OECD Regional Development Papers No. 28

Monitoring land use in cities using satellite imagery and deep learning

Alexandre Banquet,
Paul Delbouve,
Michiel Daams,
Paolo Veneri

https://dx.doi.org/10.1787/dc8e85d5-en

OECD

OECD Regional Development Papers

Monitoring land use in cities using satellite imagery and deep learning

By: Alexandre Banquet*, Paul Delbouve*, Michiel N. Daams♣ and Paolo Veneri*

Over time, cities expand their physical footprint on land and new cities emerge. The shape of the built environment can affect several domains which are policy relevant, such as carbon emissions, housing affordability, infrastructure costs, and access to services. This study lays a methodological basis for the monitoring and consistent comparison of land use across OECD cities. An advanced form of deep learning, namely the U-Net model, is used to classify land cover and land use in EC-ESA satellite imagery for 2021. This complements conventional statistical data by monitoring large surfaces of land efficiently and in near real-time. In specific, following the availability of detailed data for model training, built-up areas in residential or business-related use are mapped and analysed for 687 European metropolitan areas, as a case application. Recent urban expansion's speed and shape are explored, as well as the potential for assessing land use in cities beyond Europe.

JEL codes: R14, R31
Keywords: Land Use Monitoring – Machine Learning – Satellite Imagery – Economic Indicators – Cities

ABOUT THE OECD

The OECD is a multi-disciplinary inter-governmental organisation of 38 member countries which engages in its work an increasing number of non-members from all regions of the world. The Organisation's core mission today is to help governments work together towards a stronger, cleaner, fairer global economy. Through its network of 250 specialised committees and working groups, the OECD provides a setting where governments compare policy experiences, seek answers to common problems, identify good practice, and co-ordinate domestic and international policies. More information available: www.oecd.org.

ABOUT OECD REGIONAL DEVELOPMENT PAPERS

Papers from the Centre for Entrepreneurship, SMEs, Regions and Cities of the OECD cover a full range of topics including regional statistics and analysis, urban governance and economics, rural governance and economics, and multi-level governance. Depending on the programme of work, the papers can cover specific topics such as regional innovation and networks, sustainable development, the determinants of regional growth or fiscal consolidation at the subnational level. OECD Regional Development Papers are published on http://www.oecd.org/cfe/regional-policy.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Centre for Entrepreneurship, SMEs, Regions and Cities, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This paper is authorised for publication by Lamia Kamal-Chaoui, Director, Centre for Entrepreneurship, SMEs, Regions and Cities, OECD.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

This publication was produced with the financial support of the European Union. Its contents are the sole responsibility of European Commission, Directorate-General for Regional and Urban Policy and do not necessarily reflect the views of the European Union.

Acknowledgements

This project was financially supported by the European Commission, Directorate-General for Regional and Urban Policy. Authors are grateful to Rudiger Ahrend (OECD), Philip McCann (University of Sheffield), as well as Lewis Dijkstra and Joachim Maes at the European Commission for discussions and their insights.

* OECD Centre for Entrepreneurship, SMEs, Regions and Cities

♣ Department of Economic Geography, University of Groningen, Groningen, The Netherlands

Table of contents

Tables

Figures

Boxes

Introduction

The global population increasingly concentrates in cities. To accommodate population growth, new cities emerge from smaller towns and existing cities expand or densify their physical imprint on land. The spatial dimension of these imprints, by built-up areas, may shape economic processes as well as the resilience and sustainability of daily life in cities, as the United Nations 2030 Sustainable Development Goals (UN, 2015) underline (SDG 11 on sustainable cities and communities).

The shape of cities, notably in terms of built-up area, potentially affects many processes taking place in cities, such as mobility, resource consumption, access to services, the cost of infrastructure provision or the ease of social and business interactions. This underlines a need for timely information on how land is used in cities, and how much area is allocated to key uses such as residential or business-related.

A timely monitoring of trends and patterns in such key urban land uses, however, is largely missing at the global level. Relatively well known is how much of global land cover is 'urban' in general, for example from the European Commission's Global Human Settlement Layer's 'built-up grid' (Corbane et al., 2021) or maps by Copernicus (Buchhorn et al., 2020) or ESRI (Karra et al., 2021). Also relatively well known from such existing map products is how the built-up areas of cities have grown every several years or decades. However, little is known about the extent to which urban land is used for residential or business-related purposes. Consequently, the economic sources of urban land use remain unclear – and crucially, any timely statistics from national administrations are either absent or rely on varying land use definitions.

In response, this study develops an approach to monitoring key types of *urban* land use in OECD cities on an internationally consistent basis. As a case application, this paper maps residential as well as business-related built-up areas, as uniquely recent as per 2021, for 687 European cities. To ensure the consistent comparability of results across cities, a definition of metropolitan area boundaries (see Dijkstra, Poelman, and Veneri, 2019) developed jointly by the OECD and the European Commission is adopted.

This exercise requires a mapping effort that covers a land surface of 1 million square kilometres. This scale of analysis is achieved efficiently by using recent advancements in space technology and machine learning for image processing. Satellite imagery, which unlike conventional statistical datasets are available on a near real-time basis, are sourced from EC-ESA Sentinel satellites and then analysed using deep learning.

A deep learning model, in this work the well-established U-Net, can be 'trained' to assign areas of land cover to distinct land use types based on information on optical reflectance (Sentinel-2 imagery) and radar pulses (Sentinel-1 imagery). Such information signals what materials cover a land surface. The U-Net model learns to independently reproduce pre-existent, in part manually generated, maps of urban land use, to then track a set of targeted land uses in more recent satellite imagery. In the images, the deep learning model can accurately categorize coherent patterns of pixels (see, e.g., Sirko et al., 2021). This makes the U-Net particularly useful for tracking land use in satellite images of cities, to the extent that people and firms, as well as zoning governance, allocate distinct economic activities to separate plots of land.

The results for 2021 quantify how the use of land in cities varies substantially. The average amount of built-up area per urban inhabitant, whether in a residential use or in an industrial or commercial use, varies by

up to a fivefold between countries. Large variation stems from how compact or spread-out land use is in the commuting zones that surround cities. Further results explore the speed and shape of urban expansion over 2018-21, and show the model's potential for application to more OECD countries.

Overall, this study's approach complements geospatial monitoring by governmental agencies, as well as emerging private initiatives. The case application's inclusion of small and medium-sized cities is also academically, as most studies focus on the largest of cities (Reba and Seto, 2020). Importantly, the approach developed in this paper supports monitoring early signals of economic growth in cities.

1 Conceptual Choices in Classifying Urban Land from Satellite Imagery

At first glance, patterns in urban land use are easily observed from a satellite image. However, using satellite imagery to infer such patterns both accurately and consistently requires a study design that integrates several key conceptual choices.

1.1. Land Cover or Land Use: 'What Categories are Targeted?'

A long-acknowledged but important first choice regards whether the study's objective is to observe 'land cover' or 'land use' (see, e.g., Anderson et al., 1976). Whereas land cover refers to the physical matter which covers the land surface, for instance vegetation or artificial structures, land use instead refers to human activities which put land to a particular use. In studies of land cover, urban land is commonly considered as a single class, separately from land cover classes that are of a natural or agricultural character (for example, the Copernicus Global Land Cover product as described in Buchhorn et al., 2020). While such distinct land cover classes may largely cohere with general types of land use, there are subtle differences and interrelations that matter.

A key difference is that urban land cover can be associated with various uses of land – with residential use being the most widespread, besides commercial, recreational, or infrastructural uses. In order to distinguish between such types of urban land use, however, and to do so 'from above', the targeted land use types should be associated with distinct 'spectral signatures'. In other words, can land use types be clearly separated from each other in remotely sensed imagery, based on the reflectance and dimensions of associated materials such as buildings and other structures? This is plausible, in particular due to the spatial separation of distinct land uses in OECD countries which results from zoning policies (OECD, 2017). More complex patterns of land use, however, such as mixed-use, or the provision of housing in transformed office buildings, may be observed less effectively than singular (and more prevalent) uses.

With this in mind, this project focuses on classifying key forms of land use within urban areas, which are dominantly present on particular plots of land. These key uses of urban land will be separated from each other as well as from agricultural land use and natural land cover. Observing land use as well as land cover within a single classification scheme is appropriate, as long as these can be separated well. However, care should be taken as land 'use' and 'cover' are not necessarily mutually exclusive. This involves a subtle interrelation between land use and land cover that brings up the next conceptual issue.

1.2. Spatial Unit of Classification: 'What Makes for a Coherent Patch of Land Use?'

The spatial unit of classification concerns whether a classification scheme adopts a broad or a narrow definition of what land cover makes part of a particular land use. For instance, does vegetated land cover classify as residential land use when it is part of a garden, but not when it is located in an isolated area?

In operational terms, the question here is whether land uses will be classified based on information at the level of pixels or at the level of groups of pixels that together resemble 'objects' (e.g., a residential plot of land), which may bundle various land cover types. In the present study, the object-based approach is used, as it aligns well with the aim to distinguish urban land uses that are dominantly and coherently present on particular plots of land. This choice also has the benefit of mitigating possible statistically unstable assignment of land use classes to pixels that are isolated amongst a patch of another use or inhibit idiosyncratic reflectance values.

These considerations apply specifically to land use classification, which may bundle distinct but coherent land cover classes, rather than the classification of individual land cover classes (e.g., forested land), which may simply be adjacent to each other to form a coherent patch. In the case of individual land cover classes, the spatial unit of classification relates mostly to the resolution at which land cover is captured. Resolution, however, plays a broader role, also in the classification of land use.

1.3. Imagery Resolution: 'How Granular is Granular Enough?'

What resolution of satellite imagery is appropriate to use follows from the study's classification objectives. While it may be intuitive to assume that a higher, more granular, resolution is always preferable, this is not necessarily always the case. This could apply, for example, when classifying images into coherent patches of particular land uses, as per an 'object-based' approach. The observation of the smallest elements of built-up areas in highly-detailed imagery might direct a model away from the coherence of elements that would signal a commercial or residential use. Conclusive theoretical guidance, however, is mostly absent.[1]

Nevertheless, given that urban areas may bundle highly heterogeneous built features, the observed imagery should have a resolution below the dimensions of those features. This then allows the targeted land uses to be distinguished from each other. A higher resolution, however, also implies a more heterogeneous spatial separation of distinct forms of land cover (e.g., an open space enclosed by an apartment complex). This means that the choice of resolution and spatial unit of assessment (individual pixels or bundles of pixels as coherent urban 'objects' or sites) are interrelated.

Box 1. Illustration of Sentinel Satellite Imagery's Ability to Capture Urban Patterns

The present study observes publicly available imagery in which urban land use and land cover types can be distinguished effectively, at a high resolution and potentially in any OECD city. This innovative imagery is sourced from Copernicus, the Earth Observation programme that is jointly coordinated and managed by the European Commission and the European Space Agency. Detailed information regarding the imagery data will be provided in Section 3.3. However, a start can be made to illustrate the imagery's ability to distinguish between various urban land uses in Figure 1.

[1] A practical consideration is that imagery at very high (sub-meter) resolutions may come not only at licensing costs but, due to the size of such data, also at considerable computational (time) costs, in particular for large-scale studies.

Figure 1. Information captured by high-resolution imagery for the city of Luxembourg.

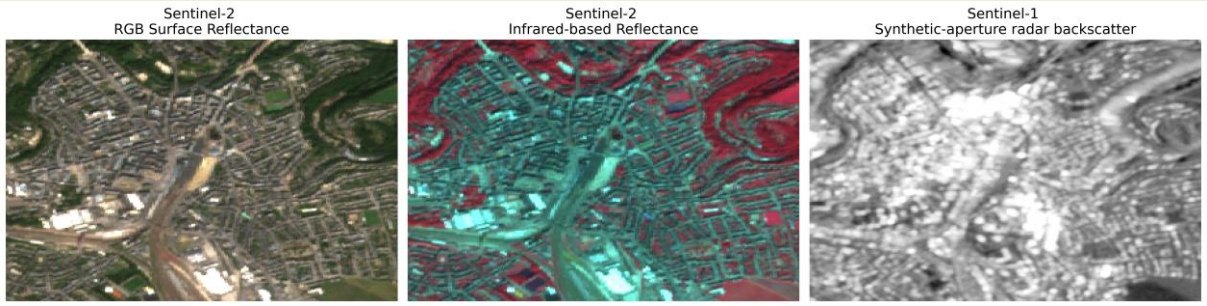


Figure 1, Panel A, shows the city of Luxembourg in the high-resolution satellite imagery which captures urban area at the scale of 10m x 10m pixels. The colour for each pixel here reflects composite of reds, greens, and blues as also the human eye could observe. However, the imagery also captures, as Section 3.3 will further clarify, information about near-infrared values (Panel B) as well as radar backscatter (Panel C), which are also both measured at a 10-meter resolution. This combination of different sorts of remotely sensed information offers a rich basis for separating various classes of urban land cover and land use as this study aims to do. Residential areas can be distinguished from industrial sites, or from urban green spaces or agricultural land. However, the subtle complexity of urban areas, as Figure 1 also illustrates, underlines the need for advanced models to adequately categorize, or 'segment', image pixels into the correct classes.

2 Methodology

2.1. Satellite Image Segmentation through Deep Learning

To track urban land use in satellite imagery, individual pixels in an observed image need to be assigned to a corresponding land use or land cover category. This process is referred to as semantic image segmentation. Semantic image segmentation is a common technique used in computer vision, which is a field of work on artificial intelligence that extracts high-level information from images.[2] A general example of the method of segmenting, or partitioning, an image's pixels into specific classes, as widely applied across different industrial and academic activities, is given in Figure 2.

Figure 2. General example of image segmentation into object-categories in a deep learning context.

Note: Figure adapted from Jeong, Yoon, and Park (2018).

This study applies deep learning to achieve image segmentation efficiently and at scale on satellite imagery of metropolitan areas in OECD countries. Deep learning is a sub-field of machine learning, which consists of learning complex representations with different levels of abstraction from large amounts of data based on computational models called neural networks.[3] These methods have in recent years dramatically improved the state-of-the-art in speech recognition, object recognition, and many other domains such as drug discovery, genomics, or autonomous driving.

One of the most state-of-the art models, namely U-Net, was introduced in 2015 in a biomedical context (Olaf Ronneberger, 2015), and has since been applied to various other problems, including the segmentation of satellite imagery (e.g., Sirko et al., 2021). U-Net has been applied to the segmentation of

[2] Common methods in computer vision other than image segmentation are instance segmentation, which flags objects in images such that these can be counted as individual instances, and image classification, which assigns a class to an image as a whole given its content. In this paper 'image segmentation' and 'image classification' may be used interchangeably, although it is consistently the same process of image segmentation which is referred to.

[3] A neural network is composed of different layers of artificial neurons, connected to one another. The first model was proposed by (Rosenblatt, 1958). Since, different neural network architectures have been developed: first feed-forward neural networks in which information only moves in one direction; as well as convolutional neural networks (CNN) (LeCun et al., 1990) which are better able to capture spatial and temporal dependencies and thus to process images, videos, speech and audio; and recurrent neural networks (RNN) for sequential data (LeCun, Bengio, and Clinton, 2015).

distinct land uses, be it limited and not always in an urban context (Solórzano et al., 2021; Giang et al., 2020; Zhang, Liu, and Wang, 2017).[4]

The U-Net model's architecture is presented in Figure 3. The architecture consists of two paths following a symmetrical U-shape. The first path consists of a typical convolutional network architecture and captures context, while the second path enables precise localization of patterns.[5]

Figure 3. Model structure of the U-Network for image segmentation (Ronneberger, 2015).

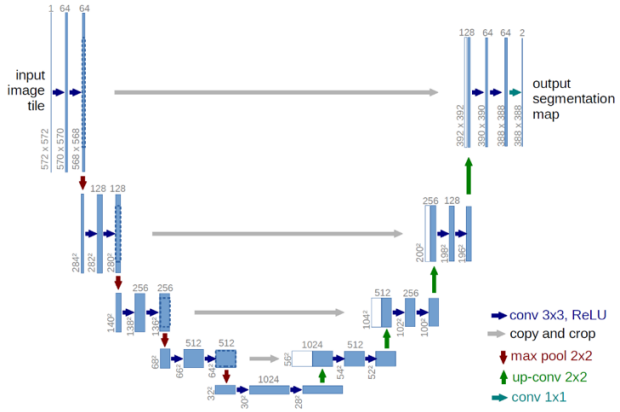


Figure 4 shows the pipeline to segment satellite images. For each pixel, the same set of image bands is observed. The model returns as output a probability tensor. This tensor consists of a 3-dimensional array that for each pixel gives the probability that it belongs to a specific class. Finally, the output classification mask is obtained by taking the class with the highest probability observed for a pixel across the distinct class-layers in the previous step. The model is trained in a supervised manner, meaning that input image patches, of size 160x160x7, are paired with ground truth land use masks (of size 160x160x1) to learn from.

Before training the model, for every city, the set of image patches that cover the metropolitan area is split into train, validation, and test sets using split ratios of 60%, 20%, and 20%, respectively.[6] These allow the

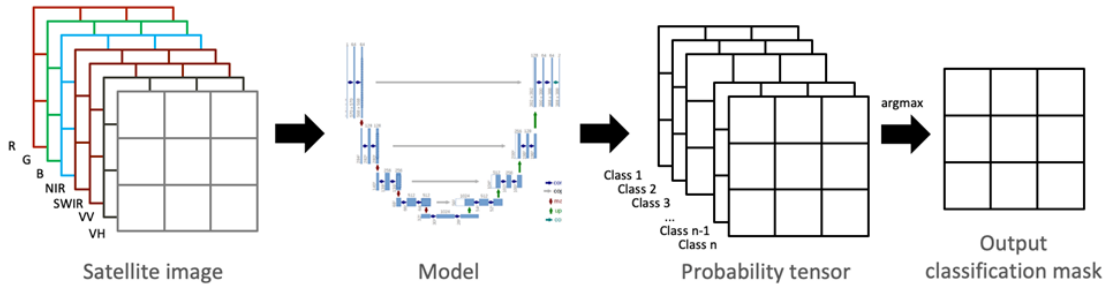
[4] Other variants of U-Net have been used for image segmentation, such as (Iglovikov and Shvets, 2018) which combines U-Net with a VGG11 encoder. Other deep learning models for semantic image segmentation have been developed and applied to remote sensing data, such as Fully Convolutional Networks (FCNs) (Long, Shellhamer, and Darrell 2014), as an extension of image classification models AlexNet (Krizhevsky, Sutskever, and Hinton, 2012), VGGNet (Simonyan and Zisserman, 2015), and GoogLeNet (Szegedy et al., 2014) for image semantic segmentation; or more recently SegNet (Badrinarayanan, Kendall, and Cipolla, 2017), DeepLab (Chen et al., 2018), and DenseNet (Gao Huang et al., 2016).

[5] The contraction path follows the classical pattern of a convolutional network: 3x3 convolutions, followed by a ReLU function and a 2x2 max pooling operation with stride of 2 for down-sampling. At each step of down-sampling, the number of feature channels is doubled. Every step in the expansion path consists of an up-sampling of the feature map followed by a 2x2 up-convolution, which halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU function. At the final layer, a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers. After each layer, a dropout of 0.25 is applied.

[6] Unlike many remote sensing studies which implement sample-based designs to obtain training and evaluation data, ground truth data are observed for nearly each of the observed pixels (see Section 4.2). This implicitly follows the good practice of using a probabilistic design for the inclusion of particular pixels in the training, test, and evaluation sets, as otherwise sample stratification by class would achieve. This mitigates possible bias in the mapped areas as the targeted land use classes are represented in the estimation data largely in line with their actual spatial prevalence.

model to learn from the training set what imagery-values to associate with what class, to then tune its predictive ability on a validation set. The validation set is used at the end of each epoch during the training to assess the model performance and avoid overfitting. An early stopping criterion tracks at the end of each epoch if the loss (sparse categorical cross-entropy) on the validation set reaches a minimum. Finally, accuracy is evaluated on an independent test set.

Figure 4. Empirical pipeline overview.



As such, accuracy evaluation comes from estimates on data values that are out-of-sample: the model has not yet seen these data during the training and tuning process.[7] In order to increase the transparency of deep learning model generation, and to allow for replication, it is recorded which image patches are allocated to the train, test, and validation sets.

This project is implemented in Python using the TensorFlow Deep Learning framework. The U-Net implementation was inspired by the repository of Kumar (2018). A proof-of-concept was generated on a subset of cities in Belgium using a server equipped with a T4 GPU, 2 vCPU, 25 GB of RAM. The analysis was then scaled to cover 687 cities across European OECD countries using Microsoft Azure compute clusters equipped with a T4 GPU, 16 vCPU, and 110 Gb of RAM.

2.2. Evaluation of Model Performance in Classifying Urban Land

Model performance should be quantified at the spatial scale at which land use or land cover is analysed, such that the accuracy of resultant insights can be understood. Both the scope of the analysis and its policy context may guide whether accuracy should be reported for the entire study area, its sub-regions, or individual cities. This issue will be discussed in more detail in section 4.2 before providing estimates of land use areas at national and sub-national scales. Next to be considered here are the operational details of this study's approach to model evaluation.

In specific, the performance of the image segmentation classifier, the U-Net, will be evaluated using standard metrics. Each of these metrics can be computed from an error matrix, which is also commonly referred to as a confusion matrix. A confusion matrix compares for each pixel the predicted land use class, in the test set of image patches (see Section 2.1), with the reference class. This means that the matrix gives count values of the number of pixels that are correctly assigned to a particular land use class, or incorrectly attributed to another land use class. Importantly, the assumption here is that the reference classification is of superior quality, as compared to the map classification that results from the model's prediction, and so offers a ground truth to benchmark against. The ground truth data observed in this study will be described in Section 3.2, and metrics of model performance will be introduced and discussed along with the results in Section 4.1.

[7] A common concern is spatial autocorrelation. However, as the patches are of 1.6 km dimensions, cross-patch correlations of land use in the training and test and validation sets are mitigated (as compared to common approaches where *xy*-location based land use samples may be located within a few meters or pixels from each other).

3 Data and Study Area

3.1. Observed OECD Metropolitan Areas

Our study area encloses metropolitan areas across the 27 European OECD countries[8]. Within these countries, the boundaries of metropolitan areas are defined by Functional Urban Areas (FUAs). The FUAs definition is a joint effort by the OECD and the European Commission to define metropolitan areas in a way that is consistent across countries. In particular, the boundaries of FUAs enclose an observed city, or a set of cities in the case of a polycentric metropolitan area, which is densely populated as well as the wider zone from which a substantial share of local residents commutes to the city (i.e., commuting zone). This offers an integrated spatial economic definition of cities and metropolitan areas that is optimized for the international comparisons, including in terms of city size. Across the 27 European OECD countries, 687 FUAs are observed, for each of which this study aims to classify urban land use using the deep learning-based segmentation model. As such, the model will be applied to a set of metropolitan areas, which covers a total surface of 992,680 square kilometres, spread across a variety of urban landscapes in terms of economic development structures, urban planning styles, and climate zones.

3.2. Ground Truth Data on Urban Land Classification

Ground truth data, which serve as input for the training of the deep learning model, and which also provide reference in the evaluation of model accuracy, are sourced from the Copernicus Urban Atlas project. These data offer a comprehensive classification of land use and land cover, and are based on granular vectors. In specific, any patches of land use that are at least 0.25 hectare in size are mapped granularly. The observed Urban Atlas maps are for the year 2018, the latest available version of these maps. Coverage is provided for FUAs across Europe.[9] An illustration of the classification, for the FUA of Luxembourg, is provided in Figure 5. By overlaying these data with satellite imagery, for each satellite image pixel the associated land use or land cover class according to the Urban Atlas map can be observed. The Urban Atlas map's resolution is matched with the satellite imagery's 10m x 10m resolution to allow for straightforward assessment of agreement between predicted and 'true' land use classes. The full set of Urban Atlas data for European OECD countries is used in the labelling of the training, validation, and test sets of satellite images. [10]

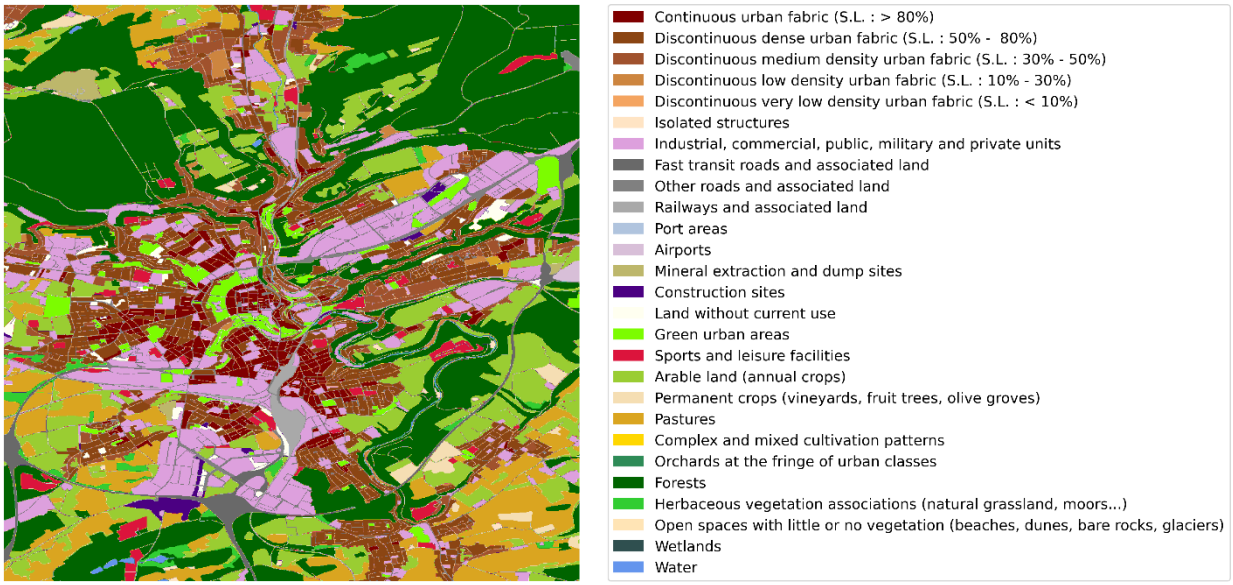
[8] The 27 European OECD countries are Portugal, Spain, Italy, France, Belgium, the Netherlands, Luxembourg, Switzerland, the UK, Ireland, Germany, Austria, Slovak Republic, Czech Republic, Denmark, Norway, Finland, Sweden, Iceland, Slovenia, Hungary, Poland, Estonia, Latvia, Lithuania, Greece, and Turkey.

[9] The Urban Atlas dataset as a whole ranges beyond OECD countries, and covers 788 FUAs across EU27 and EFTA countries, the West Balkans, Turkey, and the UK.

[10] For a subset of OECD-EC FUAs (n = 60), the Urban Atlas does not provide coverage. In those cases, imagery patches for those (typically small) cities are omitted from the modelling process, but city-specific area estimates are provided. The area estimates now are adjusting not for map biases observed at the city level but at the national level.

The Urban Atlas dataset allows to overcome one of the main bottlenecks of satellite image segmentation, which is the lack of granular reference data for model training and accuracy evaluation. The underlying assumption is that the Urban Atlas provides a 'ground truth' and so by definition is more accurate than the predictions which will result from the deep learning model. This is a reasonable assumption, as the Urban Atlas is based on imagery at a resolution higher than this study's imagery and, more importantly, was largely created via manual interpretation (Montero et al. 2014). Such processes of human visual judgment of 'what is what exactly' are challenging to replicate in a fully automated way. At the same time, in some instances it is plausible that the deep learning model will offer a more granular classification of land (from the map user's perspective). To ensure that ground truth data are consistently of a higher quality than predicted maps, a common approach is to generate ground truth data manually, typically for a limited number of (a few hundred or a few thousand) sample locations (see, e.g., Curtis et al., 2018; Olofsson et al., 2014; Zhu et al., 2016). Such manual efforts are high-cost, notably when it comes to the object-based labelling of images rather than the labelling of subsample of individual pixels. This represents a major barrier to studying urban areas both at granular scale and at a high level of quantified accuracy, which would be required to inform local policy with confidence in the estimates.

Figure 5. Urban Atlas disaggregated definition of land use (and land cover) in the city of Luxembourg.



For this paper's purposes, Urban Atlas maps offer a highly relevant, comprehensive, and granular definition of land use and land cover classes. Importantly, accuracy estimates for model performance and area estimates can, despite pan-European coverage, be reported at the policy-relevant unit of individual cities.

However, before proceeding, the Urban Atlas maps for each FUA is aggregated into six main classes of (non-)built-up areas. This simplification[11] allows the deep learning model to more effectively learn and predict key types of urban land use, but still offers a level of disaggregation of built-up area that is complementary to existing large-scale land cover products such as the JRC Global Human Settlement Layer (Corbane et al., 2021) or ESRI's Global Land Cover map (Karra et al., 2021) do.[12] In the remote sensing literature, observing 5 to 10 distinct classes rather than more disaggregated classes (recall Figure

[11] For instance, without such aggregation of land use classes, a particular plot of land cover which could be considered as a park could be observed as a mixture of lower-level land use classes such as a playground or recreational forest.

[12] Initial classification outcomes for different levels of Urban Atlas category-aggregation are presented in the Appendix.

5) is common (Zhu et al., 2016). The resultant and final 'ground truth' definition of urban land use and land cover classes is visualized in Figure 6, again for the city of Luxembourg. The underlying Urban Atlas sub-categories to which the aggregate classes can be traced back are defined in Table 1.

Table 1 also presents for each observed land use class its relative share in the total area of European FUAs. This shows that residential land use is typically the largest class, although the most prevalent classes in the study area are open space and agricultural land use. The reason for this is that FUAs include commuting zones, which mostly consist of built-up areas of low density which are surrounded by open and agricultural land. This 'imbalance' in the area shares of the distinct land use classes in the accuracy assessment has implications for the interpretation of the accuracy of predicted maps. For instance, if one large class such as agricultural land is particularly well predicted, a poor performance of a class such as residential built-up area, which is overall smaller in area but more central to this study's urban setting, would remain hidden in an overall accuracy measure (see Equation 1 in Section 5.1). The reason for this is the relatively limited areal weight of any errors that are associated with this smaller class in the generated map. Therefore, the analysis will instead mostly rely on class-specific accuracy measures, which convey more targeted accuracy information.

Figure 6. The final ground truth categories of land use (and land cover) illustrated for Luxembourg FUA.

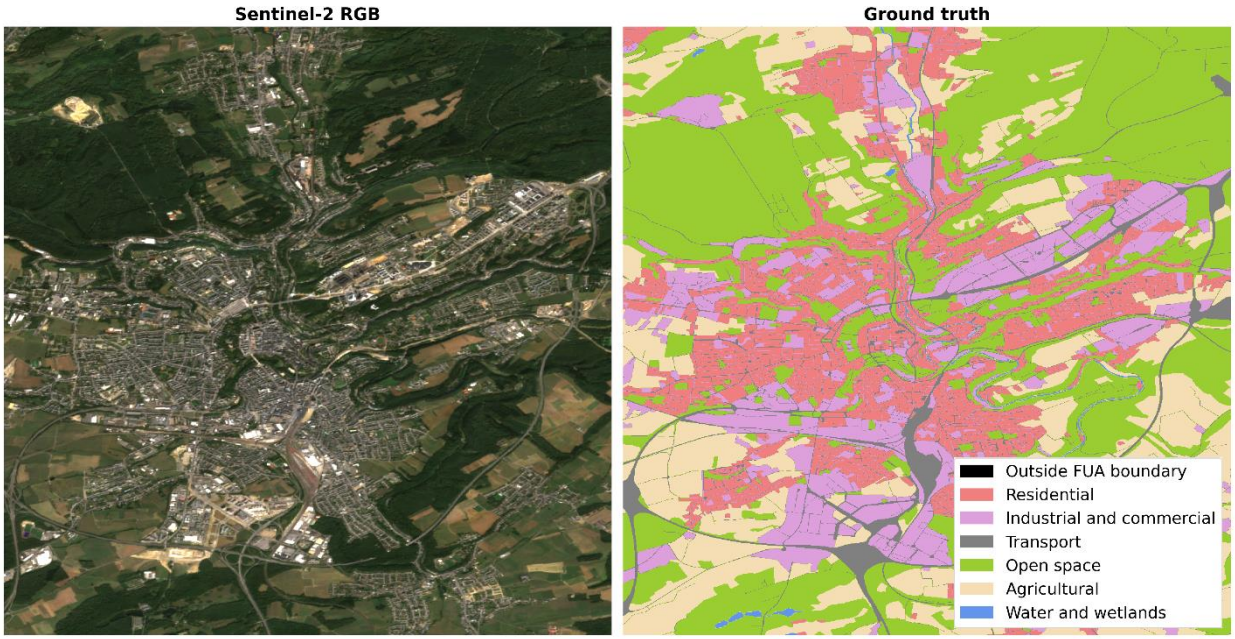


Table 1. Definition of observed urban land use (and land cover) classes in the 2018 reference data.

Observed Class	Original Sub-Category in Urban Atlas	Mean (% of area)	St. dev. (% of area)
Residential	Discontinuous and continuous urban fabrics, Isolated structures	7.8	5.3
Industrial and commercial	Industrial, commercial, public, military, and private units, mineral extraction and dump sites, construction sites, land without current use	4.1	3.3
Transport infrastructure	Fast transit roads, other roads, railways, port, airports	2.7	1.6
Open space	Forests, herbaceous areas, open space without vegetation (beaches, bare land), green urban areas, sports and leisure facilities	33.8	19.3
Agricultural	Arable land, permanent crops, pastures, complex and mixed cultivation, orchards	48.8	19.1
Water and wetlands	Water, wetlands	2.7	4.3

Note: mean and standard deviation values are for the shares of observed land use classes across each of the observed individual cities (N = 687). The land use shares are obtained from Urban Atlas reference data for 2018. Area shares that are disaggregated by land use class can be obtained from Table 7 in the Appendix.

3.3. High-Resolution Satellite Imagery

High-resolution imagery (10m x 10m) is publicly provided by the Copernicus Program. In specific, images are sourced from the European Commission and European Space Agency's Sentinel-2 and Sentinel-1 satellite constellations. Sentinel-2 'optical' sensing instruments measure radiations reflected or emitted by observed objects or landscapes. This process passively captures scenery as also the human eye would see it, as well as more, by exploiting waves on the electromagnetic spectrum beyond visible light, such as infrared radiations. Sentinel-1 'synthetic aperture radar' sensing instruments, on the other hand, actively emit radio signal impulses. When such signal impulse meets an obstacle, it scatters back to the sensor to some degree. Based on amount and travel time, it is then possible to estimate how far away the obstacle is, which provides information on the shape and depth of urban landscape configurations.

Combining Sentinel-1 and Sentinel-2 imagery allows us to observe a comprehensive set of 'image bands', or layers. Each of these image bands, as listed in Table 2, has specific qualities relevant to image segmentation. In the Sentinel-2 imagery, selected bands capture: blue, green, red, near-infrared (NIR), and short-wave infrared (SWIR1 and SWIR2) values.[13]

Blue, green, and red reflectance captures what human sight also observes. NIR, however, is not observed by the human eye and is particularly useful at capturing vegetated land cover, so that in image classification vegetation can be separated effectively from land that is in residential or commercial or industrial use. Every form of matter with a temperature above absolute zero (-273.15°C) emits infrared radiation according to its temperature. SWIR bands, on the other hand, contribute to the separation of water bodies from 'dry' land and help to identify various kinds of open space given soil moisture and vegetation cover.

[13] Using the Sentinel-2 bands, also various indices that measure land cover by vegetation (NDVI), built-up area (NDBI), or water (NDWI) were computed. However, these indices were not selected for the main analysis as their contribution to model performance was limited. This limited contribution to model performance is likely due to the ability of deep learning models to compile relevant combinations of spectral information without necessarily requiring a priori guidance as offered by pre-specified land cover indices. Also the use of the gradient on specific bands has been explored.

Table 2. Sentinel constellation technical properties.

	Sentinel-1	Sentinel-2
Sensor component	C-band Synthetic Aperture Radar	Multi Spectral Instrument
Launch data	April 2014	June 2015
Spatial resolution	10 m: All bands	10 m: Bands 2, 3, 4, and 8 20 m: Bands 5, 6, 7, 8A, 11, and 12 60 m: Bands 1, 9, and 10
Bands	HH: Single co-polarization, horizontal transmit / horizontal receive HH+HV: Dual-band cross-polarization, horizontal transmit/vertical receive HV: Partial dual, HV only VV: Single co-polarization, vertical transmit / vertical receive VV+VH: Dual-band cross-polarization, vertical transmit / horizontal receive VH: Partial dual, VH only	Band 1: Coastal aerosol Band 2: Blue Band 3: Green Band 4: Red Band 5: Vegetation Red Edge 1 Band 6: Vegetation Red Edge 2 Band 7: Vegetation Red Edge 3 Band 8: Near Infra-Red (NIR) Band 8A: Narrow NIR Band 9: Water vapor Band 11: Short-Wave Infrared 1 (SWIR1) Band 12: Short-Wave Infrared 2 (SWIR 2)
Revisit time	6 days with 2 satellites	5 days with 2 satellites

The optical bands are complemented by the radar-based bands which come from Sentinel-1 imagery. In specific, in Sentinel-1 imagery the VV (vertical transmit and vertical receive) and VH (vertical transmit and horizontal receive) bands are selected. The VV band and, in particular, the VH band together add further information on the depth of land configurations and so, to some extent, built-up area density (Li et al., 2020).

To pre-process and obtain the selected imagery products, for each FUA, Google Earth Engine is used. Earth Engine, as described in Gorelick et al. (2017), is a cloud-based platform that offers the massive computational power that data acquisition for this study's continental-scale analysis requires.

More specifically, this study processes all individual Sentinel-1 and Sentinel-2 (Level-2A) images available for 2018 that cover, or partially cover, European FUAs.[14] Sentinel-2 images are filtered based on a 60%-threshold regarding total cloud cover in the observed image following Braaten (2021). In addition, clouds and cloud shadows are filtered internally in images using a cloud probability dataset. For Sentinel-1 images, such cloud-based filtering procedures are not required. This is because radar instruments, unlike optical instruments, are able to collect informative imagery in all weather conditions and both by day or night. Each Sentinel-1 image, however, was pre-processed to remove thermal noise, radiometric calibration, and terrain correction (orthorectification), followed by a correction of outcome values to decibels via log-scaling. Finally, from the selected and pre-processed images, for each pixel, and for each image band separately, the observed band's median value is obtained. This effectively removes the possible influence of clouds, haze, and shadows. In Polar regions, median values are based on the May-October period, to avoid influence from snow. As a result, for each FUA an image is observed in which urban areas can be tracked clearly and consistently.

[14] An alternative approach to observing year-round imagery, is to select images for the 'green season', as that could possibly help to even better separate vegetation from built-up structures (Gong et al. 2019). However, this would introduce complicating assumptions on the varying timing of the green season across years and regions, and would increase the influence of outliers in pixel values as fewer images would be possible to observe for a given city, in particular for cities that are frequently covered by clouds. For these reasons, and to ensure consistent image selection also in years beyond those covered in the current study, year-round imagery is used.

4 Results

4.1. Deep Learning Model Evaluation

4.1.1. Visual Assessment of Mapping Accuracy for European Metropolitan Areas

This section examines the deep learning model's performance at classifying urban land uses, in particular for built-up areas. Before turning to the results for relevant evaluation metrics, the estimated maps for several OECD Functional Urban Areas (FUAs) can be examined qualitatively. Figure 7 maps the estimated map classifications for subareas of Berlin, Amsterdam, and Vienna (middle panels). Also shown is the underlying satellite imagery (left-hand side panels, only RGB bands) from which land use is estimated, as well as the 'ground truth' map classification (right-hand side panels) that the model uses to associate patterns in data-values for image pixels with particular (built-up) land use types.

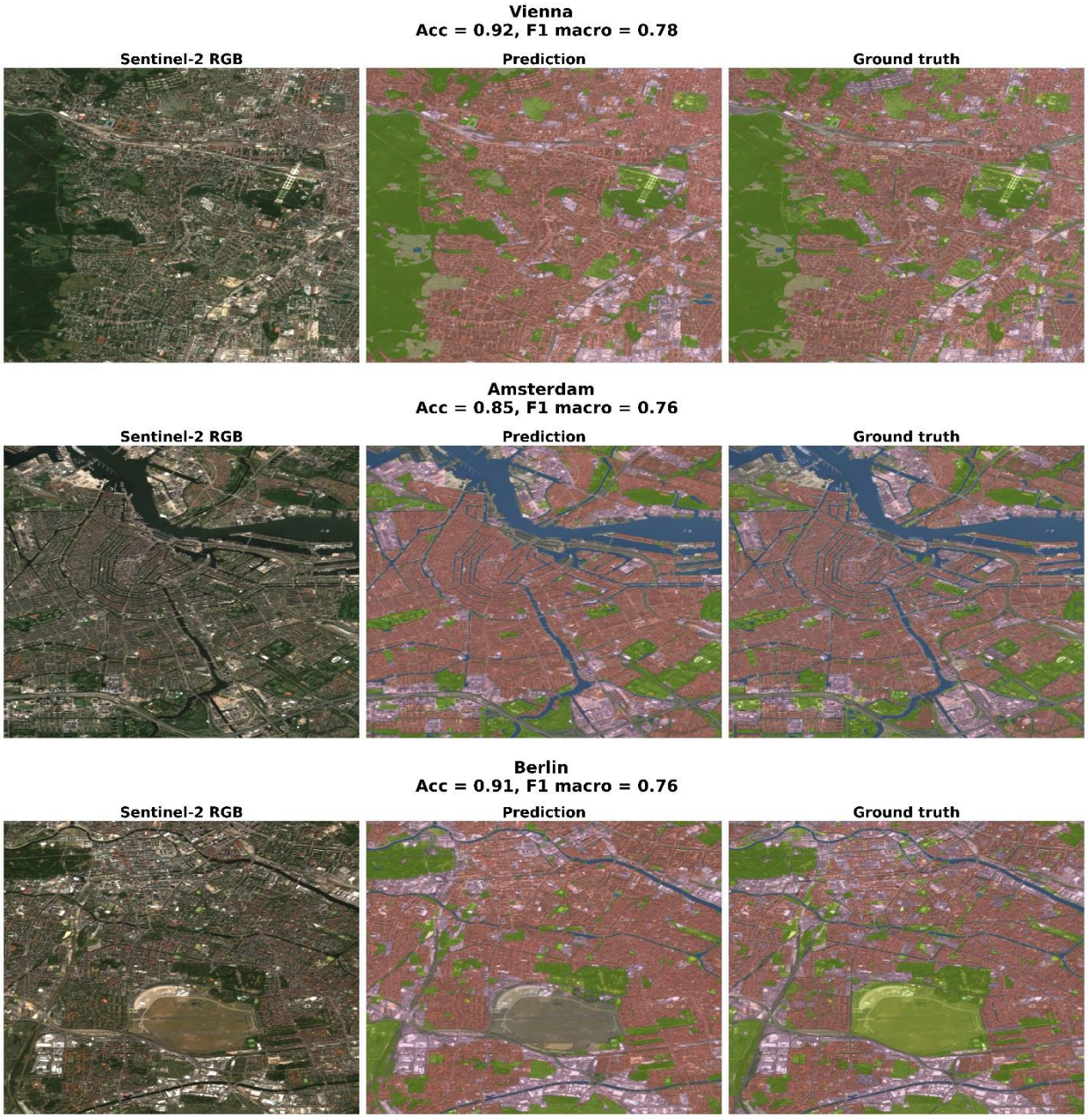
Based on the figure, a first observation is that the U-Net can be trained to recognize and then classify key types of urban land in a way that is largely accurate. Vegetated areas are well-separated from built-up areas, and within built-up areas, residential and industrial or commercial areas are distinguished clearly. For the FUA of Berlin, the model predicts the large park as a transportation infrastructure instead of an open space. This park corresponds to Tempelhof, which used to be an airport, and still comprises landing strips and terminals. This is a compelling example of land use that is challenging for the model to capture.

4.1.2. Assessing Accuracy at the Level of the European Study Area

To move beyond visual impressions, the model's performance will now be quantified. A starting point is Figure 8, which shows a confusion matrix. The confusion matrix is obtained from pixel-wise comparison, on the test set, of the predicted class against the true class according to the Urban Atlas reference data. This confusion matrix is normalised according to the true labels. This means, for example, that 84% of the pixels corresponding to residential areas in the reference data were correctly predicted by the model. For most of the targeted types of urban land use the model performs well. This provides a first indication that combining Sentinel-2 with Sentinel-1 imagery allows for a precise segmentation of economically relevant urban land use classes across the continental study area.[15]

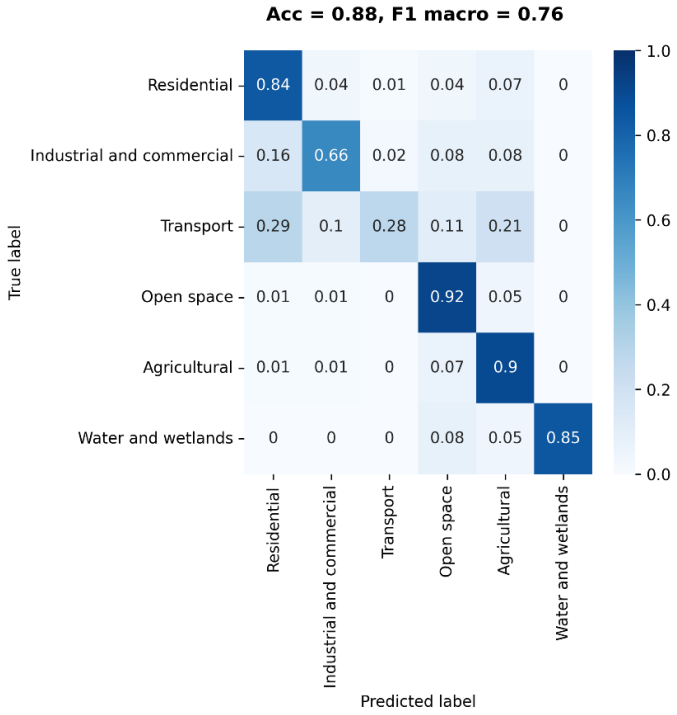
[15] To ascertain that the sample size of 1.6 km x 1.6 km imagery patches is sufficient, and to clarify how the model would perform on smaller samples, the main model is re-trained and re-estimated for subsamples of the full dataset. In specific, models were estimated after retraining on random samples of 20%, 40%, 60%, and 80% of the full training sample. The test set of images, for which accuracy metrics are calculated, was held constant. Resultant balanced accuracy values, for various sample sizes, showed a range of only 10 percentage-points. This suggests that also small sample sizes may achieve decent overall map accuracies, although accuracies specific to land use types may vary.

Figure 7. Maps of predicted land uses in subareas of several OECD-EC metropolitan areas (FUAs).



From confusion matrices, standard model evaluation metrics can be computed, whose values all range from 0 to 1. A first evaluation metric indicates that the predictions of land use classes, as pooled across the observed 687 European FUAs, have an overall accuracy of 0.88 (see Figure 8). This means that 88% of all evaluated pixels are correctly predicted.

Figure 8. Confusion matrix (normalised) for the test set of images for European OECD FUAs.



Note: The proportions in the confusion matrix are normalised according to rows, that is, by the total number of true label observations. The diagonal elements correspond to recall metrics for each class.

In more detail, overall accuracy (*OA*) is perhaps the most intuitive performance measure, being defined as the fraction of correct pixel predictions among the total number of observations[16]:

$$OA = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (1)$$

$$= \frac{Correctly\ Classified\ Pixels}{All\ Pixels}$$

The overall accuracy metric should, however, be interpreted with care as the model pertains to multiple land use classes. As such, overall accuracy does not account for class imbalance. The accuracy could be particularly high with a model that performs poorly on particularly low-prevalence classes (e.g., transport infrastructure), which may be hard to learn and detect. Indeed, highly prevalent classes (e.g., agricultural land), which are easier to detect, have more weight in the data, and thus tend to drive high accuracy values. This may partially explain the high level of overall accuracy of the current predictions. Other useful metrics include the precision and recall. Precision reflects the fraction of true positives amongst the elements predicted as positive by the model:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

[16] A true positive [true negative] is an outcome where the model correctly predicts the positive class [another class]. A false positive [false negative] is an outcome where the model incorrectly [not] assigns a pixel to the observed class.

As such, the 'precision' score captures the share in all pixels assigned to an observed land use class that should indeed be assigned to that class according to the ground-truth (Urban Atlas-based) reference data. The precision metric thus signals whether an observed class is over-predicted.

On the other hand, the 'recall' score reflects the fraction of true positives amongst all the pixels for an observed land use class that should have been predicted positive according to the ground-truth class labels:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3}$$

As such, recall signals whether an observed class is under-predicted: is all land of a given type retrieved by the model, or have some pixels erroneously been assigned to another land use type? As the confusion matrix in Figure 8 was normalised according to the true label, the values reported in the diagonal correspond to the recall scores for each class.

Out of the two metrics of precision and recall, the F1-score can be defined as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

It accounts for both false positives and false negatives by providing a composite signal of whether the classifier over-predicts and under-predicts. This metric ranges from 0 to 1. A value of 1 corresponds to a model that perfectly predicts each observation as the correct class, and a value of 0 to a model unable to predict any pixel as the correct class.

In the case of multi-class classification, these metrics can be adapted as:

- *Macro metrics*: metric computed class by class and then averaged.
- *Micro metrics*: metric computed for each pixel no matter its class.

The main difference between macro and micro metrics is that macro weighs each class equally whereas micro weighs each sample equally. For the recall metric, this would give for example:

$$Recall_{micro} = \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} TP_i + FN_i} \tag{5}$$

$$Recall_{macro} = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FN_i} \tag{6}$$

Macro-metrics are in general impacted by class imbalance, as individual metrics are computed class by class and then the same weight is given to each class before averaging. Particularly under-represented classes in the dataset, that are harder to detect and to learn, can consequently lead to a low macro metric.

Table 3 shows the F1-score, recall and precision obtained for each class. These metrics are high for residential areas, agricultural areas, water, wetlands, and open space; and slightly lower for industrial and commercial areas. Transportation networks are, instead, under-predicted, with a recall score of 0.27. The precision for this class is however much higher, showing that the pixels predicted by the model as transport are well predicted. A reason for this is that the associated land-based features (e.g., roads in dense urban settings) are, by nature, spatially woven-in between neighbouring land uses, which invites under-prediction

for this class despite the imagery's high resolution.[17] This is confirmed by the precision obtained for the residential class, which is lower than the recall, suggesting that the model tends to over-predict this class. At the macro-level, the F1-score is 0.76, which indicates a good overall model performance.

Table 3. F1-score, precision and recall for each land use class and at the macro level.

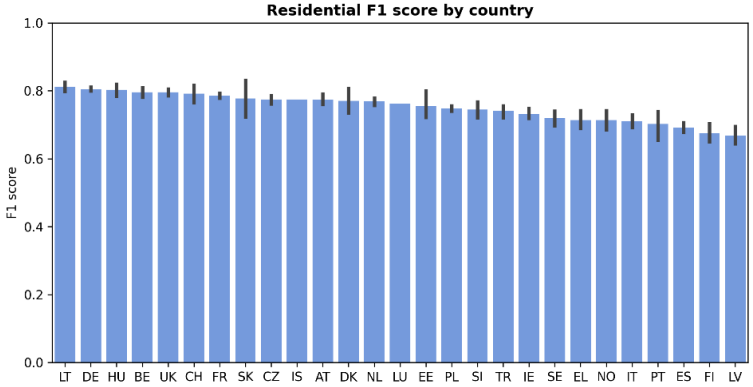
Observed Class	F1-score	Precision	Recall
Residential	0.78	0.72	0.84
Industrial and commercial	0.66	0.67	0.66
Transport infrastructure	0.38	0.62	0.27
Open space	0.90	0.88	0.92
Agricultural	0.91	0.93	0.90
Water and wetlands	0.89	0.93	0.85
Macro	**0.76**	**0.79**	**0.74**

4.1.3. Assessing Accuracy at the Level of Countries

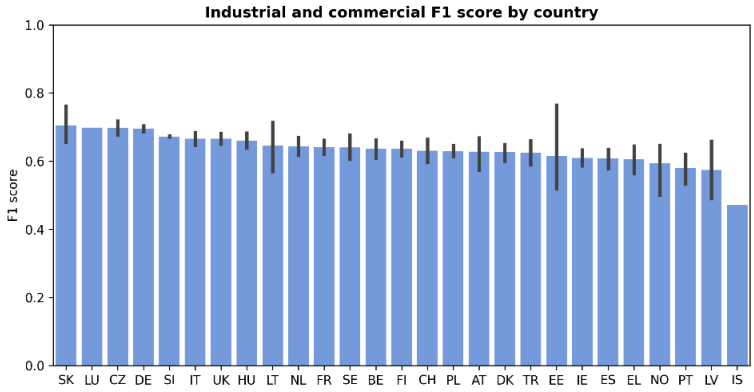
In addition to the model's general performance at classifying urban land, its performance can also be evaluated at the scale of countries. In doing so, the focus is on the two types of urban land that are key to the analysis, residential land and industrial or commercial land. For both uses of land, Figure 9 shows an F1-score (the harmonized mean of precision and recall, which conservatively weighs towards the lower value of those observed for precision and recall) for each of the observed OECD countries. F1-scores vary from 0.67 to 0.81 for residential built-up areas and from 0.47 to 0.70 for industrial and commercial built-up areas.

Figure 9. Average F1 scores by country for the classification of residential land use (top panel) and the industrial or commercial built-up areas (bottom panel).

Bars indicate variation in scores at the city-level



[17] Probability values underlying the assignment to particular classes (see the discussion of the U-Net's probability tensor in Section 3.1), however, suggest that road detection could be enhanced. This, however, is beyond the scope of the current paper where the transport class is mostly included to mitigate assignment of infrastructure-related pixels, despite their limited weight in the data, to other classes that involve developed land, as these are spatially correlated.

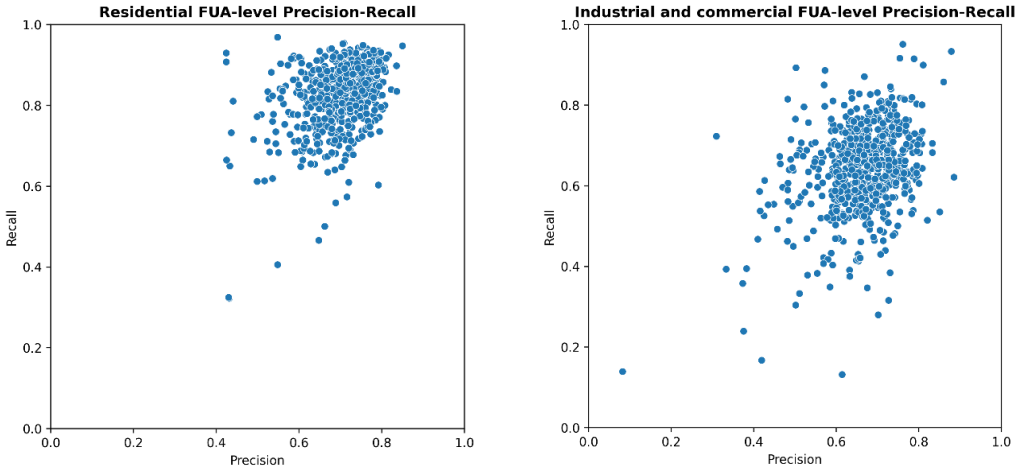


Furthermore, the modest range of error bars in Figure 9 indicates that, internal to countries, the accuracy of residential and industrial or commercial built-up area classification is relatively constant across cities. A few cross-city accuracy ranges that are particularly high are observed for both the observed land use types in the Northern countries of Norway, Latvia, Estonia, and Lithuania. Generally, F1-scores appear to be higher for countries with a mostly temperate climate, although climate plays a limited role in overall model performance.[18]

4.1.4. Assessing Accuracy at the Level of Functional Urban Areas

Next, as FUAs represent a policy-relevant scale of analysis, it is necessary to consider how classification accuracy varies at the level of individual FUAs. To this end, Figure 10 plots FUA-level precision and recall outcomes against each other. In both of Figure 10's panels, most FUAs are positioned in the top-right quadrant, which reflects relatively accurate classifications.

Figure 10. Precision-recall accuracy metrics for individual FUAs by key urban land use class.



More precisely, most cities show recall and precision values above 0.7 and 0.6, with distributions that lean towards values of 0.95 and 0.8 for built-up area in a residential or in an industrial or commercial land use, respectively. Also noticeable is the dispersion around these values, with accuracies being very low for a few cities. Spatial variation in classification accuracy is natural to any map product and thus relevant to quantify.

[18] U-Net models were initially also estimated by climate zone, but performance was similar as for the pooled model.

4.1.5. Urban Shape Influences the Accuracy of Land Use Classification

Due to the consistent nature of the OECD-EC definition of metropolitan boundaries by FUAs, this study is uniquely able to examine whether the physical shape of cities may influence how accurately land uses are estimated. Is land use estimated similarly well in cities where the density of built-up area is higher or lower?

We now turn to assess FUA-level classification accuracy for built-up area in a residential or in an industrial or commercial land use, for different levels of built-up area density. Density, in this relatively 'physical' land cover related exercise, is for simplicity defined as the ratio of built-up area (the combined areas of residential, industrial, and commercial land) divided by the total area (including non-built-up area) of the observed city (the city, or cities, not the entire FUA's surface, is considered here to maximize variation in density).[19]

A priori, the relationship between built-up area density and land use classification accuracy is ambiguous. A high level of built-up area density can signal urban complexity, as land uses may spatially blend together. However, high built-up area density may also reflect zoning policies in which distinct uses of land are tightly separated and thus potentially easier to separate from each other using computer vision.

A low level of built-up area density may imply that land use zones are less cohesive in terms of the (developed and non-developed) land cover mix. Moreover, in satellite imagery for dispersed urban areas, built structures might be more easily obfuscated by features such as leafy trees.

More definitive insights follow from Figure 11. Figure 11 plots, at the level of individual FUAs, the F1-score for built-up areas in a residential land use (left panel) and for built-up areas in an industrial or commercial land use (right panel) against the measure of each city's built-up area density.[20] This shows a clear positive relationship, with a moderate slope, between the level of built-up area density and classification accuracy.

For residential land use, this positive relationship is steeper than for industrial or commercial land. Potential explanations may relate to how (de-)centrally these land uses are generally allocated within cities as well as to how well urban features are captured in satellite imagery of a particular resolution at different urban density compositions. [21]

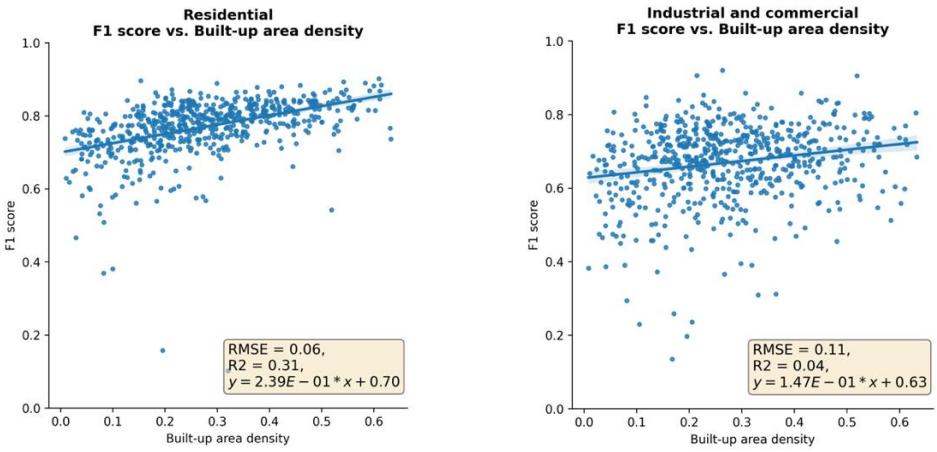
The new insights regarding how classification accuracy varies with the density of urban built-up areas address a relevant knowledge gap in the scientific literature on the remote sensing of cities (Reba and Seto, 2020).

[19] Cities, as concentrations of human activities, are by nature closely related to the density of built-up area, whereas their commuting zones are defined by the movement of people, and lesser so by a spatial concentration of built space. Commuting zones may more so reflect a mix of decentral open landscapes and urban concentrations –of sizes below the threshold for being an urban centre– whereas FUA city boundaries tend to enclose a density built-up area that is relatively high (relative to that FUA overall level of concentration or dispersion of people and firms). As FUA cores are defined based on a minimum threshold of population density (Dijkstra et al. 2019), cross-city variation in built-up area density can be compared consistently in this exercise. Although the measure ignores vertical density, a lack of open space between typical urban land uses offers a general signal of compact urban form.

[20] Recall that the F1 score is the harmonized mean of precision and recall, and so provides a composite signal of whether the model overpredicts and underpredicts an observed land cover class.

[21] In this context, it can also be observed that the U-NET model performs better for residential and industrial or commercial land use classes in FUA cores, or cities, than in commuting zones (see Annex D for results). Moreover, non-built-up classes are better segmented in commuting zones. This further illustrates that the shape of cities may influence mapping accuracy. However, the overall influence is limited at the study area-level, and is mostly an issue to consider at the level of individual urban areas.

Figure 11. FUA-level accuracy of land use prediction varies with the core's built-up area density.



The results here underline the key importance of policy-relevant variation in map accuracy being quantified. Using the resultant information on accuracy, appropriate estimates of land use areas can in the Section 5.2 be produced: area estimates can now be adjusted for a known degree of bias in the underlying maps.

Box 2. Expanding the Mapping Process to Non-European OECD Countries

The U-Net deep learning model presented in this paper can be applied to any geographical area. To explore such wider application, land use maps are predicted for several OECD metropolitan areas outside of this paper's study area, including San Francisco, Seoul, Tokyo, Auckland, Santiago, Sydney, Mexico City and Bogota.

Important to note is that the model has not 'seen and learned' any urban land use patterns outside of Europe. Therefore, this exercise gives an impression of how transferable the European model's predictive ability is to urban settings in further world regions.

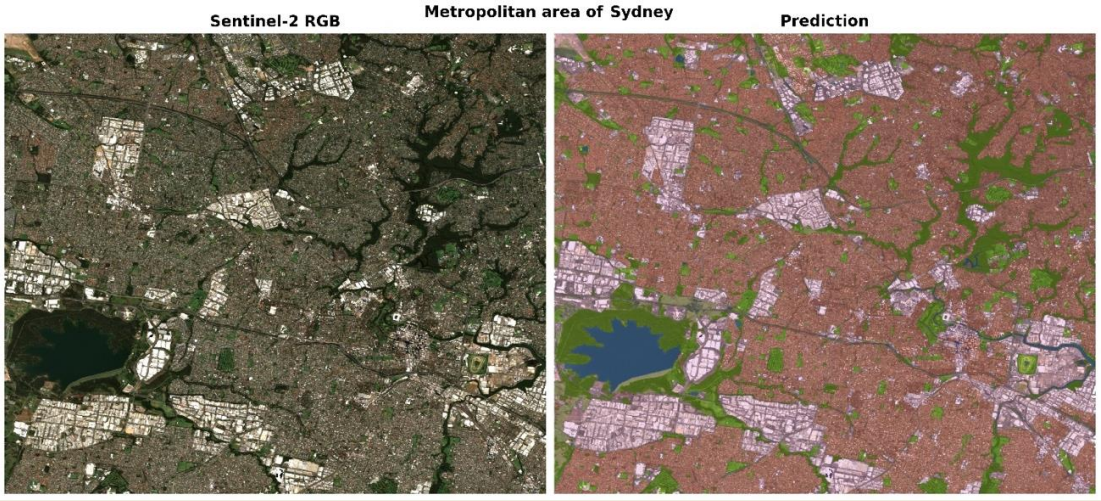
The exploratory results (see Annex F and Figure 12 below) suggest that the European U-Net model can effectively segment urban land uses also in non-European OECD countries. Maps generated for Sydney and Mexico City reveal a clear ability of the model to distinguish industrial or commercial built-up areas from residential built-up areas. Maps for San Francisco, and Bogota, for instance, identify transport infrastructure features such as ports and airports relatively well. Open (green) spaces and agricultural land are well-identified for each of the metropolitan areas observed here.

However, map predictions are now applied to satellite images that enclose a new level of cross-country and cross-continental variety in land use structures, as driven by a variety of market forces and urban planning cultures. For instance, in Seoul and Tokyo, separating residential and industrial or commercial uses of land from each other is more challenging than in Sydney or San Francisco, where dominant uses of land are generally more spatially separated.

The observations above have two main implications regarding the inclusion of further OECD countries in the mapping process. One is that the mapping objective, of what land to assign to which land use type, should be conceptually consistent across countries (Section 1). Secondly, more heterogeneity in urban land use features and patterns should be considered in the collection of training and evaluation data. In remote sensing studies, this generally involves the manual labelling of land use in satellite images. In this study's case, this may instead relate to whether readily existing ground truth maps can be obtained for OECD countries beyond Europe that are consistent with the Urban Atlas. An

overarching challenge is to achieve cost-reductions in the collection of consistent imagery labels at scale, to track urban land use accurately in more OECD countries.

Figure 12. Illustration of land use predictions for a metropolitan area (Sydney) new to the model.



4.2. Area Estimates for Land Use in European Metropolitan Areas

4.2.1. Procedure to Adjust Area Estimates for Known Map Biases

While any map that classifies urban land might appear visually compelling, an important question is to what extent the map captures the areas of the observed land use types accurately. There are two main ways in which areas can be estimated. One way is to simply count pixels associated with a particular land use. This approach, however, ignores the degree of error in the predicted map (recall the evaluation metrics in section 5.1). A preferred approach, therefore, is to report area estimates that are adjusted for the known degree of error in the mapping of specific land use classes – as quantified using reference data that is assumed to be of a superior quality (UN FAO 2016; Olofsson et al. 2014; Stehman 2013).[22] The sample-based approach is adopted in this section's analyses. In specific, the map proportion for land class *k* is estimated using the equation:

$$\hat{p}_{\cdot k} = \sum_{i=1}^{q} W_i \frac{n_{ik}}{n_{i\cdot}}. \tag{7}$$

where $\hat{p}_{\cdot k}$ denotes the area proportion for the *k*-th land use class. The area proportion estimate follows from the sum of proportions of pixels (n_{ik} / n_i) that are either correctly assigned to class *k* or 'omitted' through being incorrectly assigned to another *i*-th class while belonging to class *k* according to the

[22] It may be noted that also standard errors and confidence intervals associated with the area estimates are obtained, following equations (10) and (11) in Olofsson et al. (2014). However, reporting these explicitly, would in this study's case add limitedly insightful information. The reason for this is that the number of observed reference pixels is very large, thus compressing the confidence intervals such that these might suggest over-precision of the estimates (as compared to estimates in smaller-sample studies for which confidence interval reporting is advocated).

reference data, multiplied by the *i*-th class's total area proportion (W_i). The area proportion for the *k*-th class can then be multiplied with the total map area to obtain the estimate for the class's area.

The information that the estimation requires can be obtained from the predicted map and the associated confusion matrix. As mapping accuracy varies across countries and urban areas (recall section 4.1), area estimates are adjusted for errors at the lowest spatial scale of analysis in this study that is of clear policy relevance, the individual FUA. This also flexibly allows for area aggregations to (inter-)national scales.

4.2.2. Land Use Area Estimates for 2021 Across Metropolitan Areas

Using the deep-learning model's land use predictions for metropolitan areas in Europe, as defined by OECD-EC functional urban areas (FUAs), and after adjusting these estimates for known mapping errors[23], a variety of key urban land use patterns can now be examined on a consistent basis.

Amongst the 992,680 km^2 of observed urban surface, estimates suggest that 7.5% of land surface is allocated to a residential use (see Table 4). Residential built-up area, as such, is the largest of the 'typical' urban of urban land in terms of areal cover. In cities (densely populated cores), the share of residential built-up area is 12%, which is approximately two times the amount of land allocated to an industrial or commercial use, and roughly three times the land surface that is in an infrastructural use. Open spaces, which may reflect green spaces as well as recreational spaces and bare land, and agricultural land together represent more than two-thirds of the total observed metropolitan surface.[24]

Important in the areal distribution of land use in FUAs is the role of commuting zones, where, in relative terms, less land is used for residential or industrial and commercial purposes than in cities. In absolute terms, however, most of the land that is covered by built-up area in a residential, industrial, or commercial use is found within commuting zones. This means that, across European FUAs, locations where the use of built-up areas tends to be relatively spread-out have the most weight in the overall physical, or built-up, 'footprint' that FUAs have on land.

Table 4. Estimated FUA land area in 2021, by type of use (within European OECD countries).

	Area (km^2)			Area (%)		
	City	Commuting zone	FUA	City	Commuting zone	FUA
Transport	8,506	17,078	25,585	3.9	2.2	2.6
Water and wetlands	8,848	19,964	28,812	4.1	2.6	2.9
Industrial and commercial	13,476	21,331	34,808	6.2	2.7	3.5
Residential	25,880	48,170	74,049	12.0	6.2	7.5
Open space	78,080	284,725	362,804	36.2	36.7	36.5
Agricultural	81,153	385,422	466,575	37.6	49.6	47.0

It should be noted that comparing FUAs based on areal shares of different land uses relative to the total land area of the FUA or city has two main drawbacks. First, the comparison would be highly dependent on the areal size of the local units used to define the FUA or city. Second, the resultant share is likely to only increase as FUA or city boundaries are held constant. Therefore, the analysis below relies on measures of the areas of distinct types of land on a per capita basis, based on the number of FUA or city inhabitants.

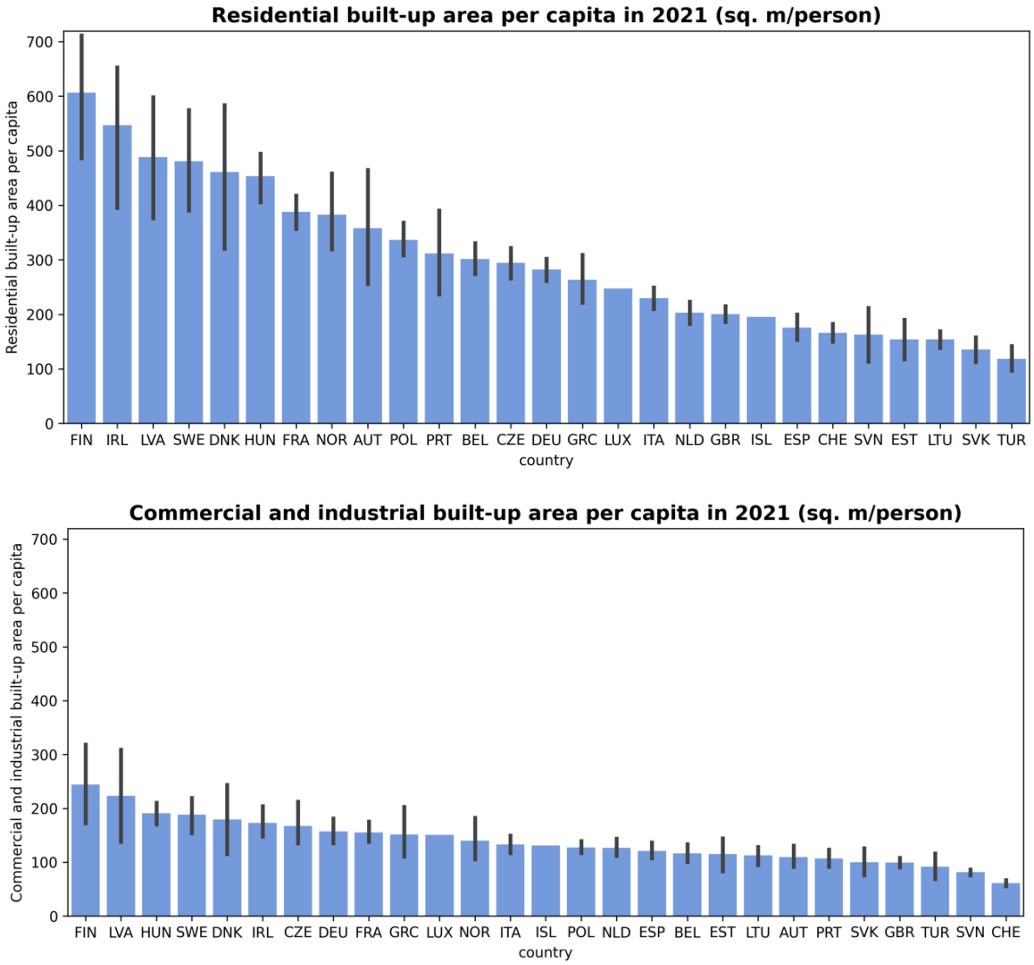
Figure 13 shows for each of the observed countries the average built-up area per capita across FUAs, as differentiated for key uses of urban land. In specific, built-up area per capita is shown for residential as well

[23] Adjustment follows the procedure in Section 5.2A and assumes that mapping errors in 2018 data apply also in 2021.

[24] Water bodies are only partially captured in the FUAs definition because of definitional reasons (see Dijkstra et al. 2012) and therefore have a limited weight in the estimated urban land use areas, despite many cities being close to water.

as for industrial or commercial uses. For each country, the thin lines inside the bars indicate variation of built-up area per capita, in the observed use, amongst FUAs.[25]

Figure 13. Urban built-up area per capita (2021), by country and land use type.



Between the 27 countries observed in Figure 13, the range of built-up area per capita is substantial. For residential land use, the lowest amount of built-up area per inhabitant across FUAs is about 120 square meters (Turkey) whereas the highest built-up area per capita is close to 600 square meters per inhabitant (Finland).[26] Between these two outer values, the distribution across countries is relatively gradual. This shows that although within Europe there is noteworthy variation in how much land per inhabitant is in residential use, there is no general divide between countries.

However, on a country-by-country basis, it can now be observed that the residential built-up area per capita in Ireland's FUAs is, on average, about twice that of FUAs in Germany, or that such per capita outcomes are smaller for FUAs in Spain and Estonia than for FUAs in countries that are widely known for their compact cities, such as the Netherlands and the United Kingdom. Largely similar patterns are observed, in the lower panel of Figure 13, for industrial or commercial land use, also in terms of how countries rank.

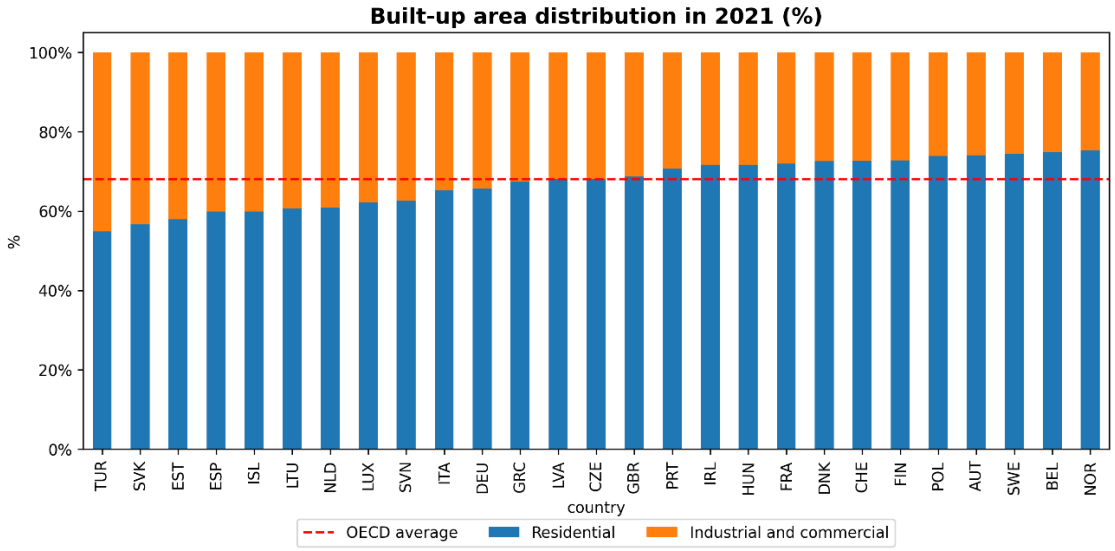
[25] For Iceland, no line is plotted as only one FUA is observed and so is represented by the country-level estimate.

[26] It may be recalled that these per capita built-up areas do not capture internal floor space but any land, including gardens, that is associated with a residential use.

Noteworthy, in Figure 13, is that within several countries built-up area per capita varies relatively widely between FUAs, as the lines plotted within the bars show. For these countries, this observation of relatively wide cross-FUA variation in built-up area per capita holds for both industrial or commercial use and residential land use. This hints at regional variation in planning cultures or land market structures that are consistent across uses. However, in some countries variation in built-up area per capita is particularly wide either for residential built-up areas or for industrial and commercial built-up areas.

Figure 14 gives a further impression of the extent to which FUAs in different countries are relatively residential, or relatively industrial and commercial, in terms of their built-up area composition. This shows that countries including Turkey, Slovakia, Estonia as well as Iceland and Spain are characterized by relatively industrial and commercial FUA-surfaces. The surfaces of FUAs in Norway, Austria, Belgium or Poland, on the other hand, are relatively more of a residential nature. Latvia, the Czech Republic, and the United Kingdom are positioned at a ratio around the OECD average (for the observed European FUAs).

Figure 14. Urban built-up area distribution (2021) over residential and industrial or commercial uses.



In Figure 15 and Figure 16, built-up area per capita is now further disaggregated. First by cities versus commuting zones and subsequently by FUA-size category. This is necessary to appropriately interpret the built-up area per capita, in the light of standard economic explanations of how intensively or extensively land is used (see, for instance, Alonso [1960], Bertaud and Renaud [1998], or Evans [2008]).

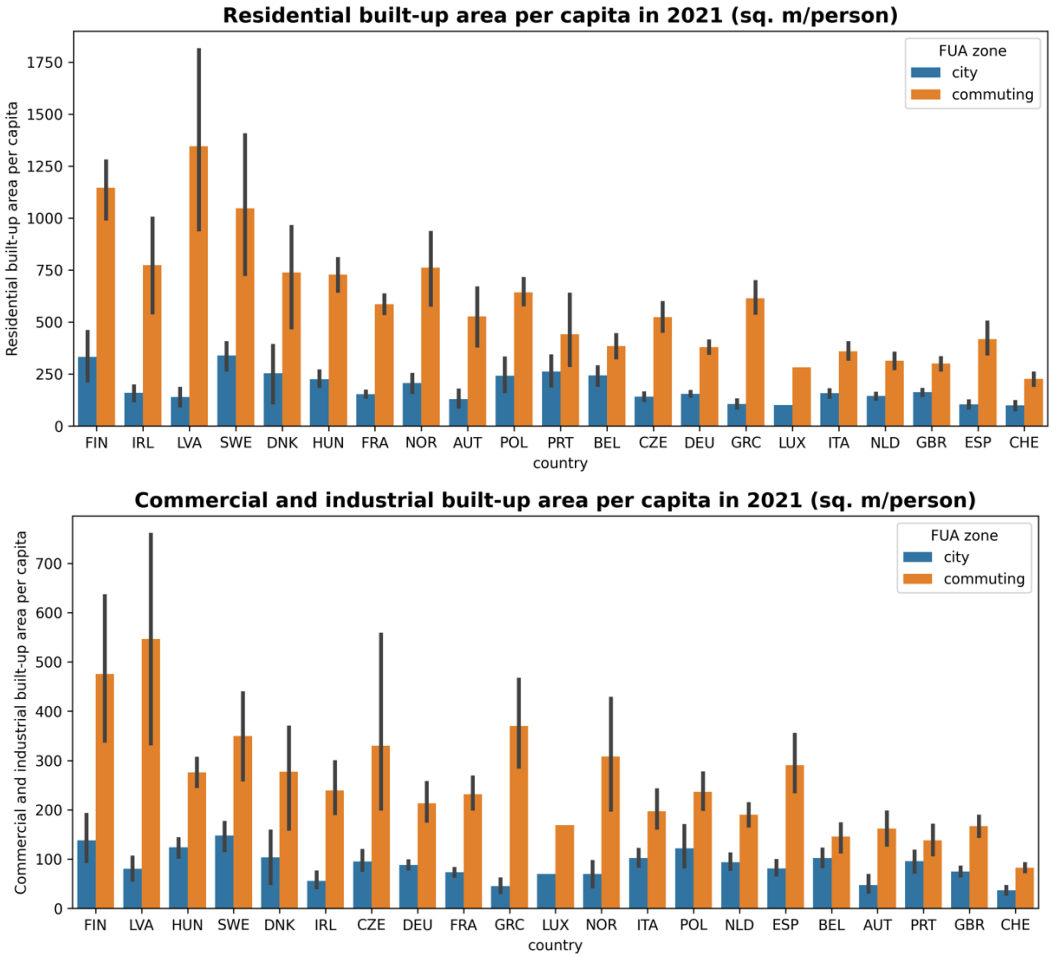
Figure 15 shows how commuting zones drive up the average built-up area per capita for FUAs (compare Figure 13).[27] The reason for this is that land values in commuting zones tend to be relatively low, in comparison to land values in cities as there competition for space is higher. Therefore, land in commuting zones tends to be used extensively (in low densities). This underlines the importance of using the FUA-based distinction between densely populated cities and their commuting zones in examining built-up area.

Several of the countries (e.g., Ireland and Latvia) for which in Figure 13 some of the largest built-up areas per capita were observed (based on the surface and population of entire FUAs), are now shown in Figure 15 to actually have cities with relatively low built-up area per capita; this means that in those countries, the relatively large FUA-wide built-up area per capita outcomes are primarily driven by spread-out land use in

[27] For some countries no commuting zones are observed. For Iceland, the commuting zone's built-up area is omitted as population estimates from the JRC GHSL-POP layer appeared to be inconsistent with satellite-based observations.

commuting zones (rather than in cities). Noteworthy is that large relative gaps in built-up area per capita outcomes for cities and commuting zones are observed not only for countries with the largest of FUA-wide built-up areas per capita. Similar relative gaps are observed for countries of various levels of built-up area per capita, including Norway, Greece, and Spain.

Figure 15. Urban built-up area disaggregated by the FUA-definition of cities and commuting zones.



Next, in line with theoretical expectations, Figure 16 shows that built-up area per capita tends to be smaller for larger FUAs. The reason for this is that competition for space by households and firms leads land to be used more intensively in larger cities, which are thus associated with relatively compact built-up areas. However, also amongst the largest of FUAs there is substantial variation in residential or commercial and industrial built-up area per capita.

Figure 16. Urban built-up area by Functional Urban Area size-category.

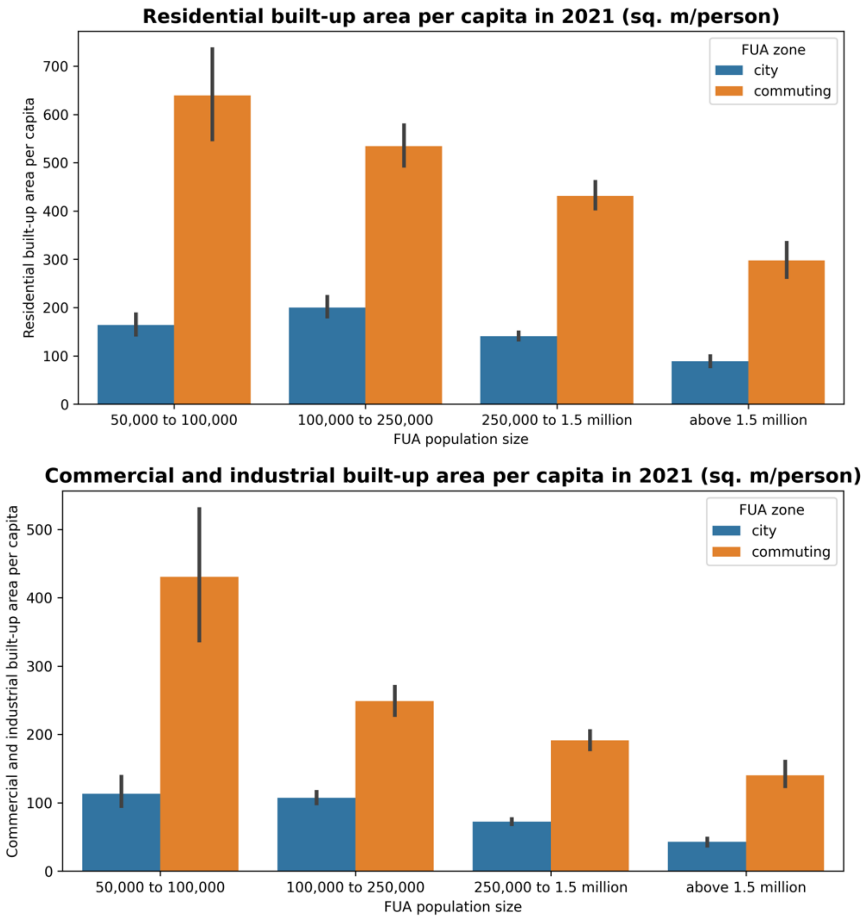


Table 5 and Table 6 show the large metropolitan FUAs that in 2021 rank in the separate top-10's in terms of smallest and largest built-up areas per capita (by land use type), respectively. These results suggest a North-South divide amongst large metropolitan FUAs (n = 41), where Southern FUAs tend to have smaller built-up areas per capita than Northern FUAs. This observation holds for both residential and industrial or commercial land uses. Outcomes for all European FUAs are mapped Figure 17. Similar top-10 rankings and maps, but based on land use at the scale of cities instead of metropolitan areas as a whole may be observed in Annex E.

Table 5. Top-10 metropolitan areas (cities and commuting zones) with the smallest built-up area per capita, by land use type.

Rank	Metropolitan Area	Country	Residential Built-up area per capita (m^2)		Rank	FUA	Country	Industrial or Commercial Built-up area per capita (m^2)
1	Istanbul	Turkey	28.7		1	Istanbul	Turkey	19.6
2	Izmir	Turkey	37.4		2	Athens	Greece	28.3
3	Bursa	Turkey	45.4		3	Bursa	Turkey	34.5
4	Gaziantep	Turkey	58.8		4	London	United Kingdom	35.7
5	Ankara	Turkey	60.6		5	Izmir	Turkey	38.0
6	Barcelona	Spain	76.0		6	Barcelona	Spain	43..6
7	Madrid	Spain	83.6		7	Naples	Italy	48.5
8	Valencia	Spain	103.3		8	Ankara	Turkey	54.2
9	Athens	Greece	109.9		9	Paris	France	58.0
10	Milan	Italy	114.2		10	Madrid	Spain	61.3

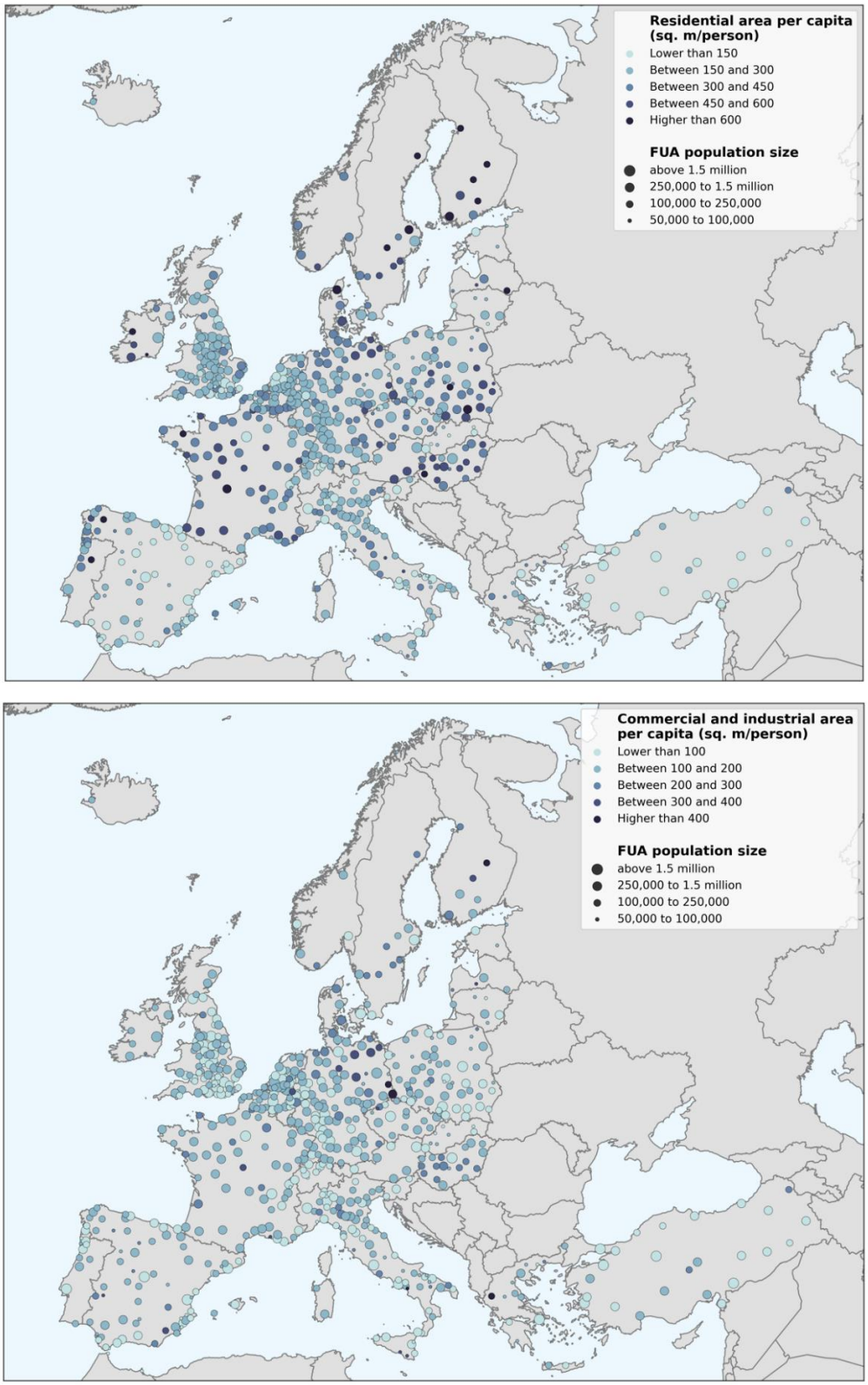
Notes: Metropolitan areas are in this table defined by the subset of Functional Urban Areas that categorize as large metropolitan (n = 41).

Table 6. Top-10 metropolitan areas (cities and commuting zones) with the largest built-up area per capita, by land use type.

Rank	Metropolitan Area	Country	Residential Built-up area per capita (m^2)		Rank	FUA	Country	Industrial or Commercial Built-up area per capita (m^2)
1	Warsaw	Poland	297.8		1	Dublin	Ireland	133.9
2	Stockholm	Sweden	295.4		2	Berlin	Germany	121.1
3	Brussels	Belgium	279.7		3	Rotterdam	Netherlands	101.1
4	Dublin	Ireland	271.0		4	Katowice	Poland	99.0
5	Copenhagen	Denmark	256.4		5	Vienna	Austria	98.0
6	Budapest	Hungary	250.3		6	Glasgow	United Kingdom	96.8
7	Lyon	France	241.4		7	Copenhagen	Denmark	94.1
8	Hamburg	Germany	226.3		8	Hamburg	Germany	92.4
9	Prague	Czech R.	221.5		9	Prague	Czech R.	90.9
10	Vienna	Austria	216.8		10	Budapest	Hungary	86.0

Notes: Metropolitan areas are in this table defined by the subset of Functional Urban Areas that categorize as large metropolitan (n = 41).

Figure 17. Built-up area per capita (2021) in European FUAs, by land use type.



Box 3. Timely Tracking of Urban Expansion (And Its Speed and Shape)

The U-Net deep learning model can also be applied to track changes in land use over time. A first illustration can be given below. First, a methodological note is provided, and then followed by an application of tracking and categorizing urban expansion across large European metropolitan areas.

Changes could be tracked using several methodological approaches. For instance, change could be detected by comparing each satellite image pixel and estimate which land use is present there with the highest probability for 2018 and 2021 imagery separately, and then compare classifications. This, however, is a noisy procedure. Consider a pixel that could potentially be classified as one of two likely classes, as these classes have very close probabilities in the output probability tensor for two points in time. This pixel could then be classified differently, even if there has been no land use change, if ground reflectance changes only slightly over time. Some of such noise may be removed from the land use change estimation by applying a sieving operation, but that would be a limited solution.

A second, more stable, method is to compare the probability tensors for the two points in time i and f, and to compute for class k the probability difference $p_{k,f} - p_{k,i}$. If this difference is above a certain threshold, this is taken as an indication that there is an expansion of class k. Figure 18 shows such probability-difference based signals of expansion for a sub-centre within the FUA of Dublin (Ireland), as an illustration. This illustration suggests a relatively large-scale and spatially concentrated expansion of residential areas in the observed settlement.

Figure 18. Urban expansion (2018-21) tracked in a settlement within the Dublin metropolitan area, based on probability-differences in the model's assignment of land use types to satellite image pixels.

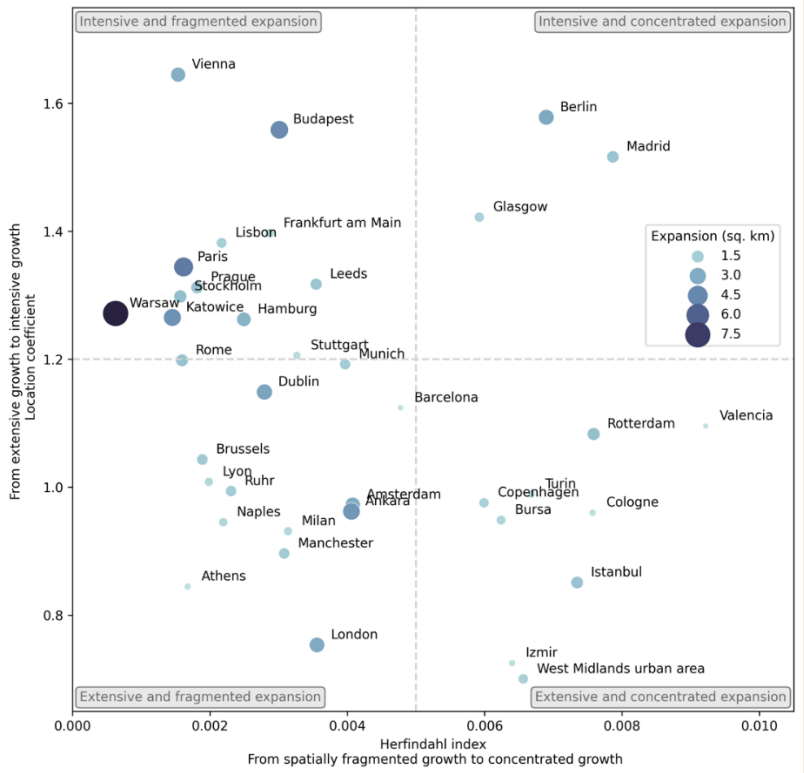
Systematic analysis of land use change, however, requires reference data for both the initial and the final period of observation, which in the case of Urban Atlas is unavailable for the present study. Also,

to appropriately quantify accuracy, evaluation procedures should be tailored to account for the potential subtleties of changes in spectral profiles when urban land is converted from one use to another. Another possible source of temporal variation in mapping accuracy could be changes in the luminosity or contrast of images. Detailed analysis of land use change would thus warrant additional accuracy assessment.

In Figure 19, European metropolitan areas are illustratively categorized in terms of the *speed* and *shape* of urban expansion. The figure's *y*-axis plots whether expansion of built-up areas takes place in FUA sub-areas that are intensively or extensively developed. This *signals* whether a FUA moves towards compaction or dispersion. The underlying measure captures for each individual expansion-site (red areas in Panel A) the share of residential, industrial, and commercial land, within a 1-km radius, amongst all developable land, and divides this share by such ratio captured at FUA-level; finally, site-specific values are averaged at the FUA-level. The *x*-axis plots a Herfindahl index, which captures the degree to which FUA urban expansion is concentrated in few or in many individual sites. The speed of expansion is measured from the expansion's areal surface relative to the total surface of the observed FUA.

Based on the metrics outlined above, the shape and speed of urban expansion can be consistently compared across OECD metropolitan areas in a single overview. For instance, urban expansion in Vienna is relatively intensive, taking place close to or in existing built-up areas, yet fragmented across many sites, whereas Istanbul's urban surface expanded more extensively but across relatively few sites.

Figure 19. The speed and shape of the urban expansion (2018-21) of OECD metropolitan areas.



Conclusion

Whether today's cities become more compact or more spread out will have a lasting influence on the economic efficiency, resilience, and sustainability of life in cities (OECD, 2018a; 2018b; 2019). Many processes regarding mobility, carbon emissions, resource consumption, housing affordability, people's access to services, infrastructure costs, or the ease of social and business interactions, may be successfully or less successfully facilitated by a metropolitan area's physical shape and expansion. This highlights the importance of this paper for the timely monitoring of urban land use across OECD countries.

This study developed an efficient method for the near real-time monitoring of land use in OECD cities, on an internationally consistent basis. As a case application, land use in 687 European metropolitan areas was examined for the year 2021. The large spatial scope of the study was supported by combining imagery from an innovative constellation of satellites, operated by the European Commission and ESA, with an established machine learning model for processing such imagery. In specific, a deep learning model in a U-Net architecture was used to map key economic land uses in cities.

Overall, the model was found to detect and distinguish residential and industrial or commercial uses of land at a high level of accuracy. As any mapping exercise involves some degree of error, the accuracy of results was quantified at the level of individual metropolitan areas. Such reporting ensures that analytical insights can be appropriately obtained at a policy-relevant scale of analysis. This exercise further showed that the accuracy of urban land use mapping clearly varies with built-up area density, which has not been accounted for yet in other studies.

The findings quantify how the compactness of land use in cities varies strongly across European countries. Across the observed countries, whether measured from residential or from industrial and commercial uses, the built-up area per urban inhabitant varies between countries by up to a fivefold. Large variation stems from how compact or spread-out the use of (built-up) land is in the commuting zones that surround central cities. A related finding is that in some countries the industrial or commercial use of land has noticeably more weight in the metropolitan land surface than in other countries.

As an extension of the main analysis, the speed and shape of urban expansion over 2018-2021 was explored. These results showed major differences, even within Europe's low-population growth environment. Metropolitan areas grew inward or outward to a varying extent and, similarly, substantial differences are observed in whether growth is concentrated in few locations or fragmented over many sites across the metropolitan area's surface. This raises the question of how the development of built-up areas in European metropolitan areas might compare to other world regions where urban growth rates may vary even more. The current model, which was trained solely on European data, was qualitatively indicated to transfer well to other world regions. Monitoring land use also for non-European countries would require an efficient and harmonized approach to collecting further data for model training and evaluation purposes. In this light, the present model may offer a methodological basis for indicators that are even more centred on the monitoring of specific forms of economic development in cities, to capture early signals of growth or change.

References

Alonso, W. (1960), “A theory of the urban land market”, *Papers in Regional Science* 6 (1), 149–157.

Anderson J. R., Hardy E. E., Roach J. T., Witmer R. E. (1976), “*A land use land covers classification system for use with remote sensor data*”, Geological Survey Professional Paper 964.

Badrinarayanan, V., Kendall, A. and R. Cipolla (2017), “SegNet: A deep convolutional encoder-decoder architecture for image segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 2481-2495.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A. and E. F. Wood (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Nature: Scientific Data*, 5 (1), 1–12.

Bertaud, A., and B. Renaud (1997), “Socialist cities without land markets”, *Journal of Urban Economics* 41 (1), 137–151.

Braaten, J. (2021), “Sentinel-2 Cloud Masking with s2cloudless”, https://github.com/google/earthengine-community/blob/master/tutorials/sentinel-2-s2cloudless/index.ipynb.

Buchhorn, M. et al. (2020), Copernicus Global Land Service: Land Cover 100m: Collection 3: Globe (V3.0.1).

Chen, L. et al. (2018), “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 834–848.

Corbane, C. et al. (2020), “Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery”, Neural Computing and Applications 33, 6697–6720.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., and M. C. Hansen (2018), Classifying drivers of global forest loss. *Science* 361 (6407), 1108–1111.

Dijkstra, L., Poelman, H., and P. Veneri (2019), “*The EU-OECD definition of a functional urban area*”, OECD Regional Development Papers. Paris: OECD Publishing.

Evans, A.W. (2008), “*Economics, Real Estate and the Supply of Land*”, John Wiley & Sons.

Food and Agriculture Organization of the United Nations (2016), “*Map Accuracy Assessment and Area Estimation: A Practical Guide*”. Rome: FAO.

Gbodjo, Y. et al. (2020), “Object-Based Multi-Temporal and Multi-Source Land Cover Mapping Leveraging Hierarchical Class Relationships”, Remote Sensing 12.

Giang, T. et al. (2020), “U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV imagery”, *IEEE Access* 8, 186257–186273.

Gong, P., Li, X., and W. Zhang (2019), “40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing”, *Science Bulletin*, 64 (11), 756-763.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and R. Moore (2017), “Google Earth Engine: Planetary-scale geospatial analysis for everyone” *Remote Sensing of Environment* 202, 18–27.

Helber, P. et al. (2017), “EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification”, CoRR.

Huang, G. et al. (2016), “Densely Connected Convolutional Networks”, CoRR.

Iglovikov, V. and A. Shvets (2018), “TernausNet: U-Net with VGG11 Encoder Pre-Trained on imagenet for Image Segmentation”, CoRR.

Jeong, J., T. Yoon and J. Park (2018), “Towards a meaningful 3D map using a 3D lidar and a camera”, *Sensors* 18, 2571.

Karra, K. et al. (2021), “*Global land use / land cover with Sentinel 2 and deep learning*”, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 4704–4707.

Krizhevsky, A., I. Sutskever and G. Hinton (2012), "*ImageNet classification with deep convolutional neural networks*", NeurIPS Proceedings.

Kumar, S. (2018), *Deep U-Net for satellite image segmentation*, https://github.com/reachsumit/deep-unet-for-satellite-image-segmentation.

LeCun, Y. et al. (1990), "*Handwritten digit recognition with a back-propagation network*", Advances in Neural Information Processing Systems, Vol. 2.

LeCun, Y., Bengio, Y., and G. Clinton (2015), "Deep learning", *Nature* 521, 436–444.

Li, X., Zhou, Y., Gong, P., Seto, K. C., and N. Clinton (2020), "Developing a method to estimate building height from Sentinel-1 data", *Remote Sensing of Environment* 240, 111705.

Long, J., E. Shelhamer and T. Darrell (2014), "Fully convolutional networks for semantic segmentation", CoRR.

Montero, E., Van Wolvelaer, J. and A. Garzón (2014), "The European Urban Atlas", In: *Land use and land cover mapping in Europe* (115–124). Dordrecht: Springer.

OECD (2017), "*The Governance of Land Use in OECD Countries*", Paris: OECD Publishing.

OECD (2018), "*Rethinking Urban Sprawl: Moving Towards Sustainable Cities*", OECD Publishing, Paris.

OECD (2018), "Climate-resilient infrastructure", *OECD Environment Policy Papers*, No. 14, OECD Publishing, Paris, https://doi.org/10.1787/4fdf9eaf-en.

OECD (2019), *Accelerating Climate Action: Refocusing Policies through a Well-being Lens*, OECD Publishing, Paris, https://doi.org/10.1787/2f4c8c9a-en.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and M. A. Wulder (2014), "Good practices for estimating area and assessing accuracy of land change." *Remote Sensing of Environment* 148, 42–57.

Ronneberger, P.O. (2015), "U-Net: Convolutional networks for biomedical image segmentation", CoRR.

Reba, M., and K. Seto (2020), "A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change", *Remote Sensing of Environment* 242, 111739.

Rosenblatt, F. (1958), "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological Review* 65, 386–408.

Simonyan, K., and A. Zisserman (2015), "*Very deep convolutional networks for large-scale image recognition*", International Conference on Learning Representations.

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., and J. Quinn (2021), "*Continental-scale building detection from high resolution satellite imagery",* Google Research.

Solórzano, J. et al. (2021), "Land use land cover classification with U-Net: Advantages of combining Sentinel-1 and Sentinel-2 imagery", *Remote Sensing* 13, 3600.

Srinivasan, V., Zhang, D., and M. Rezaee (n.d.), "*Land use / land cover classification using ResNet50*".

Stehman, S. V. (2013). "Estimating area from an accuracy assessment error matrix". *Remote Sensing of Environment* 132, 202–211.

Szegedy, C. et al. (2014), "Going deeper with convolutions", CoRR.

Liu, X. Z. (2018), "Recent progress in semantic image segmentation", Artificial Intelligence Review.

UN General Assembly, *Transforming our world: the 2030 Agenda for Sustainable Development*, 21 October 2015, A/RES/70/1, available at: https://www.refworld.org/docid/57b6e3e44.html

Zhang, Z., Liu, Q., and Y. Wang (2017), "Road extraction by deep residual U-Net", *IEEE Geoscience and Remote Sensing Letters*.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., ... and R. F. Auch (2016), "Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative", *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 206–221.

Annex A. Alternative Groupings of Urban Atlas Categories and Associated Confusion Matrices

Table A A.1. Full Urban Atlas typology and relative coverage (%) by FUA in 2018.

Urban Atlas 2018 Typology	Mean	St. dev.
Continuous urban fabric (S.L. : > 80%)	0.9	1.1
Discontinuous dense urban fabric (S.L. : 50% - 80%)	2.3	2.5
Discontinuous medium density urban fabric (S.L. : 30% - 50%)	1.8	1.8
Discontinuous low density urban fabric (S.L. : 10% - 30%)	1.3	1.2
Discontinuous very low density urban fabric (S.L. : < 10%)	0.9	1.1
Isolated structures	0.7	0.6
Industrial, commercial, public, military and private units	3.3	3.0
Fast transit roads and associated land	0.2	0.2
Other roads and associated land	2.0	1.1
Railways and associated land	0.2	0.2
Port areas	0.1	0.4
Airports	0.2	0.4
Mineral extraction and dump sites	0.4	0.7
Construction sites	0.1	0.1
Land without current use	0.2	0.3
Green urban areas	0.6	0.8
Sports and leisure facilities	0.8	0.8
Arable land (annual crops)	28.6	19.0
Permanent crops (vineyards, fruit trees, olive groves)	3.2	9.0
Pastures	17.0	13.7
Complex and mixed cultivation patterns	0.1	0.8
Orchards at the fringe of urban classes	0.0	0.0
Forests	23.3	17.8
Herbaceous vegetation associations (natural grassland, moors...)	8.1	13.8
Open spaces with little or no vegetation (beaches, dunes, bare rocks, glaciers)	0.9	4.2
Wetlands	0.5	1.4
Water	2.2	3.8

Figure A A.1. Confusion matrix for the full set of Urban Atlas categories

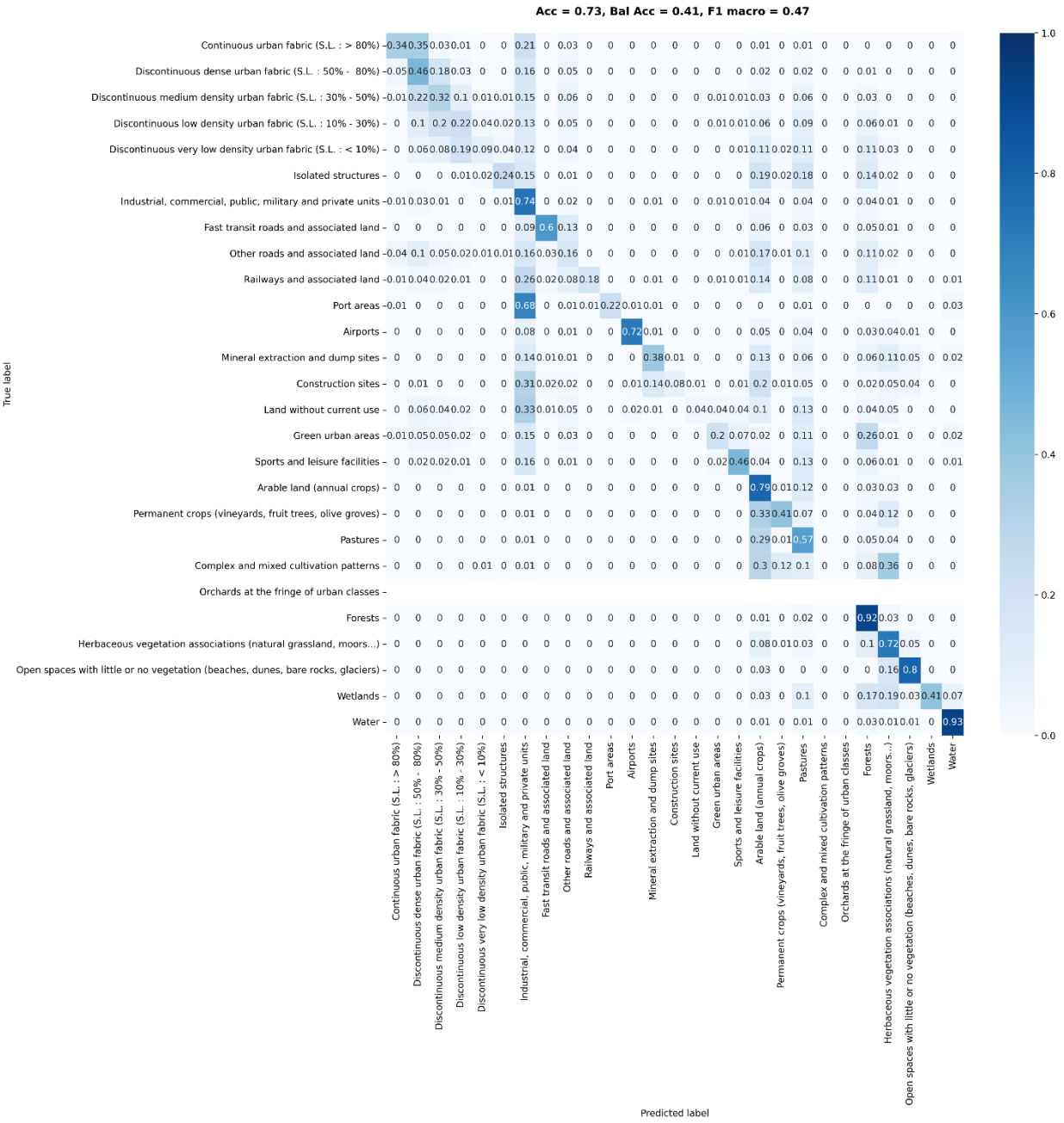


Table A A.2. Aggregation of Urban Atlas categories.

Grouped class	Original class
Urban fabric (low density)	Discontinuous low density urban fabric (S.L.: 10% - 30%), Discontinuous very low density urban fabric (S.L.: < 10%), Isolated structures
Urban fabric (medium density)	Discontinuous dense urban fabric (S.L.: 50% - 80%), Discontinuous medium density urban fabric (S.L.: 30% - 50%)
Urban fabric (high density)	Continuous urban fabric (S.L.: > 80%)
Agricultural	Arable land, Permanent crops, Pastures, Complex and mixed cultivation patterns, Orchards
Water, wetlands	Water, wetlands
Roads and railways	Fast transit roads, railways, other roads, and associated land
Airports	Airports
Ports	Ports
Industrial, commercial, public, military and private units	Industrial, commercial, public, military, and private units
Mine, dump, and construction sites	Mineral extraction and dump sites, Construction sites, Land without current use
Artificial vegetated areas (non-agricultural)	Green urban areas, Sports facilities
Herbaceous area	Herbaceous areas
Forests	Forests
Open spaces	Open space without vegetation (beaches),

Figure A A.2. Confusion matrix for aggregated Urban Atlas typology.

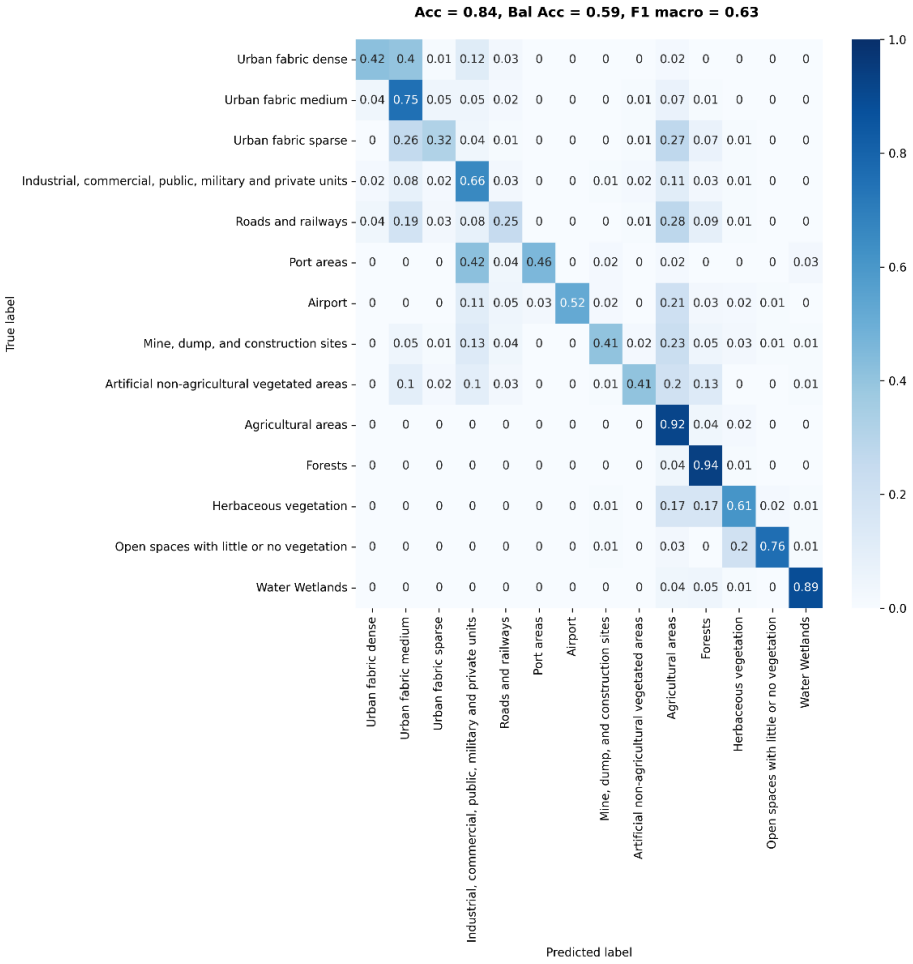
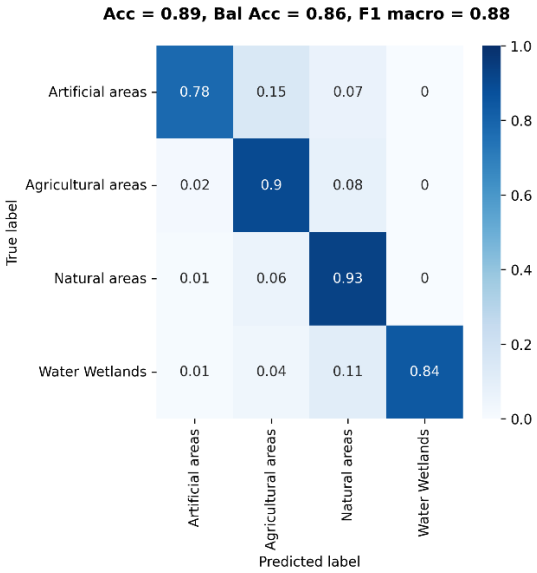


Table A A.3. Further aggregation of Urban Atlas categories.

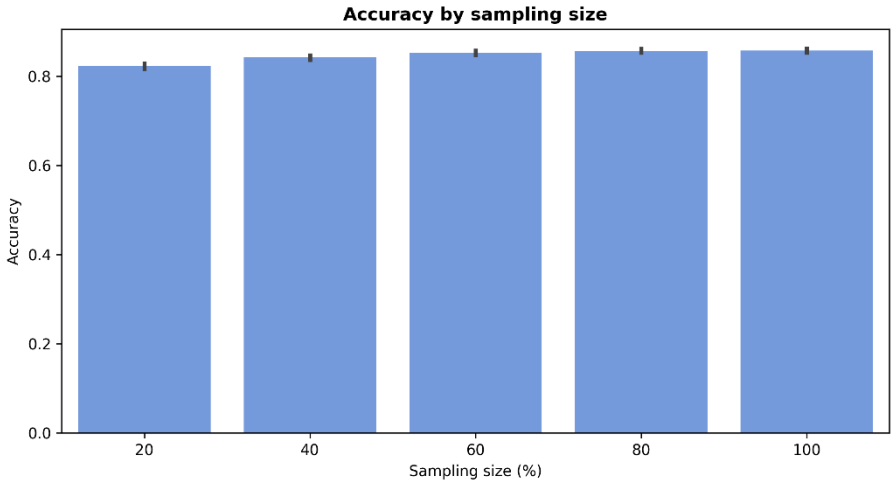
Grouped Class	Original class
Artificial areas	Urban fabric, Industrial, commercial, public, military, and private units, roads, railways, ports, airports, mineral extraction and dump sites, construction sites, lands without current use, green urban areas, sport facilities
Natural areas	Forests, herbaceous areas, open space without vegetation (beaches)
Agricultural areas	Arable land, Permanent crops, Pastures, Complex and mixed cultivation patterns, Orchards
Water, wetlands	Water, wetlands

Figure A A.3. Confusion matrix for further-aggregated Urban Atlas typology



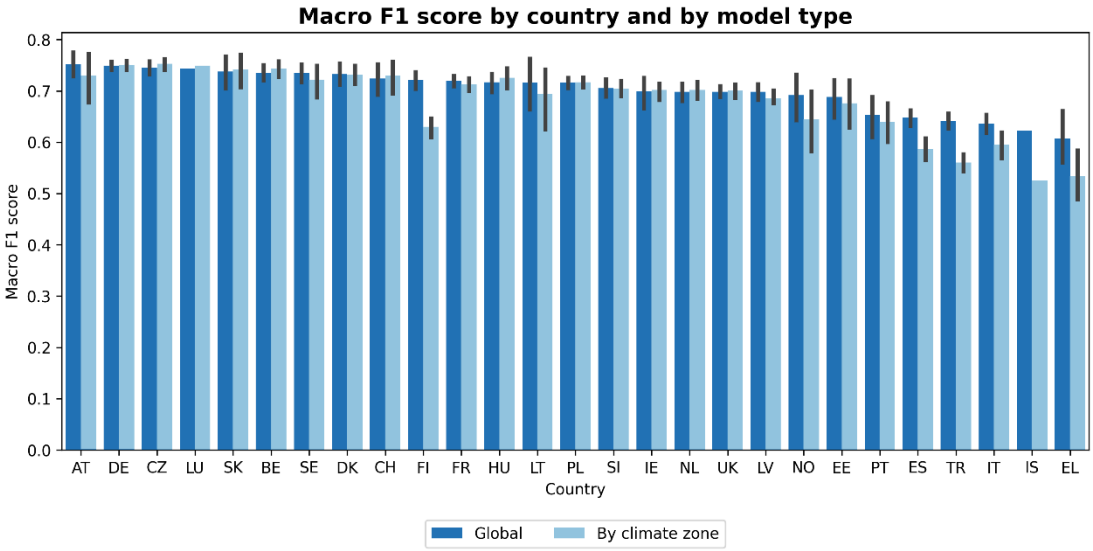
Annex B. Prediction Accuracy by Sample Size

Figure A B.1. Influence of the train set size on the model performance as obtained from the mean of FUA-by-FUA balanced accuracy values.



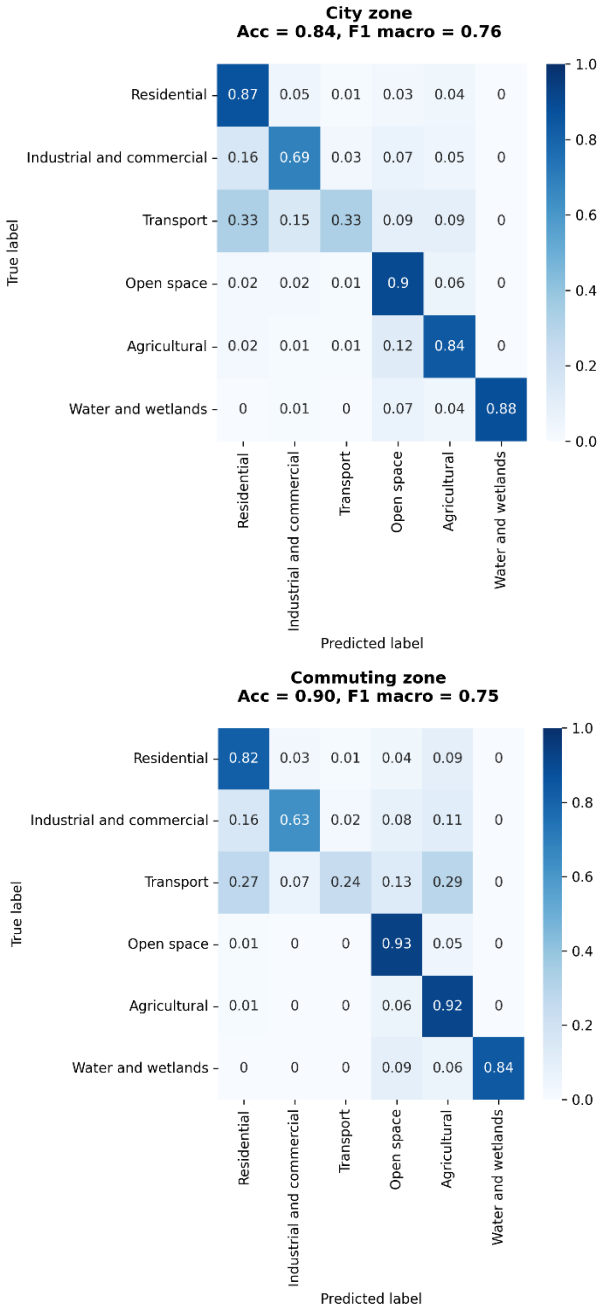
Annex C. Results for Models by Climate Zone

Figure A C.1. Results for a 'global' U-Net using all training data and for U-Nets trained by climate zone.



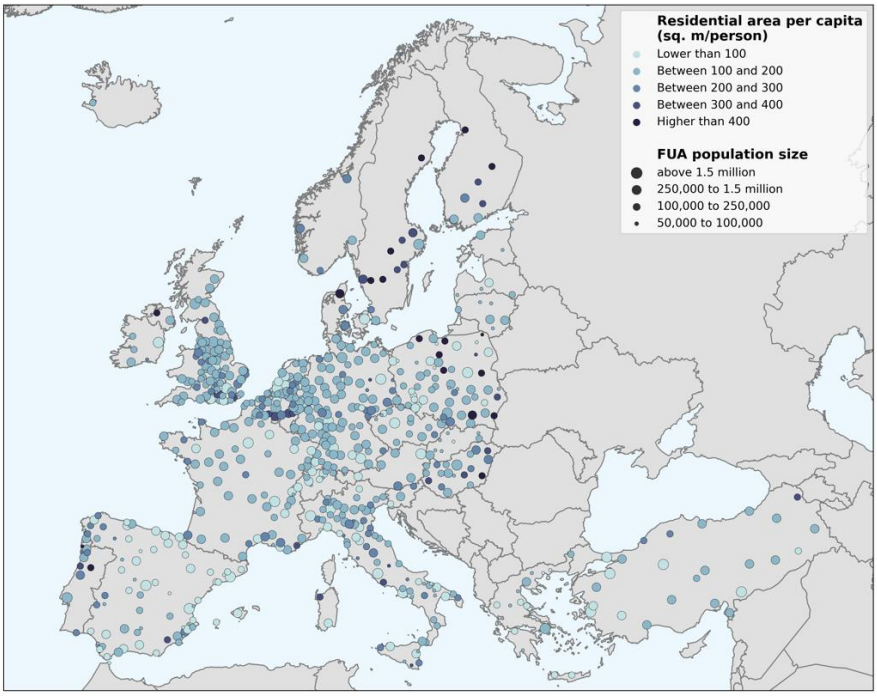
Annex D. Does Accuracy Vary Within FUAs?

Figure A D.1. Normalised confusion matrices obtained for FUA cities (top) and commuting zones (bottom)



Annex E. Maps of built-up area per capita by city

Figure A E.1. Urban area per capita by city, 2021.



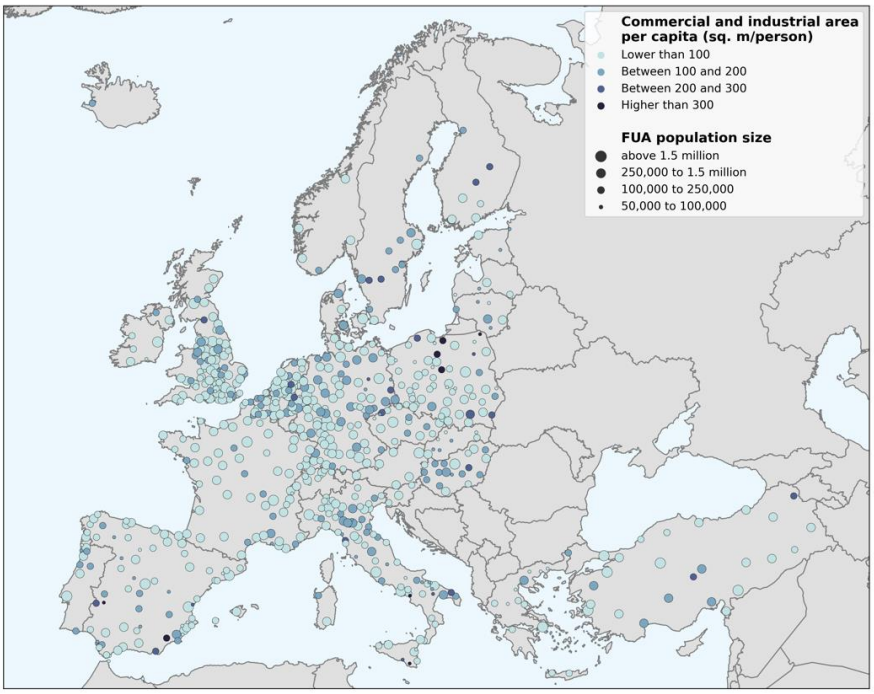


Table A E.1. Top-10 cities with the smallest urban built-up area (relatively compact urban form) by land use type.

Rank	Metropolitan Area	Country	Residential Built-up area per capita (m^2)	Rank	FUA	Country	Industrial or Commercial Built-up area per capita (m^2)
1	Istanbul	Turkey	28.7	1	Istanbul	Turkey	19.6
2	Barcelona	Spain	35.6	2	Copenhagen	Denmark	20.9
3	Izmir	Turkey	37.4	3	Barcelona	Spain	22.9
4	Madrid	Spain	40.3	4	London	United Kingdom	25.0
5	Valencia	Spain	43.9	5	Vienna	Austria	25.9
6	Bursa	Turkey	45.4	6	Munich	Germany	26.6
7	Turin	Italy	51.0	7	Valencia	Spain	27.4
8	Gaziantep	Turkey	58.8	8	Athens	Greece	28.3
9	Ankara	Turkey	60.6	9	Paris	France	28.5
10	Copenhagen	Denmark	63.1	10	Warsaw	Poland	30.5

Note: Metropolitan areas are in this table defined by the subset of Functional Urban Areas that categorize as large metropolitan (n = 41).

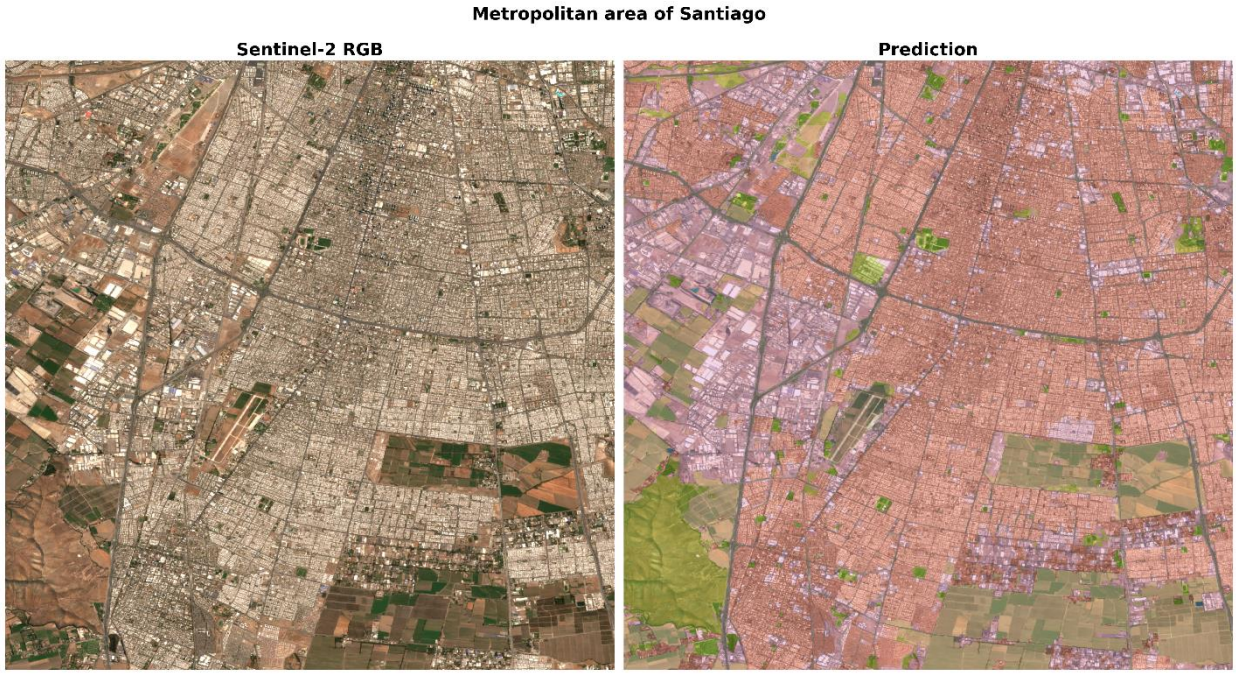
Table A E.2. Top-10 cities with the largest urban built-up area (relatively dispersed urban form) by land use type.

Rank	Metropolitan Area	Country	Residential Built-up area per capita (m^2)	Rank	FUA	Country	Industrial or Commercial Built-up area per capita (m^2)
1	Brussels	Belgium	199.1	1	Gaziantep	Turkey	79.1
2	Leeds	United Kingdom	145.0	2	Glasgow	United Kingdom	75.5
3	Ruhr	Germany	139.4	3	Katowice	Poland	68.8
4	Glasgow	United Kingdom	135.8	4	Rotterdam	Netherlands	63.4
5	West Midlands	United Kingdom	135.6	5	Brussels	Belgium	63.3
6	Manchester	United Kingdom	122.6	6	Ruhr	Germany	68.0
7	Hamburg	Germany	121.6	7	Leeds	United Kingdom	61.8
8	Athens	Greece	109.9	8	Düsseldorf	Germany	57.8
9	Budapest	Hungary	106.9	9	Amsterdam	Netherlands	57.8
10	Stockholm	Sweden	104.7	10	Cologne	Germany	56.0

Note: Metropolitan areas are in this table defined by the subset of Functional Urban Areas that categorize as large metropolitan (n = 41).

Annex F. Estimated Land Use Maps for OECD Metropolitan Areas Beyond Europe

Figure A F.1. Satellite images and estimated land uses for selected non-European metropolitan areas.



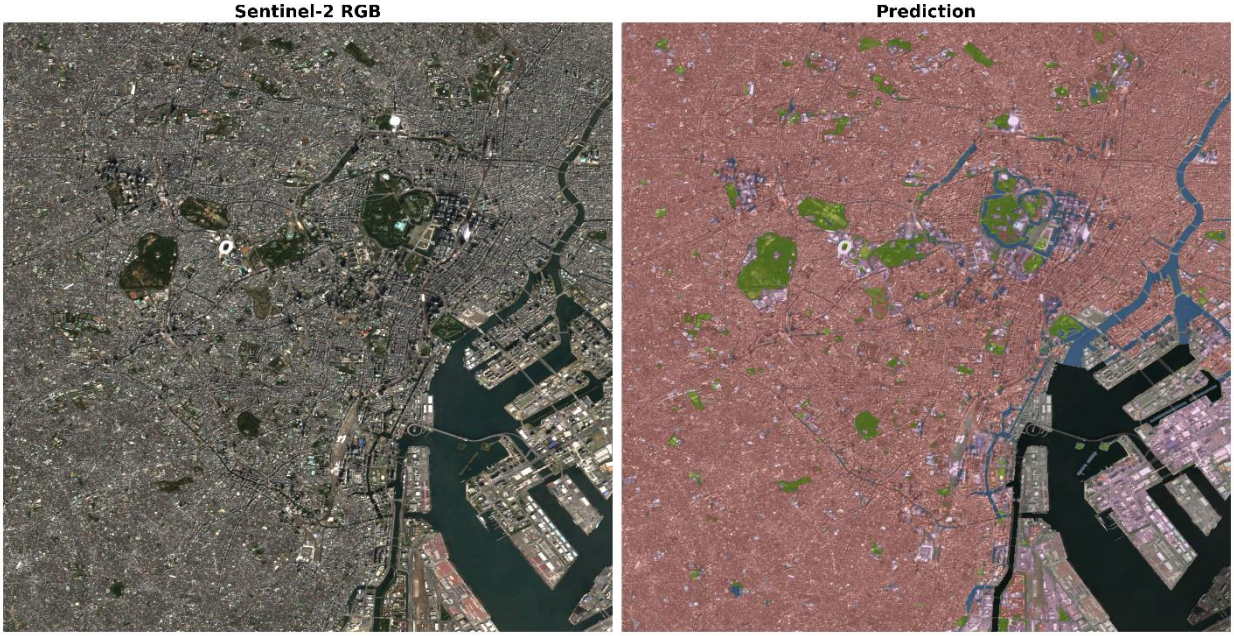

Metropolitan area of San Francisco
Sentinel-2 RGB
Prediction
Metropolitan area of Seoul
Sentinel-2 RGB
Prediction

Metropolitan area of Mexico City

Sentinel-2 RGB | **Prediction**

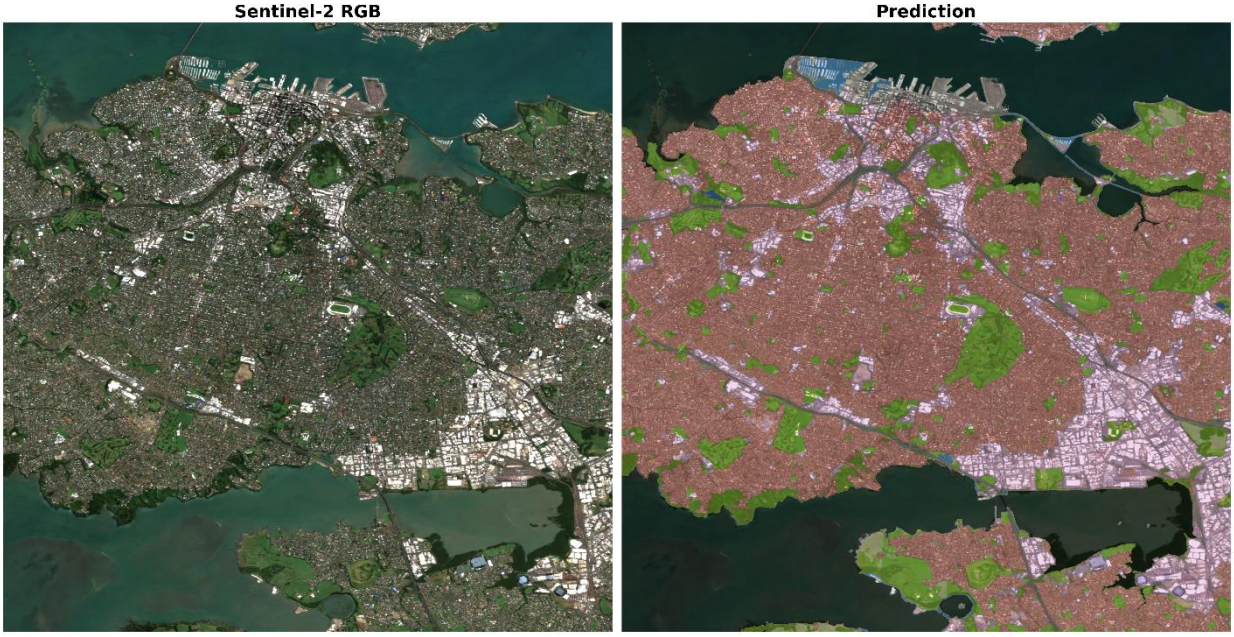
Metropolitan area of Tokyo

Sentinel-2 RGB | **Prediction**



Metropolitan area of Auckland

Sentinel-2 RGB **Prediction**



Metropolitan area of Bogota D.C.

Sentinel-2 RGB **Prediction**