Change Detection in Arctic Lakes with the NASA Arctic and Boreal Vulnerability Experiment (ABoVE) Project

Carley Fredrickson
Atmospheric Sciences
University of Washington
Seattle, WA, USA
cfred@uw.edu

Paige Lavin
School of Oceanography
University of Washington
Seattle, WA, USA
pdlogan@uw.edu

Marta Wolfshorndl
School of Oceanography
University of Washington
Seattle, WA, USA
martaw@uw.edu

ABSTRACT

NASA Arctic and Boreal Vulnerability Experiment (ABoVE) project researchers want to study changes in individual Arctic lakes over the past 35 years with Landsat satellite data. Unfortunately, heavy cloud cover obscures the surface of these lakes from satellites during 60% of the year. Careful filtering must be applied in order to remove these cloudy scenes while not removing any especially bright land. The Landsat data files are also quite large, requiring cloud computing resources such as Google Earth Engine to view them, remove clouds, and zoom to lake level. To make these data more accessible and to facilitate quick comparisons between lakes, we developed a data visualization framework that allows our technical end-users to explore these data for seven Arctic lakes over time. This tool can easily be scaled up in the future to include any other lakes of interest to the broader, technical audience of the ABoVE science team.

KEYWORDS

Arctic lakes; Google Earth Engine; Landsat; data exploration tool

ACM Reference format:

Fredrickson, Carley, Lavin, Paige, Wolfshorndl, Marta. 2019. Change Detection in Arctic Lakes with the NASA Arctic and Boreal Vulnerability Experiment (ABoVE) Project. In *2019 CSE 512: Data Visualization course, June 11, 2019, Seattle, WA, USA*. 5 pages. https://cse512-19s.github.io/FP-Change-Detection-in-Arctic-Lakes/

INTRODUCTION

Researchers involved in the NASA Arctic and Boreal Vulnerability Experiment (ABoVE) project want to be able to quickly observe how individual Arctic lakes have changed over the past few decades. Arctic lakes are a critical but understudied component of high-latitude carbon budgets. Since the Arctic region is warming twice as fast as the global average, these lakes are experiencing unprecedented warming rates (Serreze and Barry, 2011). The Butman group at the University of Washington aims to assess the carbon sequestration potential and greenhouse gas fluxes from these freshwater systems, and their changes under a rapidly changing climate, through a combination of field and satellite measurements. In addition to interannual changes in climate, shorter time scale events can also cause significant changes in lakes. For example, field data from the LandCarbon program in 2012 (Zhu and Reed, 2012) found that streams can suddenly release large amounts of carbon after heavy precipitation events.

Since field work in this remote region is difficult and costly, these scientists would like to be able to use satellite imagery to assess changes in these lakes. However, there are over 24,000 lakes in Alaska alone so hosting all of these data in a single visualization, while also allowing drill-down to individual lakes, is tricky. While Landsat satellites typically collect an image over an individual lake every fifteen days, more than 60% of these images in this region are unusable due to heavy cloud cover obscuring the lake's surface. This means that careful filtering must be applied in order to remove these cloudy scenes while not removing any especially bright (e.g., snow-covered) land in the process. Additionally, using satellite measurements to analyze inland waters such as lakes and rivers is a relatively nascent field given the computational power needed to carefully calibrate and correct satellite imagery to accurately resolve these optically complex and spatially small-scale features. The data files from Landsat are also quite large, requiring cloud computing such as Google Earth Engine (GEE) if one wants to view them at any reasonable speed. Given the scripting required to pare down to the cloud-free images and zoom in on a lake of interest with GEE, the ABoVE researchers have found it difficult to examine changes in individual lakes and differences between lakes using these satellite data as quickly as they would like. In particular, difficulties arise because GEE does not

CSE 512, June 11, 2019, Seattle, WA, USA.

provide an easy way to visually compare satellite imagery from different lakes over time. This led us to develop a data visualization framework to allow for quick exploration of seven lakes of interest to our external collaborator, which she will be able to then easily scale up to include any of the thousands of lakes of interest to the broader, technical audience of the ABoVE science team. The seven lakes we focused on are Boot Lake, Canvasback Lake, Greenpepper Lake, Ninemile Lake, Scoter Lake, and Yukon Flats 17 and 20 (YF17 and YF20).

RELATED WORK

Inland water remote sensing is an emerging field in which many large challenges, including appropriate atmospheric correction, still remain (Kuhn et al., 2019). While much work has been done recently to investigate global changes in ocean and terrestrial systems, our ability to observe the earth at the small spatial scales relevant to lakes, rivers and streams has only recently come online (Mouw et al., 2015). This has opened the door for new science advances in our understanding of how freshwater ecosystems are responding to human activities.

Many readers are likely familiar with the Google Earth Timelapse tool, which enables quick, high-quality viewing of time-lapsed satellite imagery for a given location. However, it is unsuitable for our technical audience who mask out clouds and generate annual composites using different methods than the Google Earth Timelapse team and who want to be able to inspect the time series of the individual bands of the satellite for sub-annual timescale events. Our visualization is truly the first of its kind for enabling rapid scientific analysis of satellite data to detect changes in these small-scale Arctic lakes under these parameters and we look forward to seeing the insights our technical audience will be able to generate moving forward.

METHODS

Production of the Timelapse Videos

Google Earth Engine, a cloud-based platform for planetary-scale environmental data analysis, was the platform of choice to efficiently and accurately create timelapse videos of our seven case study lakes. To make the timelapse videos, we began by collecting all available satellite imagery taken over this region by Landsat 4, 5, 7, and 8 from 1984 to 2019 (approximately 840 satellite images). These images were taken roughly every fifteen days at 30-meter resolution. We first applied a cloud mask to these images. This resulted in a reduction of the dataset to mainly spring and summer images due to the high prevalence of clouds in the fall and winter. Data from 2019 was ultimately not included in the final dataset, as the only data available was from winter and early spring, when clouds obscure the lakes. We corrected the Landsat 8 images using a standard method to facilitate better comparison with previous satellites (Roy et al., 2016). For each year, the median value of each pixel of the lake was calculated across the set of masked and corrected satellite images to create an annual composite image. The 35 annual composite images were brightened to appropriate viewing levels and stitched together into a video at 1 frame per second. This frame rate was chosen because 1 frame per half-second was too short for the user to fully take in each annual image and 1 frame every two seconds made it more difficult to compare between successive years.

Production of the Time Series Charts

After all of the satellite images were masked and corrected, median values of the red, green, blue, and near infrared satellite bands were calculated for the lake area in each image, as well as the number of pixels that were in each image. The datasets for each lake were exported as .csv files. The time series data were plotted in Observable notebooks using D3 with each satellite band (red, green, blue, or infrared) plotted separately for each lake. The median band value for a single image represents a single time point on these scatter plots with data points appearing approximately every fifteen days in years with consistent satellite coverage. The annual median over these individual images was calculated and plotted on top of the individual image data as filled circles connected by a line to better show the trend over time for each band. The y axis of each plot was fixed to better enable comparisons between datasets from different lakes. A function to find the difference in these annual median values between two separate lakes was created so users would be able to examine differences for each band between lakes within our visualization. These differences were also plotted as filled circles connected by a line so it would be clear to users that they were seeing the difference in annual median lake intensity of each band over time. The y axis was allowed to scale to the data for each new plot, as the differences in median intensity were quite different between lakes. A solid line to mark zero on the y-axis was included to better orient the viewer to whether the changes were positive or negative. A tooltip was added to each plot revealing the year and intensity or difference in intensity depending on the plot.

Development of the Data Exploration Tool

The layout and interactions of our data exploration tool were developed using HTML and JavaScript. Outside libraries and scripts that were used include the Leaflet library, the zoomhome Leaflet plugin, the MiniMap Leaflet plugin, and the D3 library. Leaflet is an open-source JavaScript library for interactive maps and D3 is a JavaScript library for visualizing data with HTML, SVG, and CSS. Notebooks containing our time series charts from Observable were embedded into the web page. Coordination between all areas of the tool were made possible by event listeners.

RESULTS

The data exploration tool created in this project consists of three parts: the dataset summary section (Fig. 1), the main single lake detail section (Fig. 2), and the lake comparison section (Fig. 3).

Dataset Summary

To orient the user to the satellite imagery dataset we visualized, static charts characterizing the data were created and embedded into our webpage (Fig. 1). The charts are ordered to tell a story about the data as the information becomes more detailed: by year, by season, and then by individual lake and satellite band. The first chart depicts how many cloud-free pixels there are in the region of interest year by year. This is intended to show the user how the usable data is distributed over time. A height encoding in the bar chart makes comparison between years natural. The second chart depicts how many cloud-free pixels there are in each season over all the years, illustrating how most of the data is from the summer months when there are fewer clouds. Another bar chart height encoding is used for the same reasoning as the first figure. The third chart depicts the normalized median pixel value over all the images for each lake and for each satellite band. The values are encoded by the radius of the circles to enable quick qualitative comparisons between lakes or bands and color-coded for easy recognition of what band is being displayed.

Single Lake Detail View

There are three areas for user interaction in this section: the map, the timelapse video, and the time series charts (Fig. 2). The map and timelapse video are on the left side of the user's screen while the time series charts are all on the right side of the screen. This divides the screen into two equal areas. A zoomed-out inset map marking the area being observed around the lakes was connected to the main map so that they zoomed and panned at the same time. A reset button was added to center the main map on the lake area in case the user wants to return to their initial view. The main map has lake data embedded into it as blue pins. When the user selects a pin, a popup opens displaying lake details and properties, and the timelapse video below is updated to display the lake the user selected. Additionally, all the plots on the right update to the user-selected lake and the difference plots in the section below are updated so that the user-selected lake is Lake #1. Since the median pixel values for each band for one lake are similar, we made four charts vertically stacked on top of each other for the user to easily scan and compare. Each satellite image's pixel values for each channel is plotted as a lighter-colored open circle on a time series. Additionally, the lake's annual median pixel value is plotted as a connected scatter graph with closed circles in a darker color than the individual data points.

Lake Comparison

In the last section of the tool, the user is encouraged to compare the annual median pixel values in all four satellite bands for the user-selected lake from the map above with those of any of the other lakes via radio button selection across the top of this section (Fig. 3). The charts are similar to the ones in the single lake section except now the difference in annual median pixel value is plotted.

Case Studies: Fires, Floods, and Algal Blooms

To evaluate our tool, our collaborator helped us focus on seven interesting lakes in a small region of Alaska. We were able to observe a wide variety of processes in these lakes using our tool. For example, in the timelapse video for Canvasback Lake, damage from a wildfire can be identified by the scorched tan and red earth around the lake's edge. This video also shows large increases in lake area due to anomalously high rain years in 2006 and 2013. Some other interesting examples include: bright green algal blooms in the timelapse video for YF17 in 1989 and 2009, and a brown plume of sediment from a flooding creek that enters the left side of the YF20 in 2015 in the timelapse video. Since everything updates with the click of a map pin, the user is in full control and has all the information they need for future analysis and research questions.

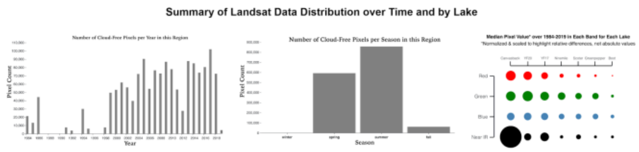

Figure 1: The dataset summary section with plots of annual and seasonal cloud-free pixel counts in this region and overall median comparisons between lakes for the different bands.

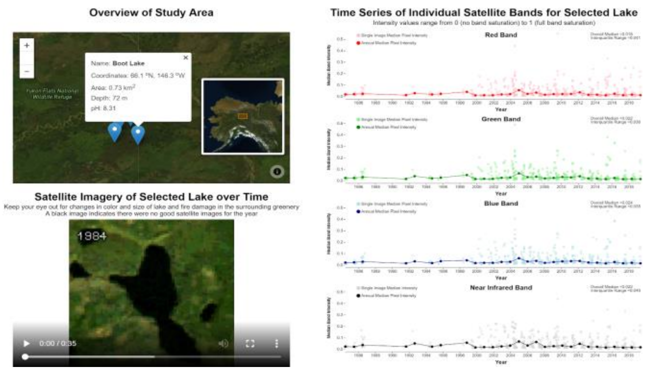

Figure 2: The main section of our data visualization tool showing an example of a selection of a lake and the corresponding video and time series plots for that lake.

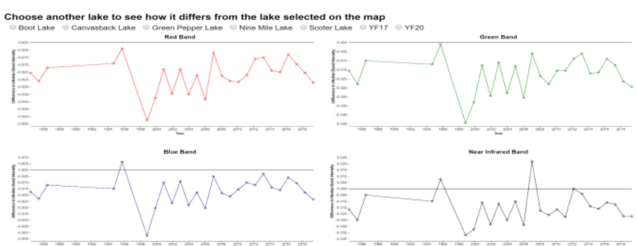

Figure 3: The lake comparison section showing the difference in annual median band intensity between two selected lakes. The radio buttons can be used to select which lake to subtract.

DISCUSSION

Using our new data visualization tool, our collaborator and her colleagues at the University of Washington have already been able to begin generating new insights about the seven lakes we highlighted. For example, many of the algal blooms and flood events for individual lakes that we described above in the Results section had never before been observed for these lakes and are of great interest to our audience. Many researchers who investigate these lakes focus their efforts on on-the-ground field campaigns and have never before seen overhead views of the lakes from satellite imagery. They found the videos remarkable as they were able to finally see data supporting full-lake trends they had gotten a sense of from their many years of field measurements but not been able to fully capture in their point measurements.

Our collaborator also shared with us that the timelapse videos will be especially helpful to her and her USGS and NASA colleagues during their outreach efforts to the broader public. In particular, she is looking forward to using it for several upcoming events including Evening Talks at the Olympic National Resource Center, guest lectures at local Seattle public school, and blog posts for the Freshwater Initiative.

The time series plots of band intensity over time are also proving interesting to our collaborator and her colleagues, as they are able to look for trends in annual median lake intensity for the first time (e.g., are any of these lakes getting greener over time?). They can also see if the number of extreme color-change events (those with band values significantly above the median) has changed over time. Our collaborator is already planning new ways to dig further into the data based on explorations using our visualization tool. Finally, our collaborator is able to directly compare band intensity between different lakes for the first time which helped her to explore some outstanding question she has about variability across the entire region over time related to different modes of interannual climate variability (e.g., the Pacific Decadal Oscillation).

FUTURE WORK

Currently, our tool highlights the seven Alaskan lakes our collaborator focuses on for her specific research. However, many of her collaborators in the NASA ABoVE project are interested in investigating the hydrologic system of the entire Arctic, which includes hundreds of thousands of lakes. Moving forward, we hope to fine-tune the back-end of the visualization so it can readily display any number of lakes a user is interested in.

There are two other refinements to our visualization that our collaborator noted would be helpful for some of her group's science questions. One would be to swap the Landsat imagery from 2015 onward with Sentinel-2, which records data at 10 meter spatial resolution (three times finer than Landsat 2), to better resolve changes in these small-scale lakes. The other would be to assemble seasonal composites for the videos instead of annual ones to be able to better pick out changes from wetter or drier seasons. The time series plots could correspondingly bin individual images by season instead of year, which would help to highlight differences between the summer and spring band intensities.

ACKNOWLEDGMENTS

The authors would like to thank our external collaborator Catherine Kuhn for proposing this project, giving us a crash-course in Google Earth Engine, and for many wonderful interactions over the course of this project. We also wish to thank Sherry and Dominik for their feedback during last week's design review. Lastly, we would like to thank Jeff for imbuing us with so much new data visualization knowledge over the course of the quarter.

REFERENCES

[1] Serreze, M. C., and R. G. Barry (2011), Processes and impacts of Arctic amplification: A research synthesis, Glob. Planet. Change, 77(1–2), 85–96, doi:10.1016/J.GLOPLACHA.2011.03.004.

[2] Zhu, Z.-L., and B. C. Reed (2012), *Baseline and projected future carbon storage and greenhouse-gas fluxes in ecosystems of the Western United States*. doi: 10.3133/pp1797.

[3] Kuhn, Catherine, et al. (2019), "Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity." Remote Sensing of Environment 224: 104-118.

[4] Mouw, Colleen B., et al. (2015), "Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions." *Remote sensing of environment* 160: 15-30.

[5] Roy et al., 2016, Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, Remote Sensing of Environment, 185, 57-70, doi:10.1016/j.rse.2015.12.024.