Quantification of existing rooftop PV hourly generation capacity and validation against measurement data

Alina Walch¹, Martin Rüdisüli², Roberto Castello¹, Jean-Louis Scartezzini¹

¹ Solar Energy and Building Physics Laboratory, EPFL, Lausanne, Switzerland
² Urban Energy Systems Laboratory, EMPA, Dübendorf, Switzerland

Corresponding author: alina.walch@epfl.ch

Abstract. The spatio-temporal patterns of the existing rooftop photovoltaic (PV) production can provide valuable insights for the design of effective strategies for PV integration at large scale. In this work, we quantify the hourly production of existing rooftop PV installations by combining two large-scale methods, for detecting rooftop solar panels and for estimating their hourly PV generation, which are both applicable at the Swiss national scale. A validation against measured data of PV installations from 16 roofs in the Swiss Canton of Aargau shows that the existing PV area is detected accurately, while the hourly profiles overestimate the PV production, with higher errors in winter than in summer. These errors result in a mean overestimation of the annual PV production by 16%, assuming a triangular panel placement at low tilt (15°) on flat roofs.

1. Introduction
Solar photovoltaics (PV) are considered as one of the key technologies in Switzerland’s renewable energy strategy, which aims for an electricity production of 34 TWh by 2050 [1]. Achieving this ambitious goal requires the large-scale deployment of PV panels, particularly on rooftops. To design effective pathways towards an electricity supply with a high share of PV, we hence need to not only identify roofs with a high potential, but also assess the spatial and temporal patterns of the electricity production of existing PV capacities. These may indicate (i) socio-economic factors that support PV deployment, (ii) areas with recent growth in PV and (iii) current and potential future challenges of distributed renewable energy generation for the energy system. Despite some estimation of the total installed PV capacities in the country, their spatial and temporal patterns are to date unknown.

In this work, we propose an approach for quantifying the hourly production of existing rooftop PV installations, which is scalable to the national level for Switzerland, and validate the results against measurements. The approach combines, for the first time, a detection of the size and location of existing rooftop solar panels [2] with the estimation of hourly PV electricity generation [3]. We further expand the method from [3] to compare two scenarios for PV panel placement on flat roofs, as shedded rows or as east-west facing triangular rows. The estimated PV area and hourly production are compared to real electricity generation data from 16 roofs with 31 individual roof parts in the Swiss Canton of Aargau (AG). Despite the relatively small sample size, the validation indicates systematic differences between the estimation and the actual generation from rarely available measurement data. By combining two methods that are only addressed separately in the literature [4,5], this work can be used to quantify the current PV production at a high spatio-temporal resolution anywhere in Switzerland, whereby the validation results may serve as a basis to quantify the uncertainty related to the hourly PV yield.
2. Methods and data

2.1. Size and location of existing PV installations
The size and location of existing PV installations is estimated through a Machine Learning (ML)-based framework, which is used to compile a rooftop solar registry of Switzerland [2]. It combines high resolution images at 0.25 m, collected during a multi-year aerial campaign by the Swiss Federal Office of Topography (swisstopo) [6], and a Convolutional Neural Network (CNN) for pixel-wise semantic image segmentation to locate and quantify the size of rooftop solar installations. The strength of this method resides in the combination of transfer learning with an encoder-decoder architecture (U-Net), which allows to drastically reduce the number of labelled images used for training the algorithm. The use of geospatial tools for post-processing further reduces the number of false positives by eliminating predictions outside of rooftops, by aggregating the detected pixels to installations and by attributing them to the existing building stock. The framework hence allows to detect even small installations (< 30 kWp) which are not necessarily recorded by accredited certification bodies such as Pronovo AG or the Swiss Federal Inspectorate for Heavy Current Installations (ESTI).

As the aerial images for the measured buildings in the Canton of Aargau date from 2016, only installations older than this date are considered in the validation of the PV installation size.

2.2. Hourly rooftop PV electricity generation profiles
The hourly rooftop PV yield of the existing panels is estimated using the method proposed by Walch et al. [3], which consists of three steps: First, the physical potential is obtained from the incoming direct and global horizontal solar radiation and the surface reflectance (albedo). To be comparable to the validation data, we use the recorded hourly solar radiation from satellites during 2018 and 2019 [7].

Second, the geographical potential is obtained by multiplying the size of the PV installation (see Section 2.1) with the hourly rooftop tilted radiation. The tilted radiation estimation accounts for roof tilt and orientation, for shading effects and for sky visibility (see [3] for details). As PV panels on flat roofs are installed in various configurations, we model two scenarios for flat roofs: (1) In the “base” scenario, panels on fully flat surfaces (0°) are placed in south-facing shed rows, tilted at 30° and spaced apart by one projected panel width (~ 87 cm). This PV placement aims to maximize the electricity yield per panel with a realistic spacing between adjacent panel rows. (2) For practical reasons and to maximize the used PV surface, existing panels are often placed in adjacent and alternating, (approximately) east and west-facing rows at lower tilt. The “EW” scenario hence models panels on flat roofs as alternating east and west-facing rows at 15° tilt, whereby all roofs with tilt angles below 10° are defined as flat.

Finally, the technical potential is obtained by multiplying the geographical potential with the panel and inverter efficiency reported in [3], which assumes that all PV panels are mono-crystalline panels. In contrast to [3], we do not account for additional losses due to soiling, wiring, system availability etc., as most analysed systems are new and we consider these losses to be negligible. For the comparison with the real data, the electricity yield is normalized by the PV installation size of each sample roof.

2.3. Measurement data
The measurement data used for the comparison has been provided by IBW, the local utility of Wohlen (AG), which owns about a dozen PV installations placed on own and customer roofs. Five of them were installed before 2016, thus their installed PV area can be compared against the CNN-based PV detection described in Section 2.1 (referred to as “CNN algorithm”). IBW also provides 15 hourly PV generation profiles (including third-party installations) of 2018 and 2019, which are used to validate the hourly PV profiles (see Section 2.2). To increase the number of roofs for validating the CNN algorithm, the actual PV area on five of these third-party PV installation sites is quantified visually from aerial images of swisstopo, by counting the number of panels and multiplying them by 1.6 m², which is a typical size of PV panels [3]. Thus, 16 roofs with PV installations are considered in total, consisting of 31 individual surfaces. Table 1 summaries the most relevant characteristics of these roofs (and roof parts) as well as their use in this study. For reasons of confidentiality, no further information (e.g. address) can be given.
Table 1. Characteristics of the roofs (and roof parts) used in this study. CNN: Used for validation of the CNN algorithm (Section 2.1). Profile: Used to validate modelled hourly PV generation profile (Section 2.2). Source area: Sources of PV area, provided by IBW\(^a\)/owner or derived from swisstopo\(^b\).

| ID | Part | Tilt [°] | Orientation [°] | Total area [m\(^2\)] | PV mount | Source area | CNN | Profile |
|----|------|---------|----------------|----------------------|----------|-------------|-----|---------|
| 5  | a/b/c/d | 36/35/36/35 | 40/-140/40/-140 | 1391 | pitched | IBW / swisstopo | X | X |
| 6  | 8     | 45      |                | 314      | flat     | IBW         | X | X |
| 8  | a/b/c | 12 / 13 / 11 | -20/160/-20 | 634      | pitched | IBW         | X | X |
| 9  | a/b/c/d | 19/18/29/9 | 50/-130/140/140 | 781 | pitched | IBW         | X | X |
| 10 | 0     |         |                | 619      | triangular | IBW     | X | X |
| 11 | a/b   | 34/18   | -3/-3         | 446      | pitched | IBW         | X | X |
| 16 | 0     |         |                | 577      | shed     | IBW         | X | X |
| 18 | 7     | 135     |                | 1288     | flat     | IBW         | X | X |
| 19 | 0     |         |                | 628      | triangular | IBW     | X | X |
| 20 | a/b/c | 20/20/24 | 129/-51/-170  | 1160     | pitched | IBW         | X | X |
| 22 | a/b/c/d | 6/6/8/8 | 44/-136/44/-136 | 4404 | flat     | swisstopo   | X | X |
| 26 | 5     | -118    |                | 1829     | flat     | Owner       | X | X |
| 29 | 30    | 10      |                | 321      | pitched | swisstopo   | X | X |
| 32 | 3     | -15     |                | 25943    | triangular | swisstopo | X | X |
| 35 | a/b   | 4/6     | 160/160        | 470      | flat     | swisstopo   | X | X |
| 36 | 0     |         |                | 1883     | triangular | swisstopo | X | X |

\(^a\) https://www.ibw.ag/dynasite.cfm?dsmid=507635, \(^b\) https://map.geo.admin.ch/

3. Results and discussion

3.1. Size and location of existing PV installations

A qualitative comparison of the PV area detected by the CNN algorithm and the real area shows that the CNN algorithm performs very well in most cases (8/10 roofs), as shown in Figure 1b and c for roof IDs 10 and 36 (both flat). In two cases, it fails to detect the full installation (see Figure 1a for ID 9), partly due to the high reflection on the aerial images (\(g_{\text{n}}\)), which is one of the limitations of the CNN approach, and partly to the model sensitivity, which is estimated to be 84% for the class of pixels belonging to a PV installation on the image [2]. Further known limitations include the differentiation between solar thermal and PV panels and the detection of building-integrated PV, which are not present in the training dataset and could hence not be assessed by the CNN algorithm.

The above results are mostly confirmed when comparing estimated and real installed areas, shown in Table 2 as percentage of the total roof area for all roof IDs. This percentage represents the roof area that is not or cannot be covered by PV either due to obstructing superstructures and/or any other reasons (such as static, economic, design, etc.). Except for IDs 5 and 9, where the CNN algorithm cannot detect a significant part of the PV panels (see Figure 1), there is a good match between the actual and detected PV area (< 15% deviation). Out of those, the cases with a high deviation (ID 22, 29, 36) occur for roofs where the PV area has been derived visually from aerial images, which may lead to discrepancies if the actual surface per panel is not equal to 1.6 m\(^2\) (see Section 2.3). While these results are not representative due to the small sample containing only large installations (> 30 kW\(_p\)), they suggest that the PV detection algorithm underestimates the real PV area by 3% of the roof surface on average, with 18% deviation.

A comparison of the two PV scenarios for flat roofs (see Table 1) shows that in most cases, PV panels on roofs with tilt angles below 5° are installed in a triangular arrangement, corresponding to the EW scenario from Section 2.2. On those roofs with tilts of 5°-10°, panels are mostly installed flat on the roof (corresponding to the base scenario). Only one roof uses a shed arrangement, which describes PV panels placed in individual rows of tilted panels (base scenario). The validation results suggest that the EW scenario represents a more accurate approach for estimating the real PV yield of flat roofs.
mean absolute percentage error (wMAPE) unaccounted for phenomena in the model such as snow cover on PV panels, more extensive near-object shading, etc. Based on these patterns, a better calibration of the model can be conducted.

To assess the quality of the modelled hourly PV generation profiles, the normalized (scaled to kWh/m²) hourly PV generation profiles from actual measurements and modelled by the base and EW scenario are compared in Figure 2 for a selection of two flat (ID 19, 36, EW scenario) and one pitched (ID 9) roofs for five consecutive days in winter and summer in 2018. For pitched roofs, the base and EW scenarios are identical (see Section 2.2). From a visual inspection, the EW scenario matches the real profiles more accurately, which is expected for triangularly mounted panels, confirming that the lower tilt of 15° is a more realistic assumption for flat roofs. Moreover, the match of the modeled profiles is generally better in summer than in winter. This is in particular the case for entirely sunny (clear-sky) days, while for cloudy or intermittently sunny days both models become less accurate. This may be due to discrepancies between the actual solar radiation and the satellite data, caused by local cloud coverage. Further differences may arise for example from inaccuracies in the PV geometries or from unexpected shading.

To see diurnal patterns, hourly differences of the actual and modelled (only EW) profiles are shown in Figure 3 as boxplots for five selected roofs per season. To avoid large outliers, hours with missing values have been discarded. Figure 3 shows no systematic bias for some roofs (e.g. ID 26), while for other roofs the model generally underestimates (e.g. ID 8) or overestimates (e.g. ID 16) the actual PV generation. For ID 16, this overestimation takes place more prominently in the morning hours and levels off towards the evening. An opposite pattern can be observed for ID 22, where the model underestimates the actual generation in the morning and overestimates generation in the afternoon (and evening). Outliers are typically large in winter, even though absolute solar irradiance and consequently the average (median) differences are smaller in winter. This is in line with Figure 2, where in some days in winter, the modelled and the actual generation are off by a relatively large amount. This may be due to unaccounted for phenomena in the model such as snow cover on PV panels, more extensive near-object shading, etc. Based on these patterns, a better calibration of the model can be conducted.

A quantitative comparison of the actual and modelled (normalized) PV generation, using a weighted mean absolute percentage error (wMAPE), is shown in Table 3 for hourly (top) and daily (bottom) PV

Table 2. Percentage of the total roof area covered by PV for the actual and detected area (using CNN). Satellite images of roofs (ID 9, 10, 36) are also depicted in Fig. 1. Sources for the actual PV area are IBW and visual estimations from swisstopo aerial images (italic values).

| Roof ID | < 10° (flat) | > 10° (pitched) | mean ± std |
|---------|--------------|-----------------|------------|
|         | 6 10 22 32 35 36 | 5 8 9 29 |           |
| Actual area (%) | 70 38 60 74 57 55 | 74 76 92 77 | 67 ± 16 |
| Detected area (%) | 71 36 69 76 58 69 | 56 74 46 88 | 64 ± 15 |
| Difference (%) | +1 -2 +9 +2 +1 +14 | -19 -2 -46 +11 | -3 ±18 |

3.2. Hourly profiles of existing PV installations

To assess the quality of the modelled hourly PV generation profiles, the normalized (scaled to kWh/m²) hourly PV generation profiles from actual measurements and modelled by the base and EW scenario are compared in Figure 2 for a selection of two flat (ID 19, 36, EW scenario) and one pitched (ID 9) roofs for five consecutive days in winter and summer in 2018. For pitched roofs, the base and EW scenarios are identical (see Section 2.2). From a visual inspection, the EW scenario matches the real profiles more accurately, which is expected for triangularly mounted panels, confirming that the lower tilt of 15° is a more realistic assumption for flat roofs. Moreover, the match of the modeled profiles is generally better in summer than in winter. This is in particular the case for entirely sunny (clear-sky) days, while for cloudy or intermittently sunny days both models become less accurate. This may be due to discrepancies between the actual solar radiation and the satellite data, caused by local cloud coverage. Further differences may arise for example from inaccuracies in the PV geometries or from unexpected shading.

To see diurnal patterns, hourly differences of the actual and modelled (only EW) profiles are shown in Figure 3 as boxplots for five selected roofs per season. To avoid large outliers, hours with missing values have been discarded. Figure 3 shows no systematic bias for some roofs (e.g. ID 26), while for other roofs the model generally underestimates (e.g. ID 8) or overestimates (e.g. ID 16) the actual PV generation. For ID 16, this overestimation takes place more prominently in the morning hours and levels off towards the evening. An opposite pattern can be observed for ID 22, where the model underestimates the actual generation in the morning and overestimates generation in the afternoon (and evening). Outliers are typically large in winter, even though absolute solar irradiance and consequently the average (median) differences are smaller in winter. This is in line with Figure 2, where in some days in winter, the modelled and the actual generation are off by a relatively large amount. This may be due to unaccounted for phenomena in the model such as snow cover on PV panels, more extensive near-object shading, etc. Based on these patterns, a better calibration of the model can be conducted.

A quantitative comparison of the actual and modelled (normalized) PV generation, using a weighted mean absolute percentage error (wMAPE), is shown in Table 3 for hourly (top) and daily (bottom) PV
yields in 2018 and 2019 per season, for base and EW scenarios. The weighting uses measured absolute PV generation of the whole year to give higher weights to deviations when PV generation is large. For better representativity, missing values and night hours have been discarded. In agreement with Figure 2, the EW scenario generally outperforms the base scenario for flat roofs. The wMAPE is substantially higher in winter with a daily mean of 31% (EW), while in other seasons its daily mean is of 14%-21% and 18%-31% for the EW and base scenarios, respectively. Hourly wMAPEs are generally 10%-20% higher than the daily values, as small temporal shifts between the production profiles can be balanced across the day. The yearly production of the estimation lies on average 11% above the measured values (EW), possibly due to different panel efficiencies, unaccounted losses, or overestimated solar radiation.

The overestimation of the annual normalized PV production also dominates the error in the absolute annual PV yield (in kWh), which lies at 16% above the measured data, computed across 9 roofs where CNN area and hourly PV estimates are available. This error is dominated by deviations in the PV yield for large flat roofs, which reinforces the importance of assessing different scenarios for PV on flat roofs.

**Figure 2.** Comparison of the normalized actual (yellow areas) and modelled PV profiles (base as dashed lines, EW as solid lines) for two flat (ID 19, 36) and one pitched (ID 9) roof for five days in 2018.

**Figure 3.** Boxplots of the hourly differences between modelled (only EW scenario) and the actual PV generation on five roofs (ID) divided by season. Outliers due to missing data have been discarded.
Table 3. Weighted mean absolute percentage error (wMAPE, in %) between measured and modelled hourly PV profiles of roofs (ID) per season. Errors for flat roofs are shown for both scenarios, as base/EW. The column “all” shows the mean of all roofs (for both scenarios of flat roofs).

| Roof ID | Hourly | Daily |
|---------|--------|-------|
|         | < 10° (flat: base/EW for > 5% difference) | > 10° (pitched) | all |
| WIN    | 136/46 | 93/35 | 104/37 | 127/45 | 84 | 83 | 88 | 82 | 73 | 70/50 |
| SPR    | 48/22  | 28/18 | 23/18  | 34/18  | 53 | 28 | 22 | 23 | 18 | 28 | 32/26 |
| SUM    | 39/20  | 19/14 | 21/18  | 26/13  | 49 | 24 | 18 | 21 | 14 | 22 | 27/23 |
| FAL    | 80/27  | 49/21 | 28/18  | 63/23  | 71 | 45 | 23 | 27 | 20 | 42 | 43/31 |
| Year   | 56/24  | 33/18 | 24/18  | 41/18  | 57 | 33 | 22 | 25 | 18 | 33 | 35/27 |

|         | 10     | 16     | 18     | 19     | 22     | 26     | 32     | 35     | 36     | 5     | 8     | 9     | 11     | 20     | 29     |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| WIN    | 126/36 | 82/25  | 31/26  | 31/27  | 27/47  | 61/116 | 71     | 74     | 27     | 30     | 30     | 61     | 48/31  |
| SPR    | 41/14  | 20/12  | 16/27  | 11/15  | 11/44  | 22/47  | 48     | 22     | 11     | 16     | 13     | 20     | 19/14  |
| SUM    | 33/15  | 13/9   | 16/18  | 13/8   | 8/80   | 10/37  | 46     | 20     | 8      | 16     | 9      | 17     | 18/14  |
| FAL    | 75/21  | 43/13  | 22/57  | 15/21  | 13/40  | 31/78  | 66     | 39     | 11     | 20     | 14     | 36     | 31/19  |
| Year   | 50/17  | 26/12  | 18/33  | 11/16  | 12/58  | 22/54  | 52     | 27     | 11     | 18     | 13     | 26     | 28/21  |

4. Conclusion
This work presents an approach to estimate the spatio-temporal patterns of rooftop PV generation from current installations by combining existing methods scalable to the Swiss national level. A comparison to measured data of 16 roofs shows that the PV area is identified accurately in most cases, while the real production is overestimated, with higher errors in winter than in summer. The modelled annual PV yield lies 16% above the measured data, assuming a triangular PV placement at 15° tilt on flat roofs. The errors are to be treated with caution as the sample size is relatively small and there are several unknowns.

As the detection of the existing PV areas and the estimation of the hourly potential electricity generation from PV are large-scale methods, the proposed approach can be used to estimate the existing PV production for the entire Switzerland. These roofs can hence be excluded from future PV potential studies and to quantify remaining exploitable surfaces on already occupied roofs, which may play a key role in future PV integration strategies.

Acknowledgements
The authors thank IBW for providing the measured PV data. This research is supported by the Swiss National Science Foundation (SNSF) under the National Research Program 75 (Big Data).

References
[1] BFE 2020 Energieperspektiven 2050+. Zusammenfassung der wichtigsten Ergebnisse. (Bern, Switzerland: Bundesamt für Energie BFE)
[2] Castello R, Roquette S, Eguerra M, Guerra A and Scartezzini J-L 2019 Deep learning in the built environment: automatic detection of rooftop solar panels using Convolutional Neural Networks J. Phys.: Conf. Ser. 1343 012034
[3] Walch A, Castello R, Mohajeri N and Scartezzini J-L 2020 Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty Applied Energy 262 114404
[4] Buffat R, Grassi S and Raubal M 2018 A scalable method for estimating rooftop solar irradiation potential over large regions Applied Energy 216 389–401
[5] Yu J, Wang Z, Majumdar A and Rajagopal R 2018 DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States Joule 2 2605–17
[6] Swiss Federal Institute of Topography swisstopo SWISSIMAGE 25
[7] Stöckli R 2013 Daily, monthly and yearly satellite-based global radiation (Federal Office of Meteorology and Climatology MeteoSwiss)