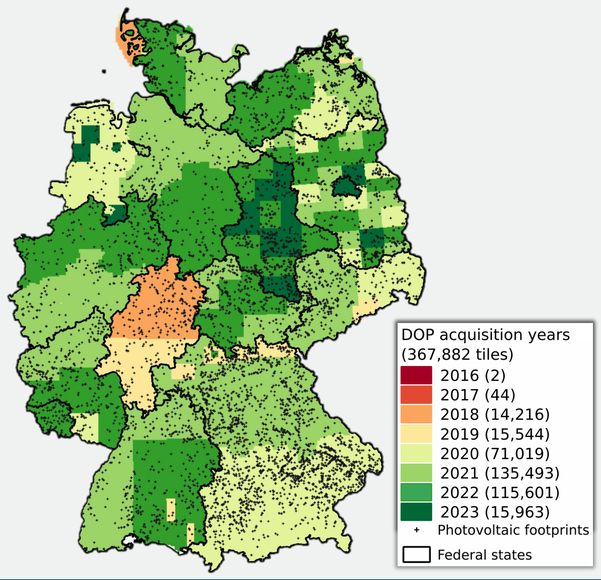
Creating a Dataset of Spatial Parameters of Ground-Mounted Photovoltaic Systems Utilising Orthophotos and the Segment Anything Model

Albert Johannes[1]; Schymik Chantal[1]; Gärtner Philipp[1]; Klingner Stephan[1]; Siegismund Jan[1]; Wehner Claudius[1]; Bunzel Katja[2]; Kempt Jeremias[2]; Höfer Rene[2]

[1] German Environment Agency, [2] Federal Agency for Nature Conservation, Germany

Motivation and data situation

- Renewable Energy Sources Act (EEG 2023) plans to increase Germany's photovoltaic systems capacity from 81.7 GW in 2023 to 400 GW by 2040
- lack of precise, up-to-date spatial data on photovoltaic installations complicates the assessment of ongoing developments and required readjustment of regulatory frameworks
- Marktstammdatenregister (central, freely accessible registry of all energy-related installations in Germany) only provides point location data for photovoltaic systems, thus lacks detailed spatial information, like ground coverage ratio of the photovoltaic module rows, their spacing and their exact orientation
- this study addresses this gap by aiming to extract detailed spatial parameters for all ground-mounted photovoltaic (GMPV) systems in Germany



Data sources

1. **GMPV footprints**: open-access dataset provided by Manske (2025) with manually digitised outlines of 8,789 GMPV footprints based on the Marktstammdatenregister in Germany as of 4th January 2024
2. **Digital orthophotos**: distortion-free aerial imagery with 0.2 m ground resolution covering Germany with acquisition dates varying by federal state and spectral channels RGB and NIR
3. **Photovoltaic parks**: 6,479 park enclosing polygons derived from OpenSteetMap using photovoltaic-specific tags, representing structural or natural boundaries

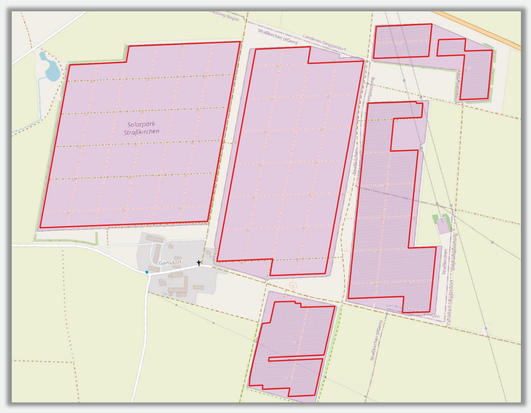
Methodology

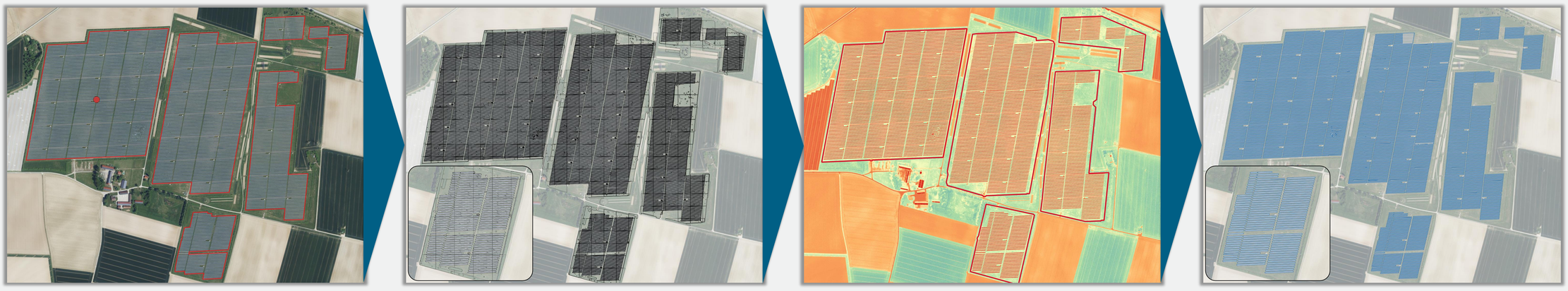
Photovoltaic module rows

- grid of overlapping RGB DOP image chips (227,000) around every photovoltaic footprint
- Segment Anything Model (SAM) for zero-shot image segmentation with pre-trained ViT-H backbone and tuned hyperparameters was applied on image chips
- execution in parallelised batch processing on HPC cluster with NVIDIA Tesla A30 GPUs
- calculation of spectral and geometric features to filter out unwanted outlier polygons
- clustering with DBSCAN algorithm based on spectral features
- separation of photovoltaic module row clusters from irrelevant noise cluster and background clusters through spectral and topographic filtering
- final data product enriched with the attributes width, length, area, rectangularity, orientation, nearest neighbour distance and the DOP acquisition date of each module row
- visual interpretation based evaluation of the data products showed that over 71 % of the photovoltaic footprints have satisfactory to flawless module row polygons

Photovoltaic parks

- park boundaries were obtained using the OpenStreetMap Overpass API
- identification of suited database tags for photovoltaic parks in comprehensive exploratory manner and manual inspection
- filtering and refinement process to ensure data quality and relevance
- sample evaluation showed, that 76 % of polygons are suitable for depicting park boundaries

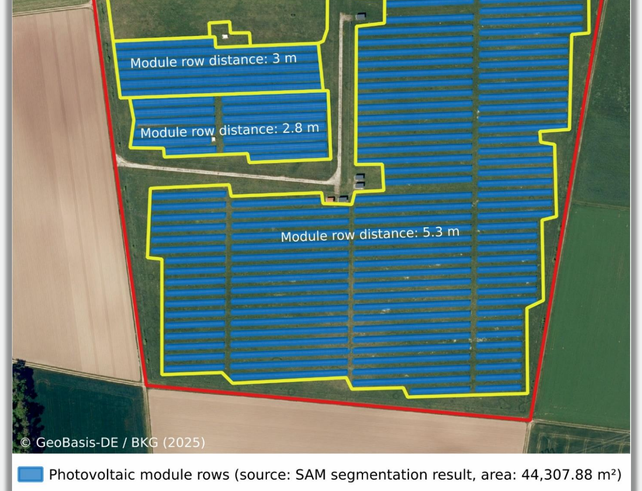




Results

Geospatial data products freely available on Zenodo:

1. **Photovoltaic system footprints**
2. **Photovoltaic module rows**
3. **Photovoltaic parks**



Dataset

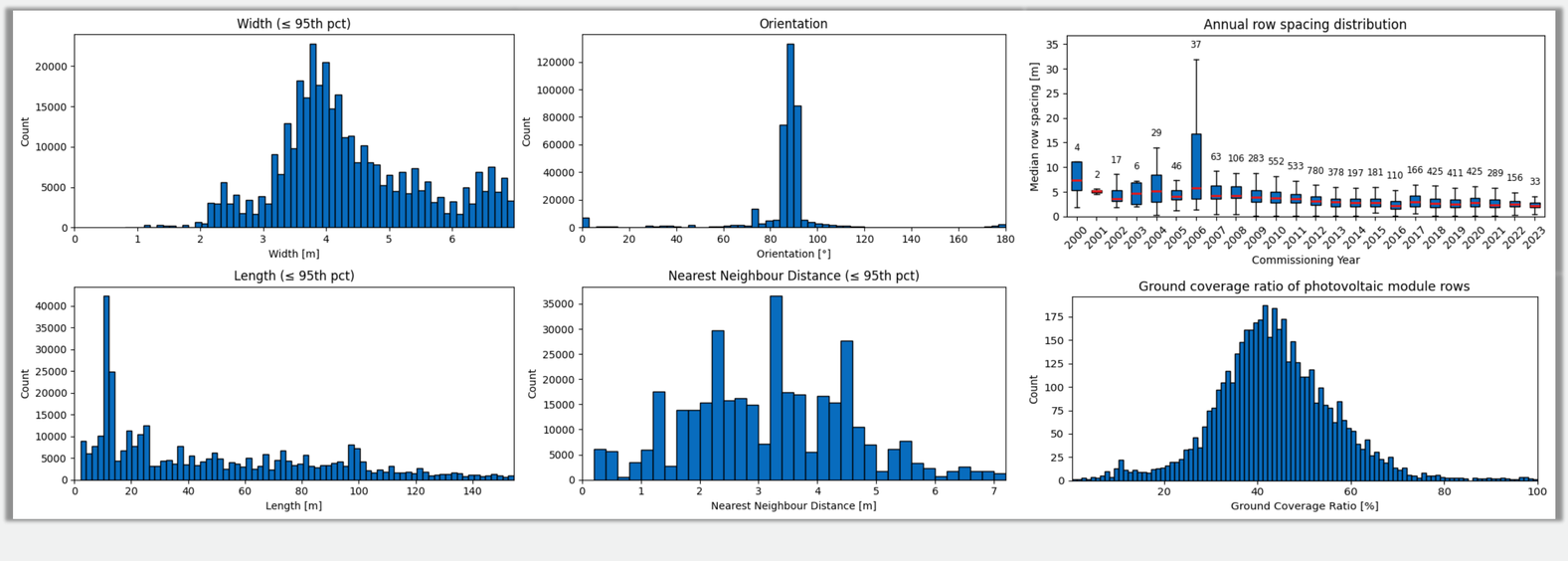
Albert, J., Gärtner, P., Schymik, C., Wehner, C., Siegismund, J., & Klingner, S. (2025). Ground-Mounted Photovoltaic Systems in Germany (Version V20250509) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15387100

Conclusion

- unsupervised extraction of photovoltaic module rows from DOP and calculation of detailed spatial parameters for GMPV systems is possible
- methodological approach can be adapted and utilized in other remote sensing applications that lack training data
- employed technical pipeline can be expanded for application in diverse geographical regions

Outlook

- creation of computer vision-readable training dataset
- train a more efficient supervised model like YOLO for detection of module rows
- deploying pretrained model on a regular basis to facilitate improved monitoring of system parameter developments



Federal Ministry
for the Environment, Climate Action,
Nature Conservation and Nuclear Safety

Get in contact
ki-anwendungslabor@uba.de

Umwelt Bundesamt

AI Lab
Application Lab for Artificial Intelligence and Big Data