Copernicus Master in Digital Earth

Student: Emanuel Goulart Farias

Student number: 12413874

Big Earth Data Concepts

Semantic Austrian Data Cube

Semantic Data Cube

The semantic data cube is an operational product derived from satellite imagery. The product is accessible as a web application and facilitates queries to extract meaning from EO (Earth Observation) observations due to its semantic nature, where each pixel is semantically enriched based on the spectral signature model-based tree.

Objective

The objective of this project is creating a knowledge base model to research different regions of interest. Each inference is arbitrarily chosen to represent a good application of the knowledge base model.

Knowledgebase Model

For my knowledge base model, I decided to investigate the presence of vegetation and snow, trying to observe snow-vegetation dynamics. This could serve as a proxy to understand vegetation ecology's diversity distribution, considering that vegetation is semantically understood as spectral signatures represented by pixels classified as labels, due to the likely correspondence of that signal being representative of vegetation, water, bare soil, etc.

The knowledge base model starts by defining how vegetation could be represented semantically. In the database, it is possible to access different collections containing earth observation data. Vegetation is selected within the collection "appearance" and includes labels: Strong Vegetation High NIR, Low NIR, Average Vegetation High NIR, and Low NIR.

To complete the knowledge base, there are two more defined concepts: Cloud and Snow. Cloud is defined semantically in the atmosphere collection, rather than appearance, and is described by the label "Cloud." Snow is semantically described by the collection appearance, with labels: Snow or Water Ice, Snow Shadow.

The figure 1 show the concepts of water in a user-friendly display

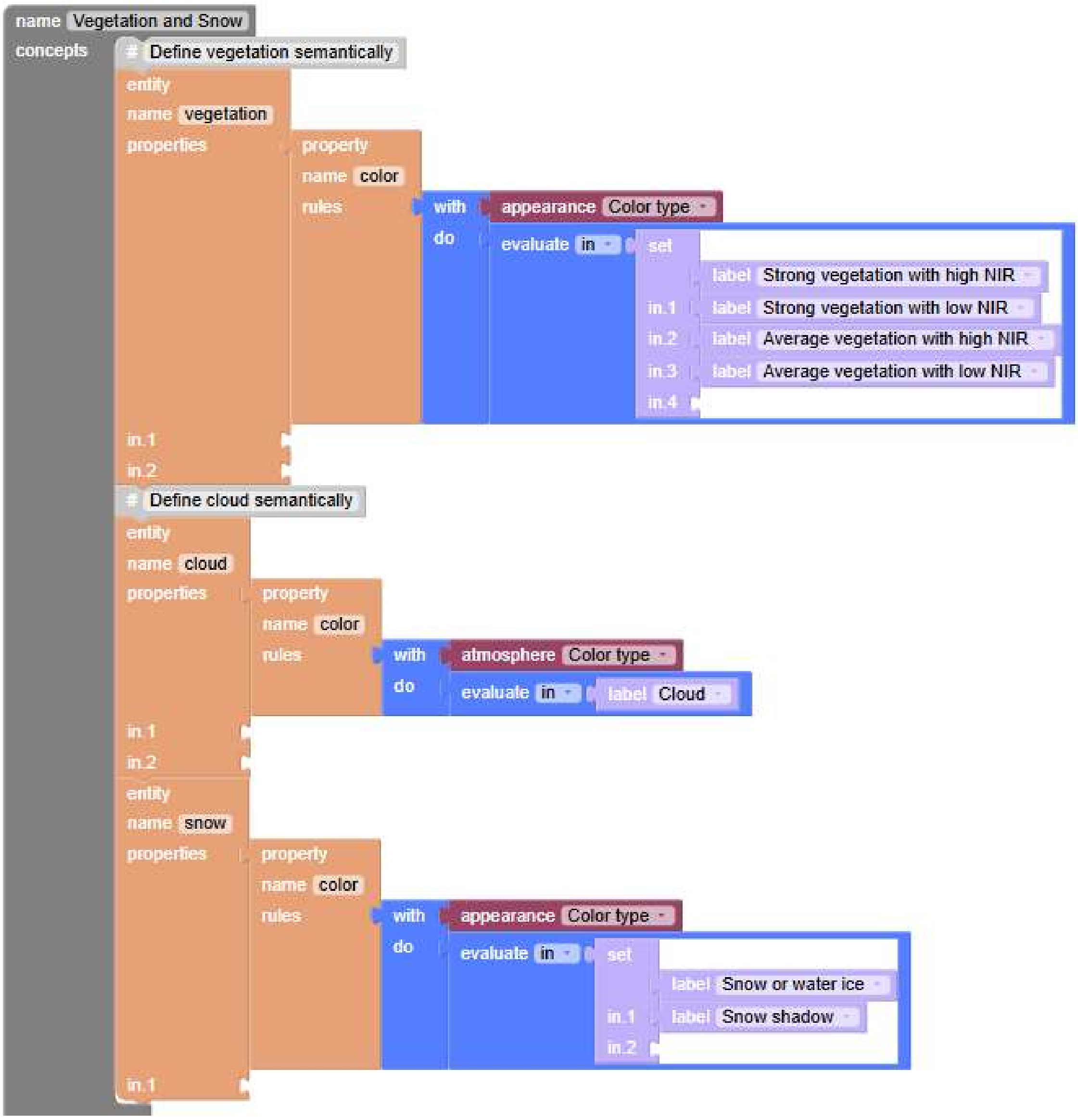


Figure 1 –Vegetation, Snow and Cloud semantically defined.

Subsequently, the query creates a result object by plotting the snow percentage of each pixel through the time dimension. A time-series plot contains the percentage of pixels observed as vegetation and snow, reduced in the space dimension, for each observed image, and a double-check plot containing cloud observations.

The result can be visualized in the figure 2:

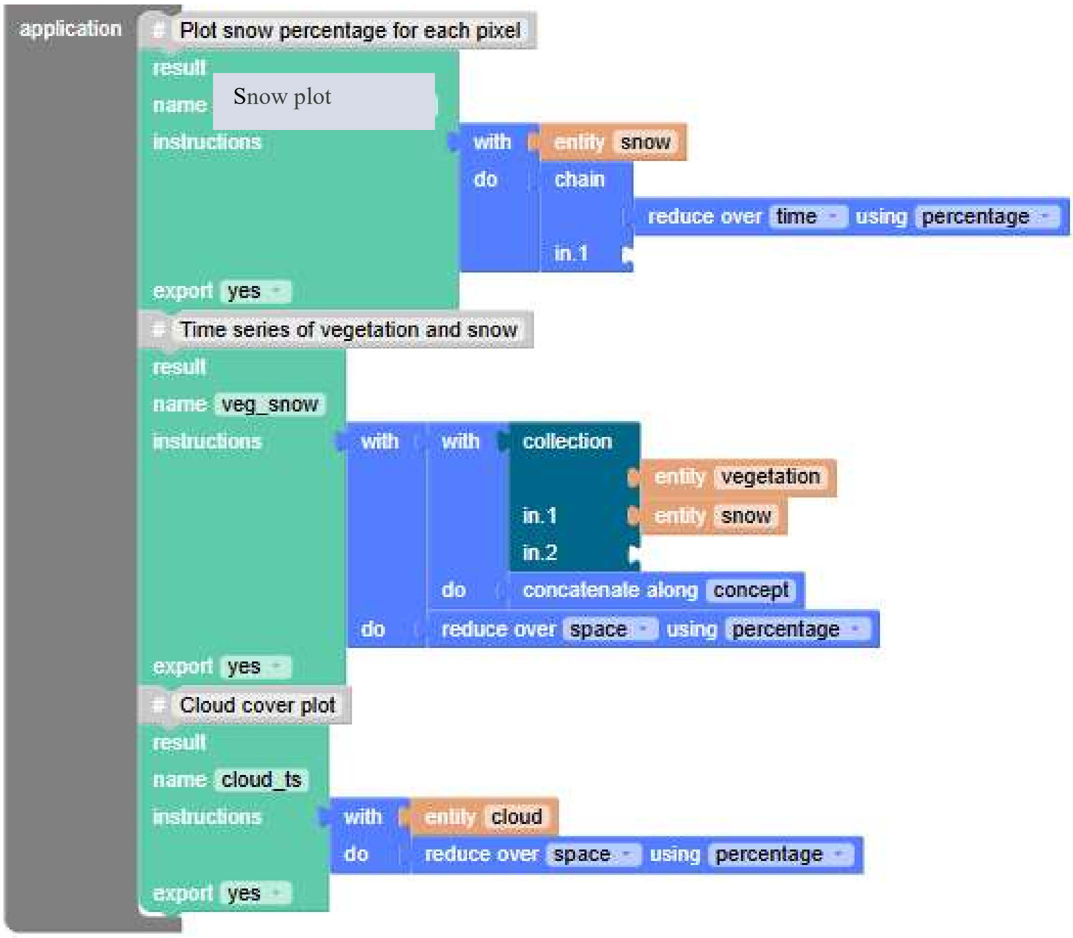


Figure 2 - Result part of the knowledgebase model.

Factbase

To analyze the knowledge base model in different regions of interest, four areas of interest were chosen. The temporal subset was chosen from 01/03/2023 until 30/11/2023, representing the extension of snow cover, excluding the colder period of winter when snow extension increases. The images are of four mountains close to Salzburg.

The images are on four mountains close to Salzburg.

Inference

Four inferences were created, and the results are described in the next section of this report.

Results

1. Near Zell Am See lake

The first area of interest is in Salzburg, a mountain close the Zell Am See lake. The region can be visualized on the figure 3.

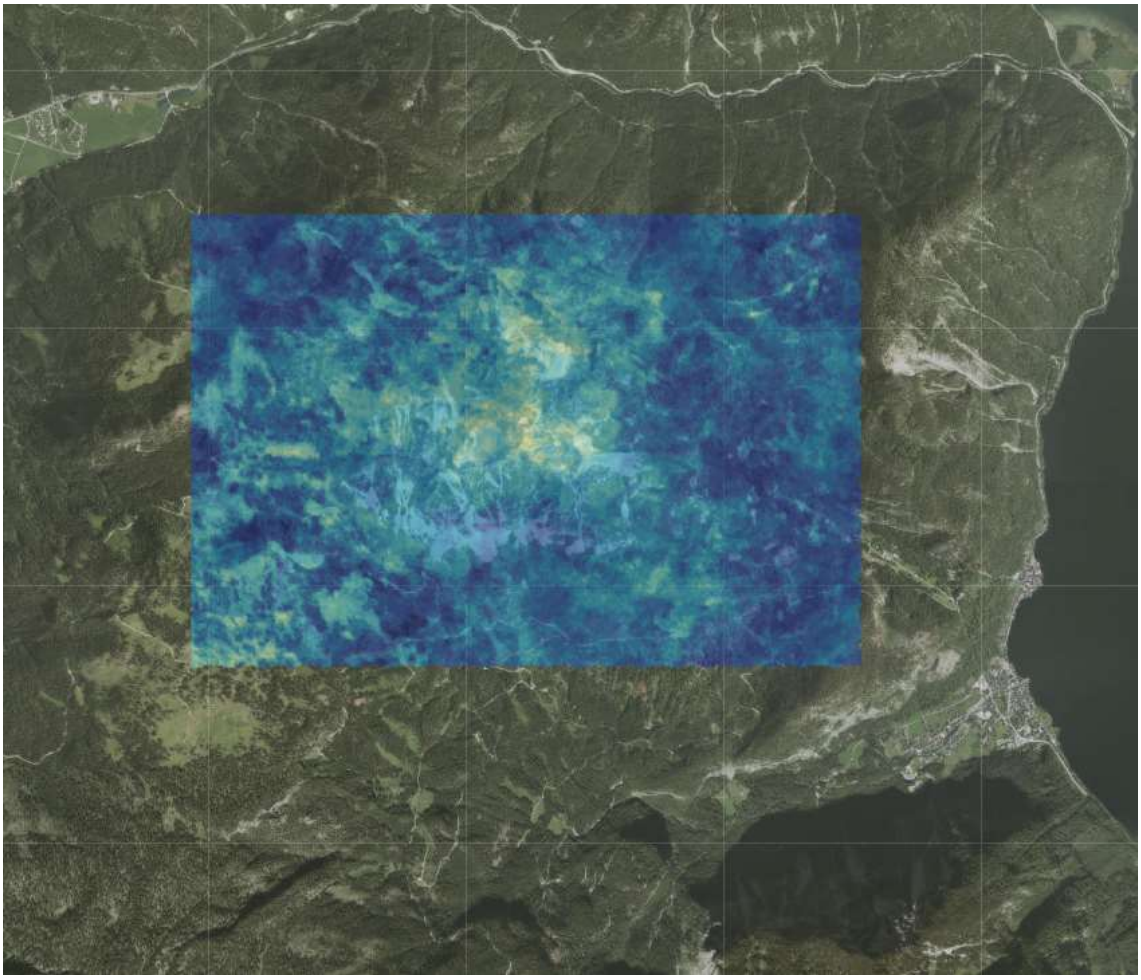


Figure 3 - Peak of a mountain.

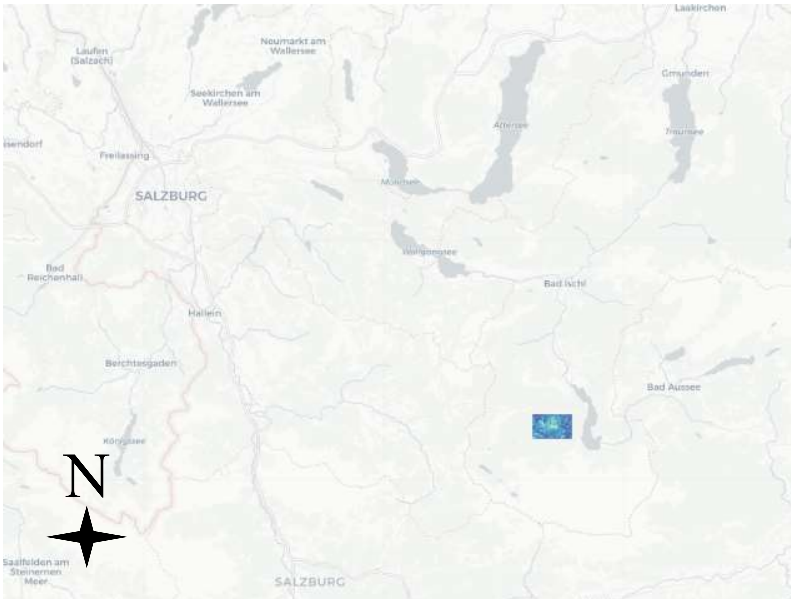


Figure 4 - Region of the interest near Salzburg

The results are available on the figure 5 and 6:

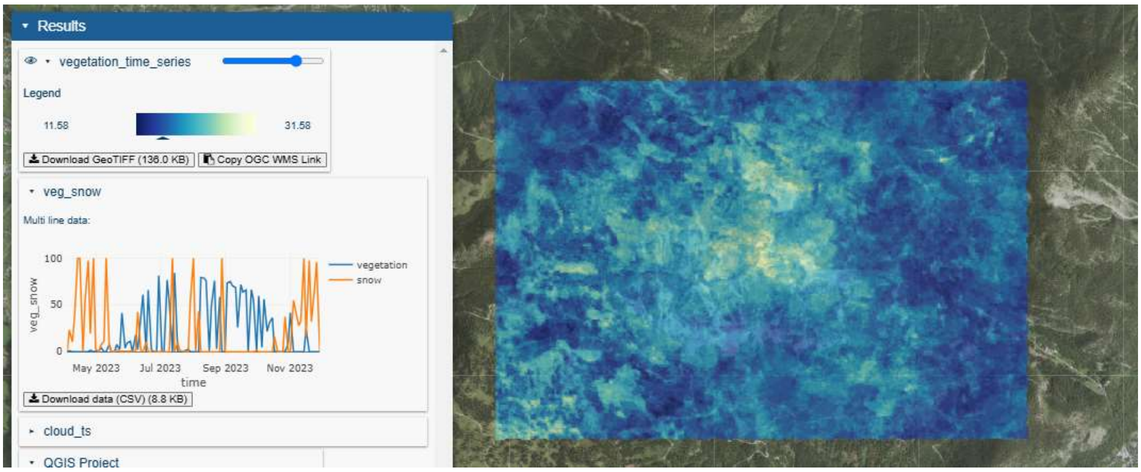


Figure 5 - Observation of snow's percentage for the assigned period.

The graph can be visualized on the figure 4.

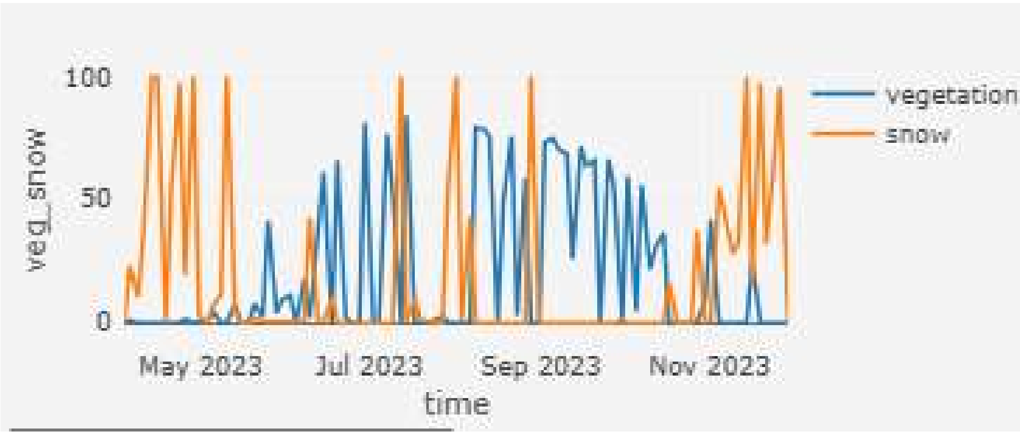


Figure 6 - Plot of time-series vegetation- snow

t is interesting to observe in the graphic that the peaks of snow percentage in the image correspond with almost no pixels classified as vegetation. The greenness state of vegetation starts to increase at the beginning of summer and decreases with the start of spring.

2. **Mosaukopf** peak – 1122m

At a peak with 1122 meters, Mosaukopf is located and analysed on the figure 7.

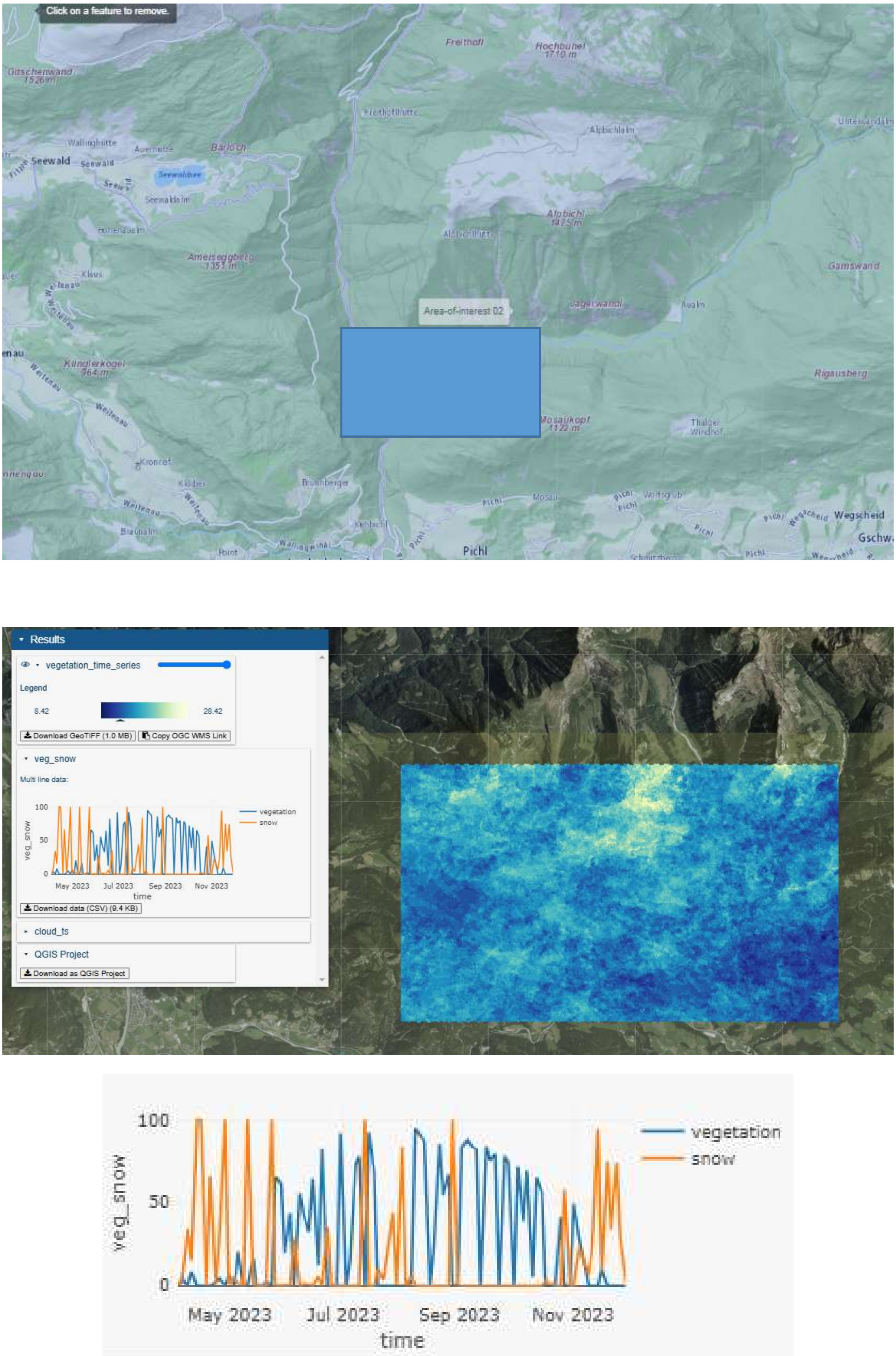


Figure 7 - Time Series - Vegetation and Snow.

At the altitude of the peak, under 1500 meters, only a few pixels are represented as being snow at least ~30 percent of the time. This pattern of snow can be visualized in the gradient of the altitude. As the altitude increases, so does the snow percentage of classified pixels. Figure 8 highlights this observation.

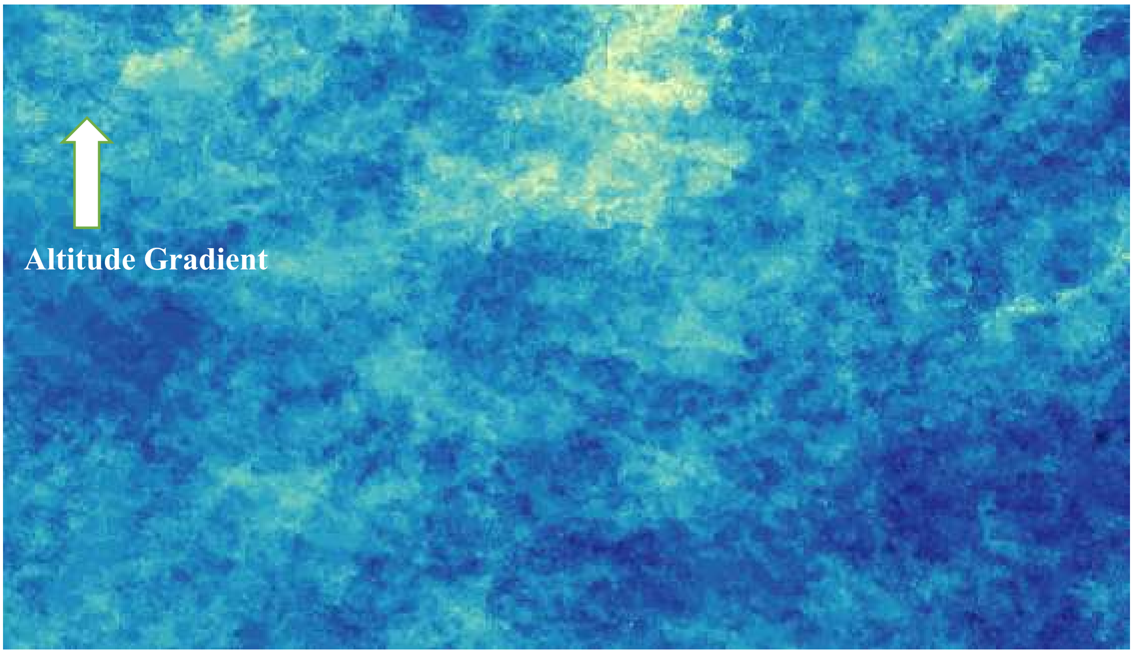
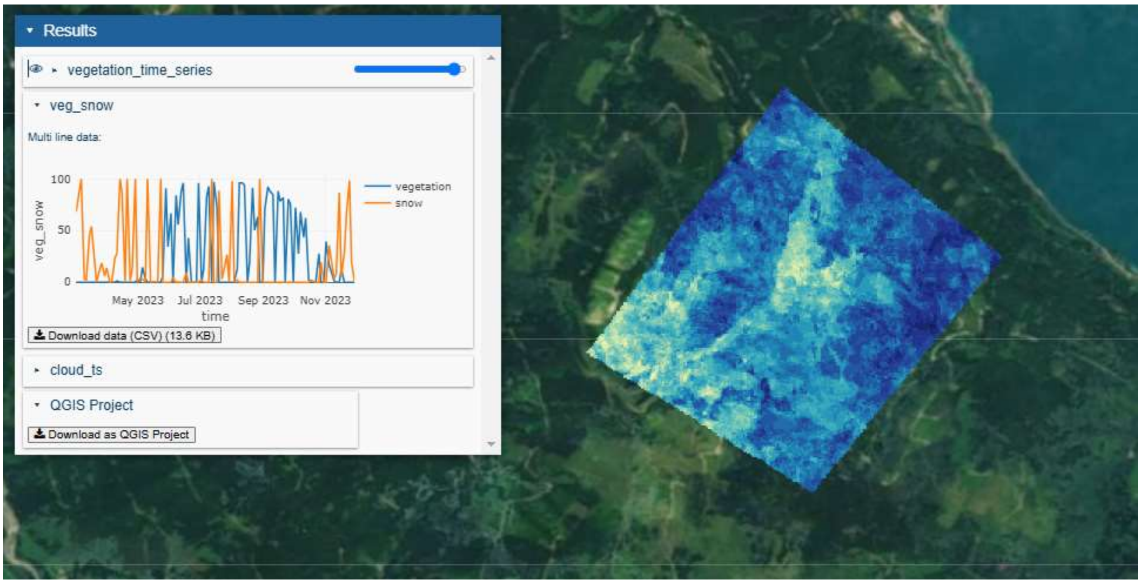


Figure 8 - Pattern of snow percentage of pixels increasing at the higher altitude.

3. Zwolferhorn – Peak at 1520m

Zwolferhorn peak is above 1500 meters and the analysis can be visualized in the figure 9.



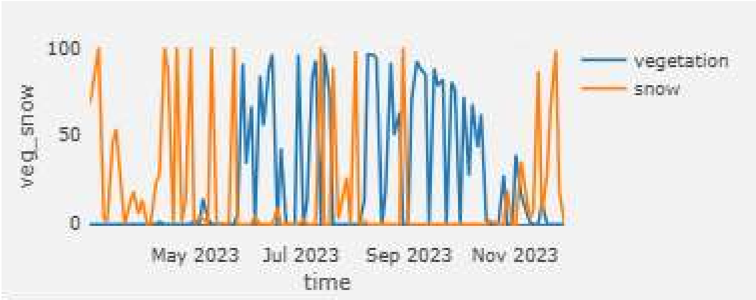


Figure 9 - Observation of the mountain and time series.

For this peak, it is interesting to observe the snow distribution, particularly at the top, which is in the middle of the area of interest.

4. Bleikogel – Peak at 2420 meters.

The mountain is southeast of Salzburg city and has a peak of 2420 meters above sea level. The data cube inference can be visualized in Figure 11.

Figure 10 - Location of Bleikogel.

The data cube inference can be visualized on the figure 11.

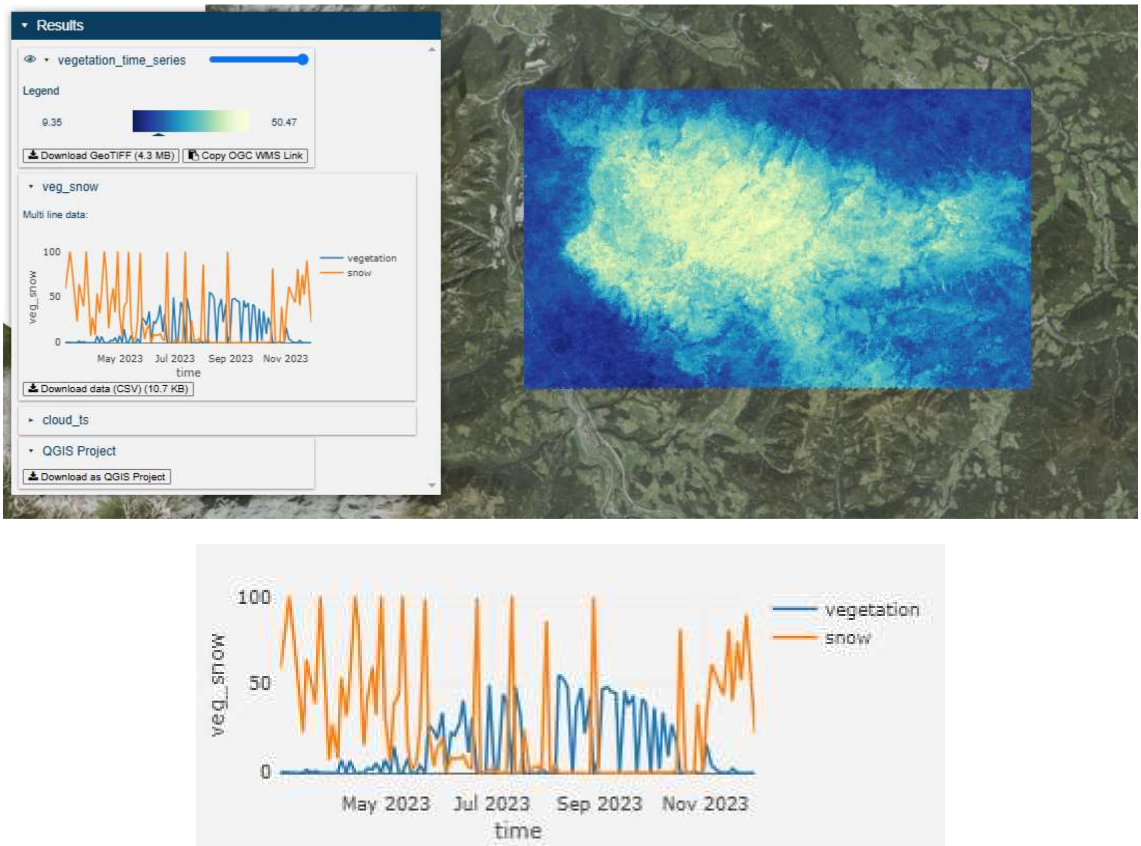


Figure 11 - Map and Time series.

The reduction over time creates a very interesting map, showing a relatively high percentage of pixels classified as snow, with a maximum of 50% for the centered regions including the peak at 2420 meters. The vegetation at the peak of summer points to only a maximum of 50% of pixels classified as such, which likely describes an exposure of rocks rather than vegetation during warmer periods. However, the highest peak analyzed so far clearly shows a greater distribution of snow compared to lower peaks.

Conclusion:

The Semantic Data Cube project presents a significant opportunity to explore earth observation analysis through time and space dimensions, facilitating the analysis of land-use and land-cover change over time, and illustrating periods and patterns captured by the time dimension and quantitatively interpreted by percentage ranges. The ability to select pixels per label classification is much more feasible and widely applicable than using ranges of spectral values that depend on expert knowledge to make assumptions about labels.