Cleaning Corrupted Data through Bayesian Interpolation

Hoang Long Nguyen

21 May 2021

1 Research Question

Initially our research question was broadly formulated as whether we could find an effective method to detect cloud corruption in satellite images. The initial proposal included the use of Bayesian Interpolation, however a more simple and reliable has been found and thus pursued.

2 Available data

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a sensor launched by the National Aeronautics and Space Administration (NASA) on board of two satellites (Terra and Aqua) that captures the entire earth every 1 - 2 days. By processing the raw satellite image, the NASA also publishes a cloud mask that aligns with the original MODIS data (MOD021KM). The mask has the same dimensions as the satellite image and includes information about the impact of clouds on each pixel of the raw footage. The same MODIS data is also being processed by the MODVOLC algorithm to detect high thermal activity around volcanoes [2]. Hence, the cloud mask could be used to mark entire dates and even pixels in the MODVOLC database as cloud corrupted or not. However, because of its coarse gridding and the data encompassing the entire globe, the cloud mask for a specific date and area of interest has a size of roughly 15 megabytes. This makes it too expensive for thousands of hot spot alerts, scattered all over the world, in terms of downloading, post-processing and memory costs. Therefore, we have implemented a solution that adapts the idea of cloud masks in a more simple and scalable way by only analyzing the MODIS data for the immediate area around volcano apertures.

NASA's worldview snapshots [1] is a web interface that allows us to download light-weight information about the MODIS image, namely date, satellite, and Band 7, Band 2 Band 1 values of the on-board MODIS sensor. The latter measurements capture the radiance at a wavelength of 2155 nanometers, 876 nanometers and 670 nanometers respectively and are encoded in the red, green and blue channels of a regular RGB image file.

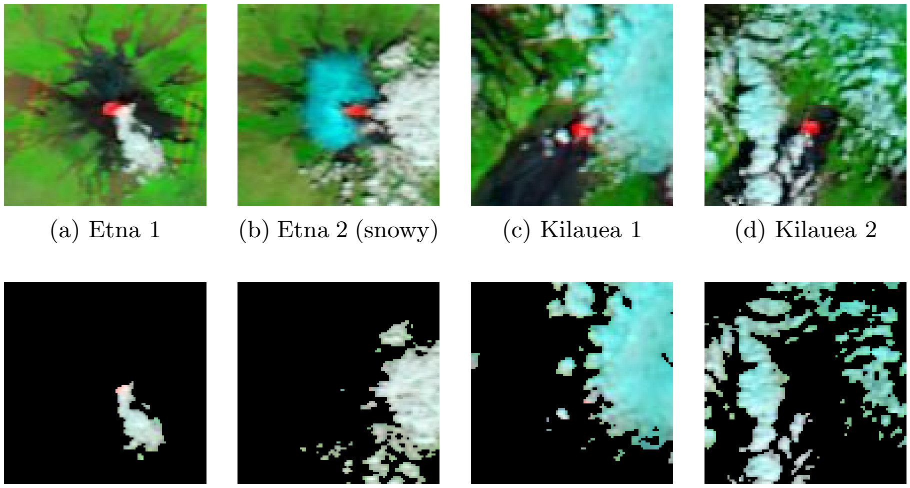
3 Methods

Since those images are exactly the same, that are being used by the MODVOLC algorithm, our proposed method is optimal for dealing with cloud corruption in volcanic hot spot alerts. We first draw a bounding box around the center of a volcano depending on its aperture size. For all dates where at least one hot spot alert has been issued in that area, we download the corresponding MODIS image for the exact same location as well. By analyzing the composition of the RGB values of the image, we can reliably detect cloud pixels and even distinguish them from ice, snow, vegetation and other terrain. As described by the data publishers [1], clouds appear in white to sky blue because of their high reflectance with respect to all given Bands 7, 2 and 1. Snow behaves similarly but additionally has lower values for Band 7, the red channel of the pixel, which is why it appears in a very bright turquoise color.

By setting strict color thresholds, we generate a cloud mask that simply indicates whether a pixel has been corrupted by a cloud or not. We further propose to use the fraction of corrupted pixels within a satellite image of the volcano to indicate the degree (or binary state) of how (or whether) clouds impacted the results from the MODVOLC database.

4 Results

Figure 1 shows examples of satellite images and our generated cloud masks for Etna and Kīlauea. We like to point out that our method is even able to distinguish between snow or ice on the ground and ice clouds with an accuracy and resolution sufficient for our purposes of identifying cloud corruption near volcanoes. The entire costs for downloading and processing data for the time period between 2000 and 2021 for Etna (more than 1500 images) and Kilauea (more than 2000 images) were less than 2 minutes and 65 megabytes each on a 2.7GHz dual-core Intel Core i5 processor with 8GB RAM.



(a) Etna 1 (b) Etna 2 (snowy) (c) Kilauea 1 (d) Kilauea 2

(e) Etna 1 mask (f) Etna 2 mask (g) Kilauea 1 mask (h) Kilauea 2 mask

Figure 1: Examples of MODIS RGB-images and respective cloud masks computed by our solution. Non-cloud pixels are colored in black.

5 Conclusion

We can answer our proposed question in the affirmative, as we have presented an effective approach that can detect cloud corruption in satellite images. With respect to volcanic research, our proposed solution can create cloud masks and cloud corruption indicators for MODVOLC data, using the same satellite images the MODVOLC algorithm uses. The solution is appropriate for flexible and large-scale use (different volcanoes, geographic environments, apertures sizes, areas of interest) and robust to parameter changes, since the filter is strict and thus serves as a conservative lower limit of pixels occupied by clouds.

References

[1] NASA. Worldview Snapshots FAQ. https://earthdata.nasa.gov/faq/worldview-snapshots-faq. [Online; last accessed 07-May-2021].

[2] Robert Wright, Luke Flynn, Harold Garbeil, Andrew Harris, and Eric Pilger. Automated volcanic eruption detection using modis. *Remote Sensing of Environment*, 82:135–155, 09 2002.