The Construction of Knowledge and Classification of Nature

Cloud identification as a lens to scientific inquiry

"The sciences do not speak of the world but, rather, construct representations that seem always to push it away, but also bring it closer" [1]

Laurel Garrett
Lewis & Clark College
ENVS 400 Spring 2016
ds.lclark.edu/laurelgarrett

Introduction

While science strives to practice objectivity, scientific findings are reliant on a network of actors, human and otherwise, that have biases. I argue that it is important to examine where bias comes from and how it factors into the credibility of scientific findings. Throughout this research I ask- *How is knowledge constructed in science and how is its validity assessed?* To examine how knowledge is constructed in science, I have immersed myself in the production process of automated cloud classification. In addition to the technical methodology, I identify personal and instrumental biases, assumptions, theoretical frameworks, and the network of actors that support this scientific research.

Theoretical frameworks

I use two theoretical frameworks in examining the production process.

1. **HYBRIDITY** comes from Bruno Latour who argues modernity attempts to purify Nature and Culture into a dichotomy. In actuality, they are an inseparable hybrid. In science, this hybrid implies knowledge is co-created from material reality (nature) and human constructed frameworks (culture).
2. **DISTANCIATION** was defined by Anthony Giddens in the context of modernity. Distanciation takes place, the physical setting and social context, and turns it into space, an independent and empty dimension. This best understood in the context of standardized measurements, which allow for place and time to be represented independent of location. In theory, this distancing allows scientists to be more objective with their observations.

Why study clouds?

"clouds and aerosols continue to contribute the largest uncertainty to estimates and interpretations of the Earth's changing energy budget" [2]

Clouds are a pivotal component of both weather and climate. As well as an integral part of the hydrologic cycle, clouds regulate the temperature of the Earth and atmosphere by both keeping the Earth insulated as well as reflecting solar radiation. However, these processes vary dramatically with the type of cloud.

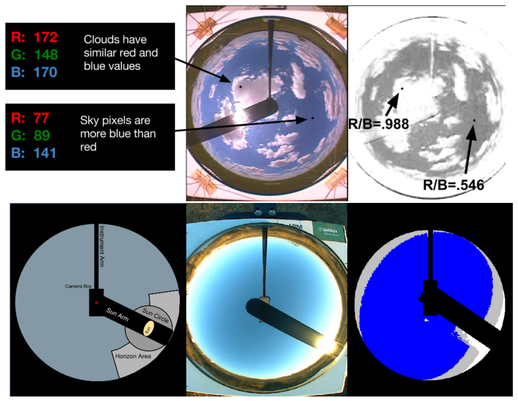
However, cloud type is a very subjective measurement that requires skilled human observers. Now, human observers are being replaced my instruments. The Total Sky Imager (TSI) shows a similar picture of what a human would see and has the potential to record cloud type. This TSI has been taking a photo of the sky every 30 seconds for the past 16 years. Through image processing, I will attempt to automatically classify these images for the type of cloud present.

Technical Methodology

1. Manual Classification

Three researchers manually classified almost 3000 images. A consensus set with rain, mixed cloud, and unsure cases removed was used for further processing.

2. Preprocessing



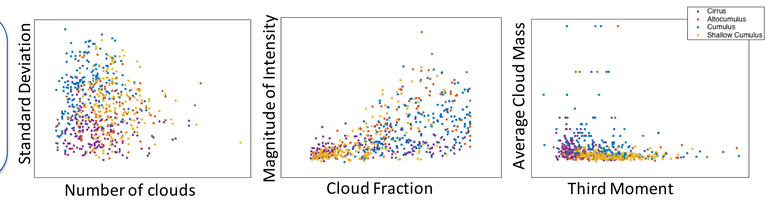
Each pixel in the sky images has a red, blue, and green component which combine to create colors. We divide the red by the blue channel to pull out cloud features. We also apply a mask so we only count the pixels that are part of the sky.

Horizon brightness is interpreted as clouds when thresholding.

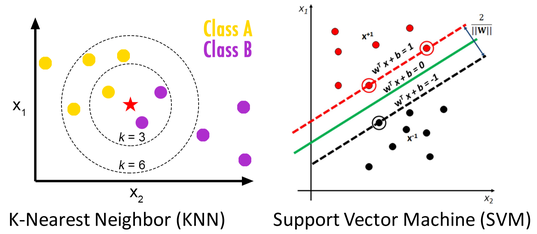
3. Calculate Statistics

Using the pixel values in an image, we can calculate statistics on the images.

Comparison of Statistics from Sky Images with Mixed Clouds

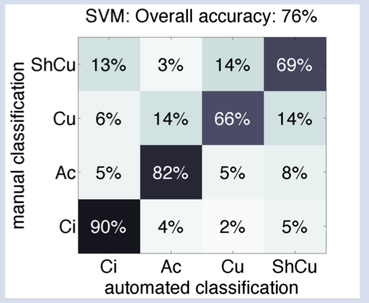


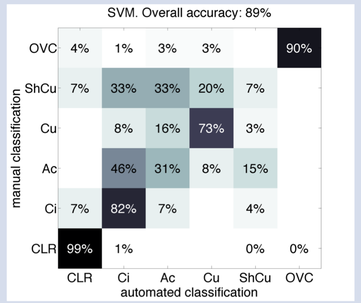
4. Automatically Classify



We tested two classifiers which both use a training set consisting of images with a known cloud type in order to find the cloud type in an unknown image.

Results





At 89%, the SVM had highest agreement with the manually classified images. However, this decreased to 76%, when clear and overcast cases are taken out. The success of the classifier is based on a human's ability to recognize the cloud type from a photograph. The agreement of the three manual classifiers was around 75%. Furthermore, the consensus set of images only includes high-quality single cloud type images. If the sky images were chosen at random, we would not expect as high of agreement.

Reflections on producing knowledge

I rely on a broad network of interdisciplinary actors including

- Atmospheric Radiation Measurement facility and DOE for funding
- Previous cloud classification scholars
- Computational power for massive amount of data
- Matlab computing environment
- Machine learning for classifiers

The construction of knowledge relies on a process of reduction and representation

- Within my research this chain goes from physical sky --> images --> statistics --> classification --> communication

Distanciation from material world simultaneously brings understanding closer

- Tradeoffs and compromises in chain of representations

"what we lose in matter through successive reductions, we regain a hundredfold in branching off to other forms that reductions- written, calculated, and archival- made possible" [3]

Conclusion

While science relies on methods that distantiate in order to reduce subjectivity and bias, individual particularities and specificities of place, the researcher, the processing, and the instrument still exist. The constructed frameworks that scientific methods utilize meld with the real world to create a hybrid understanding. The process of distanciation ends up bringing knowledge closer. Through these successive representations of the "real" world, we learn more than we could with just our own senses.

1. Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies.* Harvard University Press. 1999. *Pandora's Hope: Essays on the Reality of Science Studies.* Harvard University Press.
2. Boucher et al. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* 573.
3. Same as 1.