RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

A census from heaven

Georganos, Stefanos; Hafner, Sebastian; Kuffer, Monika; Linard, Catherine; Ban, Yifang

Link to publication

A census from heaven: Unraveling the potential of deep learning and Earth Observation for intra-urban population mapping in data scarce environments

Stefanos Georganos [a,*], Sebastian Hafner [a], Monika Kuffer [b], Catherine Linard [c,d], Yifang Ban [a]

[a] *Division of Geoinformatics, KTH Royal Institute of Technology, Stockholm 10044, Sweden*
[b] *Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7514 AE Enschede, The Netherlands*
[c] *Institute of Life, Earth and Environment, Université de Namur, Namur, Belgium*
[d] *Department of Geography, Université de Namur, Namur, Belgium*

ARTICLE INFO

ABSTRACT

Urban population distribution maps are vital elements for monitoring the Sustainable Development Goals, appropriately allocating resources such as vaccination campaigns, and facilitating evidence-based decision making. Typically, population distribution maps are derived from census data from the region of interest. Nevertheless, in several low- and middle-income countries, census information may be unreliable, outdated or unsuitable for spatial analysis at the intra-urban level, which poses severe limitations in the development of urban population maps of adequate quality. To address these shortcomings, we deploy a novel framework utilizing multisource Earth Observation (EO) information such as Sentinel-2 and very-high-resolution Pleiades imagery, openly available building footprint datasets, and deep learning (DL) architectures, providing end-to-end solutions to the production of high quality intra-urban population distribution maps in data scarce contexts. Using several case studies in Sub-Saharan Africa, namely Dakar (Senegal), Nairobi (Kenya) and Dar es Salaam (Tanzania), our results emphasize that the combination of DL and EO data is very potent and can successfully capture relationships between the retrieved image features and population counts at fine spatial resolutions (100 meter). Moreover, for the first time, we used state-of-the-art domain adaptation methods to predict population distributions in Dar es Salaam and Nairobi (R^2 = 0.39, 0.60) that did not require national census or survey data from Kenya or Tanzania, but only a sample of training locations from Dakar. The DL architecture is based on a modified ResNet-18 model with dual-streams to analyze multi-modal data. Our findings have strong implications for the development of a new generation of urban population products that are an output of end-to-end solutions, can be updated frequently and rely completely on open data.

1. Introduction

Spatially detailed population information are necessary requisites for a wide range of applications related to sustainability and planning, epidemiology, natural hazards (population at risk) and crucial elements for the monitoring of Sustainable Development Goals (SDG) (Wardrop et al., 2018). Population distribution information is essential for planning the appropriate allocation of resources, with recent examples being vaccination plans during the Ebola outbreak in West Africa (Wardrop et al., 2018) or for the COVID-19 pandemic currently ravaging the globe. However, their quality in data scarce environments such as in low- and middle-income countries is often unreliable both in terms of spatiotemporal consistency and granularity (Grippa et al., 2019b). The disparaging effects of this data gaps are most evident in Sub-Saharan Africa (SSA) where census data are not easily accessible, often outdated or not available at spatial levels

that allow for in-depth analyses (Grippa, 2019; Linard et al., 2012). For instance, Fig. 1 illustrates the number of years since the last census amongst African countries using 2019 as the reference date. Several countries have not conducted a census since more than 15 years which complicates evidence-based decision making (Wardrop et al., 2018).

International efforts such as WorldPoP (Tatem, 2017), GHS-PoP (Freire and Halkia, 2014) and LandScan (Dobson et al., 2000) have helped mitigate this gap by providing openly accessible, global population distribution products at relatively high spatial resolutions (100 m–1 km) (Chen et al., 2020). Nonetheless, their quality with respect to the intra-urban level is limited, as they were mostly designed for large scale analysis (i.e., global or national level). At the same time, SSA is facing a rapid urbanization shift with current estimates placing more than 60% of the African population in cities by 2050. This has led to the proliferation of deprived neighborhoods that often lack basic
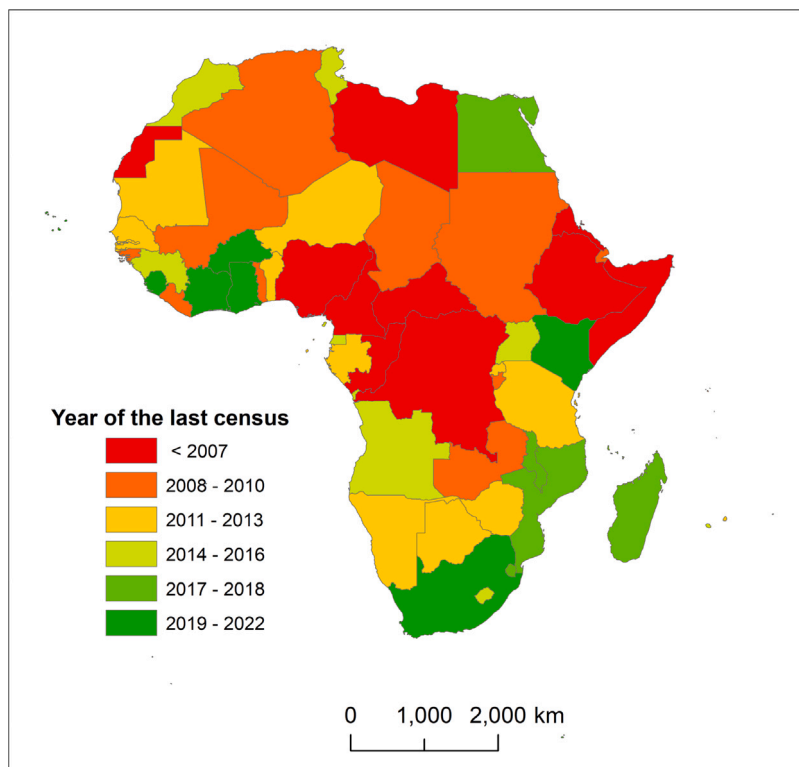
Fig. 1. Number of years since the last census in African countries using 2022 as the reference date.

services and amenities such as adequate open space and access to clean water. As recent research has shown, deprived urban communities are vastly underestimated in current global population products (Thomson et al., 2021), which severely hinders efforts to address the needs of urban residents and enhance evidence-based policy making. Thus, the need to better represent the urban population both in terms of accuracy and spatial detail is imperative.

Unfortunately, mapping population distribution at a fine resolution, intra-urban level is still vastly under-researched (Thomson et al., 2020). In the past decade, machine learning (ML) techniques have been consistently deployed to model population distribution as a function of Earth Observation (EO) information (Stevens et al., 2015). Arguably, the most common approach to do so is through top-down disaggregation of census data (Leyk et al., 2019) where ML algorithms are used to learn the relationship between satellite extracted information (the land cover) and population density. Alternatively, bottom-up approaches rely on geostatistical models between spatial layers and micro-census surveys to extrapolate population counts across the desired area (Neal et al., 2021; Wardrop et al., 2018). In both cases, a major limitation of the current state-of-the-art in urban population mapping in data-scarce environments such as SSA, is the absence of models which are (i) generalizable and transferable, (ii) least dependent on census data and (iii) do not require the continuous collection of accurate land-use, land-cover and ancillary data to perform the census disaggregation. Indeed, although recent work has demonstrated remarkable achievements in intra-urban population mapping in SSA cities (Grippa et al., 2019b), challenges remain, such as the lack of efforts to transfer population models from an urban area to another, the reliance in proprietary and costly satellite imagery, and the spatio-temporal restrictions of censuses or the costly burden of conducting local surveys. Moreover, current ML methods may not be adequately tailored for spatial data and as such, a knowledge gap in the field is evident.

Emerging from the field of computer vision and image processing, deep learning (DL) has shown a fascinating efficiency in almost all tasks it has set its claws on, particularly on issues related to semantic segmentation, object detection and scene classification (Ma et al., 2019),

through the use of several architectures, particularly of Convolutional Neural Networks (CNN). CNN automatically extract a range of simple to complex representational features directly from the raw image data, alleviating the need for the handcrafting of representational features which is common in standard ML methods (Krizhevsky et al., 2012). Surprisingly, there have been few, significant and largescale applications of the use of CNN for population inference from remotely sensed images. Robinson et al. (2017) used a standard CNN network, Landsat data and census population counts to predict population distribution across the United States at a roughly 1-km spatial resolution, with results well-aligning with reference data, showing strong advantages over traditional methods. Tiecke et al. (2017) used CNN architectures to first detect settlements in remotely sensed imagery (Facebook's High Resolution Settlement Layer) and then distribute population at the global level via ML methods, in a two-step procedure, with encouraging results. Nonetheless, the focus of their validation was on the accuracy for detecting buildings and not on the quality of the population distribution. Recently, Huang et al. (2021) demonstrated the capabilities of DL to map population patterns in two cities in the United States through a variety of different state-of-the-art architectures by training on existing global population grids. Consequently, while showing some preliminary findings, the field of DL-based population prediction through satellite images is widely open for exploration, particularly when it comes to end to end solutions and investigating different EO data as inputs to the models. Equally importantly, DL approaches have not been adjusted to the peculiarities of SSA cities (i.e., limited or absence of training data).

As such, the overarching aim of this research is to investigate and propose a framework to efficiently couple DL and EO information for the task of intra-urban population mapping in data scarce environments. This aim can be partitioned down to the following multifaceted set of objectives:

1. We investigate the potential of existing gridded urban population data derived from sub-meter, very-high-resolution (VHR) satellite imagery to serve as training data in SSA cities. This will allow for

Table 1
Satellite and population data used in the study.

City	Earth observation data		Population data		Building footprints
	Sentinel-2 (acq. date)	Pleiades (acq. date)	Grid level (date and # cells)	Census level (date and # units)	
Dakar	2015-11-25	2015-07-07	2013, 20,372 (100-m grid) (Grippa et al., 2019a,b)	2013, 677 units (ANSD (Agence Nationale de la Statistique et de la Démographie), 2013)	Google building footprints (Sirko et al., 2021)
Nairobi	2019-02-03	No data	No data	2019, 111 units (Kenya National Bureau of Statistics, 2019)	
Dar es Salaam	2016-07-10	No data	No data	2012 units, 79 units (National Bureau Statistics Office (NBS), 2012)	

the development of significantly more accurate training data than existing global population products as well as a more adequate source of validation.

2. We evaluate the potential of state-of-the-art DL approaches to extract population counts at the grid level using as input multi-source satellite imagery.

3. We unravel the potential of publicly available EO data such as Sentinel-2 imagery to map urban population patterns in SSA cities through a systematic comparison against VHR imagery (Pleiades) and assess the contribution of geospatial ancillary data such as Google's building footprint dataset.

4. For the first time, we explore the transferability potential of models trained in Sentinel-2 images from a single city to accurately predict population distribution in other SSA cities where no training data are used, through tailored domain adaptation (DA) frameworks.

2. Case studies and data

As a proof of concept, we involve data from a set of SSA cities – Dakar (Senegal), Nairobi (Kenya) and Dar es Salaam (Tanzania). Dakar is used as the flagship of this research due to the vast amount of multi-source datasets and high-quality census data available, while the other two cities are used to evaluate the transferability experiments. All cities have exhibited strong urban transformation changes in the last decades and provide diversity with respect to building patterns and urban morphology. They are all heavily populated cities with intrinsic morphological variations. For instance, Dar es Salaam is mostly informally built, while in Nairobi, more than 50% of the population resides in about 6% of the built-up area, in slumlike conditions (Abascal et al., 2022a). They are also climatically diverse, ranging from the Sahelian semi-arid climate of Dakar, the sub-tropical conditions of Nairobi, to the tropical climate of Dar es Salaam.

Table 1 documents the data used in this study. To investigate the contribution of different EO information, we used both Sentinel-2 MSI and VHR Pleiades imagery in Dakar (Fig. 2, panels a, b). The Sentinel-2 MSI mission acquires optical images at 13 spectral bands with various spatial resolutions (10–60 m). As our aim is intra-urban population modeling and mapping, we made use of the 10 m bands (blue, green, red and near-infrared) which are the best capable to discriminate between variations of the urban fabric. Sentinel-2 images are available in Google Earth Engine (GEE), a cloud-based platform for geospatial data analysis (Gorelick et al., 2017), as analysis-read datacubes composed of ortho-corrected images scaled by a factor of 10,000 (UTM projection). The VHR Pleiades imagery was acquired at a 50 cm spatial resolution and the visible bands of the electromagnetic spectrum were used.

To investigate the transferability potential, we used Sentinel-2 imagery in Nairobi and Dar es Salaam (Fig. 3). For Dakar and Nairobi the top-of-atmosphere Sentinel-2 image (Level-1 A) with the lowest cloud coverage for the respective year was downloaded from GEE. Due to the lack of cloud-free images, a cloud-free composite was generated for Dar Es Salaam, using cloud probability information retrieved via the Sentinel Hub's cloud detector for Sentinel-2 imagery.[1]

In all case studies, we considered building footprints as an additional input to the experiments. Buildings are an essential residential population mapping variable, and for modeling purposes, spatially invariant across domains. The source of the building footprints is the Open Buildings dataset of Google (Sirko et al., 2021). The dataset currently covers 64% of the African continent with over 500 million building footprints produced (inference date between 2018–2021, according to Google Earth's historical archive).

With respect to the population information, a fine scale census dataset was available in Dakar (Fig. 2, panel d). To create training and validation data that are suitable for DL frameworks, we acquired a publicly available product that is derived from the same census, using VHR land use and land cover to distribute population counts, which is of unparalleled quality (R^2 values of over 0.80 at the neighborhood block level) at a 100-m spatial resolution (Grippa et al., 2019b). In that way, we could train and validate the models both at the grid and census level. The gridded map is modeled, taking under account the increased population density of deprived urban areas so it can truly capture intra-urban heterogeneity. Fig. 3 illustrates the images, building footprints and census data used for the other two cities. The small temporal mismatch between census and image data (up to 3 years) is considered acceptable as shown in previous studies employing such data (Georganos et al., 2021b; Grippa et al., 2019b,a; CIESIN, 2005).

3. Methods

3.1. Deep learning framework

We explore four DL approaches that use different input data in combination with three model architectures. An overview of the model architectures is given in Fig. 4. All models are based on residual neural networks, commonly known as ResNets (He et al., 2016), which have been identified as a promising architecture for EO data-based population mapping (Zhuang et al., 2021; Huang et al., 2021). In the following two subsections, the architectures of the fusion model (Fig. 4.b) and the domain adaptation model (Fig. 4.c) are described.

3.1.1. ResNet-18 architecture
The original ResNet architecture, specifically ResNet-18, is shown in Fig. 4.a. It features 17 convolutional layers with 3×3 kernels, followed by an average pooling layer. After the first layer, the network is split into four blocks, each consisting of four convolutional layers. Starting with 64 filters, the number of filters is doubled at the first layer of each consecutive block, while the size of the feature map is halved. Residual shortcut connections are inserted throughout the network to avoid the vanishing gradient problem during network training (He et al., 2016).

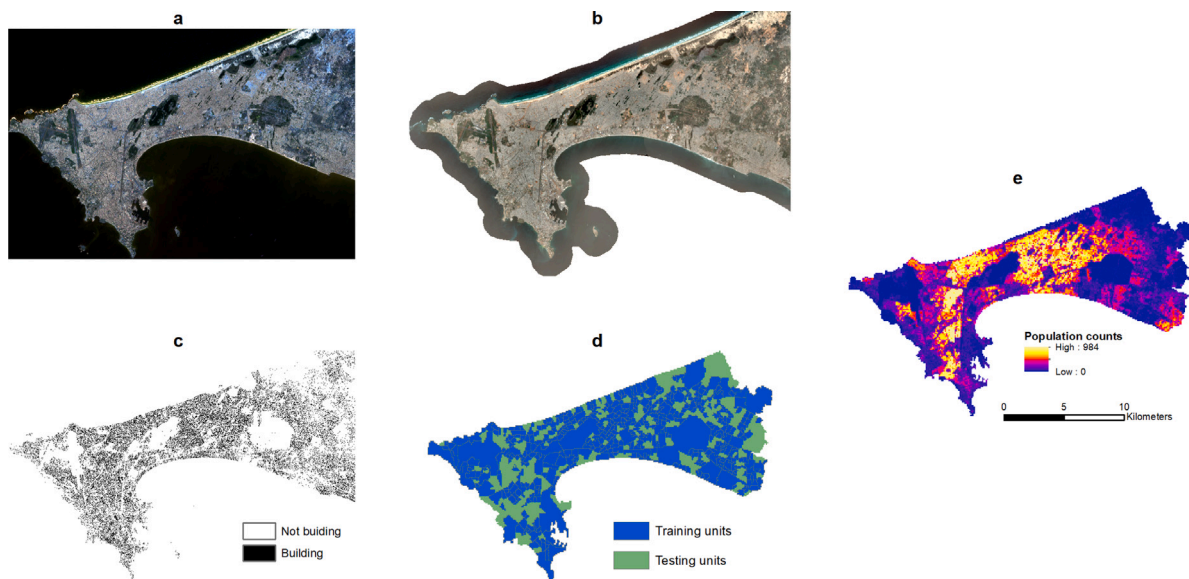[1] https://github.com/sentinel-hub/sentinel2-cloud-detector.

Fig. 2. Data used in Dakar, Senegal. (a) Sentinel-2 truecolour composite, (b) Pleiades truecolor composite, (c) Building footprints, (d) census administrative units and (e) gridded population map at 100 m resolution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
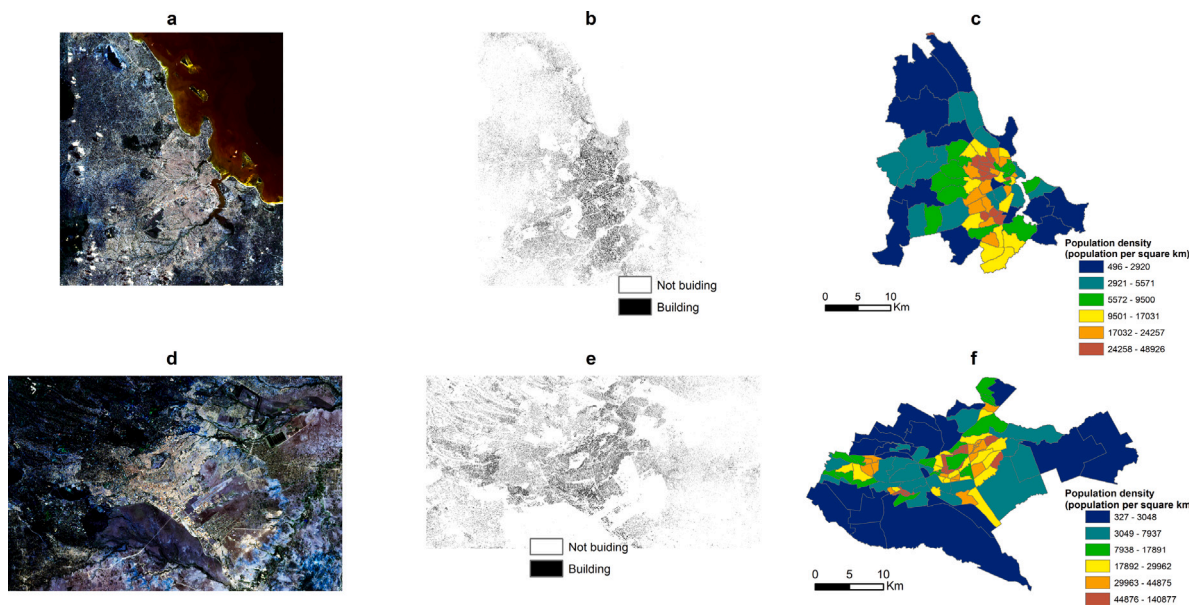

Fig. 3. Data used in Dar es Salaam, (a) Sentinel-2 truecolour image, (b) building footprints, (c) population density at the census level. Similarly, for Nairobi, (d) Sentinel-2 truecolour composite, (e) building footprints and (f) population density at the census level.

Most off-the-shelf CNN architectures were designed for RGB images having 3 input channels. However, due to the use of different inputs with varying channel numbers (i.e., 3 channels for VHR imagery, 4 channels for Sentinel-2 MSI imagery and 1 channel for building footprints), the first layer of the ResNet architecture was modified to accommodate different input sizes. For single channel inputs, the first layer was replaced with a 3 × 3 conv layer with 1 input channel and 64 output channels, using He initialization to generate the initial weights (He et al., 2015). The same was done for 4 input channels by changing the input channel number accordingly. However, He initialization was only used to initialize the weights for the fourth channel, while the pretrained weights for RBG imagery were preserved for the first three channels. Lastly, the features extracted by the ResNet encoder are converted to a population prediction *p* using a fully connected layer, followed by the ReLu activation function to prevent negative predictions. L2 loss is used to train the network by minimizing the error

corresponding to the sum of the squared difference between the true population value *y* and the predicted value *p*, defined as follows:

$$\mathcal{L} = \sum_{i=1}^{n}(y_i - p_i)^2 \tag{1}$$

where *n* is the number of samples.

3.1.2. Data fusion model

For the fusion of Sentinel-2 MSI data and building footprint data, a dual-stream model consisting of two ResNet-18 networks was employed (Fig. 4.c). The ResNet networks process the different inputs separately, before concatenating the extracted features at the decision level (i.e., decision level fusion). Previous studies have shown that this is an effective architectural design for the joint use of multi-modal data (Hafner et al., 2022b; Hu et al., 2019). The final prediction is
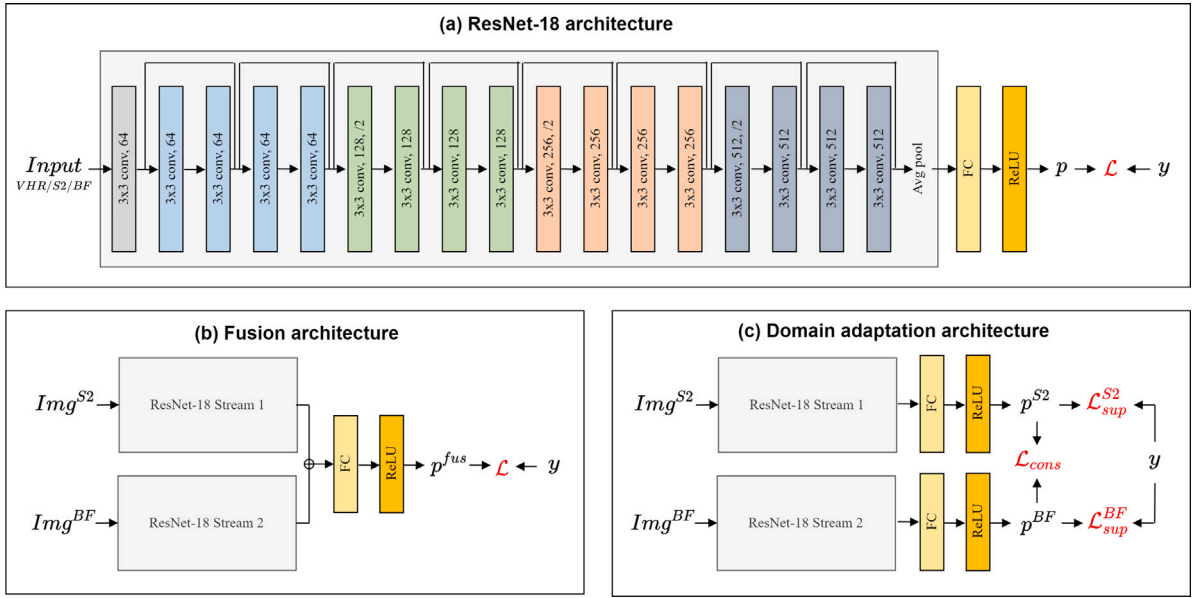
Fig. 4. Overview of the architectures of (a) the modified ResNet-18 model, (b) the dual-stream fusion model and (c) the domain adaptation model.

obtain from the concatenated features via a trainable 1×1 convolution layer followed by the ReLu activation function. This dual-stream network is trained in a fully supervised fashion using L2 loss (Eq. (2)).

3.1.3. Domain adaptation model

Domain Adaptation (DA) techniques aim to transfer a model from its training area (i.e., source domain) to new areas (i.e., target domain), where the data in the new areas is following a different underlying data distributions (i.e., domain shifts). To overcome domain shifts, the remote sensing community has developed a variety of DA techniques (Tuia et al., 2016). Consistency regularization enforces that perturbations of a sample should not significantly change the model output (Laine and Aila, 2016; Sajjadi et al., 2016). Inspired by the recent success of consistency regularization for DA (French et al., 2017; Cui et al., 2019), recent research proposed a new DA technique that leverages unlabeled satellite data acquired by two different sensors, by employing a separate network stream for each data modality and then encouraging consistent predictions across them (Hafner et al., 2022a). Therefore, multi-modal consistency regularization holds great potential to overcome domain shifts in remotely sensed data.

The proposed DA architecture is almost identical to the architecture used for data fusion (Fig. 4.c). It also consists of two ResNet-18 streams that process both data modalities separately. However, unlike the fusion approach, the extracted feature maps are not fused but population predictions are directly obtained from the Sentinel-2 MSI stream (p^{S2}) and the building footprints stream (p^{BF}). To train the model, a twofold loss function consisting of a supervised loss ($\mathcal{L}_{sup}$) for labeled samples and a consistency loss ($\mathcal{L}_{cons}$) for unlabeled samples was constructed. The loss function is defined as follows:

$$\mathcal{L}_{sample} = \begin{cases} \mathcal{L}(p^{S2}, y) + \mathcal{L}(p^{BF}, y), & \text{if } y \text{ exists} \\ \frac{1}{2} \cdot \mathcal{L}(p^{S2}, p^{BF}), & \text{otherwise} \end{cases} \quad (2)$$

The supervised loss function is composed of two sub-terms, measuring the similarity between the two population predictions (p^{s2} and p^{bf}) and the population label (y), where similarity for both sub-terms was measured using L2 loss (Eq. (1)).

To adapt the model to the target domain, the proposed DA approach leverages unlabeled data using consistency regularization. Consistency regularization is typically implemented with a loss term measuring the similarity between predictions obtained from different augmentations of the same unlabeled sample (Oliver et al., 2018). The consistency

loss term is then added to the supervised loss (Bachman et al., 2014). However, we apply a consistency loss to predictions of the sub-networks as proposed in the DA framework for multi-modal data (Hafner et al., 2022a). As a result, inconsistencies between the predicted population from the Sentinel-2 and building footprint data are penalized during training. The underlying idea of this approach is that building footprints can be used to adapt the Sentinel-2 sub-network of the model to the target domain by training both sub-networks to produce similar population predictions from unlabeled data in the target domain area. Once the model is trained, only the prediction from the Sentinel-2 stream is used for inference. Thus, the model does not require building footprints data at the deployment stage.

3.2. Experimental setup

All models were trained for 20 epochs with a batch size of 8. To accelerate training, AdamW (Loshchilov and Hutter, 2018) was employed as optimizer with an initial learning rate of 10^{-4}. As input, patches of size 100 x 100 m were used, resulting in an input dimension of 200 x 200 pixels for the VHR imagery. The Sentinel-2 imagery was upsampled to the same dimension using nearest neighbor interpolation, and the building footprints were rasterized to a spatial resolution of 0.5 m, also resulting in an input size of 200 x 200 pixels. Two data augmentation operations were employed during model training, namely rotations and flips, in order to enhance the training dataset by generating more variant versions. While the label remains unchanged, a rotation randomly rotates images by an angle of $k \cdot 90°$, where $k \in \{0, 1, 2, 3\}$, and a flip horizontally or vertically flips images with a probability of 50%. We implemented everything in Python using Facebook's machine learning framework PyTorch (Paszke et al., 2019) and trained the networks on a Nvidia Titan Xp graphics card. Code is available at https://github.com/SebastianHafner/DDA_PopulationMapping.

3.3. Capturing the intra-urban morphology

The quality of urban population maps is a reflection of the degree they are able to capture morphological variations of the urban forms such as land-use. For instance, deprived urban areas (also known as slums) exhibit higher population densities than planned neighborhoods (Klemmer et al., 2020; Kuffer et al., 2020, 2016; Thomson et al., 2021). Moreover, robust population models at fine geographical

scales, should be able to capture non-residential built-up regions to satisfactory degrees such as industrial, commercial and administrative regions (Grippa et al., 2019b). To assess this aspect, we make use of various products for the three case studies. For Dakar and Dar es Salaam, we use documented land-use maps, which is produced at the street-block level using VHR land cover and elevation data and are publicly available (Grippa et al., 2018; Grippa and Georganos, 2018; Georganos, 2020). In Nairobi, we made use of two datasets, (i) a public land use product developed from the Spatial Information Design Lab, University of Columbia (Williams et al., 2014) and a shapefile of deprived residential areas developed by Spatial Collective and was recently used to characterize deprivation profiles in Nairobi (Georganos et al., 2021a). We selected areas larger than 1 hectare (to exclude blocks smaller than the resolution of the population grids) for 3 classes, namely "Deprived urban area", "Planned Residential" and "Administrative, Commercial and Industrial" (ACI) and extracted the population density of the different models. In the case of Dar es Salaam and Nairobi, we also incorporated WorldPop data as an external comparative source.

3.4. Accuracy metrics

In Dakar, we used 70% of the census units for training and 30% for testing (Fig. 2.d). In Nairobi and Dar es Salaam, we used all available census units for testing. To evaluate the predictions we make use of three commonly employed metrics in population studies, namely the Mean Absolute Error (MAE; Eq. (3)), the Root Mean Squared Error (RMSE; Eq. (4)) and the coefficient of determination (R^2; Eq. (5)), Linard et al. (2012) and Georganos et al. (2021b). Additionally, to provide information at the city level for Dar es Salaam and Nairobi, we computed the ratio between the total sums of predicted and census population (Pop ratio) as an indicator of practical use of these models. A value of one implies that the model corrected predicted the total population counts, while values lower and higher than 1 indicate under- and overestimation, respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{d_i - f_i}{\sigma_i}\right)^2} \tag{3}$$

$$\text{MAE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - x_i| \tag{4}$$

where x_i, n is the sample size and y_i is the inference.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{5}$$

where SS_{res} refers to the residual sum of squares and SS_{tot} to the total variability of the data.

4. Results

4.1. Training

Fig. 5 illustrates the training and testing loss curves for the DL models. Additionally, Subfigure f illustrates the loss curve for the consistency term used for DA. Losses for all models converged within the first 5 epochs of training. Furthermore, performances on the test set remained stable towards the end of training (i.e., no overfitting occurred). The consistency loss sharply increased at the beginning of training, indicating strong disagreement between the population predictions of the two sub-networks. However, after the initial training phase, the consistency loss started to decrease, showing that building footprints were successfully employed to adapt the Sentinel-2 sub-network to the target domain.

Table 2
MSE, MAE and R^2 values at the grid and census levels in Dakar.

		Pleiades	Sentinel-2	Building footprints	Sentinel-2 + Building footprints
Grid level	RMSE	82.6	92.9	112.6	92.7
	MAE	49.6	58.4	70.1	58.0
	R^2	0.79	0.71	0.59	0.73
Census level	RMSE	1773.1	2197.7	2482.2	2182.7
	MAE	1061.7	1313.9	1472.3	1269.7
	R^2	0.84	0.67	0.59	0.73

4.1.1. Comparative analysis in Dakar

Fig. 6 illustrates the predictions of the DL models in Dakar on the test samples at the grid level. Not surprisingly, using VHR imagery as input produced the most accurate population predictions with a MAE of 49.6 and an RMSE of 82.6. Notably, models using Sentinel-2 imagery performed remarkably well (MAE = 58.4, RMSE = 92.9), highlighting its potential for retrieving fine-scale, intra-urban predictions. Interestingly, simply using building footprints as an input produces satisfactory results with a MAE of 70.1 and an RMSE of 112.6. Combining features from Sentinel-2 and building footprints exhibits slightly better results than their individual use.

Aggregating the results at the census level can also be informative and provides stronger validation as the census population counts are official data. Fig. 7 exhibits the same type of results but aggregated at the census level. In a similar fashion, VHR-based predictions are the most accurate (MAE = 1061.7, RMSE = 1773.1). However, in this case, the merits of combining Sentinel-2 imagery with building footprints are more evident. This is most apparent when predicting low population counts as the overdispersion is clearly reduced compared to single building footprint or Sentinel-2 models. The results of both rounds of comparative experiments are shown in Table 2.

4.1.2. Capturing the intra-urban morphology in Dakar

Fig. 8 exhibits the boxplots depicting the population density distribution of each of the selected land use categories in Dakar. It is evident that there is a clear separation with respect to population density between the three classes. Encouragingly, all DL-based models appear to capture these differences. For instance, using the VHR model, which is the best performing model, deprived areas demonstrated significantly higher means (mean = 342.54 people per hectare) than planned residential areas (mean = 155.04 people per hectare). Notably, the average density within the ACI class was particularly low (average 78.23 people per hectare) compared to the residential classes, indicating that the models are able distinguish urban morphological variations. The rest of the models follow similar patterns, a finding that suggests that Sentinel-2 data can capture land-use differences and their impact on population density in a satisfactory manner.

4.2. Transferability experiments

Using training data from Dakar, we apply the Sentinel-2 and building footprint models in Dar es Salaam and Nairobi to predict population counts at a 100-m resolution. To validate the predictions, we aggregate them at the finest census level available in both cities.

4.2.1. Nairobi

Fig. 9 illustrates the results in Nairobi. It is evident that when using Sentinel-2 as input, the model is unable to capture the intrinsically different populations distributions of Nairobi (MAE = 35 488, RMSE = 51 579), even when considering the addition of building footprints (MAE = 32 503, RMSE = 46 089). The results drastically change when we apply the consistency-based DA method where the model is also exposed to unlabeled image and building footprint data in other cities.
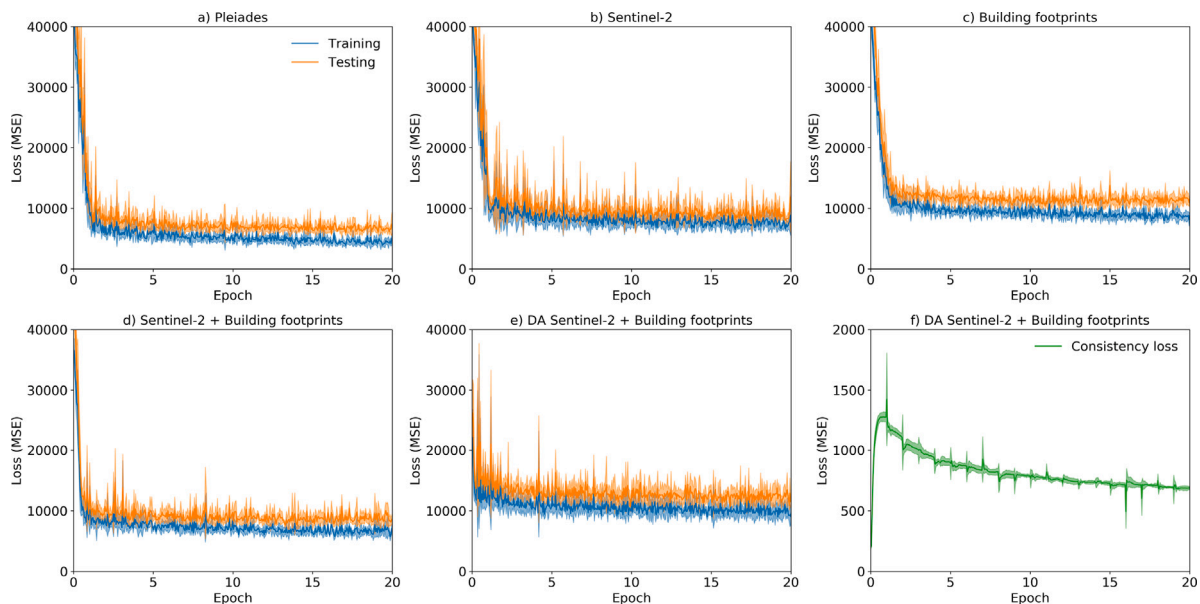
Fig. 5. Training and testing curves for different models: (a) Pleiades, (b) Sentinel-2, (c) Building footprints, (d) Sentinel-2 and building footprints and (e) Sentinel-2 and building footprints with domain adaptation. (f) shows the consistency loss curve of the domain adaptation model.
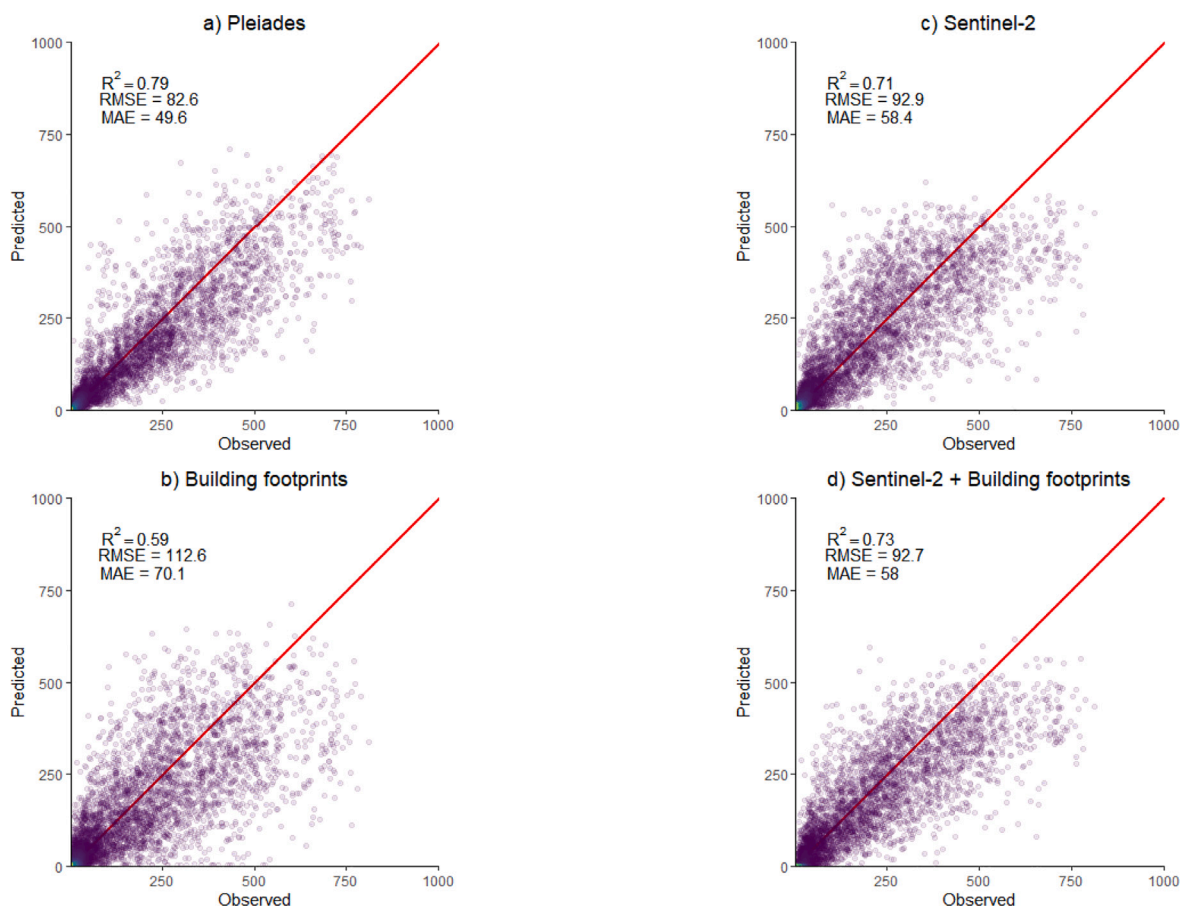

Fig. 6. DL model predictions at the grid level on the test dataset using different types of image inputs. (a) Pleiades RGB images, (b) Sentinel-2, (c) Building footprints and (d) Sentinel-2 and building footprints. The red line indicates the 1 to 1 axis.

For instance, the DA-based approach consistently produced lower error metrics (MAE = 23 562 and RMSE = 35 117) against the non-DA models. Additionally, the inference in the DA approach only uses Sentinel-2 imagery as the building footprints are utilized only during the training stage. Fig. 10 illustrates the gridded population products of the various experiments in Nairobi. The spatial patterns appear more realistic and aligned with the census population information in the approaches involving DA.
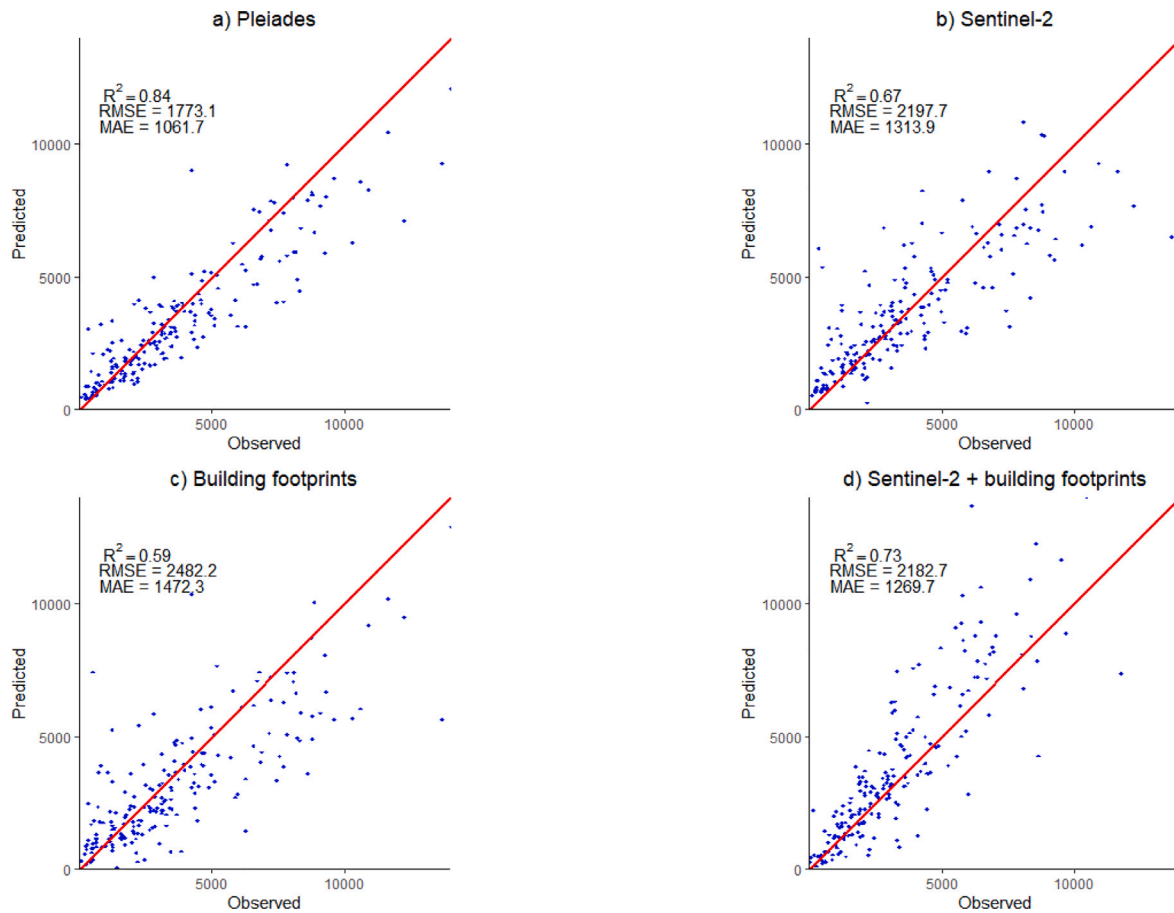
Fig. 7. DL model predictions aggregated at the census level, on a test dataset using different types of image inputs. (a) Pleiades RGB images, (b) Sentinel-2, (c) Building footprints and (d) Sentinel-2 and building footprints. The red line indicates the 1 to 1 axis.

4.2.2. Dar es Salaam

Fig. 11 presents the result outputs for the city of Dar es Salaam. Similar with the case of Nairobi, directly applying the Dakar-trained model in Dar es Salaam was of limited merit both in the case of Sentinel-2 imagery (MAE = 35 693, RMSE = 43 270) but also when including building footprints (MAE = 29 595, RMSE = 35 147). On the other hand, the DA-based models captured the relationships between image and population information in a more realistic way with significantly lower error margins and more consistent with census data (MAE = 15 667, RMSE = 19 655). Fig. 12 illustrates the predicted gridded outputs for the different experiments where the merits of DA approaches is evidently shown, with spatial patterns aligning with census values. The complete evaluation metrics for both cities are described in Table 3. Notably, the Pop ratio values for Nairobi, indicate an underestimation of the total population by all models, with the building footprint and DA-based models being the best performing. In Dar es Salaam, all models estimated the total population in a more precise manner, with the building footprint models exhibiting a mild overestimation (Pop ratio = 1.10), while the DA approach showed a mild underestimation (Pop ratio = 0.84).

4.2.3. Capturing the intra-urban morphology in Nairobi and Dar es Salaam

As previously assessed in Dakar, Figs. 13 and 14 demonstrate the capability of the models' inferences to detect population variations with respect to land use typologies in Nairobi and Dar es Salaam. To further investigate their value for applications where no local data are available, we additionally use WorldPoP products as a benchmark. In the case of Dar es Salaam, it can be observed that the best performing DL-based models, better differentiate between residential and non residential areas. For instance, the mean population density per hectare in

the DA Sentinel-2 model for the ACI class is 54.95 while the value for planned residential areas is 131.81. The WorldPop product showcases a diminished ability to detect these intrinsic differences (mean density for ACI = 106.66, planned residential = 138.95). Moreover, all 3 products appear to correctly assign increased density to the highly dense deprived residential areas. Notably, in Nairobi, the WorldPop product exhibits a higher mean population density per hectare for non-residential areas than residential ones. The DL models appear to better discriminate between the two types of land use, although marginally.

5. Discussion

One of the bottlenecks in urban population mapping is the dependency on reliable, local census data when performing top-down disaggregation. In the case of bottom-up approaches, adequate numbers of survey (micro-census) information are necessary to develop geostatistical models which is a tedious and costly process that is rarely undertaken. While the literature has assessed the strengths and limitations of both approaches (Leyk et al., 2019; Wardrop et al., 2018), little effort has been made to assess the potential contribution of alternative approaches. Our proposed framework suggests that DL and EO data exhibit a tremendous potential in such direction. By utilizing census data from a single city of a SSA country, we were able to predict population in a satisfactory manner in cities of other SSA countries alleviating the need of local census or survey data. The success of this approach has important implications regarding the next stage of global population maps and towards the ideal scenario of end-to-end, near-real-time urban population mapping using EO information. For instance, in rapidly urbanizing countries with outdated census data or cases of rapid migration/displacement, conventional approaches are likely to fail, while DL-EO methods can be highly promising.
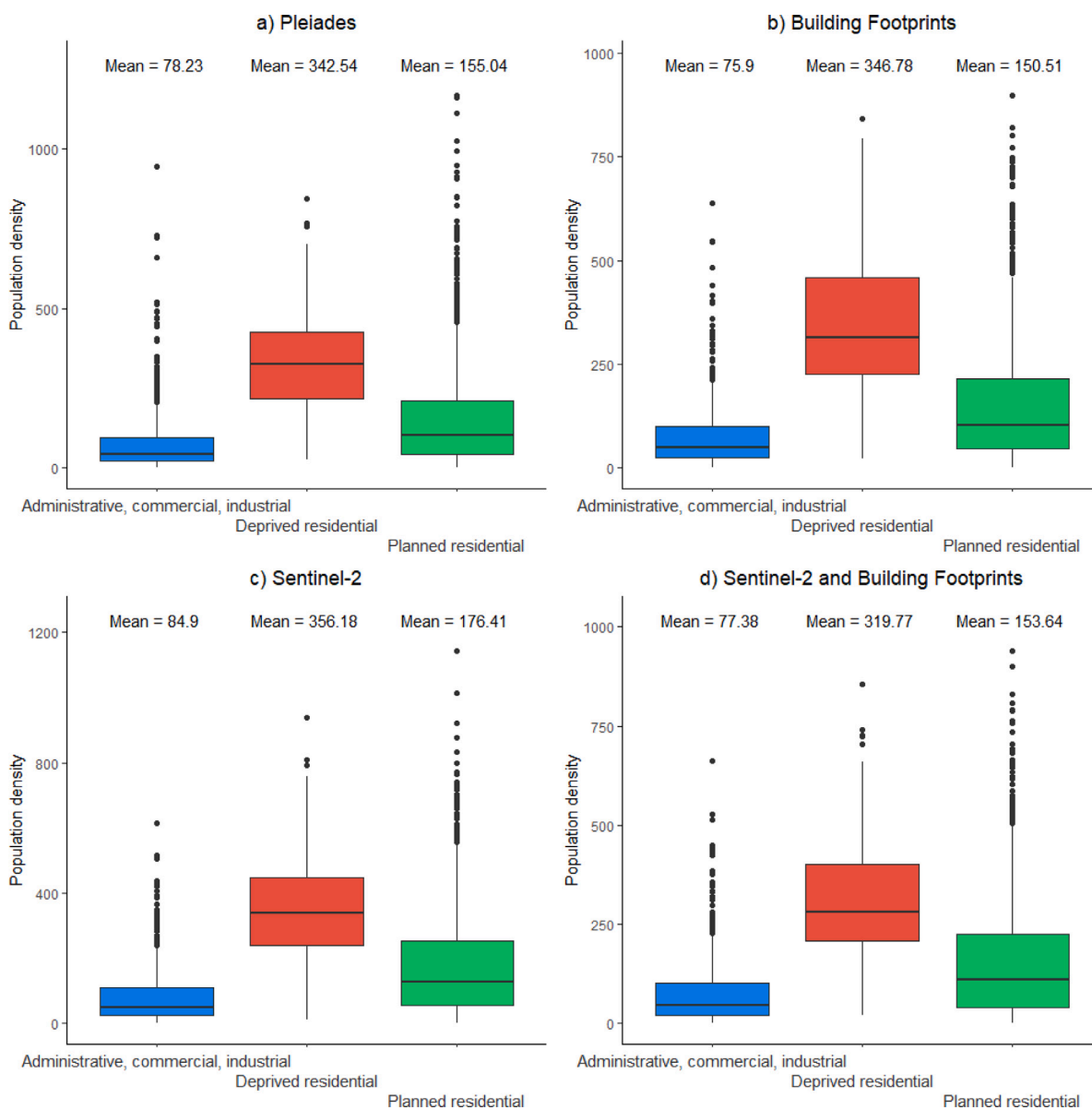
Fig. 8. Boxplots of population density of different built-up land use types in Dakar, derived from DL-based models using different inputs, (a) Pleiades, (b) building footprints, (c) Sentinel-2 and (d) Sentinel-2 and building footprints.

Table 3
Accuracy metrics at the census level in Nairobi, Kenya and Dar es Salaam, Tanzania.

		Building footprints	Sentinel-2	Sentinel-2 + Building footprints	DA Sentinel-2 + Building footprints
Nairobi	RMSE	31,046.5	51,579.5	46,089.4	35,117.3
	MAE	20,036.4	35,488.2	32,503.4	23,562.7
	R^2	0.40	0.01	0.61	0.38
	Pop ratio	0.79	0.12	0.16	0.57
Dar es Salaam	RMSE	24 671.4	43 270.1	35 147.5	19 655.9
	MAE	16 116.6	35 693.7	29 595.4	15 667.1
	R^2	0.47	0.09	0.46	0.60
	Pop ratio	1.10	0.29	0.42	0.84

5.1. When local data are available

As demonstrated in the case of Dakar, when local census information is available at a fine spatial scale, EO data with DL models can very successfully capture the population patterns with particularly low error rates and were able to discriminate between the various land-use cases (i.e., residential, industrial and deprived neighborhoods). Although in the presence of rigorous census data top-down disaggregation methods are typically used, they still require high-resolution ancillary data to drive such processes (i.e., land cover, land use) which can be a nuisance, especially when temporal aspects are considered (Tu
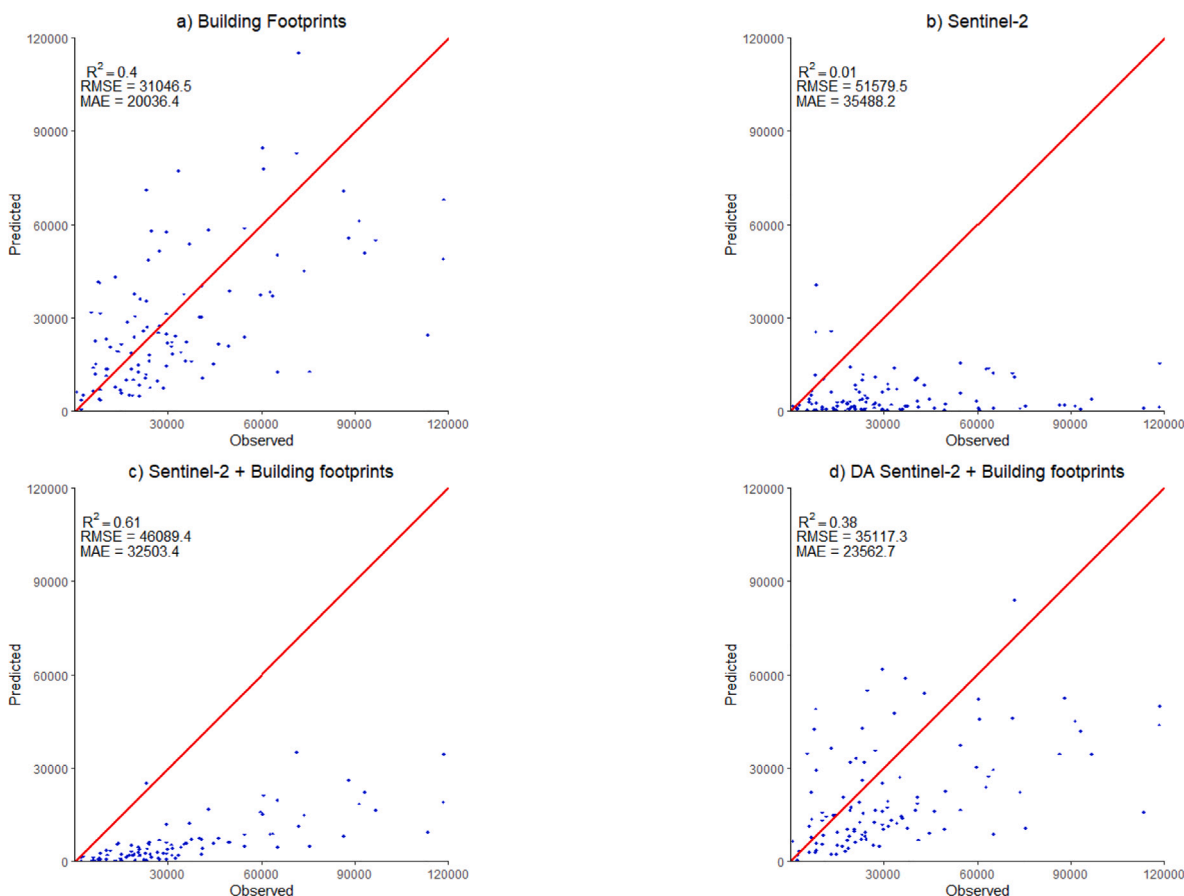
Fig. 9. DL model predictions aggregated at the census level in Nairobi using different types of image inputs and architectures. (a) Building footprints, (b) Sentinel-2, (c) Sentinel-2 and building footprints and (d) domain adaptation. In (d) the predictions are made using Sentinel-2 imagery (the building footprints are used only during the training stage).
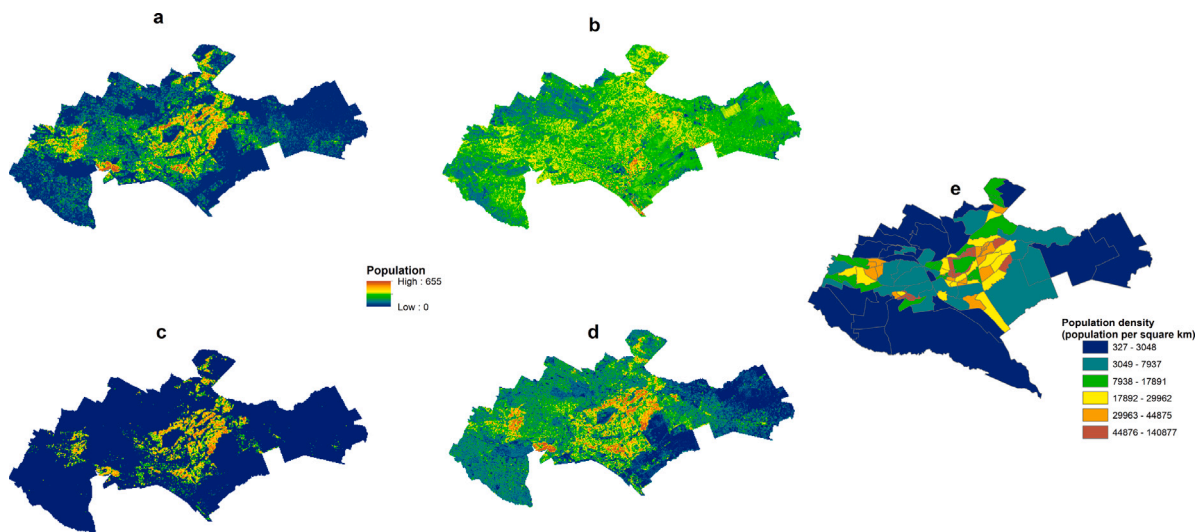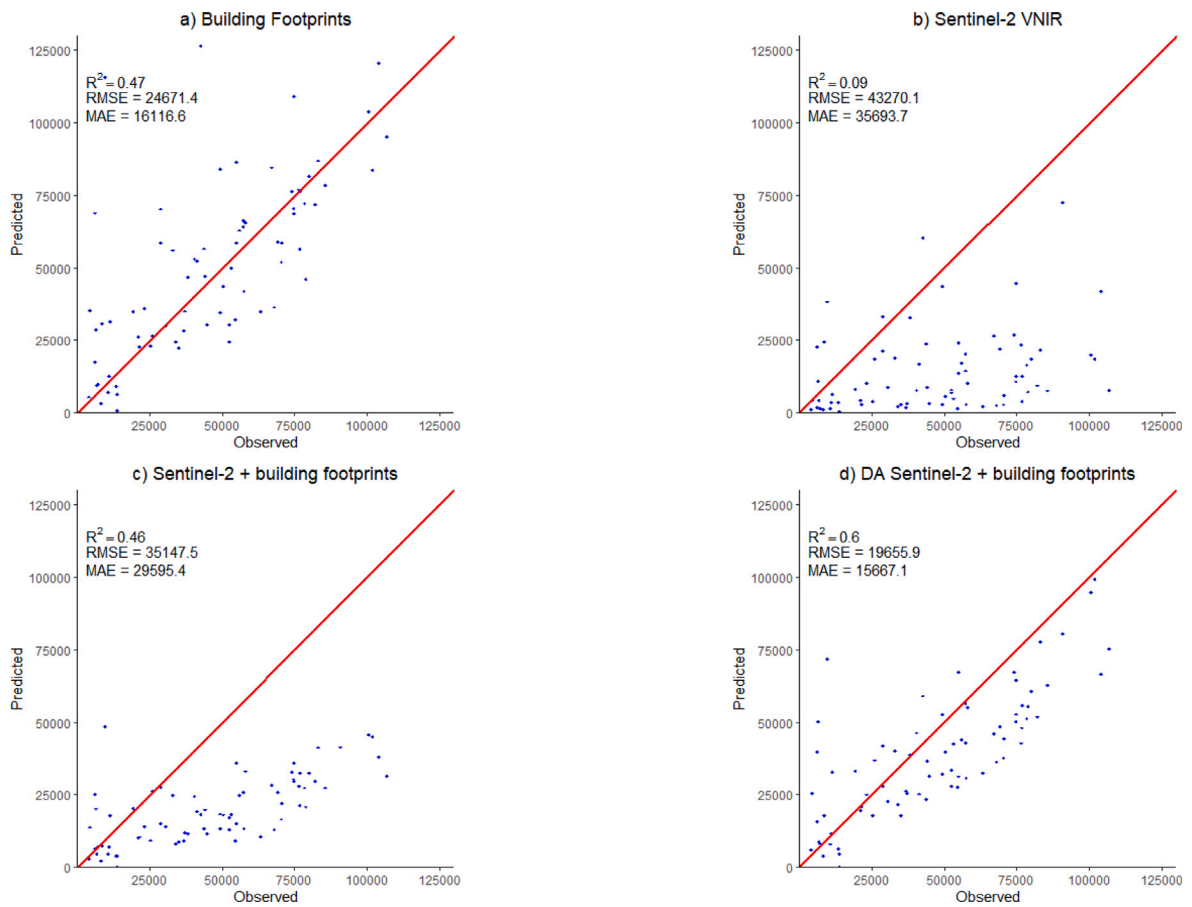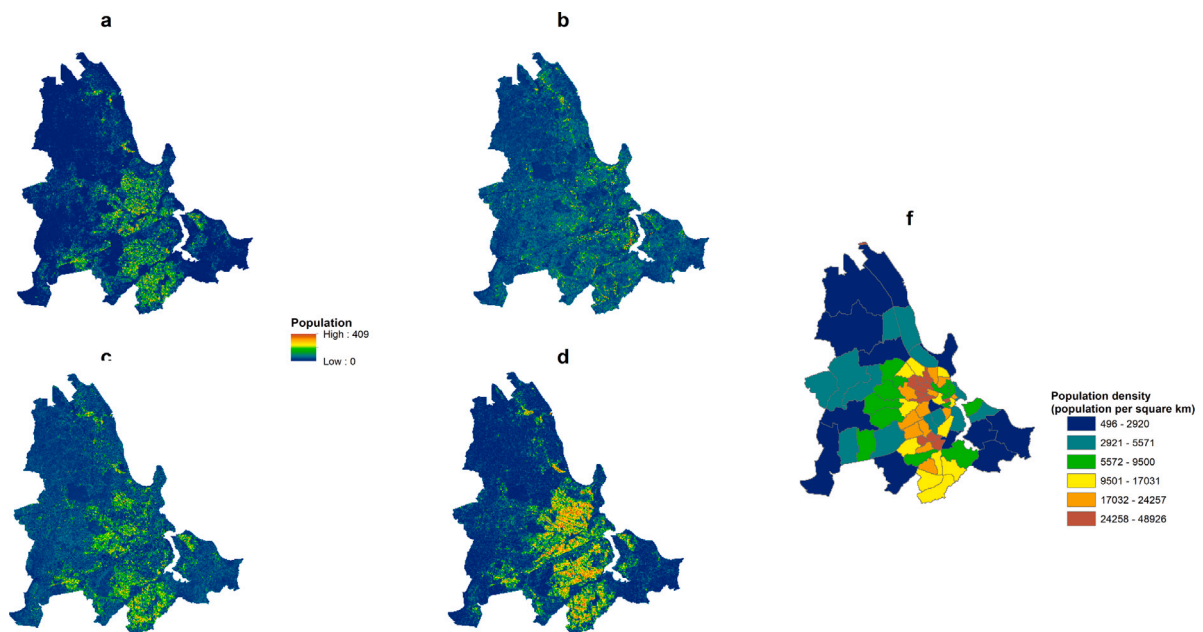

Fig. 10. Predicted population distribution maps in Nairobi, Kenya using different inputs and architectures. (a) Building footprints, (b) Sentinel-2, (c) Sentinel-2 and building footprints, (d) domain adaptation and (e) population density at the census level. In (d) the predictions are made using Sentinel-2 imagery (the building footprints are used only during the training stage).

et al., 2022). End-to-end approaches such as ours can easily be used to project population evolution across decadal periods, alleviating the need for a new census to be conducted as it only relies on EO data and available geospatial information such as building footprints, which are becoming globally available at a rapid pace since the past few years. In the case of limited survey (micro-census) data, rather than complete census data being available, the proposed method can be of benefit as it can make use of unlabeled image data to increase the quality of the inferences. Census-independent DL approaches using micro-census data were recently deployed with success (Neal et al., 2021) further fortifying our position.

Fig. 11. DL model predictions aggregated at the census level in Dar es Salaam using different types of image inputs and architectures. (a) Building footprints, (b) Sentinel-2, (c) Sentinel-2 and building footprints and (d) domain adaptation. In (d) the predictions are made using Sentinel-2 imagery (the building footprints are used only during the training stage).

Fig. 12. Predicted population distribution maps in Dar es Salaam, Tanzania using different inputs and architectures. (a) Building footprints, (b) Sentinel-2, (c) Sentinel-2 and building footprints, (d) domain adaptation and (e) population density at the census level. In (d) the predictions are made using Sentinel-2 imagery (the building footprints are used only during the training stage).
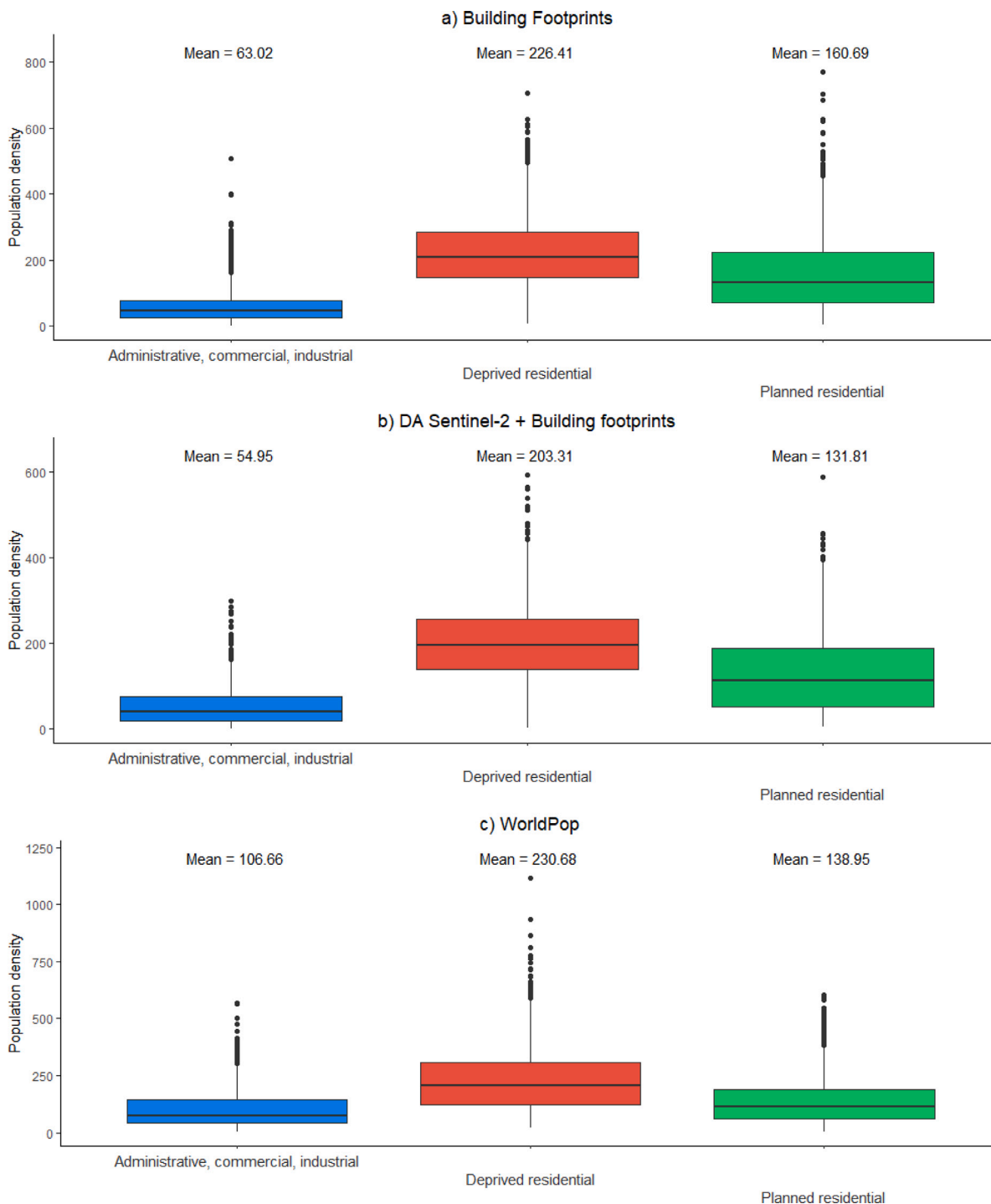
Fig. 13. Boxplots of population density of different built-up land use types in Dar es Salaam, derived from the best performing DL-based models and WorldPoP, (a) Building footprints, (b) DA Sentinel-2 and building footprints and (c) WorldPoP.

5.2. In the face of data scarcity

Without a doubt, the most impactful aspect of this work is in urban areas where neither census nor micro-census information is available. In an urban context, the underlying rational is that although variable across countries, the relationship between image features and population density is relatively stable and can be retrieved with the help of (a) unlabeled image data and (b) guidance from ancillary information such as building footprints. Both types of data investigated in this work are openly available as we made use of Sentinel-2 imagery and Google's building footprints data for the transferability experiments. Equally important, the proposed method uses Sentinel-2 data for the prediction while utilizing the building footprints only during the training stage. This provides solutions in the case where building footprints are also not available, provided the network has been adequately trained in similar areas before. Another important aspect is that the proposed framework can be used when census data are available but unreliable for a particular residential type. This is a common issue in deprived urban areas (slums), where official records can be vastly underestimated (Abascal et al., 2022b; Kuffer et al., 2020; Thomson et al., 2021). Nonetheless, we have to stress the fact that the proposed framework
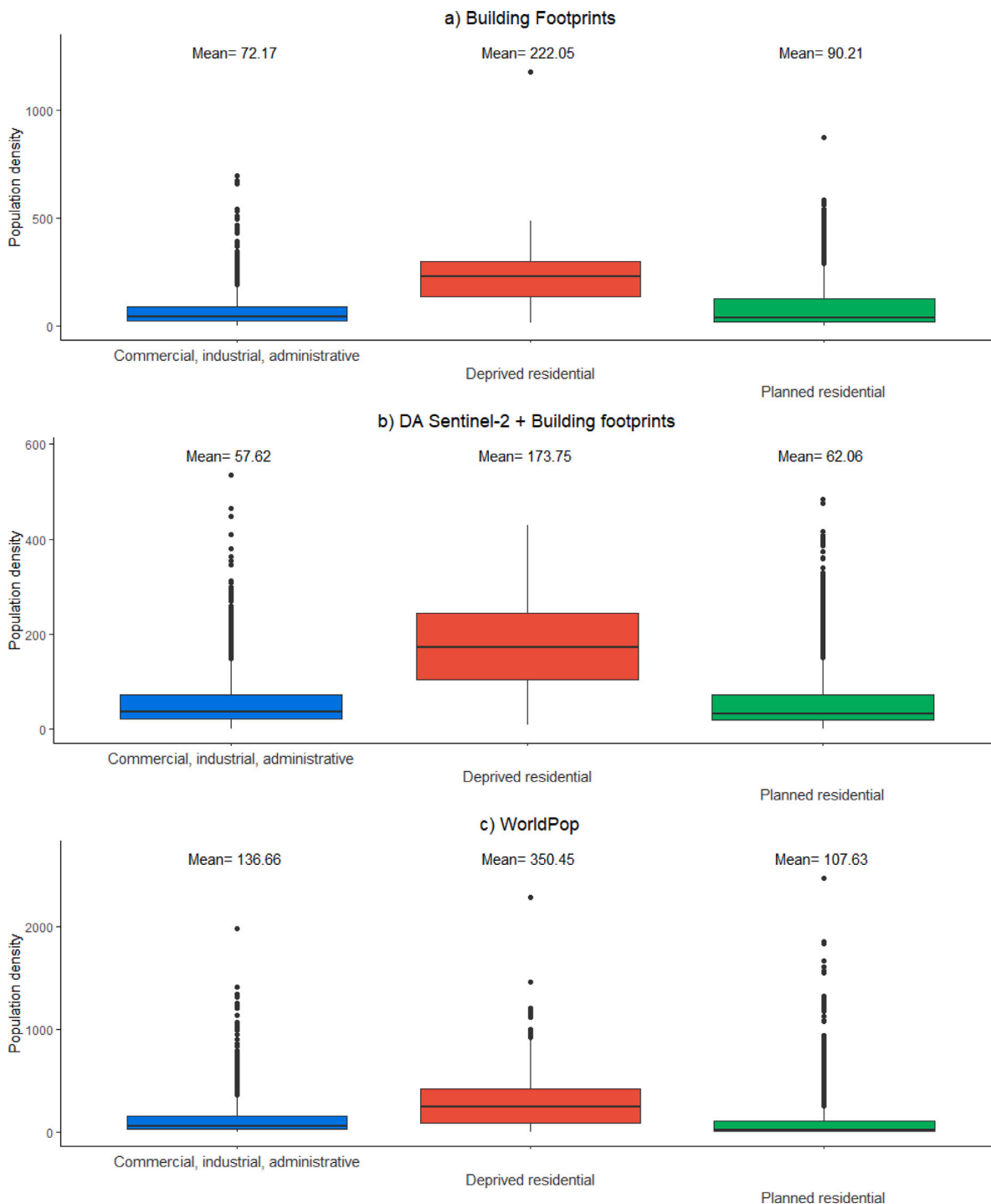
Fig. 14. Boxplots of population density of different built-up land use types in Nairobi, derived from the best performing DL-based models and WorldPoP, (a) Building footprints, (b) DA Sentinel-2 and building footprints and (c) WorldPoP.

is only but a first step in evaluating the potential of DL for high-quality population inferences. The error margins when applying the DL-EO models in areas without local training data are not marginal and can suffer from significant over- or underestimations, and should be interpreted with caution and according to the specifications of each application case. They, however, provide encouraging insights and strong evidence to pursue large-scale training (i.e., on several high-quality census/survey data) to create domain invariant population models at national/continental levels. Moreover, the best performing DL-models in Nairobi and Dar es Salaam (building footprint and DA- based) were able to better differentiate population density between the different

land-uses than WorldPop products without using local data at all. While their overall accuracy can range from moderate to satisfactory it implies that satellite features processed through DL algorithms can retrieve intrinsic properties regarding the urban form, that might be challenging to assemble otherwise. Nonetheless, DL-based models should not aim to directly replace conventional geostatistical approaches, as they are challenging to interpret, data-hungry and computationally expensive algorithms but rather compliment them. Particularly for small-case studies, it might be more parsimonious and efficient to undertake standard approaches such as disaggregation from building data or land cover maps.

5.3. Improvements in the use of EO-derived data

In future work, additional experiments can be made with respect to the input data. For instance, using multitemporal Sentinel-2 images could improve the performance by reducing seasonal effects (i.e., dry and wet season). Moreover, the potential of the cloud penetrating radar Sentinel-1 information can be of merit, particularly in tropical regions where the acquisition of cloud-free optical imagery is challenging. Notably, the availability of building footprint data at large scales is now more accessible than ever. Along with Google's dataset that was used in this study, Ecopia (Ecopia and Technologies, 2020), Microsoft (Microsoft, 2019) have provided building footprints at national or almost-continental levels in Africa with unprecedented quality and should be considered for upscaling our recommendations. Last but not least, upcoming datasets that provide three-dimensional information on building morphology (i.e., build height, volume) should be investigated as means to further improve the modeling process (Esch et al., 2022), as they can reflect key aspects of population density.

5.4. Towards the next steps in urban population mapping

Our results indicate that the next generation of urban population products should harness the full potential of deep learning and publicly available EO data, particularly in situation of dire data scarcity. Available census datasets of fine spatial scale should be exploited as they can provide the means towards generalizable population models, transcending the needs for local data availability. Domain adaptation methods are rapidly evolving but demonstrate a key element to such attempts (Tuia et al., 2016) and should be further investigated. The proposed framework should be expanded and evaluated in cities of different geographical environments, configurations and morphologies.

6. Conclusions

We present an end to end deep learning-based framework to model and map population distribution patterns in three Sub-Saharan African cities, namely Nairobi, Dar es Salaam and Dakar. We conducted experiments using very-high-resolution (Pleiades) imagery and moderate resolution Sentinel-2 imagery, along with publically available building footprint data. Our results demonstrated that population counts can be retrieved with satisfactory accuracy in situations of abundant training data and with moderate success in cases where no local data where available at all. We dealt with the issue of data scarcity by deploying a consistency-based domain adaptation approach that uses the building footprints as anchor to formulate the network weights for the satellite image part of the network. The results pave the way for consistent and accurate solutions that overcome traditional bottlenecks on the field such as the reliance in good quality land-cover and land-use data, as we demonstrated that the Earth observation-based models can understand differences in land-use such as between informal and formal settlements, as well as non-residential neighborhoods.

CRediT authorship contribution statement

Stefanos Georganos: Data curation, Conceptualization, Methodology, Writing – original draft. **Sebastian Hafner:** Methodology, Analysis, Re-reviewing & editing. **Monika Kuffer:** Internal review, Re-reviewing & editing. **Catherine Linard:** Internal review, Re-reviewing & editing. **Yifang Ban:** Supervision, Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abascal, A., Rodríguez-Carreño, I., Vanhuysse, S., Georganos, S., Sliuzas, R., Wolff, E., Kuffer, M., 2022a. Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. Comput. Environ. Urban Syst. 95, 101820.

Abascal, A., Rothwell, N., Shonowo, A., Thomson, D.R., Elias, P., Elsey, H., Yeboah, G., Kuffer, M., 2022b. "Domains of deprivation framework" for mapping slums, informal settlements, and other deprived areas in LMICs to improve urban planning and policy: A scoping review. Comput. Environ. Urban Syst. 93, 101770.

ANSD (Agence Nationale de la Statistique et de la Démographie), 2013. Rapport définitif RGPHAE-2013, 418. Available from http://www.ansd.sn/ressources/RGPHAE-2013/resultats-definitifs.htm, Dakar, Senegal.

Bachman, P., Alsharif, O., Precup, D., 2014. Learning with pseudo-ensembles. arXiv preprint arXiv:1412.4864.

Chen, R., Yan, H., Liu, F., Du, W., Yang, Y., 2020. Multiple global population datasets: Differences and spatial distribution characteristics. ISPRS Int. J. Geo-Inf. 9 (11), 637.

CIESIN, C., 2005. Gridded population of the world (GPW), v4. Available at sedac. ciesin. Columbia. edu/data/collection/gpw-v4. Accessed May 5, 2016.

Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., Ye, C., 2019. Semi-supervised brain lesion segmentation with an adapted mean teacher model. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 554–565.

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. Photogramm. Eng. Remote Sens. 66 (7), 849–857.

Ecopia, A.I., Technologies, M., 2020. Building Footprints. Ecopia.AI & Maxar Technologies.

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 Sens. Environ. 270, 112877.

Freire, S., Halkia, M., 2014. GHSL application in europe: Towards new population grids. In: European Forum for Geography and Statistics, Krakow, Poland.

French, G., Mackiewicz, M., Fisher, M., 2017. Self-ensembling for visual domain adaptation. arXiv preprint arXiv:1706.05208.

Georganos, S., 2020. Malaria in high-resolution: Modelling and mapping plasmodium falciparum parasite rate using very- high-resolution satellite derived indicators in sub-saharan African cities. http://dx.doi.org/10.5281/zenodo.3871497.

Georganos, S., Abascal, A., Kuffer, M., Wang, J., Owusu, M., Wolff, E., Vanhuysse, S., 2021a. Is it all the same? Mapping and characterizing deprived urban areas using worldview-3 superspectral imagery. a case study in Nairobi, Kenya. Remote Sens. 13 (24), 4986.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021b. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 36 (2), 121–136.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Grippa, T., 2019. Very-high resolution earth observation data and open-source solutions for mapping urban areas in sub-saharan Africa.: Implementation of an operational framework for production of geoinformation. Application on ouagadougou (burkina faso) and dakar (Senegal).

Grippa, T., Georganos, S., 2018. Dakar land use map at street block level.

Grippa, T., Georganos, S., Vanhuysse, S., Lennert, M., Mboga, N., Wolff, É., 2019a. Mapping slums and model population density using earth observation data and open source solutions. In: 2019 Joint Urban Remote Sensing Event (JURSE). IEEE, pp. 1–4.

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 Int. J. Geo-Inf. 7 (7), 246.

Grippa, T., Linard, C., Lennert, M., Georganos, S., Mboga, N., Vanhuysse, S., Gadiaga, A., Wolff, E., 2019b. Improving urban population distribution models with very-high resolution satellite information. Data 4 (1), 13.

Hafner, S., Ban, Y., Nascetti, A., 2022a. Unsupervised domain adaptation for global urban extraction using sentinel-1 SAR and sentinel-2 MSI data. Remote Sens. Environ. 280, 113192.

Hafner, S., Nascetti, A., Azizpour, H., Ban, Y., 2022b. Sentinel-1 and sentinel-2 data fusion for urban change detection using a dual stream U-net. IEEE Geosci. Remote Sens. Lett. 19, 1–5. http://dx.doi.org/10.1109/LGRS.2021.3119856.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hu, W., Patel, J.H., Robert, Z.-A., Novosad, P., Asher, S., Tang, Z., Burke, M., Lobell, D., Ermon, S., 2019. Mapping missing population in rural India: A deep learning approach with satellite imagery. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. pp. 353–359.

Huang, X., Zhu, D., Zhang, F., Liu, T., Li, X., Zou, L., 2021. Sensing population distribution from satellite imagery via deep learning: Model selection, neighboring effects, and systematic biases. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5137–5151.

Kenya National Bureau of Statistics, 2019. Kenya Population and Housing Census. Nairobi, Kenya.

Klemmer, K., Yeboah, G., de Albuquerque, J.P., Jarvis, S.A., 2020. Population mapping in informal settlements with high-resolution satellite imagery and equitable ground-truth. arXiv preprint arXiv:2009.08410.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.

Kuffer, M., Pfeffer, K., Sliuzas, R., 2016. Slums from space—15 years of slum mapping using remote sensing. Remote Sens. 8 (6), 455.

Kuffer, M., Thomson, D.R., Boo, G., Mahabir, R., Grippa, T., Vanhuysse, S., Engstrom, R., Ndugwa, R., Makau, J., Darin, E., et al., 2020. The role of earth observation in an integrated deprived area mapping "system" for low-to-middle income countries. Remote Sens. 12 (6), 982.

Laine, S., Aila, T., 2016. Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242.

Leyk, S., Gaughan, A.E., Adamo, S.B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F.R., Blankespoor, B., Frye, C., et al., 2019. The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. Earth Syst. Sci. Data 11 (3), 1385–1409.

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.

Loshchilov, I., Hutter, F., 2018. Fixing weight decay regularization in adam.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.

Microsoft, 2019. Tanzania - uganda building footprints. Microsoft.

National Bureau Statistics Office (NBS), 2012. Population and housing census: Population distribution by administrative areas. Dar es Salaam, United Republic of Tanzania.

Neal, I., Seth, S., Watmough, G., Diallo, M.S., 2021. Census-independent population estimation using representation learning. arXiv preprint arXiv:2110.02839.

Oliver, A., Odena, A., Raffel, C., Cubuk, E.D., Goodfellow, I.J., 2018. Realistic evaluation of deep semi-supervised learning algorithms. arXiv preprint arXiv:1804.09170.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703.

Robinson, C., Hohman, F., Dilkina, B., 2017. A deep learning approach for population estimation from satellite imagery. In: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities. pp. 47–54.

Sajjadi, M., Javanmardi, M., Tasdizen, T., 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. arXiv preprint arXiv:1606.04586.

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:2107.12283.

Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS One 10 (2), e0107042.

Tatem, A.J., 2017. Worldpop, open data for spatial demography. Sci. Data 4 (1), 1–4.

Thomson, D.R., Gaughan, A.E., Stevens, F.R., Yetman, G., Elias, P., Chen, R., 2021. Evaluating the accuracy of gridded population estimates in slums: A case study in Nigeria and Kenya. Urban Sci. 5 (2), 48.

Thomson, D.R., Kuffer, M., Boo, G., Hati, B., Grippa, T., Elsey, H., Linard, C., Mahabir, R., Kyobutungi, C., Maviti, J., et al., 2020. Need for an integrated deprived area "slum" mapping system (IDEAMAPS) in low-and middle-income countries (LMICs). Soc. Sci. 9 (5), 80.

Tiecke, T.G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E.B., et al., 2017. Mapping the world population one building at a time. arXiv preprint arXiv:1712.05839.

Tu, W., Liu, Z., Du, Y., Yi, J., Liang, F., Wang, N., Qian, J., Huang, S., Wang, H., 2022. An ensemble method to generate high-resolution gridded population data for China from digital footprint and ancillary geospatial data. Int. J. Appl. Earth Obs. Geoinf. 107, 102709.

Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. IEEE Geosci. Remote Sens. Mag. 4 (2), 41–57.

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. Proc. Natl. Acad. Sci. 115 (14), 3529–3537.

Williams, S., Marcello, E., Klopp, J.M., 2014. Toward open source Kenya: Creating and sharing a GIS database of nairobi. Ann. Assoc. Am. Geogr. 104 (1), 114–130.

Zhuang, H., Liu, X., Yan, Y., Ou, J., He, J., Wu, C., 2021. Mapping multi-temporal population distribution in China from 1985 to 2010 using landsat images via deep learning. Remote Sens. 13 (17), 3533.