Internship Report

Company – Ubicube GmbH

Written by

Parinda Pannoon

Copernicus Master in Digital Earth

Supervised by

Dr. Andreas Salentinig and Dr. Mostapha Harb

1 May 2023 – 31 July 2023

Internship Overview

I worked as a geospatial intern at Ubicube GmbH, Austria from 1 May 2023 to 31 July 2023. Ubicube collects and analyzes multi-resolution satellite data to extract relevant information about properties with machine learning-driven processing pipelines. They have integrated additional geodata and external services to enrich the information content and work on building footprint extraction services.

My main task was performing geospatial analysis including map visualizations and statistics. I was assigned to the projects building footprints extraction models using Machine learning and performing the analysis of flood hazard insights for properties and improving the quality of real estate valuations.

Responsibility and outcome

1. Flooded-area visualization

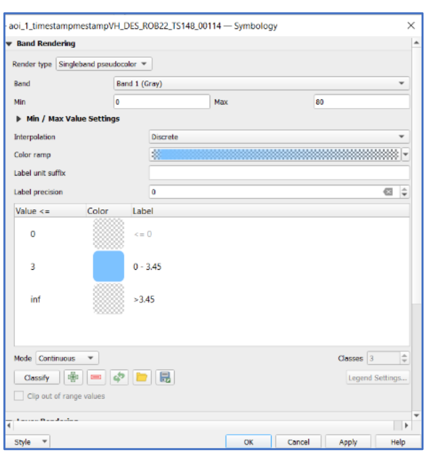
The task aimed to visualize flood masks covering land parcels from different resolution rasters, including various timesteps. And assess the risk of being flooded reoccurrence every year in Austria.

Responsibility

My role was to create flood layers that visualized flood extents and affected parcels. The processed data in QGIS typically used both low and high-resolution raster images representing flood inundation, including filtered specific timesteps. This involved setting color scales that clearly show inundated areas overlaid on base maps and other contextual layers. Single-band pseudocolor in QGIS was utilized to set interval ranges of the symbols that can be distinguished from the flood and non-flooded pixels.

Outcome

The flood extent maps were shared as QGIS layers to inform further decision-making. The results were shown as blue pixels on land parcels. As well as in low-resolution images and raster with selected significant dates that had flood occurrences. The outcome of such visualization is especially valuable for urban planners, and emergency response teams who rely on accurate, detailed data to make informed decisions.




2. Flooded-area validation using EO Browser

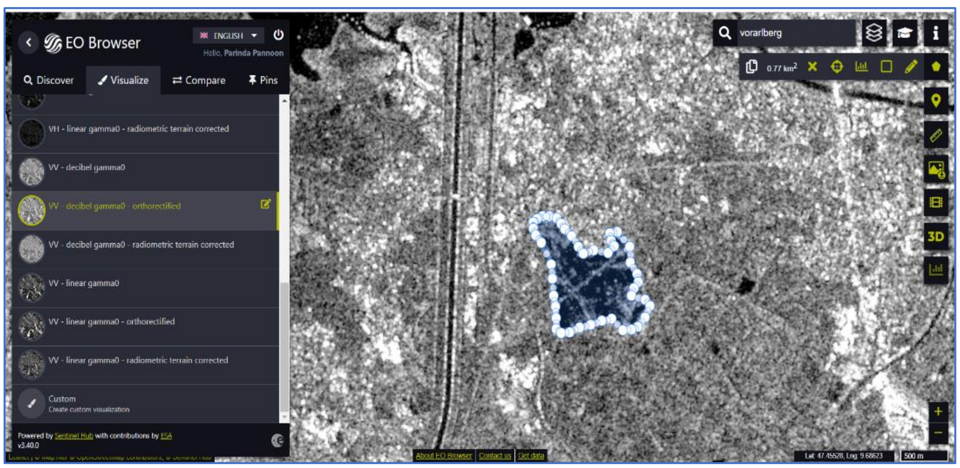
The task of flooded-area validation using the EO (Earth Observation) Browser involves the use of satellite imagery and remote sensing technology to accurately identify and analyze flooded regions. The EO Browser, a tool that allows users to access and visualize satellite data, was employed to facilitate this process. This work aimed to check the actual flooding areas based on the documents then validated them with satellite images in the browser.

Responsibility

I delineated areas that are expected to be flood zones on the EO Browser website compared with flood-recorded documents. Then I exported the geometry coordinates to process in Python with Pyshp library for converting them to shape files.

Outcome

The flood extents derived from EO Browser were transformed into shapefile for further analysis and processing in QGIS. They were compared together to see which areas have large flood extents.



3. Statistical Analysis: Property level Flood insights extraction from Earth Observation data(PROFEO)

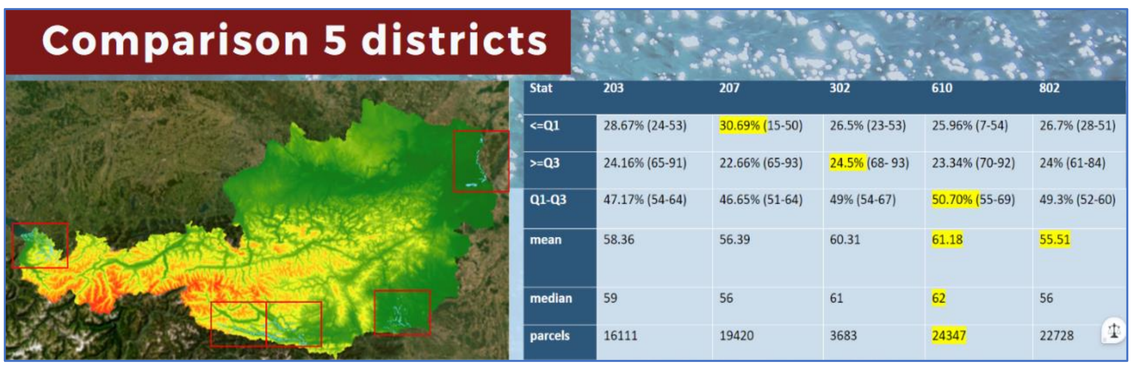
The project got flood hazard insights for arbitrary properties and improved the quality of real estate valuations which enabled new business opportunities. The main innovation of the project and current satellite images (Copernicus Sentinel) is to produce dynamic flood hazard maps that are capable of keeping up with fast-paced climate change induced variations of local flood patterns.

Responsibility

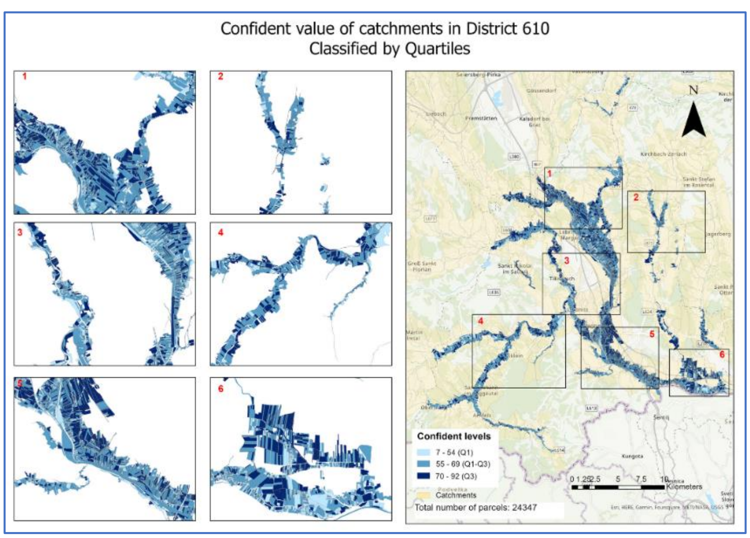
To perform a statistical analysis of confidence value, absolute value, digital elevation and demonstrate the comparative statistical analysis of each catchment district. Also, the results were reported as maps and data visualizations.

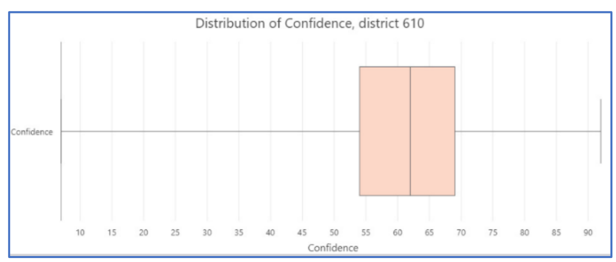
Outcome

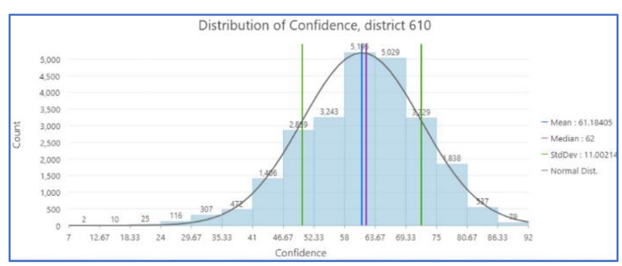
District 610 has the highest mean of confident value among the 5 districts. This is because the 610 district is in Flood-prone area. Moreover, this area is surrounded by an urban area so there is high possibility to increase runoff and reducing groundwater recharge.
The parcels of District207 have confident value between 15-50 are 30.69% which is the highest percentage of Q1 among 5 districts.
District302 has a relatively higher mean value after district610. The parcels that have confident value between 68-93 are 24.5% of all parcels which is the highest percentage of Q3 among 5 districts.
District 802 has the lowest mean of confident value among the 5 districts. The district parcels are at a high elevation. Since this catchment is mountainous, it is possible that the amount of water moves downhill, and plants can retain water into the ground more than impervious areas



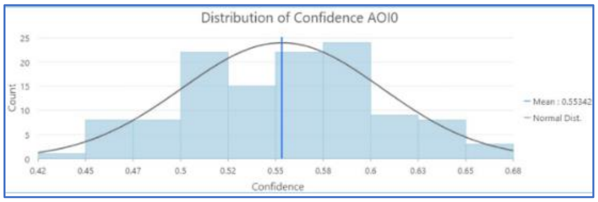
Stat	203	207	302	610	802
<=Q1	28.67% (24-53)	30.69% (15-50)	26.5% (23-53)	25.96% (7-54)	26.7% (28-51)
>=Q3	24.16% (65-91)	22.66% (65-93)	24.5% (68- 93)	23.34% (70-92)	24% (61-84)
Q1-Q3	47.17% (54-64)	46.65% (51-64)	49% (54-67)	50.70% (55-69)	49.3% (52-60)
mean	58.36	56.39	60.31	61.18	55.51
median	59	56	61	62	56
parcels	16111	19420	3683	24347	22728

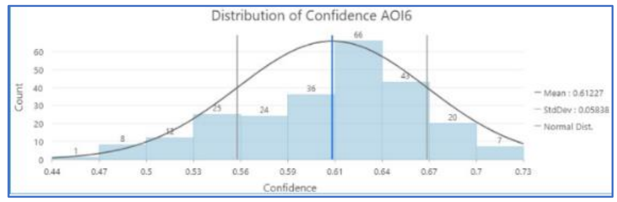


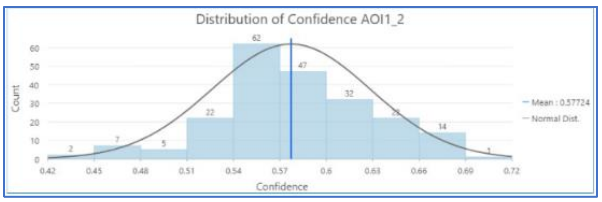


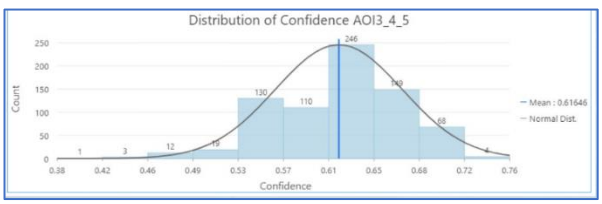


This data is a bit skewed left which means there is a concentration of high confident value on the right. So this district has the highest mean. The standard deviation of these parcels is from 50.18-72.18 or 68.42% of all parcels in the catchment. 25.96% or 6320parcels where have confident values less than equal quartile 1(7-54) 50.70% or 12344parcels within q1-q3 (55-69) 23.34% or 5683 parcels where have confident values more than quartile 3 (70-92).









According to all areas of interest, confidence values are most likely the same range (0.5-0.6). AOI3_4_5 in district610 has the highest mean of confidence (6.616) which means this study area has a quite high level of certainty that flood events occur in each parcel.

4. Building footprint analysis

The purpose of this work is to evaluate the products of building footprints and compare the correctness and suitability for use in the next step.

Responsibility

My role was to perform statistical analysis of missing buildings, outdated buildings, erroneous buildings by referencing with MAXAR satellite image. In the part of GIS analysis, I performed the symmetrical analysis to see the difference among the products which are BEV, OSM, Mircrosoft and Ubicube's building footprints.

Outcome

Ubicube's building footprints improve the results compared to -
-Microsoft footprint -19%(average)
-BEV footprint – 43% (average)
-OSM footprint – 34%(average)

Big buildings added by Ubicube compared to -
-Microsoft – 56(average)
-BEV – 68 (average)
-OSM – 40(average)

And Symmetrical difference
-Microsoft – 51 sq.m.
-BEV – 162 sq.m.
-OSM – 45 sq.m.

5. Building segmentation models

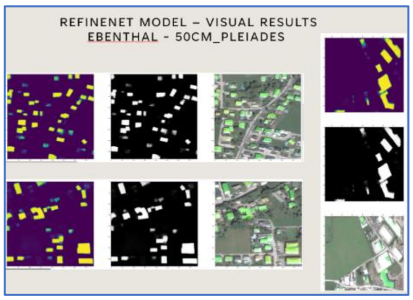
Building footprint segmentation models are specialized applications of computer vision and machine learning aimed at extracting the shapes and boundaries of buildings from aerial or satellite images. These models are crucial in urban planning, and real estate management.

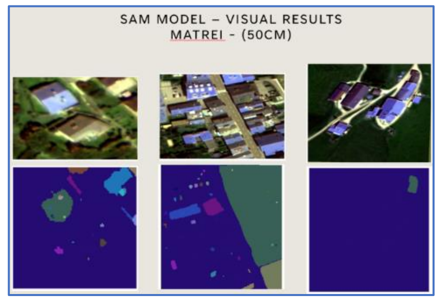
Responsibility

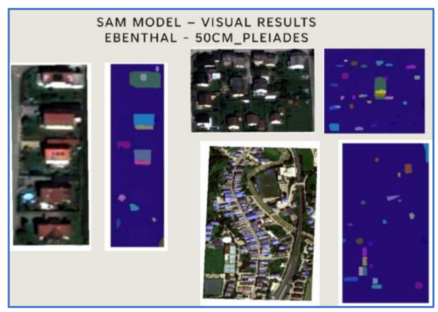
Choosing and testing an appropriate segmentation algorithm, typically a form of machine learning models such as SAM by meta and Refinenet also K-means, which are popular for their efficiency in image segmentation tasks. Then I trained the models, the model trained on the prepared dataset or satellite image Pleiades, then the buildings will be distinguished from other objects and backgrounds. Then the last step was analyzing the performance evaluation of these models compared to the reference image to see which models are the most correct.

Outcome

SAM-Segment anything model was not able to segment the building objects correctly because the segmented shapes were not correct as rectangle but it gave the random shapes. Also, small roofs were not detected in the results. For K-means clustering model can only classify 2 classes separately but the segmentation shapes were not similar to the real building footprints. The last model, Refinenet model worked quite well compared to others because the segmented results were similar shapes to the reference images even small buildings were segmented. These models will be selected for the further tasks to perform classification and labeling the building class.







Conclusion

During my three-month internship at Ubicube, I had the opportunity to immerse myself in the field of geodata science and contribute to various projects. I particularly enjoyed working on the Building footprint extraction task, which allowed me to improve my programming skills on raster images. I found many tasks that using statistical analysis is beneficial for my future career as a geospatial analyst. My supervisors were supportive, providing regular feedback and guidance in every task. Team members were always ready to help, fostering a supportive and learning-rich environment. Overall, my internship at Ubicube was a good experience and I hope to work or collaborate with the company in the future. I would recommend this internship to others seeking to enter the Earth observation field.