The Construction of Knowledge and Classification of Nature: Cloud Identification as a Lens to Scientific Inquiry

# 1 Top of Hourglass

## 1.1 Framing Question

**How is knowledge constructed in science and how is its validity assessed?** This question deals very broadly with the production of scientific knowledge and the validity of the outcomes. Meteorology and climate science are related and intersecting fields that have become increasingly important because of climate change. I plan to answer this question by examining the disciplinary history of meteorology and climate science with an emphasis on the production of knowledge. I will consider the social and political forces as well as how human observations and technology are developed and used and how that, in turn, affects our understanding. My research question asks the same question but at the level of automated cloud classification.

## 1.2 Background

### 1.2.1 Existing Studies

There is a wide variety of literature on the production of scientific knowledge. While usually science strives to be as objective as possible, many factors influence the outcomes including values, politics, place, and instrumentation. Sundberg 2005 discusses the production process in meteorology and states it is necessary to study how science is practiced and the origin of knowledge not just how it is used. Since I am talking about clouds as a hybrid object, I will introduce the nature-culture hybridity here. (Latour and Hulme)

### 1.2.2 Justification for studying meteorology and climate science

* Climate change is currently a huge factor in studying weather and climate, but historically meteorology was especially important for agriculture and mariners.
* Entanglement of meteorology and climate science and the hybridity of weather 🡪 (Hulme 2010)
* Climate models and weather forecasting: See Vilhelm Bjerknes
* Clouds as a part of both meteorology & climate science
	+ Earth’s Radiative Budget
	+ Cloud types predicting weather

### 1.2.3 Brief History of Meteorology and Climate science

Throughout these historical examples I will discuss what factors influenced the production of scientific knowledge. The accuracy and/or validity piece is interesting to consider since this is very dependent on the technology and understanding at the time.

* Ancient Meteorology
	+ Aristotle’s *Meteorologica*
	+ Greek influence up to early modern period (Jankovic 2000)
* The Enlightenment period of English meteorology
	+ Entanglement of nature and culture through studying weather.
	+ Luke Howard’s cloud classification modeled after the Linnaean system standardized measurements and observations.
	+ Significance of just studying/recording anomalous weather versus continuous observations
	+ Beginning of continuous manual observations
* A later shift to talking about processes
	+ How rain/snow/hail/clouds/other meteors are formed?
	+ Process-oriented: a shift towards modeling, and inquiry about the physical conditions corresponding to cloud creation and cloud presence. E.g. the importance of the Norwegian cyclone model.
* Local/global scale of observations and data collection.
	+ Weather observations were (and still are) very dependent on place (often rural) although with the invention of the telegraph and a network of stations, observations could be recorded throughout larger areas. These local observation of cloud type had very practical uses to predict coming weather. These networks of weather observations expanded to encompass larger areas, but simultaneously lost specific information about each place. This leads into a discussion about how technology helped expand these networks and standardize measurements of meteorological information.

### 1.2.4. How instrumentation affects scientific knowledge

* Discussion of the advent of various meteorological instruments (Atlas 1997)
	+ I will pick one instrument measuring temperature, humidity, wind, or air pressure to explore more in depth. (possibly barometer?)
	+ How photography is used for science, cloud photography
	+ How did these instruments change the production process of knowledge?
* How does the space & time sampling capabilities of our instruments affect the data we collect and the world we perceive?
	+ Spatial and temporal data gaps can skew our perception of trends. Sea surface temperature example, Stephen Warren 1980’s cloud maps with missing ocean measurement, and USSR precipitation example.

### Cloud background

* Clouds as a hybrid object- Here I will introduce examples that emphasize the nature-culture hybridity of clouds like contrails (Burkhardt 2011), ship tracks, and clouds caused by biomass burning (Tosca et al.). Talk about cloud-aerosol interactions, highlighting both natural and anthropogenic aerosols.
* Entangling different spatial scales with examples of the soda straw view of cloud research, cloud parameterization for global circulation models (Randall 2003), and global cloud maps from satellites with different grid resolution.
* Who is involved in cloud research? Who funds it?
	+ Network of authors working on cloud classification and affiliation
* Introduction of various cloud observing technology (TSI, Ceilometer, Lidar, MMCR, CCN, and cloud radars etc.) and how satellites changed our perception of clouds.
* The long term cloud record, including problems with human observation and changes in technology, is a motivating factor for automated cloud classification (Dai et al. 2006)
* Cloud-aerosol interactions and climatic forcing motivates looking at cloud type, not just cloud fraction.

# 2 Middle of Hourglass

## 2.1 Situated Context

### 2.1.1 Introduction to automated cloud classification

Automated cloud classification exists within a network of agencies and organizations, instruments and data streams, and algorithm and models. It is my intention to investigate how the forces above lead to our understanding of clouds and inform cloud classification as well as how well the automated classification replicates human observations. While I do my very technical research, I will constantly ask myself about how the tools and algorithms I am using have led to these results.

(The following points may be dispersed through my technical methodology or separate.)

* ARM, the Southern Great Plains (SGP) site, and The Total Sky Imager (1.5 pages)
	+ Funding and grants, politics
	+ Why SGP? Where else is research done and why?
	+ Yankee Environmental Systems
	+ My interaction with PNNL
* About the data (.5 pages)
	+ Open source
	+ The pitfalls, excluding data from processing
* About matlab as an instrument
* Origin and network of classifiers used
	+ SVM- matlab central and user communities
	+ KNN

## 2.2 Focus question

**Where does our understanding of cloud classification come from and how accurate is it?**

## 2.3 Methodology

* Investigation into network of tools and algorithms used for automated cloud classification. (more points mentioned in 2.1.1)
	+ Author or origins
	+ Affiliation to a company, organization
	+ Technique (like image preprocessing for R/B, R-B, or intensity image)
* Develop automated cloud classification from TSI images and evaluate how accurately the algorithm classifies clouds type.
	+ I use matlab to analyze TSI sky images from the ARM SGP site. This involves a pre-processing of images, calculating image statistics, making a training set and doing manual classification, feeding the statistics into a classifier, and then assessing how well the automatic classifier did. This process happens many times, each time trying to improve upon the previous results with new statistics.

## 2.4 Results

* Visual aid for network of tools and algorithms
* Graphs of cloud classification accuracy (human classification vs. algorithm classification) and explanations of what they mean.
* Weave together results and talk about how these tools informed my research and the forces that made it possible for me to work on cloud classification.

# 3 Bottom of Hourglass

## 3.1 Larger Implications

* Scientific facts and accuracy
	+ Depends on classification algorithms
	+ Different instruments like Ceilometer, Lidar, radar, cloud satellite observations etc. can also measure cloud cover or type but may have very different results.
		- Comparisons of cloud observing technology (Wu 2014, Boers 2010)
	+ Can be misrepresented and communicated often through a simplification of factors that influenced the production of knowledge.
	+ Problems using the TSI to classify clouds (price of instrument means sporadic sampling, does not seamlessly append to cloud record, difficulties with automatically classifying)
* Who uses this data and how?
	+ Incentives and values of scientists, media, politician, and other are very different, which leads into the rich field of communication of scientific results.
* Bring in Atlas 1997 about atmospheric data flow and assimilation
* Optimizing system of resource users to best deal with weather and climate risks (Dutton 2002), some cool CMAPS of information flow

## 3.2 Further Study

* Data communication and miscommunication