Recovery*. Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Valencia, Spain

Malawi National Statistical Office. (2018). *Population and Housing Census 2018, Series A Population Table*. Retrieved 15 November from http://www.nsomalawi.mw/images/stories/data_on_line/demography/census_2018/2018%20MPHC%20Published%20Tables/Series%20A.%20Population%20Tables.xlsx

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., & Kakarla, A. (2020). Outlining where humans live--The World Settlement Footprint 2015. *Sci Data*, *7*(242). https://doi.org/10.1038/s41597-020-00580-5

Marconcini, M., Metz, A., Zeidler, J., & Esch, T. (2015, March-April, 30-1). *Urban monitoring in support of sustainable cities*. 2015 Joint Urban Remote Sensing Event (JURSE), Lausanne, Switzerland.

Maroko, A., Maantay, J., Pérez Machado, R. P., & Barrozo, L. V. (2019). Improving population mapping and exposure assessment: three-dimensional dasymetric disaggregation in New York City and São Paulo, Brazil. *Papers in Applied Geography*, *5*(1-2), 45-57. https://doi.org/10.1080/23754931.2019.1619092

Maxar Technologies. (2020). *Building Footprints*. Retrieved 08 August from https://www.maxar.com/products/building-footprints

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, *55*(1), 31-42. https://doi.org/10.1111/0033-0124.10042

Mennis, J., & Hultgren, T. (2006). Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. *Cartography and Geographic Information Science*, *33*(3), 179-194. https://doi.org/10.1559/152304006779077309

Merkens, J.-L., & Vafeidis, A. (2018). Using Information on Settlement Patterns to Improve the Spatial Distribution of Population in Coastal Impact Assessments. *Sustainability*, *10*(9), 3170. https://doi.org/10.3390/su10093170

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology Evolution*, *12*(9), 1620-1633. https://doi.org/10.1111/2041-210X.13650

Minitab LLC. (2021). *Interpret the key results for Contour Plot*. Retrieved 08 February from https://support.minitab.com/en-us/minitab/20/help-and-how-to/graphs/contour-plot/key-results/

Mohanty, M. P., & Simonovic, S. (2021). 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

Nagle, N. N., Buttenfield, B. P., Leyk, S., & Spielman, S. (2014). Dasymetric modeling and uncertainty. *Annals of the Association of American Geographers*, *104*(1), 80-95. https://doi.org/10.1080/00045608.2013.843439

National Research Council. (1994). *Science Priorities for the Human Dimensions of Global Change*. The National Academies Press. https://doi.org/10.17226/9175

Nieves, J., Stevens, F., Gaughan, A., Linard, C., Sorichetta, A., Hornby, G., Patel, N., & Tatem, A. (2017). Examining the correlates and drivers of human population distributions across low- and middle-income countries. *Journal of the Royal Society interface*, *14*(137), 20170401. https://doi.org/10.1098/rsif.2017.0401

Nieves, J. J., Bondarenko, M., Sorichetta, A., Steele, J. E., Kerr, D., Carioli, A., Stevens, F. R., Gaughan, A. E., & Tatem, A. J. (2020a). Predicting Near-Future Built-Settlement Expansion Using Relative Changes in Small Area Populations. *Remote Sensing*, *12*(10), 1545. https://doi.org/10.3390/rs12101545

Nieves, J. J., Sorichetta, A., Linard, C., Bondarenko, M., Steele, J. E., Stevens, F. R., Gaughan, A. E., Carioli, A., Clarke, D. J., Esch, T., & Tatem, A. J. (2020b). Annually modelling built-settlements between remotely-sensed observations using relative changes in subnational populations and lights at night. *Computers, Environment and Urban Systems*, *80*, 101444. https://doi.org/doi.org/10.1016/j.compenvurbsys.2019.101444

Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., & Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics*, *6*(1), 5. https://doi.org/10.1186/1478-7954-6-5

Orynbaikyzy, A., Gessner, U., & Conrad, C. *Spatial Transferability of Random Forest Models for Crop Type Classification using Sentinel-1 and Sentinel-2.* .

Ottensmann, J. R. (2018). On population-weighted density. *SSRN*. https://doi.org/10.2139/ssrn.3119965

Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), 217-222. https://doi.org/10.1080/01431160412331269698

Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., Sorichetta, A., Zeidler, J., Kuenzer, C., & Dech, S. (2019). New perspectives for mapping global population distribution using world settlement footprint products. *Sustainability*, *11*(21). https://doi.org/10.3390/su11216056

Palacios-Lopez, D., Bachofer, F., Esch, T., Marconcini, M., MacManus, K., Sorichetta, A., Zeidler, J., Dech, S., Tatem, A. J., & Reinartz, P. (2021). High-Resolution Gridded Population Datasets: Exploring the Capabilities of the World Settlement Footprint 2019 Imperviousness Layer for the African Continent. *Remote Sensing*, *13*(6), 1142. https://doi.org/10.3390/rs13061142

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, *11*(5), 1633-1644. https://doi.org/10.5194/hess-11-1633-2007

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418-422. https://doi.org/10.1038/nature20584

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., & Syrris, V. (2016). Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. *Publications Office of the European Union*, 1-62. https://doi.org/10.2788/253582

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., & Lu, L. (2013). A global human settlement layer from optical HR/VHR RS data: concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(5), 2102-2131. https://doi.org/10.1109/JSTARS.2013.2271445

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., & Bayol, N. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communicationss*, *11*(1), 1-11. https://doi.org/10.1038/s41467-020-18321-y

POPGRID. (2021). *Leaving no one off the Map*. https://www.popgrid.org/sites/default/files/documents/popgrid-executive-summary-final.pdf

Qader, S., Lefebvre, V., Tatem, A., Pape, U., Himelein, K., Ninneman, A., Bengtsson, L., & Bird, T. (2021). Semi-automatic mapping of pre-census enumeration areas and population sampling frames. *Humanities and Social Sciences Communications*, *8*(1), 3. https://doi.org/10.1057/s41599-020-00670-0

Qi, Y., & Wu, J. (1996). Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices. *Landscape ecology*, *11*(1), 39-49. https://doi.org/10.1007/BF02087112

Qiu, Y., Zhao, X., Fan, D., Li, S., & Zhao, Y. (2022). Disaggregating population data for assessing progress of SDGs: methods and applications. *International Journal of Digital Earth*, *15*(1), 2-29. https://doi.org/10.1080/17538947.2021.2013553

Rader, B., Astley, C. M., Sewalk, K., Delamater, P. L., Cordiano, K., Wronski, L., Rivera, J. M., Hallberg, K., Pera, M. F., & Cantor, J. (2021). Spatial Accessibility Modeling of Vaccine Deserts as Barriers to Controlling SARS-CoV-2. *medRxiv*. https://doi.org/10.1101/2021.06.09.21252858

Rader, B., Nande, A., Adlam, B., Hill, A. L., Reiner, R. C., Pigott, D. M., Gutierrez, B., group, C.-d. w., Brownstein, J. S., Castro, M. C., Tian, H., Pybus, O. G., Scarpino, S. V., & Kraemer, M. U. G. (2020). Crowding and the epidemic intensity of COVID-19 transmission. *medRxiv*, 2020.2004.2015.20064980. https://doi.org/10.1101/2020.04.15.20064980

Reed, F., Gaughan, A., Stevens, F., Yetman, G., Sorichetta, A., & Tatem, A. (2018). Gridded population maps informed by different built settlement products. *Data*, *3*(3), 33. https://doi.org/10.3390/data3030033

Ritchie, H., & Roser, M. (2019). *Urbanization*. Retrieved 13 May from https://ourworldindata.org/urbanization

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-

cover classification. *ISPRS journal of photogrammetry & remote sensing*, *67*, 93-104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rosling, H., Rosling, O., & Rosling, A. (2018). *Factfulness*. Hodder & Stoughton Ltd.

Rubinyi, S., Blankespoor, B., & Hall, J. W. (2021). The utility of built environment geospatial data for high-resolution dasymetric global population modeling. *Computers, Environment & Urban Systems*, *86*, 101594. https://doi.org/10.1016/j.compenvurbsys.2021.101594

Ruiz Hernandez, I. E., & Shi, W. (2018). A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. *International Journal of Remote Sensing*, *39*(4), 1175-1198. https://doi.org/10.1080/01431161.2017.1395968

Sadahiro, Y. (2000). Accuracy of areal weighting interpolation: effects of geometrical properties of zonal systems. *Geographical review of Japan, Series B.*, *73*(1), 85-90. https://doi.org/10.4157/grj1984b.73.85

Sankoh, O. (2017). Why population-based data are crucial to achieving the Sustainable Development Goals. *International Journal of Epidemiology*, *46*(1), 4-7. https://doi.org/10.1093/ije/dyx010

Schug, F., Frantz, D., van der Linden, S., & Hostert, P. (2021). Gridded population mapping for Germany based on building density, height and type from Earth Observation data using census disaggregation and bottom-up estimates. *PLoS One*, *16*(3), e0249044. https://doi.org/10.1371/journal.pone.0249044

Scikit-learn. Machine Learning in Python. Retrieved 01 June from https://scikit-learn.org/stable/index.html

Serrano Giné, D., Russo, A., Brandajs, F., & Pérez Albert, M. Y. (2016). Characterizing European urban settlements from population data: a cartographic approach. *Cartography and Geographic Information Science*, *43*(5), 442-453. https://doi.org/10.1080/15230406.2015.1076737

Shang, S., Du, S., Du, S., & Zhu, S. (2021). Estimating building-scale population using multi-source spatial data. *Cities*, *111*, 103002. https://doi.org/10.1016/j.cities.2020.103002

Sinha, P., Gaughan, A. E., Stevens, F. R., Nieves, J. J., Sorichetta, A., & Tatem, A. J. (2019). Assessing the spatial sensitivity of a random forest model: Application in gridded population modeling. *Computers, Environment and Urban Systems*, *75*, 132-145. https://doi.org/10.1016/j.compenvurbsys.2019.01.006

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). Continental-Scale Building Detection from High Resolution Satellite Imagery. *arXiv preprint arXiv:.12283*. https://doi.org/10.48550/arXiv.2107.12283

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature communications*, *10*(1), 1-7. https://doi.org/10.1038/s41467-019-09282-y

Sorichetta, A., Bird, T. J., Ruktanonchai, N. W., zu Erbach-Schoenberg, E., Pezzulo, C., Tejedor, N., Waldock, I. C., Sadler, J. D., Garcia, A. J., Sedda, L., & Tatem, A. J. (2016). Mapping internal connectivity through human migration in malaria endemic countries. *Scientific Data*, *3*, 160066. https://doi.org/10.1038/sdata.2016.66

Sorichetta, A., Hornby, G. M., Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data*, *2*(1), 1-12. https://doi.org/10.1038/sdata.2015.45

Stéphane, D., Laurence, D., Raffaele, G., Valérie, A., & Eloise, R. (2020). Land cover maps of Antananarivo (capital of Madagascar) produced by processing multisource satellite imagery and geospatial reference data. *Data in Brief*, *31*, 105952. https://doi.org/10.1016/j.dib.2020.105952

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. (2015a). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, *10*(2). https://doi.org/10.1371/journal.pone.0107042

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015b). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, *10*(2), e0107042. https://doi.org/10.1371/journal.pone.0107042

Stevens, F. R., Gaughan, A. E., Nieves, J. J., King, A., Sorichetta, A., Linard, C., & Tatem, A. J. (2020). Comparisons of two global built area land cover datasets in methods to disaggregate human population in eleven countries from the global South. *International Journal of Digital Earth*, *13*(1), 78-100. https://doi.org/10.1080/17538947.2019.1633424

Sturrock, H. J., Woolheater, K., Bennett, A. F., Andrade-Pacheco, R., & Midekisa, A. (2018). Predicting residential structures from open source remotely enumerated data using machine learning. *PLoS One*, *13*(9), e0204399. https://doi.org/10.1371/journal.pone.0204399

Su, M. D., Lin, M. C., Hsieh, H. I., Tsai, B. W., & Lin, C. H. (2010). Multi-layer multi-class dasymetric mapping to estimate population distribution. *Sci Total Environ*, *408*(20), 4807-4816. https://doi.org/10.1016/j.scitotenv.2010.06.032

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, *22*(16), 3061-3076.

Tatem, A. J., Noor, A. M., von Hagen, C., Di Gregorio, A., & Hay, S. I. (2007). High resolution population maps for low income nations: combining land cover and census in East Africa. *PLoS One*, *2*(12), e1298. https://doi.org/10.1371/journal.pone.0001298

Taw, N. P. (2015). *The 2014 Myanmar Popualtion and Housing Census. Highlights of the Main Results*. M. o. I. a. P. Department of Population.

Tellman, B., Sullivan, J., Kuhn, C., Kettner, A., Doyle, C., Brakenridge, G., Erickson, T., & Slayback, D. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80-86. https://doi.org/10.1038/s41586-021-03695-w

Thomson, D. R., Gaughan, A. E., Stevens, F. R., Yetman, G., Elias, P., & Chen, R. (2021a). Evaluating the Accuracy of Gridded Population Estimates in Slums: A Case Study in Nigeria and Kenya. *Urban Science*, *5*(2), 48. https://doi.org/10.3390/urbansci5020048

Thomson, D. R., Leasure, D. R., Bird, T., Tzavidis, N., & Tatem, A. J. (2021b). How accurate are WorldPop-Global-Unconstrained gridded population data at the cell-level?: A

simulation analysis in urban Namibia. *Preprints*, *2021020492*. https://doi.org/10.20944/preprints202102.0492.v2

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., & Prydz, E. B. (2017). Mapping the world population one building at a time. *arXiv*, arXiv:1712.05839. https://doi.org/10.48550/arXiv.1712.05839

Tobler, W., Deichmann, U., Gottsegen, J., & Maloy, K. (1997). World population in a grid of spherical quadrilaterals. *International Journal of Population Geography*, *3*(3), 203-225.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, *74*(367), 519-530. https://doi.org/10.1080/01621459.1979.10481647

Tuholske, C., Caylor, K., Evans, T., & Avery, R. (2019). Variability in urban population distributions across Africa. *Environmental Research Letters*, *14*(8), 085009.

Tuholske, C., Gaughan, A. E., Sorichetta, A., de Sherbinin, A., Bucherie, A., Hultquist, C., Stevens, F., Kruczkiewicz, A., Huyck, C., & Yetman, G. (2021). Implications for Tracking SDG Indicator Metrics with Gridded Population Data. *Sustainability*, *13*(13), 7329. https://doi.org/10.3390/su13137329

Twinoburyo, E., Henao, L., Dushime, O., Simkoko, A., Kassa, Y., & Ndahiro, D. (2019). *Africa 2030. Sustainable Development Goals three-year reality check*. https://sdgcafrica.org/wp-content/uploads/2019/06/AFRICA-2030-SDGs-THREE-YEAR-REALITY-CHECK-REPORT.pdf

UN-Climate Change. (2022). *The Paris Agreement* https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

UN-Habitat. (2022). *New Urban Agenda*. https://unhabitat.org/about-us/new-urban-agenda

UN-Statistics Division. (2020). *Standard country or area codes for statistical use (M49)*. Retrieved 23 September from https://unstats.un.org/unsd/methodology/m49/

UN-Statistics Division. (2022). *2020 World Population Housing Census Programme*.

UN. (2018). *World Population Policies 2015: Highlights*. https://www.un.org/en/development/desa/population/publications/pdf/policy/WPP2015/WPP2015_Highlights.pdf

UN. (2019). *World Population Prospects 2019: Highlights* (ST/ESA/SER.A/423). https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf

UN. (2021). *The Sustainable Development Goals 2018. A data revolution in motion*. Retrieved 17 December from https://unstats.un.org/sdgs/report/2018/data_revolution

UN. (2022a). *Leave No One Behind*. https://unsdg.un.org/2030-agenda/universal-values/leave-no-one-behind

UN. (2022b). *Sustainable Development Goals*. Retrieved 01 December from https://sdgs.un.org/goals

UNCTAD. (2021). *Reaping the potential benefits of the African Continental Free Trade Area for Inclusive Growth* (Economic Development In Africa Report 2021, Issue. https://unctad.org/system/files/official-document/aldcafrica2021_en.pdf

UNDRR. (2015). *Sendai Frameworkd for Disaster Risk Reduction 2015-2030*. https://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf

UNESCO. (2019). *UN World Water Development Report 2019: Leaving No One Behind*. https://unesdoc.unesco.org/ark:/48223/pf0000367306

UNESCO. (2021). *Education in Africa*. Institute for Statistics. Retrieved 13 December from http://uis.unesco.org/en/topic/education-africa

UNFPA. (2012). *Conducting a Social, Demographic and Economic Survey of Afghanistan*. https://www.unfpa.org/news/conducting-social-demographic-and-economic-survey-afghanistan

UNFPA. (2014). *Population Estimation Survey 2014 for the Pre-War Regions of Somalia*. https://somalia.unfpa.org/sites/default/files/pub-pdf/Population-Estimation-Survey-of-Somalia-PESS-2013-2014.pdf

UNFPA. (2021). *The World Population Dashboard*. Retrieved 01 December from https://www.unfpa.org/data/world-population-dashboard

Ural, S., Hussain, E., & Shan, J. (2011). Building population mapping with aerial imagery and GIS data. *International Journal of Applied Earth Observation and Geoinformation*, *13*(6), 841-852. https://doi.org/10.1016/j.jag.2011.06.004

Vandeput, N. *Forecasting KPIs: RMSE, MAE, MAPE & Bias*. Retrieved 19 October from https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Wardrop, N., Jochem, W., Bird, T., Chamberlain, H., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., & Tatem, A. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3529-3537. https://doi.org/10.1073/pnas.1715305115

Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., Bhaduri, B. L., Moehl, J. J., & Reith, A. E. (2018). Census-independent population mapping in northern Nigeria. *Remote Sensing of Environment*, *204*, 786-798.

Wicaksono, P., & Aryaguna, P. A. (2020). Analyses of inter-class spectral separability and classification accuracy of benthic habitat mapping using multispectral image. *Remote Sensing Applications: Society Environment*, *19*, 100335. https://doi.org/10.1016/j.rsase.2020.100335

WorldPop. (2020). *Top-down estimation modelling: Constrained vs Unconstrained*. Retrieved 09 January from https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained

Wright, J. K. (1936). A method of mapping densities of population: With Cape Cod as an example. *Geographical Review*, *26*(1), 103-110. https://doi.org/doi.org/10.2307/209467

Xie, J., & Zhou, J. (2017). Classification of urban building type from high spatial resolution remote sensing imagery using extended MRS and soft BP network. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*, *10*(8), 3515-3528. https://doi.org/10.1109/JSTARS.2017.2686422

Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., & Wang, W. (2017a). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment & Urban Systems*, *64*, 215-228. https://doi.org/10.1016/j.compenvurbsys.2017.03.001

Zhang, X., He, G., Zhang, Z., Peng, Y., & Long, T. J. C. C. (2017b). Spectral-spatial multi-feature classification of remote sensing big data based on a random forest classifier for land cover mapping. *Cluster Computing*, *20*(3), 2311-2321. https://doi.org/10.1007/s10586-017-0950-0