AGILE: GIScience Series, 5, 30, 2024. https://doi.org/10.5194/agile-giss-5-30-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

Water Body Detection Using Sen2Cube.at and Comparison to Open Government Data - Assessing for Floating Photovoltaics

Franziska Hübl [1], Hannah Augustin [2], Martin Sudmanns [2], Dirk Tiede [2], and Johannes Scholz [2]

[1]Department of Geodesy, Graz University of Technology, Graz, Austria
[2]Department of Geoinformatics, Paris Lodron University of Salzburg, Salzburg, Austria

Correspondence: Johannes Scholz (johannes.scholz@plus.ac.at)

Abstract. Currently, energy transition is in full swing and aims to increase the production and utilization of green and sustainable energy. Especially in Central Europe, available space to create new photovoltaic power plants is limited. Spatial planners need to maintain a balance between areas for photovoltaics (PV), agriculture, forestry and settlement space. One idea in Europe is to utilize water bodies for floating PV. However, it is hard to quantify the potential area suitable for floating PV. This paper helps to answer this question by using a combined multi-criteria decision analysis and earth observation approach based on Sentinel-2 data via semantically enriched imagery available in Sen2Cube.at. The objective is to develop and test an approach to detect stable surface water that stays invariant in size over several years. We then evaluate the size of detected stable water, which is crucial for the application of floating PV, as well as the littoral zones, which are currently limiting the application of floating PV systems. The results based on Sentinel-2 time-series using Sen2Cube.at are then compared with available open government data of water body extent and littoral zones.

Keywords. sentinel-2, Sen2Cube.at, spatio-temporal analysis, water detection, floating PV

1 Introduction

Multi-criteria decision analysis (MCDA) was conducted within the scope of the PV4EAG project (Sabor et al., 2023; Rothschedl et al., 2023; Welscher et al., 2023) to define suitable water bodies for floating PV in test areas across the province of Styria, Austria. Several criteria were defined by experts and extracted from a literature review. However, the MCDA could not include all identified criteria, and the defined buffer zone between the possible PV panel substructure and water's edge was not based on current geographic information, rather was a static approximation for all standing water body types. Water level varies in many water bodies over time and this variability is a key criterion critical to evaluating the PV potential of a water body. The geographic extent of stable water over time was not incorporated into the MCDA for floating PV within the PV4EAG project because of a lack of relevant, up-to-date geographic information. This paper focuses on characterising water level stability over time and thus the definition of standing water bodies' actual littoral zone that is unsuitable for floating PV as derived from open and free optical satellite imagery.

Variability of water body extent is often poorly documented. The task of identifying water surfaces suitable for floating PV means that surface water has to be detected regularly over time, requiring repeated observations throughout the year, ideally over multiple years. The Sen2Cube.at system facilitates semantic querying of dense time series of satellite imagery thanks to a semantic Earth observation data cube (Augustin et al. (2019)). This means that satellite imagery time series can be queried for their content using application-independent spectral categories derived from a physical-model-based decision tree rather than relying on simplified reflectance thresholds to classify image content (Baraldi et al. (2010, 2013)). In this case, the Sen2Cube.at system is used to query Sentinel-2 time-series based on spectral categories tied to water observations to monitor water stability over time. The results provide a means to cross-check the static littoral zones provided as open government data originally used in the MCDA, but also provide relevant and up-to-date geographic information to characterise surface water variability over time within known water body extents.

Ideally, these results will replace the approximation of the littoral zone based on open government data used for the MCDA with up-to-date geographic information on stable water areas, refining the estimated area and location potentially suitable for floating PV.

The following research questions are to be answered:

1. Is the littoral zone assumed by experts confirmed by the chosen approach?
   (a) If not, to which degree does it differ?
   (b) What is the actual utilizable area in a water body stable enough for floating PV?
2. Are Sentinel-2 datasets suitable for this use case and what may be the limitations?

2 Methodology

2.1 Use case and outline

The idea to check and confirm the littoral zone assumed by experts came while deriving MCDA criteria from factors of influence, and considering the shortcomings of existing datasets to locate potential water bodies suitable for floating-PV. Within the scope of the PV4EAG project the overall criteria, solar potential and shadowing effects, protected areas, hazard zones, as well as PV-type specific ones, namely type and usage of the water body, and an approximated littoral zone of 7-15m (which was suggested by interviewed experts) for all standing water body types, were taken into account. Not all criteria identified for assessing geographic PV potential could be covered with available open government data, including water level variability and the actual littoral zone. These criteria indirectly take the slope angle into account and define a natural buffer between the water's edge and the possible PV panel substructure. This paper evaluates the approximated littoral zone and the inner area of water bodies using all available Sentinel-2 data over 6 years to characterise the actual dynamics of water bodies over time.

The same test areas for the PV4EAG project were used, meaning about 5% of Styria's area was selected using municipal boundaries and balancing the urban-rural proportion of Styria.

2.2 Data

Open government data on lakes and ponds was used as the basis for water body delineation and Sentinel-2 imagery was the main source for characterising water and its variability. The open government data source "Seen und Teiche Steiermark" of Styria (data.gv.at) was used as the basis dataset for standing water bodies. It consists of polygons digitized from orthoimages over multiple years, with the most recent being in 2019. Data was either provided or reprojected into EPSG:3035 (Annoni et al. (2003)), an equal-area projection that is the current standard of Statistic Austria and EU data.

The semantic enrichment available in the Sen2Cube.at system based on Sentinel-2 Level-1C (L1C) imagery were queried for water-like observations (see below). The Sen2Cube.at system is backed by an Open Data Cube implementation (Killough (2018)) that contains all available Sentinel-2 Level-1C imagery of Austria (2015-present) as well as an interpretation of each multi-spectral reflectance signature into spectral categories that have semantic associations (e.g. water, high intensity vegetation) (Sen2Cube.at (a)). Sentinel-2 imagery has a maximum revisit time of 5 days over Austria. The spectral categories included in Sen2Cube.at are generated using the satellite automatic image mapper (SIAM) (Baraldi et al. (2018b, a)). Only spectral categories per pixel over time were used in this spatio-temporal analysis.

2.3 Water Detection

The semantic Earth observation data cube of Austria, known as Sen2Cube.at (Sudmanns et al. (2021)), was used to characterise variability across each water body's extent over time, whereby water areas with low variability were assumed to have a higher potential for hosting floating PV modules. The Sen2Cube.at system enables queries of reflectance values and spectral categories together over time using the semantic querying language called *semantique* (van der Meer et al. (2022)). Semantic queries can be formulated in Python, or built using Sen2Cube.at's programming-free front-end based on a customized implementation of Blockly (Blockly) and then saved as semantic models in the Sen2Cube.at system's knowledge base (i.e. collection of semantic models) to be reused, extended or shared with other users.

Each test area was queried for water-like observations using a semantic model in Sen2Cube.at, resulting in the mode of spectral categories over time and a binary mask of water considered to be stable for comparison with the open government data geometries. While any time-frame with data can be queried, the semantic model in Sen2Cube.at was applied to all Sentinel-2 data available between 1 January 2018 and 31 December 2023, or 6 years worth of observations. A binary mask was generated by first filtering out cloud-like spectral categories and then creating a few entities: water-like categories into an entity "water", spectral categories commonly occurring just at the edge of water bodies, and spectral categories outside or excluded (i.e. vegetation, bare soil). Pixels containing at least 30 percent of observations considered water over time, but no more than 30 percent in any of the other constructed entities after cloud masking were assumed to be relatively stable. The mode result serves as a plausibility check for water-like categories over time when compared to the binary mask. The mode based on spectral category for the pixels remaining in the binary mask were all included in the water entity.

2.4 Comparison to Open Government Data and MCDA Criterion

Here we focus on the evaluation of the defined 15m buffer, assumed as the littoral zone, in order to verify the experts' opinion. A zonal statistic was calculated with the 10x10m^2 dataset generated by Sen2Cube.at on stable water occurrence using 6 years of imagery. In detail, the standing water body polygons from open government data were divided into the buffer zone and the inner area representing the potentially suitable water surface for floating PV. According to the interviewed experts, the distance from the PV panels to the water's edge should be at least 7m. Depending on the water level variability (the higher variability, the more distance needed), a buffer of 15m is a reasonable estimate for most water bodies. In order to cover that criterion and better match the pixel size of the stable water identified by Sen2Cube.at, we chose to use the 15m buffer for comparison. The resulting statistical values *count* and *sum* of stable water occurrence per pixel were then used to calculate the percentage of agreement within the chosen 15m littoral zone and the inner area of the water bodies. This means that the buffer zone and the inner water body area are validated in terms of actual water occurrence in percent, taking the total number of pixels counted within each polygon as the basis for validating the experts' recommendation.

3 Results

The open government data on water bodies were evaluated by analyzing the percentage of water detected over time based on Sentinel-2 data to separate a stable area over 6 years from more variable surface water. An example result is shown in figure 1 in form of SIAM's 33 spectral categories. The stable water binary mask is a result of merging these categories together, and is the basis for comparison to the open government data on water body geometries.

Assuming that the detected stable water correlates with the boundaries of the open government data, the overall agreement for water bodies within the test areas lies at 63.77%. This percentage represents values within the boundaries of the water body polygons and does not cover results outside of those polygons. By separating the water bodies into the inner area and outer 15m buffer, there is an overall agreement of 90.50% between the detected stable water occurrence and the inner water area without the 15m buffer. The 15m buffer zone (assumed littoral zone) has an 73.62% agreement. An example result is shown in figure 2 where the comparison of detected water and littoral zone buffer defined by experts is shown as the percentage of stable water occurrence within single water bodies. Beside the overall percentage of agreement, the frequency of water areas according to their percentage of stable water occurrence is shown in figure 3 and 4. The high frequency of low percentage of agreement (red) in figure 3 indicates

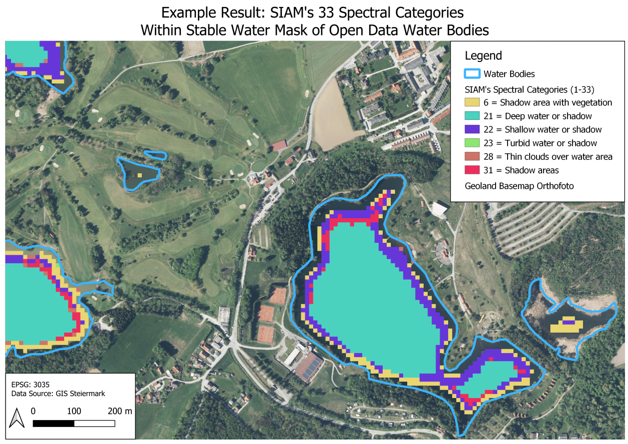

Figure 1. Example result: The water body extent provided by open government data is displayed along with the mode within this extent based on SIAM's 33 spectral categories (Sen2Cube.at (b)) using 6 years (2018-2023) of semantically enriched Sentinel-2 imagery in Sen2Cube.at

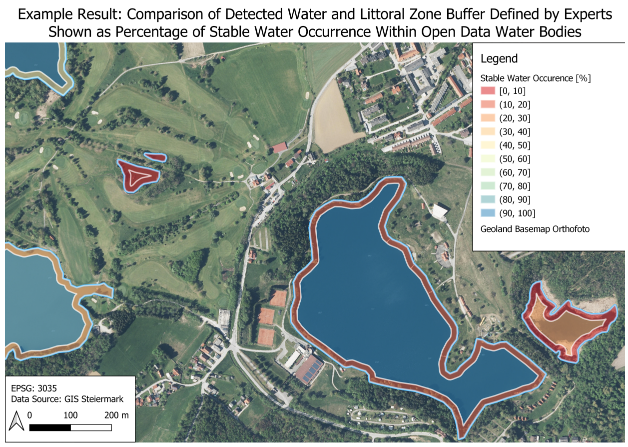

Figure 2. Example result: The stable water and littoral zone buffer defined by experts and provided as open government data is shown. The color of these areas are based on agreement in percentage of area as compared to the stable water area detected using 6 years (2018-2023) of semantically enriched Sentinel-2 imagery in Sen2Cube.at. Water-like observations detected by Sen2Cube.at outside of the open government data water body extents are not considered in this analysis.

that the buffer area is confirmed, because the stable water occurrence is very low. Higher counts for agreement in the 90-100% range can be seen for the water bodies inner area in figure 4. Here the high frequency of high percentage of agreement (blue) confirms the inner zone of stable water occurrence. Conspicuous, is that there is a high frequency of low percentage of agreement (red) too. This indicates that many water bodies delineated in the open government data may have a high variability of their water level or contain surface vegetation. The quantity of water bodies and buffer zones differ, because standing water bodies of all sizes are included in the basis dataset, which are often smaller than the buffer area of 15m, and thus have no inner area left to be analyzed.

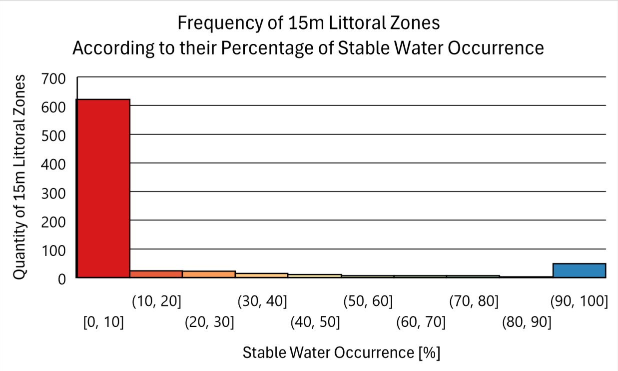

Figure 3. This histogram shows all 762 littoral zones located within the study area as defined by open government data on water bodies (15m buffer within a water body's edge) and the distribution of the percentage of stable water occurrence that the Sen2Cube.at analysis over 6 years of Sentinel-2 observations identified. In this case, the histogram shows an expected result that the majority of the littoral zones of all water bodies is not considered stable over time.

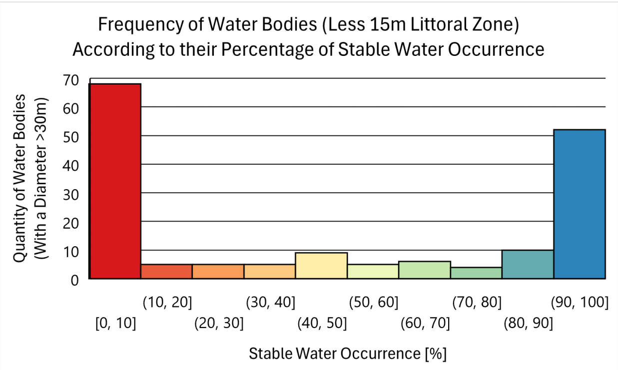

Figure 4. This histogram shows all 169 water bodies, ignoring the littoral zones (thus, only water bodies larger than the 15m buffer are included) with areas located within the study area as defined by open government data on water bodies and the distribution of the percentage of these water bodies that the Sen2Cube.at analysis over 6 years of Sentinel-2 observations identified as stable water. In this case, the histogram shows that nearly 70 water bodies identified in the open government data may not have been covered by stable surface water from 2018 through 2023, as suggested by the Sen2Cube.at results over time, with over 50 water bodies having almost complete agreement.

4 Discussion and Outlook

This method gives a good impression of how convenient it is to harness the benefits of semantically enriched Sentinel-2 data in a system like Sen2Cube.at. The given use case is a perfect playground to test the potential of such data structures by doing a query on a water body polygon level to obtain relevant changes over time. This means that fluctuations over time can be captured and represented beyond a generalised reduction over time. The need for further water body classification can be seen in figure 4 where a high frequency of water bodies excluding the 15m buffer have no stable water occurrence at all. Depending on the temporal resolution, a classification of water bodies in terms of variability of the water level, for example, over a year for seasonal changes, or weekly changes, such as detection of water reservoirs, might be possible. The results demonstrate that it is possible to support the digitizing process of water bodies by using an up-to-date stable water mask derived from Sentinel-2 data that shows the stable water body extent over time. Especially when using orthoimages as a reference, water bodies with a high proportion of surface vegetation are often not visually distinguishable from surrounding land-based vegetation. This approach could be extended with the detection of vegetation within or surrounding the water bodies by using SIAM's vegetation-like categories or other indices, like NDVI. A similar use case where this method would be of additional value is the detection of frequently flooded areas.

5 Conclusion

The Sen2Cube.at generated binary mask in combination with open government data on water bodies can clearly refine the analysis of water bodies to find potentially suitable water areas for floating PV. The results can support a new measure of estimating floating PV suitability within a given water body over time. Therefore, the results can serve as a revised criterion for a MCDA by replacing the approximated buffer zone defined by experts with water body specific spatio-temporal information on variable stable water extent. This is confirmed by the high frequency of water bodies agreeing with the stable water mask (see figure 3 and 4). The high frequency of water bodies disagreeing with the stable water mask demonstrates the need for additional information on specific water bodies for applications where surface water variability is critical. In terms of spatial resolution and observation frequency, the given Sentinel-2 data with a pixel size of $10\text{x}10\text{m}^2$ and a revisit time of at least 5 days seems sufficient for this use case. Especially with regard to economic factors for floating PV, water bodies smaller than the pixel size are of low interest. Water bodies with drastic changes in extent occurring in between Sentinel-2 observations, meaning less than 5 days, would need to be identified using other data sources. To comprehensively analyze the possible floating PV potential and cover all water bodies, the data used here could be extended by other existing open data (e.g. OpenStreetMap). Analysis could also be improved by using a buffer zone that extends outside the currently used water body boundaries to capture and characterise changes over time to water body extent and shape that exceeds existing boundaries.

References

Annoni, A., Luzet, C., Gubler, E., and Ihde, J.: Map Projections for Europe, p. 123, http://mapref.org/LinkedDocuments/MapProjectionsForEurope-EUR-20120.pdf, 2003.

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., and Baraldi, A.: Semantic Earth Observation Data Cubes, Data, 4, 102, https://doi.org/10.3390/data4030102, 2019.

Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., and Blonda, P.: Automatic Spectral-Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery—Part I: System Design and Implementation, IEEE Transactions on Geoscience and Remote Sensing, 48, 1299–1325, https://doi.org/10.1109/TGRS.2009.2032457, 2010.

Baraldi, A., Humber, M., and Boschetti, L.: Quality Assessment of Pre-Classification Maps Generated from Spaceborne/Airborne Multi-Spectral Images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification Software Products: Part 2 — Experimental Results, Remote Sensing, 5, 5209–5264, https://doi.org/10.3390/rs5105209, 2013.

Baraldi, A., Humber, M., Tiede, D., and Lang, S.: GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation level 2 product generation – Part 2: Validation, Cogent Geoscience, 4, 1467 254, https://doi.org/10.1080/23312041.2018.1467254, pMID: 30035157, 2018a.

Baraldi, A., Humber, M., Tiede, D., and Lang, S.: GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation level 2 product generation - Part 1: Theory, Cogent Geoscience, 4, 1467 357, https://doi.org/10.1080/23312041.2018.1467357, pMID: 30035156, 2018b.

Blockly: https://developers.google.com/blockly.

data.gv.at: Seen und Teiche Steiermark, https://www.data.gv.at/katalog/dataset/land-stmk_seenundteiche.

Killough, B.: Overview of the Open Data Cube Initiative, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8629–8632, https://doi.org/10.1109/IGARSS.2018.8517694, 2018.

Rothschedl, D., Welscher, F., Hübl, F., Majic, I., Giannandrea, D., Wastian, M., Scholz, J., and Popper, N.: Calculating Shadows with U-Nets for Urban Environments, in: 12th International Conference on Geographic Information Science (GIScience 2023), edited by Beecham, R., Long, J. A., Smith, D., Zhao, Q., and Wise, S., vol. 277 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 63:1–63:6, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, https://doi.org/10.4230/LIPIcs.GIScience.2023.63, 2023.

Sabor, I., Krainz, T., Loidl, M., Sumereder, C., Hübl, F., and Scholz, J.: Flächen-und Energiepotenzial von innovativen PV-Anlagen, in: Proceedings of the 13th Internationale Energiewirtschaftstagung TU Wien, https://iewt2023.eeg.tuwien.ac.at/download/contribution/fullpaper/210/210_fullpaper_20230210_151030.pdf, 2023.

Sen2Cube.at: Semantic Earth Observation Data Cube Analysis, https://www.sen2cube.at/, a.

Sen2Cube.at: Sen2Cube.at Manual - Satellite Image Automatic Mapper (SIAM), https://manual.sen2cube.at/enrichment.html#siam, b.

Sudmanns, M., Augustin, H., van der Meer, L., Baraldi, A., and Tiede, D.: The Austrian Semantic EO Data Cube Infrastructure, Remote Sensing, 13, https://doi.org/10.3390/rs13234807, 2021.

van der Meer, L., Sudmanns, M., Augustin, H., Baraldi, A., and Tiede, D.: SEMANTIC QUERYING IN EARTH OBSERVATION DATA CUBES, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W1-2022, 503–510, https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-503-2022, 2022.

Welscher, F., Majic, I., Hübl, F., Bulbul, R., and Scholz, J.: Harnessing the Sunlight on Facades - an Approach for Determining Vertical Photovoltaic Potential, in: 12th International Conference on Geographic Information Science (GIScience 2023), edited by Beecham, R., Long, J. A., Smith, D., Zhao, Q., and Wise, S., vol. 277 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 82:1–82:7, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, https://doi.org/10.4230/LIPIcs.GIScience.2023.82, 2023.

Author contributions. F.H., H.A. and M.S. carried out the experiment. F.H. and H.A. wrote the manuscript with support from M.S. . D.T. and J.S helped to supervise the project. M.S., D.T. and J.S. conceived the original idea. D.T. and J.S. supervised the project and helped to revise the manuscript.

Competing interests. No competing interests are present.

Acknowledgements. The presented results were obtained within the project PV4EAG (proj.nr.: 888491) funded by the Austrian Research Promotion Agency (FFG) http://www.ffg.at. We thank all involved project partners for their outstanding work: FH Joanneum, CAMPUS 02 Fachhochschule der Wirtschaft GmbH, dwh GmbH, Energie Agentur Steiermark gemeinnuetzige GmbH;