Solar Energy 265 (2023) 112094

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

ETHOS.PASSION: An open-source workflow for rooftop photovoltaic potential assessments from satellite imagery

Rodrigo Pueblas [a], Patrick Kuckertz [a], Jann Michael Weinand [a,*], Leander Kotzur [a], Detlef Stolten [a,b]

[a] Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research – Techno-economic Systems Analysis (IEK-3), 52425 Jülich, Germany
[b] RWTH Aachen University, Chair for Fuel Cells, Faculty of Mechanical Engineering, 52062 Aachen, Germany

ARTICLE INFO

Dataset link: https://github.com/FZJ-IEK3-VSA/PASSION

Keywords:
Rooftop photovoltaics
Potential assessment
Economic assessment
Computer vision
Deep learning
Open-source workflow

ABSTRACT

The assessment of rooftop photovoltaic potential has become increasingly accurate due to the expanding availability of satellite imagery and improvements in computer vision methods. However, the analysis of satellite imagery is impeded by a lack of transparency, reproducibility, and standardized description of the methods employed. Studies are heterogeneous, target different types of potential with redundant efforts, and are mostly not open source or use private datasets for training. With respect to the estimation of photovoltaic potential, this study proposes a conceptual frame of reference for clearly identifying tasks, their relationships, and their data. Additionally, the open-source workflow ETHOS.PASSION is introduced, which integrates the assessment of geographical, technical and economic potentials of regions under consideration along with the calculation of surface areas, orientations and slopes of individual rooftop sections. ETHOS.PASSION also includes the detection of superstructures, i.e., obstacles such as windows or existing photovoltaic installations. The novel two-look approach combines two deep learning models identifying rooftops and sections, and an additional model for the identification of superstructures. The three models show a mean Intersection Over Union between classes of 0.8478, 0.7531 and 0.4927 respectively, and more importantly display consistent results amongst randomly sampled real world images. The final results are evaluated for multiple datasets and compared against other studies, with a case study in the Aachen region of Germany being presented.

1. Introduction

Utility-scale solar photovoltaics (PV) and wind projects underpin the current growth in the use of renewable energy sources, according to the annual World Energy Outlook 2022 [1]. Deploying more renewable energy technologies also on rooftop areas is highly relevant in the context of the current global energy crisis. More self-sufficient living spaces would reduce the need for centralized suppliers and imports, and align with Europe's vision of achieving net-zero carbon emissions. Achieving net-zero emissions require for rooftop PV to be deployed at scale, being the source with the least visual impact and land sealing. In order to tap the great rooftop PV potential in the coming years, it is necessary to provide model-based estimations of this potential with a high degree of regional resolution and accuracy. Early approaches in the literature tackled the potential assessment by extrapolating the known or calculated data of a smaller region to larger areas [2,3], sometimes accounting for factors such as population density or cadastral information. Other approaches use high resolution light detection and ranging (LiDAR) data to perform micro-simulations [4], which is difficult to scale due to limited data availability. An approach that has gained popularity in recent years is the use of satellite imagery to estimate the potential. This is not a new research field, but the improvements in satellite resolution and computer vision algorithms have made it even more accessible and accurate [5–7]. Nonetheless, there is still a lack of open projects that provide reproducible results, and the need for tools that users of all technical levels can utilize is apparent. Many studies overlap and have different areas of focus, but it is currently difficult to integrate them and identify the state of the domain. A focus on creating reusable software that follows the FAIR (findability, accessibility, interoperability, and reusability) principles [8,9] is needed for the field to increase in quality.

Depending on the level of detail of the study, four types of potential can be defined [10–12]. The theoretical potential refers to the total radiation that is received in a region according to the solar incidence, without considering any other limitations. The geographical potential constraints the region of interest to the spatial fraction that is relevant to the type of study, which in this case would be all suitable building

* Corresponding author.
E-mail address: j.weinand@fz-juelich.de (J.M. Weinand).

https://doi.org/10.1016/j.solener.2023.112094
Received 7 July 2023; Received in revised form 22 September 2023; Accepted 7 October 2023
Available online 31 October 2023

rooftops in the region. The technical potential refers to the share of usable potential after taking into account all technical constraints, such as the number of panels to be installed or the estimated efficiency of the system. Finally, the economic potential translates the technical potential from power into economic terms. For this, one popular metric is the levelized cost of electricity (LCOE), which measures the average cost per unit of electricity after taking all associated parameters into account. While we use these established potential definitions, it must also be mentioned here that there are other definitions as well. For example, Bódis et al. [13] divide potential only into resource potential (comparable to theoretical potential), technical potential and economic potential. In this case, the technical potential also includes the geographical constraints.

Studies can also be divided according to the smallest spatial area to which the data is available. Low resolution studies use regional aggregated data, such as that of Mainzer et al. [14], which used data from 11594 German municipalities, Ohtake et al. [15], which used data from 50 radiance monitoring stations in Japan, or Gutschner et al. [16], which analyzed the potential of Europe at a country level resolution. High resolution studies, on the other hand, analyze individual building information, and are meant to provide a higher level of detail. These studies can obtain their data from diverse sources, such as three-dimensional scans [17], cadastere data [13], crowd-sourced data [18] or satellite data [19].

Studies reported in the literature also vary in the size of the investigated region, ranging from a group of buildings or a small district [20], to very few country scale articles [14]. Other studies propose an approach that can be applied to an entire city, such as Sampath et al. [19], which is comparable to the scope of the present work. Overall, the two main factors that can influence the potential scope of a study are data availability and computing resources.

Three different methodology classifications for PV potential assessments can be found in the literature based on different types of data [21]. *Sample methodologies* extrapolate the estimation of available rooftop areas in a small region to a larger one. *Multivariate sampling-based methodologies*, while still sample-based, look for proxy variables to draw a correlation between them and the estimation in order to account for a portion of the variance. Finally, *complete census methodologies* perform their estimations based on specific building characteristics. Castellanos et al. [22] also define three different levels to differentiate the methodologies. *Low-level approaches* assume high data homogeneity and directly use other variables such as population or building density to infer rooftop areas. *Medium-level approaches* combine this statistical data with geographical information systems (GIS) or LiDAR data sources. *High-level approaches* use advanced rooftop detection with high resolution spatial information, and aim to account for azimuth, slope, shape and other important characteristics. Finally, Song et al. [17] differentiated two main PV potential methods: on the one hand, *generalized estimations* seek to calculate the potential distribution considering the influence of environmental, economic, social, or other factors. *Detailed modeling approaches*, on the other hand, focus on GIS instead, presenting more accurate but also complex results.

The approaches based on satellite imagery are not novel, and computer vision (CV) has been used for many years. Early approaches to building extraction, like those of Saeedi and Zwick [23] or Liu et al. [24], use a set of classic image processing techniques and rules. In Singh and Banerjee [25], QGIS [26] was used to estimate the total area in a given georeferenced image. In recent years, deep learning approaches have risen in popularity: in Song et al. [17], satellite imagery was combined with digital surface model (DSM) data to estimate the potential of the Chao Yang District of Beijing, China. Lee et al. [20] proposed an approach that segments rooftop sections directly, but did not specify details with regard to their training data. Ma et al. [27] presented a multi-scale encoder–decoder network to extract buildings, but did not focus on PV potential assessments. In Phap et al. [28], the authors evaluated the technical PV potential of Hanoi, Vietnam, by combining segmentation and classification methodologies. Krapf et al. [29] presented the foundations of a segmentation work that was later further expanded [6]. The latter is an exemplary work, which is transparent in terms of both data and software. Zhong et al. [30] used a segmentation approach with interesting concepts with respect to the pre-processing of satellite images. Other studies, like those of Chen et al. [7], Jiang et al. [31], and Sun et al. [5], incorporate the semantic segmentation of satellite images as part of their workflows.

The mentioned articles like Sun et al. [5], Chen et al. [7], Phap et al. [28], Ma et al. [27] or Sampath et al. [19] all use satellite images as the data source to identify rooftops, with varying methodologies. However, there is still a lack of reproducibility and standardized descriptions in the field with only 17% of papers analyzed by Ren et al. [32] publishing their code, and most of the articles referring only vaguely to their labeled data sources. The problem of the standardized description of methods is relevant due to studies performing different analyses being constantly compared to each other. The existence of a common reference framework would ease the comparison of individual steps and how they influence the calculations.

Model architectures differ very little between studies, and the differentiating points are primarily the quality of the training data, and the re-usability of the software. In this regard, Krapf et al. [6] contributed a novel dataset, which they published along with their software project. The quality of the dataset and its documentation is noteworthy, but they do not focus on providing their own workflow for technical potential. The objective of this study is to address the lack of transparency, reusability and standardized description of methods in the field of PV potential assessments using satellite imagery. Thereby, the present paper offers the following contributions:

Standardized description of methods. We define a conceptual framework that aims to standardize the description of tasks in the research field along with their relationships to each other and to relevant data contents. It provides an implementation-independent point of reference that reduces ambiguity in academic information exchange and increases the comparability of methods, software, and data.

Open software. We provide an open-source workflow that aims to open this research field to others working in other energy systems domains. This software, called PhotovoltAic Satellite SegmentatION (PASSION),[1] which is part of the Energy Transformation patHway Optimization Suite (ETHOS), aims to be more robust, easier to use, and have a wider scope that is actually usable for researchers or users from other fields. Users can apply the software to any region for which source data is available. The current state of the project allows for city-scale estimation, with the potential of becoming even more efficient by incorporating computing parallelization and inferences.

Two-look approach. We apply a novel two-look approach, with a common U-Net [33] architecture for all of the models, which is described in Section 2.2, and separated footprint and sections segmentations achieving mean intersection over union (IoU) of 0.85, 0.75 and 0.49 respectively for the rooftop, section and superstructure segmentation.

Coherent and documented workflow. Our workflow aims to be coherent, and the information exchange between steps is well-documented with well-established standards. For this, we used Geo-referenced TIFF (GeoTIFF,[2]) for images and NetCDF [34] datasets for the rest of the results. All steps, from dataset generation to economic potential are included in the workflow with appropriate and described outputs.

This paper is structured as follows. In Section 2, a conceptual framework is proposed, and the PASSION workflow detailed. Section 3 compares the results of our methodology with open datasets and existing studies, and uses a German city as a case study for presenting the model results. Finally, Section 4 summarizes the results and discusses the potential of the developed approaches.

[1] https://github.com/FZJ-IEK3-VSA/PASSION

[2] https://www.ogc.org/standards/geotiff accessed 20/01/2023.

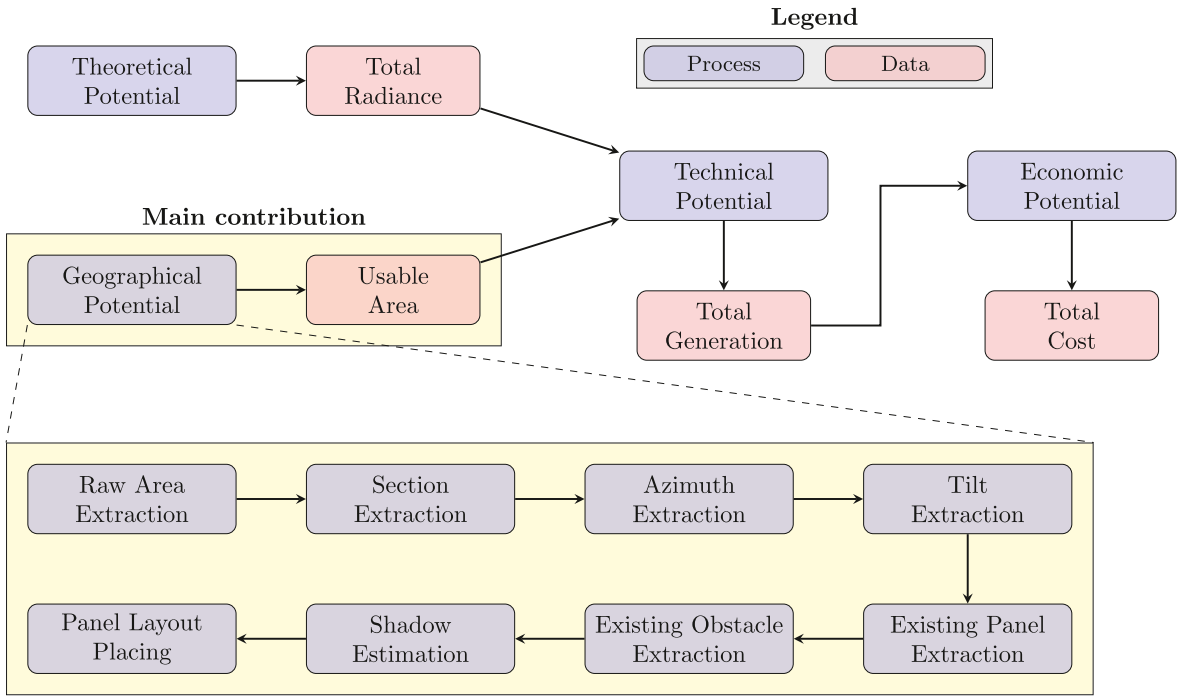


Fig. 1. Overview of different types of rooftop PV potential assessments. Other types of input data are used by different studies, but are heterogeneous and vary depending on the methodology. The calculation of the geographical potential (essentially the available area, azimuth and tilt) is the main focus of the current and comparable studies. Whether it is by inferring, analyzing the 3D data or analyzing satellite imagery, all methodologies should take into account the distribution of sections, azimuths, tilts, existing superstructures, shadowing and the final panel area, but a specific implementation may not account for one or more steps. The purpose of the framework is to be a general model capturing the nature and characteristics of the problem.

2. Methodology

In this section the open source workflow PASSION is presented. In order to describe its individual computational steps with their dependencies as well as required and resulting data sets in detail, the task of photovoltaic potential estimation is first considered at an abstract level. The conceptual framework shown in Fig. 1 formalizes the individual sub-tasks that must typically be accomplished. The implementation-independent description is intended to standardize the general understanding of the subject and serve as a basis for the identifiability and comparability of methods and software implementation scopes. In our case there is some flexibility in the order of steps, as it is not as easy to abstract the order as it is the tasks. Different potentials are usually described from broad to narrow, but computationally it is typically more efficient to only calculate the theoretical potential for the region of interest, calculated as the geographical potential. When using such conceptual frameworks, it is important for authors to map each specific step in their implementation with the corresponding reference step.

Methodologies and software packages, which can vary widely across multiple aspects such as the kind of data used, the spatial resolution of the final results or the integration of different steps, can be mapped against the proposed framework, which allows complex workflows to be broken down and individual sub-processes to be matched and contrasted. The implemented methodology of the developed PASSION workflow is depicted in Fig. 2 and is related to the conceptual framework in the text.

2.1. Data sources

A key principle of the workflow design was data availability and transparency. The approach should be generalizable to other regions of the world, although its performance may vary due to the differences with the training data set. Furthermore, equivalent data sources can be fed into any step of the system, as long as their format remain the same. In the following, we provide an explanation of the data sources used:

Regional shapefiles. In order to retrieve accurate representations of the target administrative boundaries, a small utility was integrated using Nominatim,[3] which automatically generates a shapefile based on a natural language description of a region. The alternative way of specifying a region of interest is adding the target-bounding box in latitudinale and longitudinale coordinates to the configuration (config) file. This initial information is used as part of the *satellite imagery* step of our workflow, as portrayed in Fig. 2.

Satellite imagery. High-resolution satellite input data allows not only better rooftop identification, but also a more detailed analysis at the rooftop level. It is important to note that the ground resolution of the satellite imagery should match that of the training data for the segmentation models. Some available satellite sources sorted by augmenting resolution are Landsat8,[4,5] Sentinel-2, Gaofen,[6] IKONOS,[7] QuickBird-2,[8] GeoEye-1,[9] and Worldview-3.[10] As a developer, it can be convenient to use a provider that integrates multiple data sources, such as Google Maps,[11] or Bing Maps.[12] However, one drawback of this choice is not having additional information about the images, such as

[3] https://nominatim.org/, accessed 20/01/2023.
[4] https://docs.sentinel-hub.com/api/latest/data/landsat-8/, accessed 24/01/2023.
[5] https://www.usgs.gov/landsat-missions/landsat-data-access accessed 24/01/2023.
[6] https://www.cnsa.gov.cn/n6758823/n6758838/c6808018/content.html, accessed 16/09/2023
[7] https://catalog.data.gov/dataset/ikonos-2, accessed 24/01/2023.
[8] https://earth.esa.int/eogateway/catalog/quickbird-full-archive, accessed 24/01/2023.
[9] https://earth.esa.int/eogateway/catalog/geoeye-1-full-archive-and-tasking, accessed 24/01/2023.
[10] https://earth.esa.int/eogateway/catalog/worldview-3-full-archive-and-tasking accessed 24/01/2023.
[11] https://developers.google.com/maps accessed 20/01/2023.
[12] https://www.bingmapsportal.com/, accessed 20/01/2023.

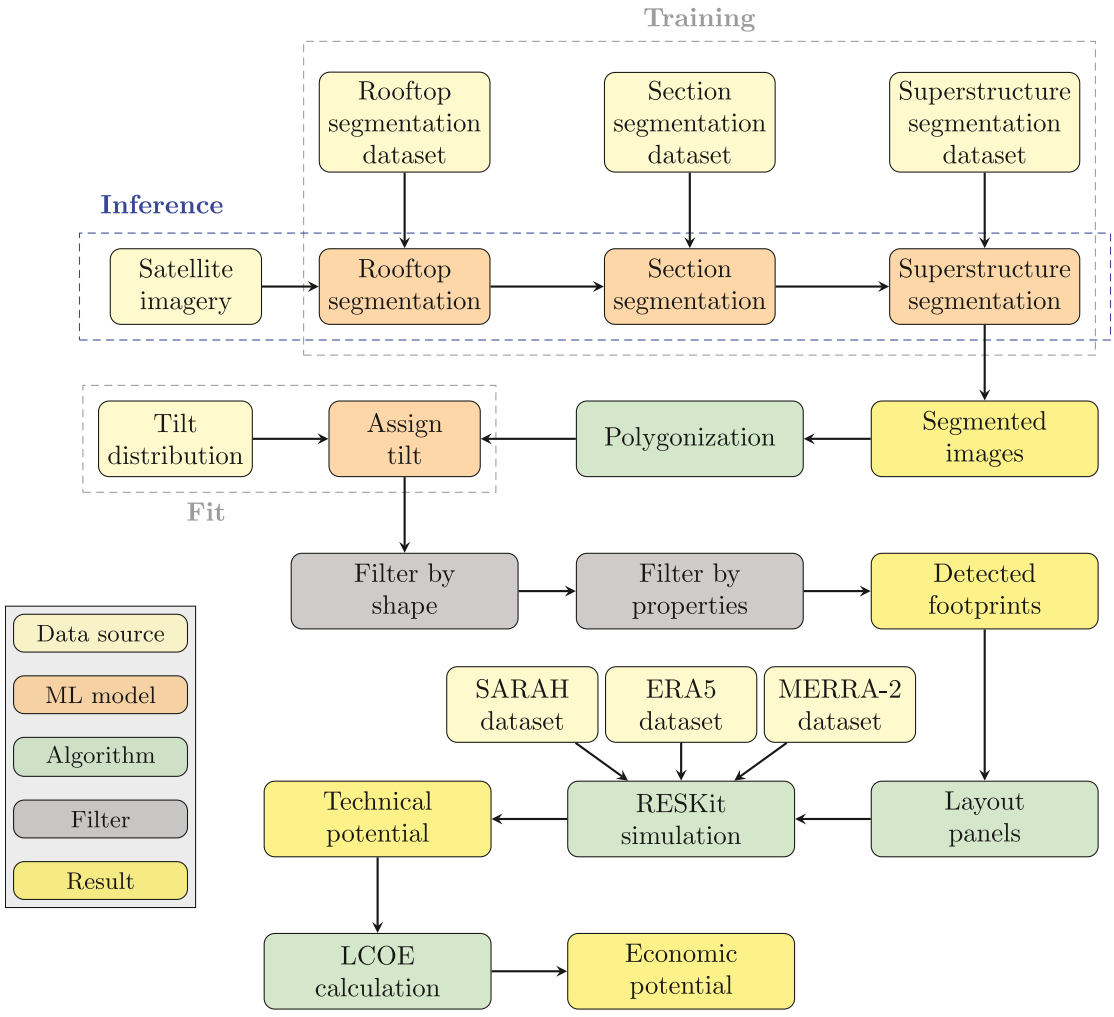


Fig. 2. Modular workflow of the PhotovoltAic Satellite SegmentatION (PASSION). Every step can be redefined, including the data, if the interface described in the source code documentation is retained. Input data (segmentation datasets, satellite imagery, tilt distribution and weather datasets) must follow the documented format, but can otherwise come from any source. Parameters such as the minimum section area, panel properties, or a polygon simplification factor can be specified by a configuration file. For the RESKit [35] simulation, the open source workflow uses the MERRA-2 [36] dataset. If the appropriate version is available, SARAH [37] and ERA5 [38] datasets can also be used.

the date or satellite's position. Both Google and Bing Maps use zoom levels to define the ground resolution with an integer starting from 0 that defines the map resolution M_r depending on the latitude (lat) according to the following formula:

$$M_r = \frac{156,543.04 \text{ m/pixel} * \cos(\text{lat})}{(2^{\text{zoom}})} \quad (1)$$

where:

zoom = Satellite zoom level, integer larger than 0.

Images can be requested at any zoom level, but the data availability is reduced for higher resolutions. Although there is no official data for global coverage per zoom level, urban areas have always been found to be available up to a zoom level of 19, or a theoretical ground resolution of 30 cm/pixel. This theoretical ground resolution is exact at the equator and tends to have an even higher resolution in other parts of the world. These images constitute the *satellite imagery* input of our workflow, as described in Fig. 2.

Labeled rooftop sections. Multiple approaches can be utilized for labeled rooftop sections. An initial approach consisted of detecting only rooftop footprints, for which a combination of AIRS (Aerial Imagery for Roof Segmentation), [39] INRIA, [40] and cleaned 3D data from the German federal state of North Rhine-Westphalia (NRW) [41] were used. The datasets vary in resolution, and were modified to meet a 30 cm/pixel ground resolution. A second methodology adapted Krapf et al.'s [6] dataset, where instead of a binary segmentation the different rooftop sections were annotated into 17 discrete orientation classes, one of these being flat rooftops. The dataset resolution was found to be similar to a zoom level of 20, or approximately 15 cm/pixel. However, we found that both data availability and computing performance would suffer for larger regions when using this resolution. Therefore, the dataset was augmented to account for both 15 cm/pixel and 30 cm/pixel. Both approaches are combined in the final methodology, in which a two-look approach is proposed (see Section 2.2), training two separate models on the INRIA and RID datasets. These datasets are termed the *rooftop segmentation dataset* and *section segmentation dataset* inputs of our workflow, as described in Fig. 2.

Labeled panels/superstructures. Our initial approach attempted to detect any superstructure using classical image processing methods. This proved not to be very general, as they could be tuned with acceptable performance for a set of data but fail when presented with new data. Therefore, we decided to use the dataset proposed by Krapf et al. [6], which labels eight different superstructure classes, most interestingly including photovoltaic panels. Table 1 compares the existing segmentation datasets for rooftops, sections and superstructures. This dataset is termed the *superstructure segmentation dataset* input of our workflow described in Fig. 2.

Crowdsourced data. Different stages of the workflow benefit from using geographic crowdsourced data [42]. Municipal administrative

Table 1
An overview of relevant segmentation datasets. During the development of this study, the AIRS dataset was taken down online, but the validation metrics have been preserved.

Dataset	Region	Resolution	Labeled classes
INRIA [40]	US and Austria	30 cm/pixel	0: background 1: rooftop
RID [6] sections	Wartenberg	15 cm/pixel	0-15: rooftop orientation in 22.5° intervals 16: flat rooftop 17: background
RID [6] superstructures	Wartenberg	15 cm/pixel	0: background 1: photovoltaic installation 2-8: other superstructures
AIRS [39]	Christchurch	7.5 cm/pixel	0: background 1: rooftop
NRW [41]	Aachen	Vector data	0: background 1: rooftop

Fig. 3. The current GUI allows the user to select the different estimated layers and a set of basemaps. In this case, the OSM [43] basemap is used. In the URL parameters, the user can specify which results to load, which variables to show and how many polygons to render.

boundaries can be extracted, the location of some existing panels found, and the open building footprints used instead of, or in addition to satellite analyses. For this, OpenStreetMap [43] (OSM) was used via the Overpass Turbo,[13] API. However, it is important to note that the quality of the annotations highly depends on the geographic region [44], and there is a lack of standardized descriptions for certain attributes, and so this data should be accounted for as partial and not exhaustive. This data can optionally be used in exchange for the *rooftop segmentation dataset* input of our workflow described in Fig. 2.

Weather data. In order to calculate the technical potential, both weather and solar radiance must be accounted for. For this, either the SARAH solar radiation [37] and ERA5 climate datasets [38] or the MERRA-2 [36] dataset can be used. These datasets are represented as the *SARAH dataset*, *ERA5 dataset* and *MERRA-2 dataset* inputs in our workflow described in Fig. 2, and either method can be used. While the used RESKit model applies these data sources, further satellite-derived irradiance databases could improve the workflow [45].

Tilt distribution. In order to account for the tilt angles, and due to the difficulty of detecting this with images, public manually labeled data from the NRW region [41] was used to estimate future predictions. This distribution is termed the *tilt distribution* input in our workflow, as depicted in Fig. 2.

2.2. Geographical potential

The focus of our study lies on the assessment of geographical rooftop PV potentials. For this, satellite images must be automatically analyzed in order to estimate areas that are suitable for panel placement, taking multiple factors into account, such as the azimuth, tilt angle, or roof obstacles. The main difference of our method compared to others is that we propose a two-look approach for detecting sections. This means that one model is trained on the task of rooftop segmentation exclusively, where more data is available, and a second is trained on the task of section segmentation. The reason for this is that we found the second model to be less robust on its own which led to a large percentage of false positives, and overestimations of the area of regions that deviate significantly from the training data. In the following, we explain the various steps of our workflow (Fig. 2).

Rooftop segmentation. The rooftop segmentation step corresponds to the *raw area extraction* in our generic workflow shown in Fig. 1. In this step, a first segmentation model performs a binary classification, where input satellite tiles of 512 × 512 pixels are predicted. This resolution allows the models to have more context while being reasonably small for our computational needs. A classical U-Net network was implemented for this task. This model follows an encoder–decoder network architecture, where the input images are first encoded multiple times, and so shrunken in size. This enables the network to learn an abstract representation of the image and to extract its fundamental features. In the same manner, this representation is decoded back to the original size, with bridged information from each encoding step, finally recreating a segmented representation of the original image that can be binary or have multiple classes. Different networks can also be used as encoders. In our case, a ResNet [46], which is a Convolutional Neural Network made out of residual blocks, is used as the backbone. The output of this stage is a list of tiles that are filtered using the predicted binary output. In some cases, tile borders can be conflicting, given that they lack spatial context. For this, a window slide stride was implemented, in which each tile is separated by 256 pixels and multiple predictions are carried out for these pixels at the borders. This model was trained on the INRIA dataset, which is augmented by randomly rotating and applying smoothing with Gaussian filters, as well as color transformations. This step could also be substituted by OSM building footprints as a baseline, but as this depends on crowdsourced data the reliability varies across regions [44].

Section segmentation. This step encapsulates both the *section extraction* and the *azimuth extraction* in Fig. 1. The same U-Net architecture as for the rooftop segmentation is used. In this instance, the model is trained on multi-class segmentation with 18 output classes: 16 azimuth classes, flat rooftops, and the background class, which encapsulates everything that the model is not trying to classify. The 512 × 512 satellite tiles are sent to this model, which segments the different rooftop sections with the predicted orientation angle. In order to reduce the number of false positives in this step, which has shown to be less robust for regions far from the training area, the portions of the segments that do not overlap with the rooftop segmentation are filtered out of the prediction. In addition, in order to reduce the number of false negatives, the rooftop segment predictions that were not estimated by the section segmentation are also included in the final estimation. We compare the evaluation of both approaches in Section 3.1.1, but this does not coincide with the performance in non-controlled real life regions. This two-look approach has been found to provide much more robust results in the wild, which refers to uncontrolled real-world imagery, than directly predicting sections, as explained in Section 3.1.2 and Fig. 6, and also than detecting the section using classical image processing techniques.

[13] overpass-turbo.eu accessed 17/01/2023.

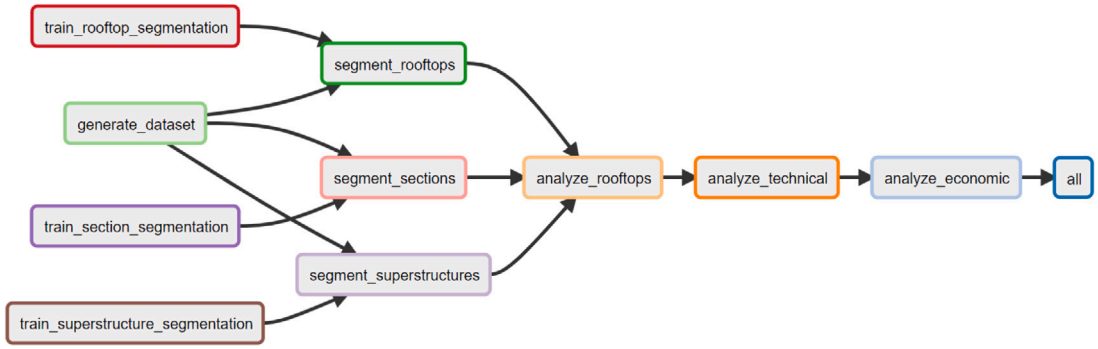


Fig. 4. Snakemake steps of PASSION. Each step is a rule defined in the Snakefile. This tool offers multiple advantages for building coherent workflows, and will only run rules that are needed to regenerate. For example, once the models are trained, Snakemake would skip them for any further analyses unless specified otherwise.

Tilt estimation. The vertical angle of the solar panel is crucial for its efficiency, and is estimated in the *tilt extraction* step in Fig. 1. Optimal tilts vary depending on the latitude, and are usually between 30 and 45°. It is challenging to accurately estimate this using computer vision techniques, especially considering that satellite images have relatively low resolutions and are captured pointing vertically down to the Earth. For this reason, the current approach uses a kernel density estimator (KDE) to estimate the probability density function of a ground truth set. For any new estimated sections, a new random tilt value is taken from the predicted density function. One factor to take into account is that this is data from the NRW region only, which probably has biases corresponding to the region. For sections estimated to be flat the panel can be installed with optimal tilt and azimuth, and so the value is set to 31°. Note that the optimal value depends on the latitude of the region, and so the method for optimally-assigning the best values can be implemented in the future.

Superstructure segmentation. This step relates to the *existing panel extraction* and *existing obstacle extraction* steps in Fig. 1. Another U-Net is trained on the RID dataset in order to segment different types of superstructures. Similarly to the section segmentation, the initial rooftop segmentation is used to filter out obstacles that are incorrectly located next to the predicted buildings, such as in the road. For the purposes of this study, the nine output classes are grouped in existing panels, obstacles, and background. This encapsulates the two steps, *existing panel extraction* and *existing obstacle extraction* of the generic workflow in Fig. 1.

Shadowing. The *shadow estimation* step of the generic workflow is important for potential assessments and represents the decrease in performance due to obstacles or other buildings creating a shadow in the potential installation. In PASSION, shadowing is accounted for with a reduction factor that can be specified by the user.

Rooftop vectorization. Image masks are one way of representing the predicted rooftops, but we need the georeferenced polygon representation for further analyses. For this, all images are first stored in a GeoTIFF format, allowing for the inclusion of metadata such as the coordinate reference system. With this information, the predicted buildings are converted from images to vectors, which is performed using an algorithm presented by Suzuki and Be [47], which is implemented in OpenCV [48]. This method follows the borders of holes in a binary image in order to identify them. This is an intermediate processing step, and part of the *section extraction*.

Polygon simplification. In order to simplify the predictions into polygons that take less storage, the Teh Chin chain approximation algorithm [49] and Ramer–Douglas–Peucker algorithm, which was first proposed in [50], are used. This would be an intermediate processing step, and part of the *section extraction*.

Area extraction and size filtering. Once these steps have been performed, the area is calculated taking into account the ground pixel resolution with respect to the latitude. Rooftops detected below a user-defined area threshold are discarded due to not being able to fit an installation. This would be an intermediate processing step, and part of the *section extraction*.

Panel layout. Finally, an algorithm for *panel layout placing* was implemented. This algorithm requires the following parameters, which must be in the same coordinate system: a target polygon, the polygon orientation, the shape of the panels, a spacing factor, a minimum distance to the border and a number of offsets. The target polygon and its orientation represent the predicted building where we will place panels, and the rest of the parameters are defined by the user as part of the configuration file. First, the algorithm shrinks the original polygon according to the user-defined minimum distance to the border. Then, a grid oriented according to the azimuth is created, with each cell's shape being the panel shape multiplied by the spacing factor, separating the panels by a typically small distance. A panel is placed in every cell that does not intersect with the shrunk original polygon, and this process is repeated by offsetting the grid n times by n/width units, and is repeated for vertical and horizontal panel placing. Finally, the placing that fits most panels is selected. The resulting data contains the geographical information (location, area, azimuth and tilt) of every estimated potential panel.

2.3. Technical and economic potential

To estimate the technical potential, the geographical information from Section 2.2, corresponding to *usable area* in Fig. 1, is transformed into energy units, with an hourly spatial resolution during an entire year. For this simulation, the open source *Renewable Energy Simulation toolkit for Python* (RESKit) [35] is applied, which is built up on PVLib [51]. Both *theoretical potential* and *technical potential* are calculated in this step. This library takes all of the mentioned information in the shape of a Python Xarray [52] and uses either the SARAH solar radiation and ERA5 climate datasets or the MERRA-2 dataset for the calculations. The data is aggregated per panel to show the total yearly generation in watt-hours. For these estimations, PVLib offers a number of predefined panel models. All of the parameters for each step are configurable by the user.

Finally, after considering all of the costs and benefits, the average cost per kilowatt-hour can be calculated during the lifespan of the system in order to obtain the *economic potential*. For this, the LCOE [53] is used:

$$\begin{aligned} \text{Lifetime Costs} &= \sum_{t=1}^{n} \frac{I_t + M_t}{(1+r)^t} \\ \text{Lifetime Benefits} &= \sum_{t=1}^{n} \frac{E_t}{(1+r)^t} \\ \text{LCOE} = \frac{\text{Lifetime Costs}}{\text{Lifetime Benefits}} &= \frac{\sum_{t=1}^{n} \frac{I_t+M_t}{(1+r)^t}}{\sum_{t=1}^{n} \frac{E_t}{(1+r)^t}} \end{aligned} \tag{2}$$

where:

n = expected lifespan of the system.
I_t = capital costs in the year t.
M_t = maintenance costs in the year t.
E_t = electrical energy generated in the year t.
r = discount rate, which compares future costs to present value based on market interest rates.

The final result of the simulation is an economic potential estimation, giving an average cost per kWh with a spatial resolution per rooftop section. This allows for many different analyses, such as setting a cost threshold per kWh to separate viable from nonviable installations, or estimating what would be the expected potential with a given investment for a region. All of this is stored as a NetCDF file with a time resolution of 30 min over a year, and a spatial resolution per estimated section.

2.4. Visualization and workflow management

In order to visualize the results, a preliminary graphical user interface (GUI) was implemented using the Python libraries Flask [54] and Folium [55]. This web interface renders the original georeferenced images on top of an OSM basemap, as well as the vectorized polygons of each step separately. Fig. 3 shows an excerpt from the interface, in which the different layers with additional information per step can be selected.

Implementing such a modular workflow like PASSION presents interesting challenges, such as keeping the file consistency between steps, easily tracing errors or selecting which steps to execute. Implementing a complex manager would be prone to inconsistencies and less scalable, so it was decided to use the Snakemake [56] workflow tool. This makes it easier to create reproducible and scalable analyses, allows for the execution of jobs in cluster environments, and orchestrates the workflows in a single human-readable file called a Snakefile. A config file can also be natively defined, allowing easier parameterization of the software. Snakemake offers multiple advantages for building coherent workflows, and will only run rules that are needed to regenerate. For example, once the models are trained, Snakemake would skip them for any further analyses unless otherwise specified. The current workflow described as Snakemake rules is shown in Fig. 4.

3. Results and discussion

This section discusses the results obtained in our experiments. In Section 3.1 the rooftop, section and superstructure segmentation models are validated across the available datasets. Section 3.2 compares the results of our approach to others found in the literature. Section 3.3 outlines a case study for the region of Aachen, Germany. Finally, Section 3.4 describes the known limitations of our approach. We validate our results for all potentials for which standard measurements are available.

3.1. Image segmentation

As part of the geographical potential, in this section the different image analysis models are evaluated against datasets external to the training data. Section 3.1.1 outlines the process used to select and train our segmentation model. In Section 3.1.2, the three models are evaluated against all available datasets for each task. Finally, Section 3.1.3 compares the results with gradual changes of the image ground resolution, showing how the performance changes for higher and lower resolution input images.

3.1.1. Model selection

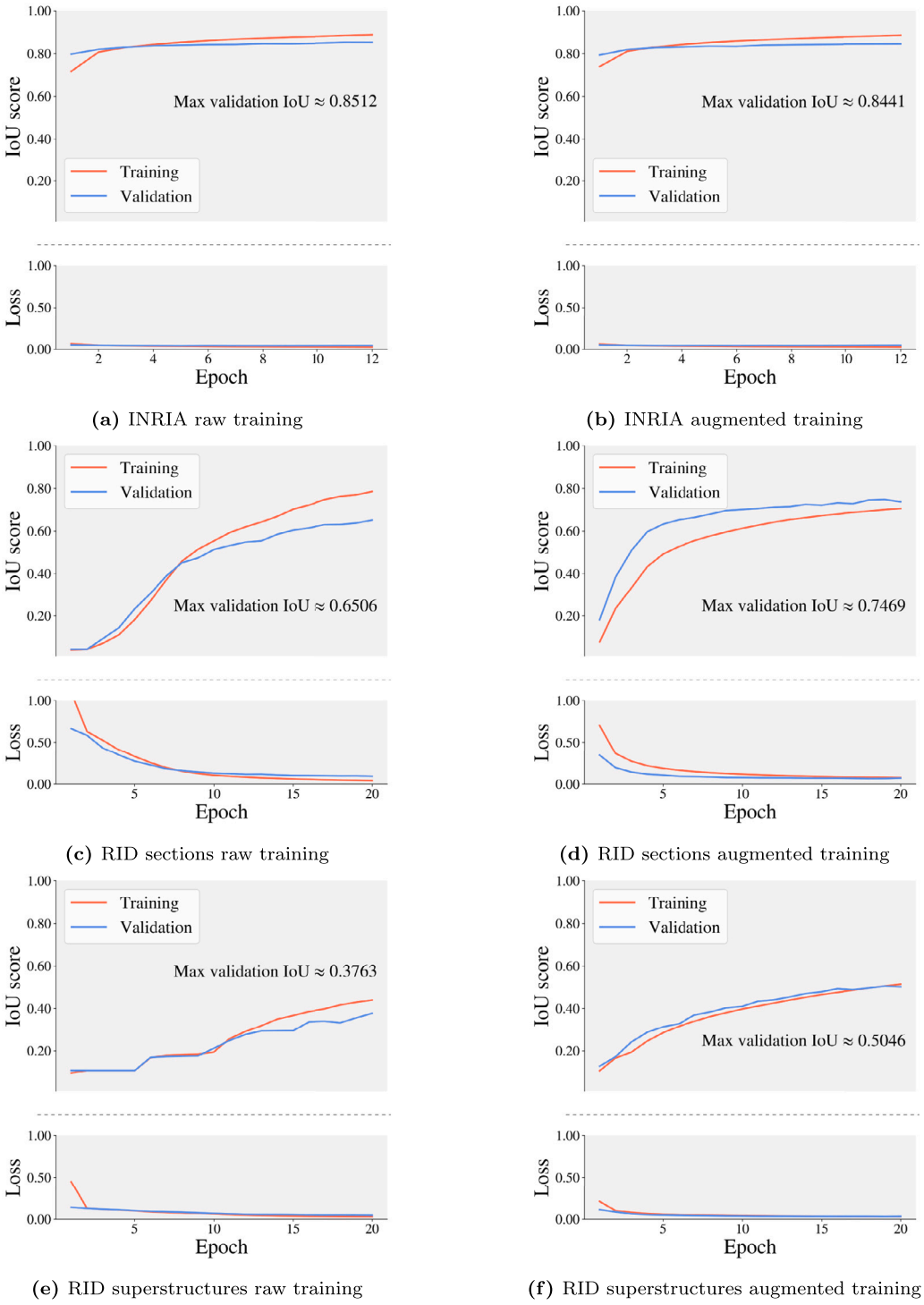
All three segmentation tasks for rooftops, sections and superstructures (Fig. 2) used the same U-Net architecture with a ResNet backbone. These tasks correspond to the *training* box in the workflow diagram in Fig. 2. As part of the training, we must use an algorithm that minimizes the error function. This is called the optimizer, and Adam [57] with a learning rate of 10^{-5} reported the best performance. For rooftop segmentation we used a batch size of 8 over 10 epochs, which means that it was iterated 10 times during training in groups of 8 images. For the section and superstructure segmentation models, 20 epochs were used. The number of output classes were specified as two, 18, and nine for rooftops, sections and superstructures respectively. Both the INRIA dataset for the rooftop segmentation and the RID dataset for the section and superstructure segmentations, were split into 80%, 10% and 10% for training, validation and test sets. Finally, all models were tested with and without data augmentation, which in this case incorporated classical computer vision transformations such as rotations in multiples of 90°, Gaussian filters, which are used for image smoothing, or slight zooms. In the case of the RID dataset, the data has a ground resolution of 15 cm/pixel, which corresponds to a zoom level of 20, and is double our target theoretical resolution of 30 cm/pixel. The dataset was resized into our target resolution, and different combinations of the data resolutions were tried. For evaluation, the IoU, [58] also known as the Jaccard Index [59], was used. This is a common evaluation metric for image segmentation in computer vision, as it measures the percentage of overlap between the target mask and the prediction output, considering the rate of false negatives and false positives. For multi-class segmentation, the mean IoU is used, which averages the individual IoUs for all classes.

Fig. 5 shows how the IoU and multi-class focal loss [58], which is a function that evaluates the performance of the model during training, evolve during training for the different models, demonstrating that the performance generally improves by using data augmentation techniques, especially for small datasets. Given its large size, and the fact that the epochs were not interrupted after a number of steps, the INRIA dataset started to perform well already after the first epoch. After the first few epochs there is negligible performance increase and it could be argued that the model slowly started to overfit. The other two models seemed to start approaching a plateau after 10 epochs.

3.1.2. Model validation

After training, all best performing models were evaluated using all available data for the same task. For rooftop segmentation, multiple sources exist and a proper assessment and comparison can be carried out. Unfortunately, this is not the case for models for section and superstructure segmentation, where the only assessment possible is to validate in the separated test set of the original RID dataset. Table 2 shows the performance of the different models against all the collected test datasets. In our abstract framework of Fig. 1, the evaluated tasks correspond to *raw area extraction* for the rooftop segmentation, *section extraction* for the section segmentation and both *existing panel extraction* and *existing obstacle extraction* for the superstructure segmentation.

Prediction samples for the datasets used in Table 2 can be found in Fig. 7 for rooftops, Fig. 8 for sections and Fig. 9 for superstructures. The rooftop segmentation model performs best in a wider range of datasets. On the other hand, the section and superstructure segmentation models can only be compared against the ground truths of the RID test set and their performance is restricted to the geographical location. This visual evaluation is also important for assessing whether the estimations are coherent, and to better understand where the models are more likely to underperform. In the case of the NRW dataset, performance was lower than for the other rooftop datasets, with a rooftop IoU of around 0.5490. After inspecting the visual results, we discovered that the NRW labeled data is not perfect itself, featuring some false negatives and overly simplified footprints, which explains these results. Overall, these results are similar to those seen in state-of-the-art research. However,



(a) INRIA raw training

(b) INRIA augmented training

(c) RID sections raw training

(d) RID sections augmented training

(e) RID superstructures raw training

(f) RID superstructures augmented training

Fig. 5. IoU and loss metrics of the models during training. Rooftop training in 5(a) and 5(b) learned the fastest and benefitted the least from augmenting the images, as the raw data (INRIA) is already large. It can be seen that after the first few epochs the model does not significantly improve and starts to overfit. Section and superstructure training from 5(c) to 5(f) significantly improve after augmenting the data (RID).

the lack of standard benchmarks in the field makes it difficult to formally compare our results with those of other studies. Most importantly, performance is very tied to geographical location in many cases, and real life results for a new region may vary even further.

The proposed two-look approach was also tested again with the RID test set. This, however, provided worse results with a section segmentation mean IoU of 0.64. Even though the results indicate that the one-look approach provides a better score in the test set, the two-look one outperforms it in the wild. This is because the RID dataset is small, and even after augmentation it fails to generalize well to other regions. With the two-look approach, rooftops are first correctly identified with a more general model, and these results are then used to filter out false positives from the section segmentation. A visual comparison of some results for the RID set can be seen in Fig. 6.

3.1.3. Multi-resolution evaluation

The availability of a very high resolution dataset makes it possible to test how our models perform across different ground resolutions

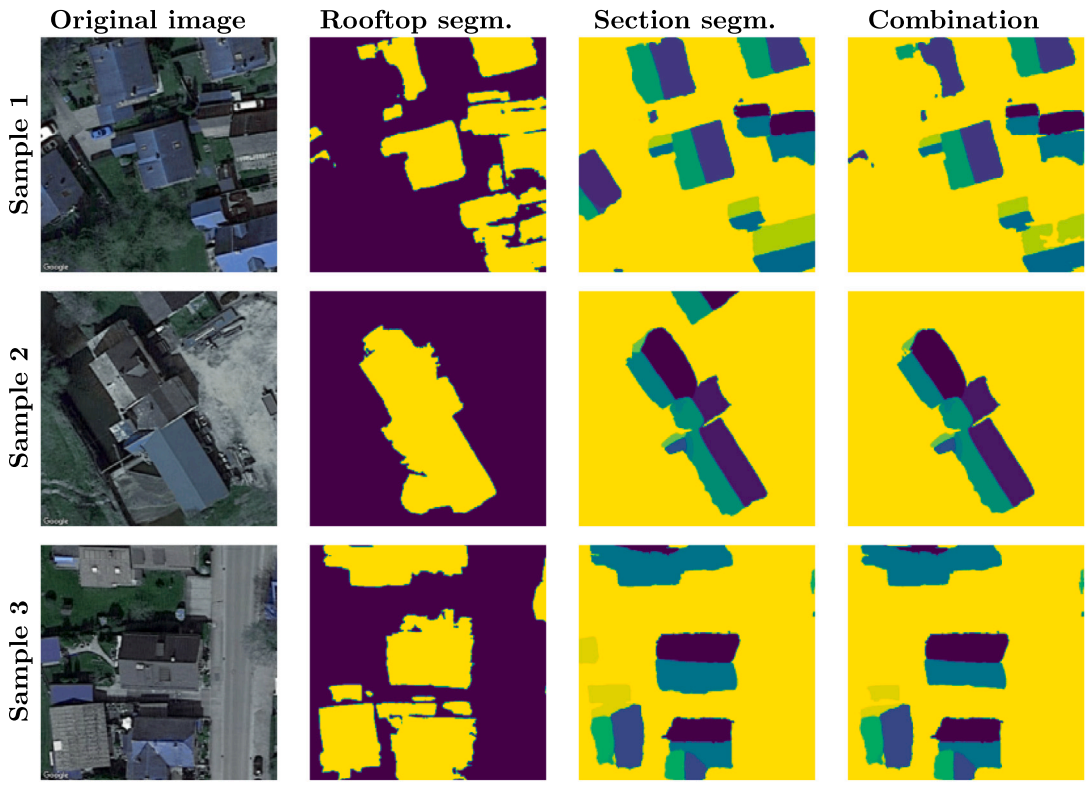


Fig. 6. In this figure, three samples from the RID [6] dataset are selected to compare the one-look and two-look approaches. The columns show, in order: the original image from Google Maps, the rooftop segmentation binary prediction, the section segmentation prediction, and the combination of the two. The results of the two-look approach in the wild have shown more coherent performance. Images in the left column correspond to samples from the RID dataset. In the second column, pixels in yellow represent a predicted rooftop. In the rest of the columns the background class is shown in yellow, and each color represents an orientation predicted by the section segmentation model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Overview of the performance of the segmentation models against all collected test datasets. The rooftop segmentation model performs best, both because of the larger training dataset and the higher complexity of the other tasks. For the section segmentation, the mean deviation metric is calculated by taking the mean orientation error of the true positively classified section pixels. The results of the superstructure segmentation model are satisfactory for the panel class, but the remaining obstacles, such as dormers or ladders, are more challenging to detect.

Rooftop dataset	Background IoU	Rooftop IoU	Mean IoU
INRIA	0.9364	0.7592	0.8478
RID	0.8789	0.6446	0.7618
AIRS	0.9777	0.7553	0.8665
NRW	0.8742	0.5490	0.7116

Section dataset	Flat IoU	Tilted IoU	Mean IoU	Mean deviation
RID	0.7718	0.7386	0.7531	16.20°

Superstructure dataset	PV panels IoU	Rest of obstacles IoU	Mean IoU
RID	0.8164	0.3767	0.4927

on the *raw area extraction* task of Fig. 1. We tested the AIRS dataset, which has a ground resolution of 7.5 cm/pixel, on different resized resolutions. Fig. 10 shows how the original model, trained with images of ground resolutions of between 20 and 30 cm/pixel, performs best around those. It is interesting to note that the model does not perform as poorly on images with higher resolutions compared to the trained data, as on images with lower resolution. In summary, the intended ground resolution that will provide the optimal results is 15 cm/pixel, but slightly lower or slightly higher resolutions will also obtain results with a close performance. In general, images with a resolution higher than 15 cm/pixel will function better than those with a lower one.

3.2. Comparison with other studies

For estimating the technical potential, which corresponds to *total generation* in our conceptual workflow, we perform micro-simulations for every predicted panel, enabling comparisons with other studies that do this with different methodologies. LiDAR analyses are generally more accurate than satellite ones, and can be a good indicator of whether or not the obtained results make sense. The chosen panel for our experiments was the LG Electronics LG370Q1C-A5, with a capacity of 370 W and surface area of 1.7×1.016 m^2, but any of the models offered by PVLib can be used. Regarding the economic parameters, we use a system expected lifespan of 25 years, a price per panel of 350 € for the LG370Q1C-A5, a maintenance cost of 1% of the initial investment per year, a yearly degradation of 0.5% per year, and a discount rate of 8%. Additionally, other installation costs such as the inverter are taken into account for the investment cost. The inverter price is defined as a ratio per installation capacity, in our case 0.2 €/W and with a lifespan of 13 years. Other initial costs are also defined, and in our case were set to 200 € per installation.

In Adjiski et al. [60], the technical rooftop PV potential is analyzed for a district of Skopje, North Macedonia. The study performs an assessment with a one meter resolution LiDAR point cloud, with which roof planes are identified and their potential simulated, taking their area, slope, morphology and shade into account. Although the article does not provide the exact coordinates of the study area, it provides a footprints image that allows it to be found manually as a region between latitudes 42°0′9″, 42°0′25″ and longitudes 21°22′30″, 21°23′1″. Fig. 11 shows the results for the study region obtained with PASSION. A detailed comparison is presented in Table 3, where the results from Adjiski et al. are compared to ours. Before filtering any rooftops, the panel area in each of the buildings totaled 228, 243 and

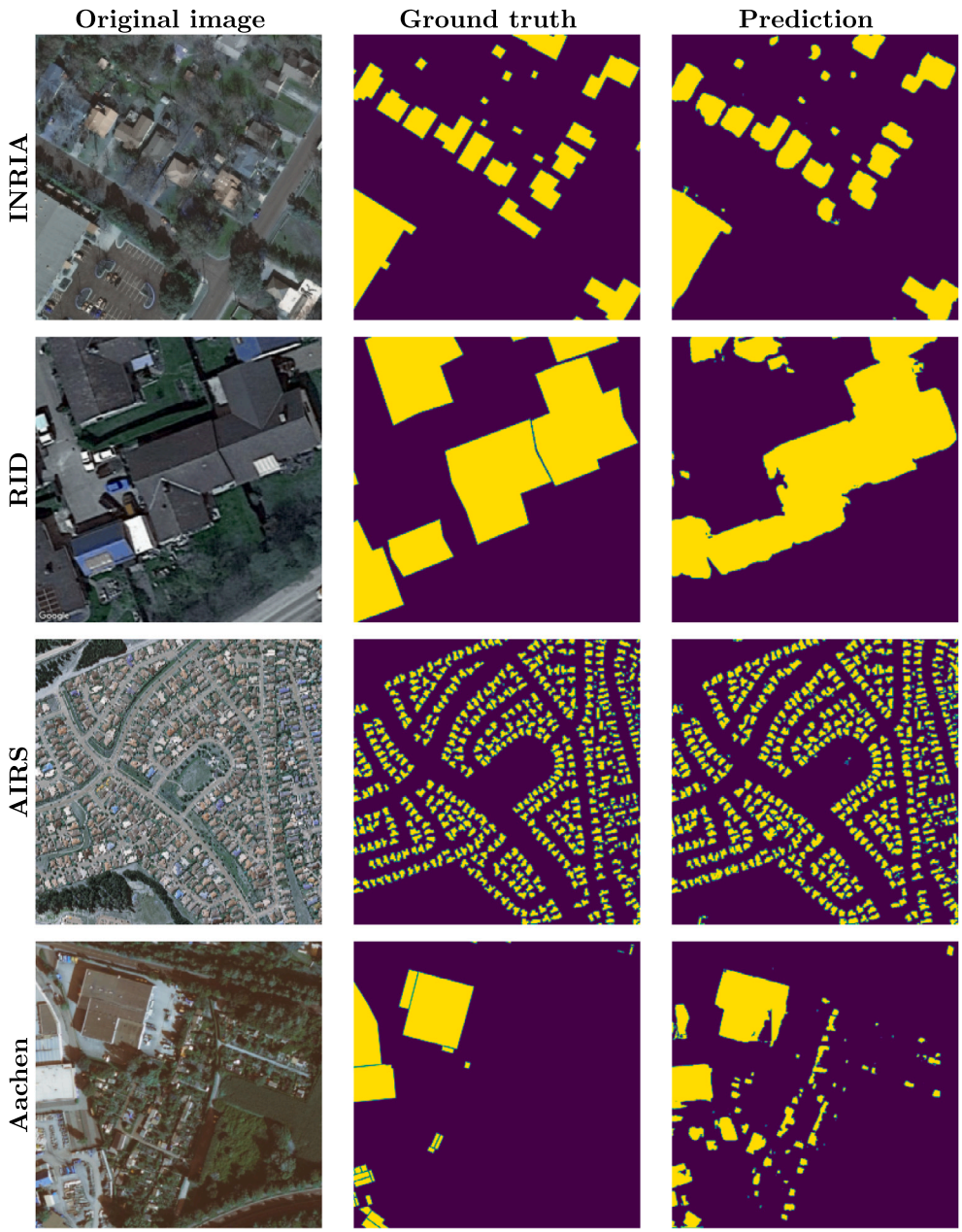


Fig. 7. Sample predictions of the rooftop segmentation model, with predicted rooftops represented in yellow. The columns show the original image, the ground truth, and the prediction, in sequence. Images in the left column show from top to bottom samples from the INRIA [40], RID [6], AIRS [39] and NRW [41] datasets respectively. The ground truth source of the NRW dataset is manually labeled, and therefore it is roughly simplified and imperfect. Our model detects smaller buildings that were not present in the original data. The source of the satellite imagery is Google Maps and Bing Maps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

233 m^2 respectively, but after applying the same filtering criteria that Adjiski et al. use, these results were reduced to 110, 108 and 233 m^2 respectively. This makes a total of 452.5 m^2, which is close to their result of 512.1 m^2, especially taking into account that parameters such as the spacing between panels, spacing from the border and the type of panel can influence the final result. Unfortunately, these parameters are not specified in the original paper.

Another example is the study by Kausika et al. [61], which analyses the entire municipality of Apeldoorn, in the Netherlands. Again, a LiDAR point cloud was used as the source data to derive a digital elevation model (DEM), and estimate the technical potential of the different rooftop sections. For the comparison with PASSION, we retrieved the shapefile of the region from OSM, and conducted a full-scale analysis. Fig. 12 shows the prediction results for the sample district used in the original paper obtained with PASSION. The results of the approach yield a yearly technical potential of 283.9 GWh. For this, rooftops are classified as unsuitable, partially suitable and optimally suitable areas, based on a south-facing orientation, a slope less than or equal to 38 degrees, and solar irradiation greater than 70% of the maximum annual radiation of the area. After applying these filters, our results exceeded these predictions at a yearly technical potential of 356.2 GWh (+ 25%). This was due to the prediction of unusually large rooftop areas in a few cases. After removing rooftop outliers that are larger than three times the standard deviation of the municipality, which account for 419 sections out of the original 27,600, the obtained prediction is 258.8 GWh (−9%), which is closer to the results of the compared paper. This demonstrates that our current models are prone to overestimating the rooftop area under certain circumstances. In addition, our panel

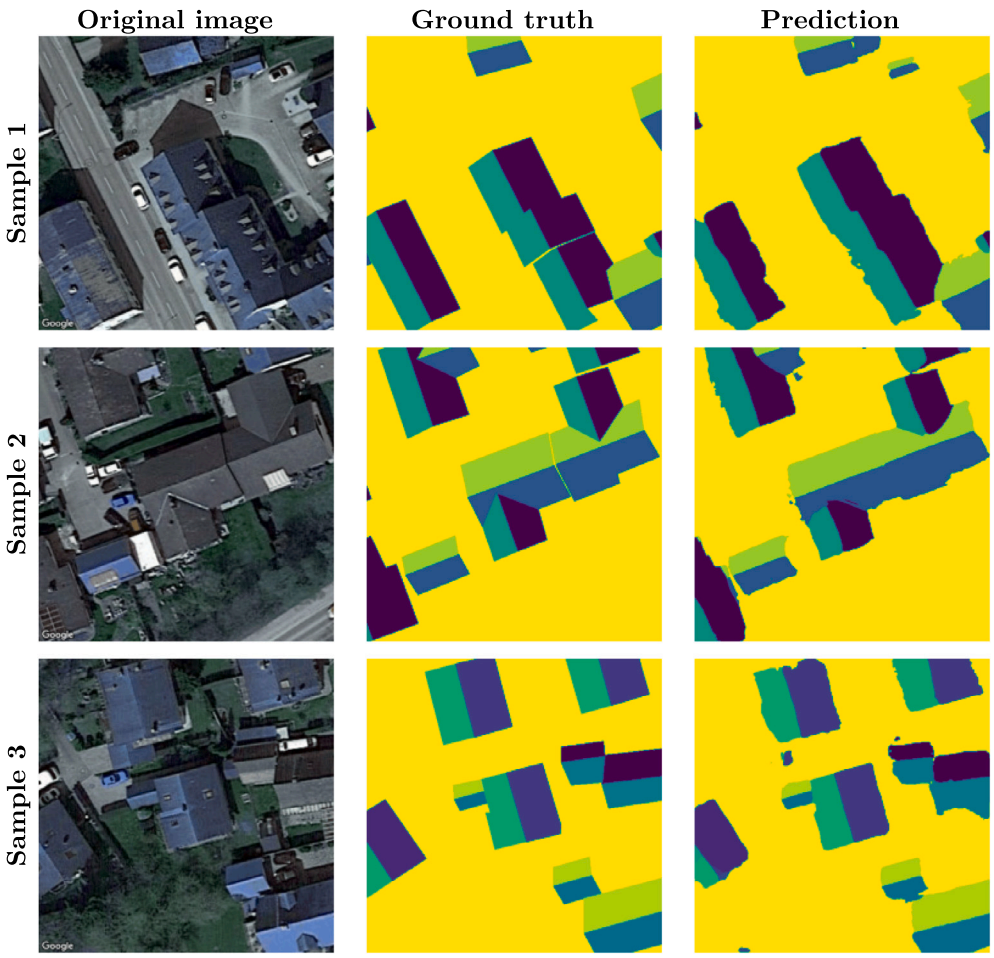


Fig. 8. Sample predictions of the section segmentation model. The columns display the original image, the ground truth, and the prediction, in sequence. The predictions show three sample images from the RID [6] dataset. The background class is shown in yellow, with each color representing an orientation predicted by the section segmentation model. The source of the satellite imagery is Google Maps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Results comparison between Adjiski et al. [60] and PASSION. In order to use the same criteria for the comparison, both the whole results (before filtering), and only the sections detected to be south-facing (after filtering), are shown. After filtering, the results shown are quite similar to those from the study.

Approach	Roof ID	Roof area (m^2)	Number of panels	Panel area (m^2)	Yearly output (MWh)
Adjiski et al. [60]	1	227	78	128.7	25.3
	2	241	82	135.3	26.9
	3	497	154	254.1	50.5
	Total	**965**	**314**	**518.1**	**102.7**
PASSION (before filtering)	1	788	132	228.0	49.0
	2	750	141	243.5	48.9
	3	590	135	233.2	65.2
	Total	**2128**	**408**	**704.7**	**163.0**
PASSION (after filtering)	1	326	64	110.5	30.1
	2	278	63	108.8	29.2
	3	590	135	233.2	65.2
	Total	**1195**	**262**	**452.5**	**124.5**

detector estimated an existing yearly production of only 0.23 MW_p installed compared to 3.4 MW_p in Kausika et al. [61]. As our focus lay on the workflow definition and geographical potential, this portion of our work is probably not yet mature and must be improved upon in future studies.

Finally, we compared the results with Mainzer et al. [18], who analyze the German municipality of Freiburg. In this study, OSM data is used in order to extract the geographical potential, and the data is not filtered by orientation or other factors. Their methodology also analyzed the extracted buildings in order to detect ridge lines to divide the rooftops into different sections. Tilted rooftops where assigned an angle from a normal distribution function with a mean value of 37°. Furthermore, the panel modules are placed using an algorithm that is not described in detail. The study also calculates the technical and economic potential. For the entire municipality of Freiburg, the final yearly technical potential is found to be 524 GWh, and the LCOE ranged from 90 to 290 €/MWh. In our initial analyses, we obtained a very different estimation, at 161 GWh for the whole municipality. After configuring PASSION to use OSM data like in Mainzer et al. [18], an underestimating result of 230 GWh was still obtained. This could

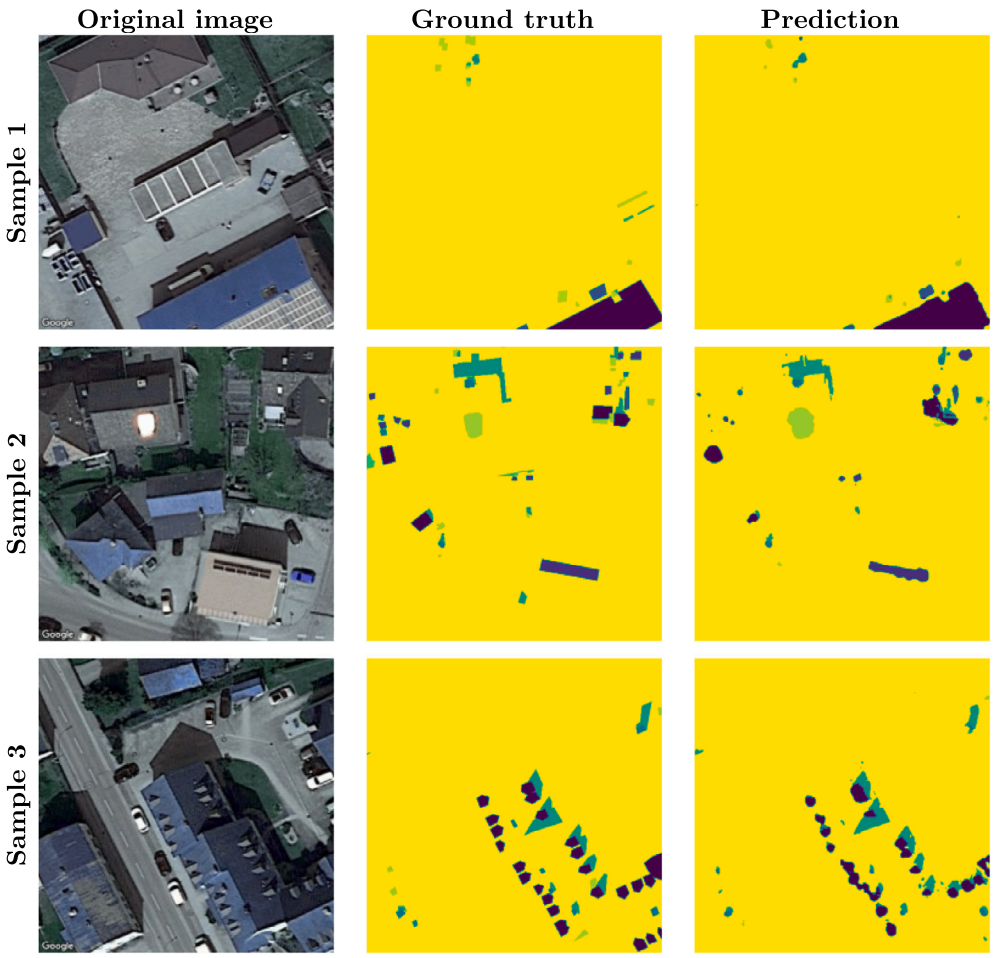


Fig. 9. Sample predictions of the superstructure segmentation model. The columns show the original image, the ground truth, and the prediction, in sequence. The predictions display three sample images from the RID [6] dataset. The background class is shown in yellow, and each different color represents a different class predicted by the superstructure segmentation model. The source of the satellite imagery is Google Maps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

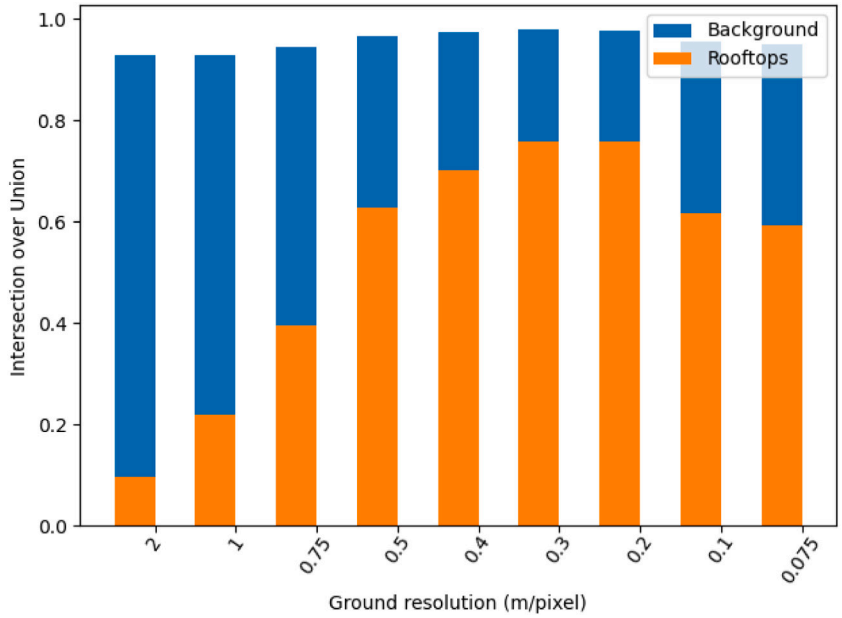


Fig. 10. The diagram shows the IoU of the rooftop segmentation model on the AIRS dataset for different ground resolutions. Test images were resized from their initial 7.5 cm/pixel resolution. Peak performance is seen at values of 20 to 30 cm/pixel, which is the resolution of the training dataset. Performance at higher resolution (i.e., 7.5 to 10 cm/pixel) is still reasonable and comparable to values at a 50 cm/pixel resolution, but once the resolution starts to decrease too greatly, at around 75 cm/pixel, the model's performance declines.

Fig. 11. District of Skopje, North Macedonia analyzed in Adjiski et al. [60]. The GUI visualizes the final sections in blue, with the sections filtered by size in red and panels in green. The buildings used for the final analysis are zoomed on the right. The background map corresponds to Bing Maps imagery. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

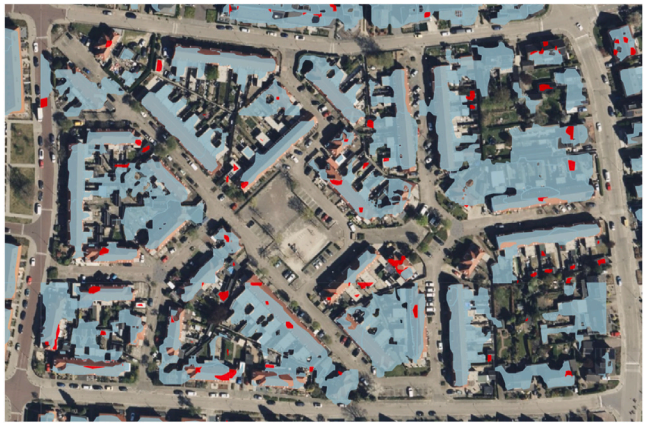


Fig. 12. A picture of the district in Apeldoorn, the Netherlands that was used in the figures of Kausika et al. [61], and the results of PASSION. Sections that were filtered due to their size are shown in red, and those that were used in the study appear in blue. The background map correspond to Bing Maps imagery. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

be explained due to a smaller panel separation. Mainzer et al. [18] propose a separation of 10 cm for slanted roofs, and a separation of twice the modules' height for flat roofs, where panels are installed at a 30° angle. In our initial configuration, we used a panel separation of 30% of the panel size, and a space to the border of the section of 1.5 m. After adjusting our parameters to a panel separation of 10% and a border spacing of 0.70 m, results became much more similar to Mainzer et al. [18]. Using the segmentation models approach we obtained a yearly technical potential of 392 GWh, and using the OSM approach we obtained a result of 557 GWh per year. Regarding our LCOE calculations, our results ranged from 58 to 249 €/MWh, which can again be explained by varying LCOE parameters.

3.3. Case study

In order to showcase the PASSION workflow, we applied it to the region of Aachen, situated in the west of the German federal state of North Rhine-Westphalia. This region, with a surface area 160.8 km^2 [62], serves as a sample of a larger municipality including its non-urban surroundings. Fig. 13 shows the administrative boundaries that were retrieved for the study. Fig. 14 depicts a sample of the detected rooftops, potential-placed panels, and existing panels in zoomed-in areas of the city center.

The techno-economic results were calculated for the entire region. First, the models predict all available sections, and the post-processing of the estimations allows for further inspection of the data. The total annual technical potential of the region was calculated to be 746.5 GWh with an installed capacity of 806.62 MW_p. After removing north-facing sections and area outliers that are more than three standard deviations from the mean, this decreases to 537.0 GWh, with an installed capacity of 578.0 MW_p. According to [63], the yearly electricity consumption in the Aachen city region was around 2930 GWh in 2018, meaning that our potential calculations would cover about 27.5%, or 19.7% (after filtering), respectively. The mean LCOE for the unfiltered data was 72.7 €/MWh, and 69.3 €/MWh (after filtering). Fig. 15 shows the distribution of potentials per rooftop section and the cost potential curve (CPC) of all the sections sorted by LCOE. Post processing enables some interesting additional data calculations, such as the allocation of a fixed budget, with which the most efficient panels can be installed using a simple greedy algorithm. In this case, an initial investment of one million € would allow a total capacity of 1.1 MW to be installed, with a yearly potential of 1.0 GWh and mean LCOE of 63.4 €/MWh.

3.4. Limitations

There is a set of limitations to the current state of the project. First, and perhaps most importantly, the model results in the wild are in some cases not very accurate. Results are expected to be more accurate in larger regions, where errors in specific buildings are balanced out. Performance varies across regions due to the nature of the training data. In the future, building more heterogeneous quality datasets may help improve the performance. Oftentimes, the rooftop segmentation model overestimates large areas by merging a group of buildings together. One workaround to limit this is to filter the buildings that are, for example, three standard deviations away from the mean, removing exceptionally large predictions, as shown in Section 3.2. Ultimately this should be targeted by either improving the performance of the model or developing an algorithm to separate these buildings. In terms of scalability, the technical potential simulation currently limits the scope of the analyses, which otherwise would be worldwide, as the SARAH dataset does not have global coverage. On the topic of superstructure detection, although the results in the test set are promising, existing panels are currently not detected in most cases for random samples in other regions. Two approaches can increase performance in the future: First, training the superstructure model with more labeled data; second, including crowdsourced panel locations in the analysis. However, we believe that the models are starting to perform well enough for certain applications in the energy system domain, where estimations for rooftop PV potential do not yet exist.

4. Summary and conclusions

With satellite availability and computer vision methods developing in recent years, the topic of rooftop photovoltaic potential estimation has the possibility of competing in accuracy with higher quality non scalable approaches like 3D analysis. For this, it is paramount that the research community has access to clearly documented datasets and software and that computational workflows and their results become transparent and reproducible. Existing workflows and approaches have not been found to be sufficiently reproducible, and this is one of the aspects where we aim to contribute.

We presented a conceptual framework for the standardized description of tasks and processes relating to photovoltaic potential estimation.

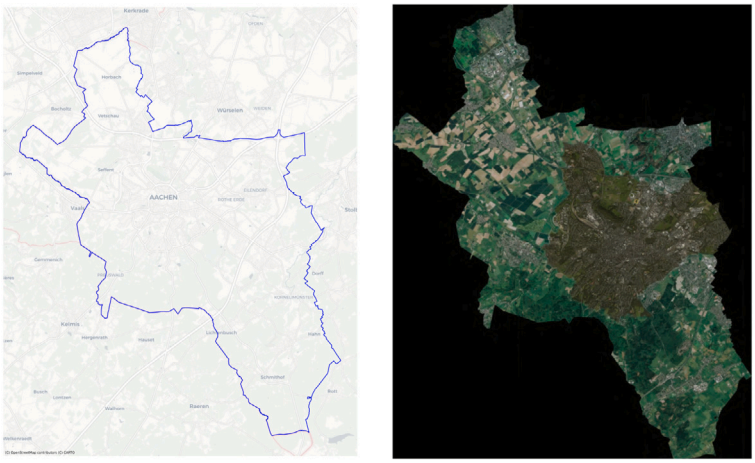


Fig. 13. On the left, the Aachen region administrative boundaries are shown. On the right, the retrieved satellite imagery from Bing Maps is shown.

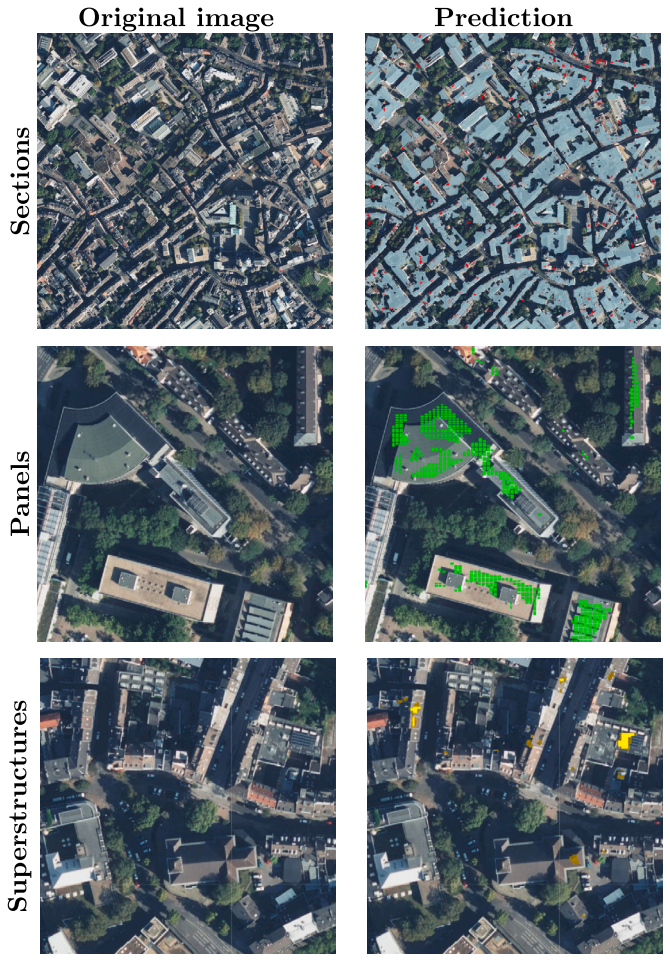


Fig. 14. PASSION sample results for specific parts of the case study of Aachen. The images in the left column represent the base map before the prediction. The first image in the right column shows predicted sections in blue and filtered sections in red. The second image shows predicted PV panels in green. The third shows superstructure predictions in brown, and predicted existing panels in yellow. The background map correspond to Bing Maps imagery. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

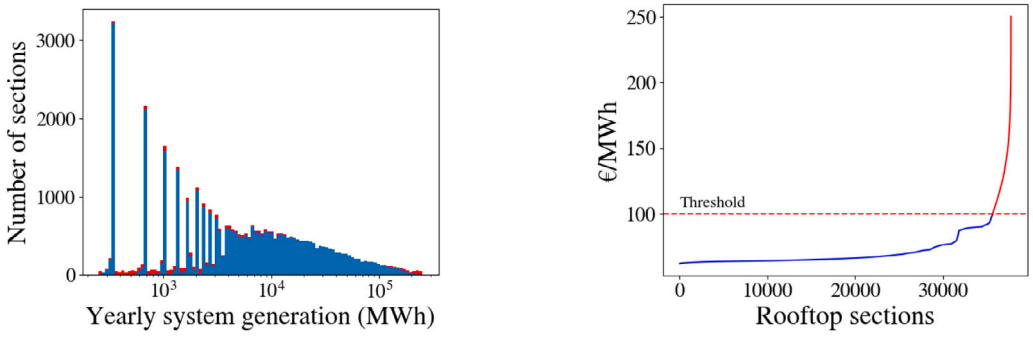
It aims to serve as a common generic reference with which ambiguity in academic information exchange can be reduced. It was utilized in this study to increase the transparency and comparability of the presented work.

Thereafter, we introduced the open-source workflow ETHOS. PASSION,[14] our approach to combining different deep learning models, thus proposing an open workflow with state-of-the-art performance. A comparison with other studies using LiDAR and crowdsourced data was carried out, with the results showing that the performance was comparable. Furthermore, an analysis of a large case study in Germany was performed, proving that the study is scalable to a city region level. When performance is further improved, assessments of broader areas at a national level will be ready to be conducted from a technical perspective. We achieved IoU's of 0.8478, 0.7531 and 0.4927 on the test sets of the rooftop, section and superstructure segmentations respectively. This is comparable to the state of the art of the field for the three tasks. Multiple publicly available datasets were evaluated for our models, showing acceptable performance. The combination of the rooftop and section segmentations improved the results in the wild, where the section segmentation model alone would not produce satisfactory results in regions far from the training data. One of the benefits of releasing an open source workflow is the chance for the community to improve it in a combined effort, rather than publishing redundant works of research. In addition, the proposed conceptual framework aims to better separate the different tasks in the field, so that experts from different fields could propose better approaches using a modular approach.

Our experiments show that the multi-level approach outperforms the single-shot approach, and combining models with different capabilities enrich the final results. It is important to note that better datasets are needed not only for training, but also to serve as a benchmark to evaluate against. On the whole, we believe that the performance of PASSION has started to become sufficiently accurate for real world, practical utilization by researchers or users from other fields. However, this can still be significantly improved, especially for the superstructure segmentation task. More novel approaches can be used for this, but ultimately the best way of improving the performance is by improving the quality of the datasets and gathering new data.

There are many directions for future work in this area, such as improving the performance of the different potential assessments through better estimation of the tilt, accounting for the shadows, detecting the division between buildings or incorporating more relevant public data sources.

[14] https://github.com/FZJ-IEK3-VSA/PASSION



(a) Distribution of section potentials

(b) Cost potential curve (CPC)

Fig. 15. Fig. 15(a) displays the distribution of potentials per section, with sections that were filtered out in red. Fig. 15(b) shows the CPC per rooftop section, after setting a threshold of 100 €/MWh. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

CRediT authorship contribution statement

Rodrigo Pueblas: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization. **Patrick Kuckertz:** Conceptualization, Investigation, Writing – review & editing, Supervision, Project administration. **Jann Michael Weinand:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration. **Leander Kotzur:** Writing – review & editing. **Detlef Stolten:** Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

The ETHOS.PASSION model workflow is published open-source on GitHub (https://github.com/FZJ-IEK3-VSA/PASSION).

Acknowledgments

The authors would like to thank the German Federal Government, the German state governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) – project number: 442146713. Furthermore, the work was supported by the Helmholtz Association, Germany as part of the program, “Energy System Design”. In addition, the authors would like to thank Shruthi Patil for her supervision of the original Master Thesis of this work, and Surya Gunti for his contributions to the model training.

References

[1] D.F. Birol, World Energy Outlook 2022, 524.

[2] L.K. Wiginton, H.T. Nguyen, J.M. Pearce, Quantifying rooftop solar photovoltaic potential for regional renewable energy policy, Comput. Environ. Urban Syst. (ISSN: 0198-9715) 34 (4) (2010) 345–357, Geospatial Cyberinfrastructure. https://www.sciencedirect.com/science/article/pii/S0198971510000025.

[3] R. McKenna, et al., Exploring trade-offs between landscape impact, land use and resource quality for onshore variable renewable energy: An application to Great Britain, Energy (ISSN: 0360-5442) 250 (2022) 123754, https://www.sciencedirect.com/science/article/pii/S0360544222006570.

[4] J.A. Jakubiec, C.F. Reinhart, A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations, Sol. Energy (ISSN: 0038-092X) 93 (2013) 127–143, https://www.sciencedirect.com/science/article/pii/S0038092X13001291.

[5] T. Sun, M. Shan, X. Rong, X. Yang, Estimating the spatial distribution of solar photovoltaic power generation potential on different types of rural rooftops using a deep learning network applied to satellite images, Appl. Energy (ISSN: 03062619) 315 (2022) 119025, https://linkinghub.elsevier.com/retrieve/pii/S0306261922004305.

[6] S. Krapf, L. Bogenrieder, F. Netzler, G. Balke, M. Lienkamp, RID—Roof information dataset for computer vision-based photovoltaic potential assessment, Remote Sens. (ISSN: 2072-4292) 14 (10) (2022) 2299, Number: 10 Publisher: Multidisciplinary Digital Publishing Institute. https://www.mdpi.com/2072-4292/14/10/2299.

[7] B. Chen, et al., An estimation framework of regional rooftop photovoltaic potential based on satellite remote sensing images, Glob. Energy Interconnect. (ISSN: 2096-5117) 5 (3) (2022) 281–292, Publisher: Elsevier.

[8] M. Wilkinson, et al., The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (2016).

[9] P. Kuckertz, et al., A Metadata-Based Ecosystem to Improve the FAIRness of Research Software, 2023.

[10] R. McKenna, et al., High-resolution large-scale onshore wind energy assessments: A review of potential definitions, methodologies and future research needs, Renew. Energy (ISSN: 0960-1481) 182 (2022) 659–684, https://www.sciencedirect.com/science/article/pii/S0960148121014841.

[11] R. McKenna, et al., On the socio-technical potential for onshore wind in Europe: A response to enevoldsen et al. (2019), energy policy, 132, 1092-1100, Energy Policy (ISSN: 0301-4215) 145 (2020) 111693, https://www.sciencedirect.com/science/article/pii/S0301421520304213.

[12] M. Hoogwijk, B. de Vries, W. Turkenburg, Assessment of the global and regional geographical, technical and economic potential of onshore wind energy, Energy Econ. (ISSN: 0140-9883) 26 (5) (2004) 889–919, https://www.sciencedirect.com/science/article/pii/S0140988304000246.

[13] K. Bódis, I. Kougias, A. Jäger-Waldau, N. Taylor, S. Szabó, A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union, Renew. Sustain. Energy Rev. (ISSN: 13640321) 114 (2019) 109309, https://linkinghub.elsevier.com/retrieve/pii/S1364032119305179.

[14] K. Mainzer, et al., A high-resolution determination of the technical potential for residential-roof-mounted photovoltaic systems in Germany, Sol. Energy (ISSN: 0038092X) 105 (2014) 715–731, https://linkinghub.elsevier.com/retrieve/pii/S0038092X14002114.

[15] H. Ohtake, et al., Estimation of satellite–derived regional photovoltaic power generation using a satellite–estimated solar radiation data, Energy Sci. Eng. (ISSN: 2050-0505, 2050-0505) 6 (5) (2018) 570–583, https://onlinelibrary.wiley.com/doi/10.1002/ese3.233.

[16] M. Gutschner, S. Nowak, D. Ruoss, P. Toggweiler, T. Schoen, Potential for building integrated photovoltaics, Int. Energy Agency Potential Build. Integr. Photovolt (2002).

[17] X. Song, et al., An approach for estimating solar photovoltaic potential based on rooftop retrieval from remote sensing images, Energies (ISSN: 1996-1073) 11 (11) (2018) 3172, http://www.mdpi.com/1996-1073/11/11/3172.

[18] K. Mainzer, S. Killinger, R. McKenna, W. Fichtner, Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques, Sol. Energy (ISSN: 0038092X) 155 (2017) 561–573, https://linkinghub.elsevier.com/retrieve/pii/S0038092X17305686.

[19] A. Sampath, P. Bijapur, A. Karanam, V. Umadevi, M. Parathodiyil, Estimation of rooftop solar energy generation using satellite image segmentation, in: Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019, 2019.

[20] S. Lee, S. Iyengar, M. Feng, P. Shenoy, S. Maji, DeepRoof: A data-driven approach for solar potential estimation using rooftop imagery, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, Anchorage AK USA, ISBN: 978-1-4503-6201-6, 2019, pp. 2105–2113, https://dl.acm.org/doi/10.1145/3292500.3330741.

[21] J. Byrne, J. Taminiau, L. Kurdgelashvili, K.N. Kim, A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of Seoul, Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 41 (2015) 830–844, https://www.sciencedirect.com/science/article/pii/S1364032114006996.

[22] S. Castellanos, D.A. Sunter, D.M. Kammen, Rooftop solar photovoltaic potential in cities: How scalable are assessment approaches? Environ. Res. Lett. (ISSN: 1748-9326) 12 (12) (2017) 125005, https://iopscience.iop.org/article/10.1088/1748-9326/aa7857.

[23] P. Saeedi, H. Zwick, Automatic building detection in aerial and satellite images, in: Robotics and Vision 2008 10th International Conference on Control, Automation, 2008, pp. 623–629.

[24] Z. Liu, S. Cui, Q. Yan, Building extraction from high resolution satellite imagery based on multi-scale image segmentation and model matching, in: 2008 International Workshop on Earth Observation and Remote Sensing Applications, 2008, pp. 1–7.

[25] R. Singh, R. Banerjee, Estimation of roof-top photovoltaic potential using satellite imagery and GIS, in: 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC), 2013, pp. 2343–2347, ISSN: 0160-8371.

[26] Willkommen beim QGIS Projekt! https://www.qgis.org/de/site/.

[27] J. Ma, et al., Building extraction of aerial images by a global and multi-scale encoder-decoder network, Remote Sens. (ISSN: 20724292) 12 (15) (2020).

[28] V.M. Phap, N.T. Thua Huong, P.T. Hanh, P. Vana Duy, D. Vana Binh, Assessment of rooftop solar power technical potential in Hanoi city, Vietnam, J. Build. Eng. (ISSN: 2352-7102) 32 (2020) 101528, Publisher: Elsevier.

[29] S. Krapf, et al., Towards scalable economic photovoltaic potential analysis using aerial images and deep learning, Energies (ISSN: 1996-1073) 14 (13) (2021) 3800, Number: 13 Publisher: Multidisciplinary Digital Publishing Institute. https://www.mdpi.com/1996-1073/14/13/3800.

[30] T. Zhong, et al., A city-scale estimation of rooftop solar photovoltaic potential based on deep learning, Appl. Energy (ISSN: 03062619) 298 (2021) 117132, https://linkinghub.elsevier.com/retrieve/pii/S0306261921005729.

[31] H. Jiang, et al., Geospatial assessment of rooftop solar photovoltaic potential using multi-source remote sensing data, Energy AI (ISSN: 2666-5468) 10 (2022) 100185, Publisher: Elsevier.

[32] S. Ren, et al., Automated extraction of energy systems information from remotely sensed data: A review and analysis, Appl. Energy (ISSN: 0306-2619) 326 (2022) 119876, Publisher: Elsevier.

[33] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015, arXiv:1505.04597 [cs]. http://arxiv.org/abs/1505.04597.

[34] R. Rew, G. Davis, NetCDF: An interface for scientific data access, IEEE Comput. Graph. Appl. (ISSN: 0272-1716) 10 (4) (1990) 76–82, http://ieeexplore.ieee.org/document/56302/.

[35] D.S. Ryberg, et al., The future of European onshore wind energy potential: Detailed distribution and simulation of advanced turbine designs, Energy (ISSN: 0360-5442) 182 (2019) 1222–1238, https://www.sciencedirect.com/science/article/pii/S0360544219311818.

[36] R. Gelaro, et al., The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Clim. (ISSN: 0894-8755) 30 (14) (2017) 5419–5454, ISSN: 1520-0442 https://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0758.1.

[37] U. Pfeifroth, et al., Surface radiation data set - Heliosat (SARAH) - Edition 2, 2017, Artwork Size: 7.1 TiB Medium: NetCDF-4 Pages: 7.1 TiB Version Number: 2.0 Type: dataset. https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V002.

[38] H. Hersbach, et al., The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. (ISSN: 0035-9009) 146 (730) (2020) 1999–2049, ISSN: 1477-870X. https://onlinelibrary.wiley.com/doi/10.1002/qj.3803.

[39] C. Chen, et al., Aerial Imagery for Roof Segmentation: A Large-Scale Dataset towards Automatic Mapping of Buildings, 2018.

[40] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark, 2017, pp. 3226–3229.

[41] 3D building models, 2023, https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/3d_gebaeudemode.

[42] A. Bandam, E. Busari, C. Syranidou, J. Linssen, D. Stolten, Classification of building types in Germany: A data-driven modeling approach, Data (ISSN: 2306-5729) 7 (4) (2022) 45, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. https://www.mdpi.com/2306-5729/7/4/45.

[43] Openstreetmap contributors, 2017, Planet dump retrieved from https://planet.osm.org, https://www.openstreetmap.org.

[44] Q. Zhou, Y. Zhang, K. Chang, M.A. Brovelli, Assessing OSM building completeness for almost 13,000 cities globally, Int. J. Digit. Earth (ISSN: 1753-8947) 15 (1) (2022) 2400–2421, Publisher: Taylor & Francis. https://www.tandfonline.com/doi/full/10.1080/17538947.2022.2159550.

[45] D. Yang, J.M. Bright, Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years, Sol. Energy (ISSN: 0038-092X) 210 (2020) 3–19, Special Issue on Grid Integration. https://www.sciencedirect.com/science/article/pii/S0038092X20303893.

[46] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, arXiv:1512.03385 [cs]. http://arxiv.org/abs/1512.03385.

[47] S. Suzuki, K. be, Topological structural analysis of digitized binary images by border following, Comput. Vis. Graph. Image Process. (ISSN: 0734-189X) 30 (1) (1985) 32–46, https://www.sciencedirect.com/science/article/pii/0734189X85900167.

[48] G. Bradski, The OpenCV library, Dr. Dobb's J. Software Tools (2000).

[49] C.-H. Teh, R.T. Chin, On the detection of dominant points on digital curve, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 859–872.

[50] R.O. Duda, Pattern Classification and Scene Analysis, Wiley, New York, ISBN: 978-0-471-22361-0, 1973, http://archive.org/details/patternclassific0000duda.

[51] W.F. Holmgren, C.W. Hansen, M.A. Mikofski, Pvlib Python: A python package for modeling solar energy systems, J. Open Source Software (ISSN: 2475-9066) 3 (29) (2018) 884, http://joss.theoj.org/papers/10.21105/joss.00884.

[52] S. Hoyer, J. Hamman, Xarray: N-D labeled arrays and datasets in Python, J. Open Res. Software (ISSN: 2049-9647) 5 (1) (2017) 10, https://openresearchsoftware.metajnl.com/article/10.5334/jors.148/.

[53] M. Papapetrou, G. Kosmadakis, in: A. Tamburini, A. Cipollina, G. Micale (Eds.), Salinity Gradient Heat Engines, in: Woodhead Publishing Series in Energy, Woodhead Publishing, ISBN: 978-0-08-102847-6, 2022, pp. 319–353, https://www.sciencedirect.com/science/article/pii/B9780081028476000006.

[54] Welcome to Flask — Flask documentation (2.2.x). https://flask.palletsprojects.com/en/2.2.x/.

[55] Folium — Folium 0.14.0 documentation. https://python-visualization.github.io/folium/.

[56] F. Mölder, K.P. Jablonski, B. Letcher, M.B. Hall, Sustainable data analysis with Snakemake | F1000Research, 2021, https://f1000research.com/articles/10-33/v1.

[57] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, arXiv:1412.6980 [cs]. http://arxiv.org/abs/1412.6980.

[58] H. Rezatofighi, et al., Generalized intersection over union: A metric and a loss for bounding box regression, 2019, arXiv:1902.09630 [cs]. http://arxiv.org/abs/1902.09630.

[59] P. Jaccard, The distribution of the Flora in the Alpine Zone.1, New Phytol. (ISSN: 1469-8137) 11 (2) (1912) 37–50, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.1912.tb05611.x.

[60] V. Adjiski, G. Kaplan, S. Mijakoski, Assessment of the solar energy potential of rooftops using LiDAR datasets and GIS based approach, Int. J. Eng. Geosci. 8 (2) (2023) 188–199, Number: 2. https://dergipark.org.tr/en/pub/ijeg/issue/73074/1112274.

[61] B.B. Kausika, et al., Bottom-up analysis of the solar photovoltaic potential for a city in the Netherlands: A working model for calculating the potential using high resolution LiDAR data, in: 2015 International Conference on Smart Cities and Green ICT Systems, SMARTGREENS, 2015, pp. 1–7.

[62] Fläche und Bevölkerungsdichte, Nordrhein-Westfalen nach Verwaltungs-bezirken, 2015 - 2017. https://www.lzg.nrw.de/00indi/0data/02/html/0200501052017.html.

[63] Regionaler dialog energiewende (render) | 2018. https://regionaler-dialog-aachen.de/2018-heute.