Condensing Massive Satellite Datasets For Rapid Interactive Analysis

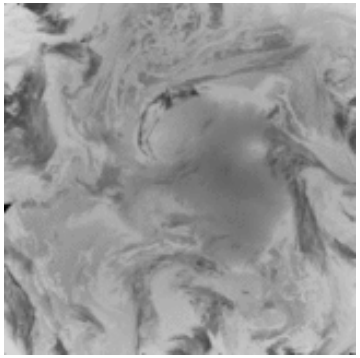
Glenn Grant

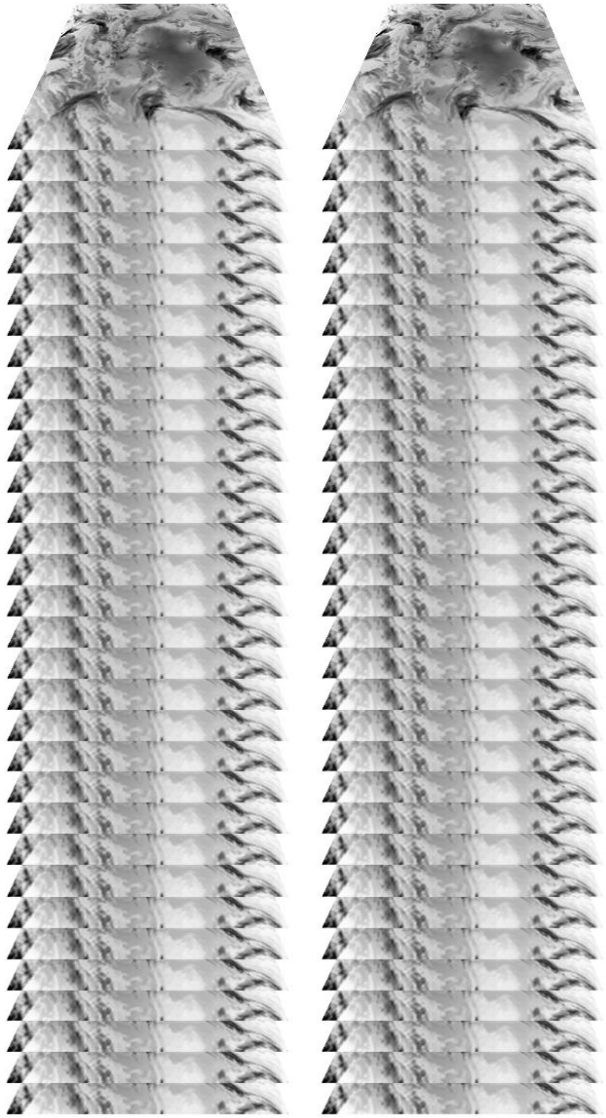
University of Colorado, Boulder

With: David Gallaher[1,2], Qin Lv[1], G. Campbell[2], Cathy Fowler[2], Qi Liu[1], Chao Chen[1], Rudolf Klucik[1], Richard McAllister[3]

1) University of Colorado, Boulder; 2) National Snow and Ice Data Center, 3) Orbital Micro Systems Inc.

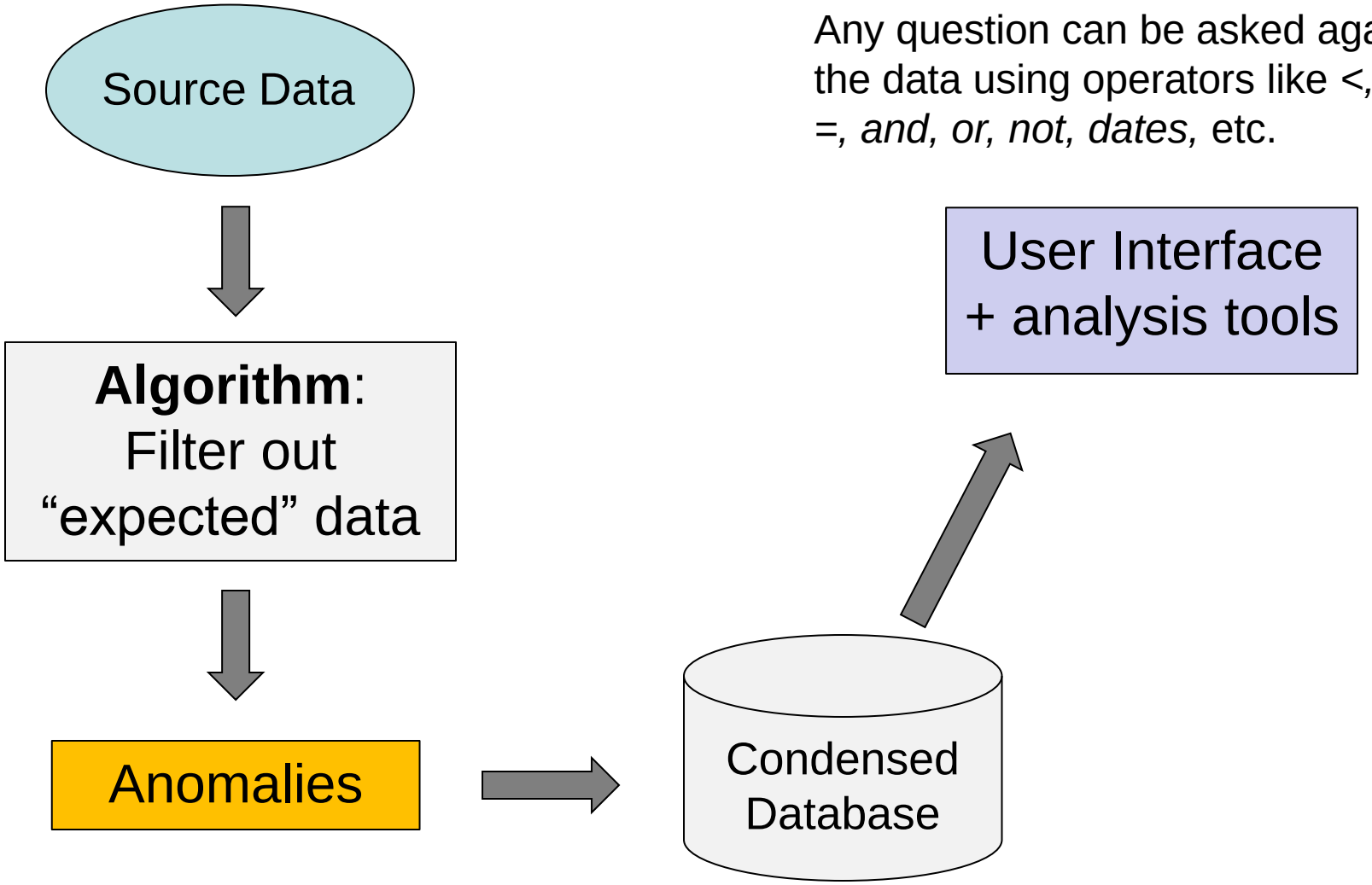
Project Overview – The Problem





Massive amounts of data accumulating –
Analysis tools have not kept up.

Project Overview - Objectives



Any question can be asked against the data using operators like <, >, =, *and, or, not, dates,* etc.

Demonstration Satellite Datasets

Antarctic Continent and Arctic Ocean

SSM/I Passive Microwave

~20 GB total (Antarctica)

25 years of daily imagery

25 km and 12.5 km pixel resolution

4 frequencies: 19, 22, 37, 85 GHz

2 polarizations most channels

AVHRR Polar Pathfinder

235 GB total (Antarctica)

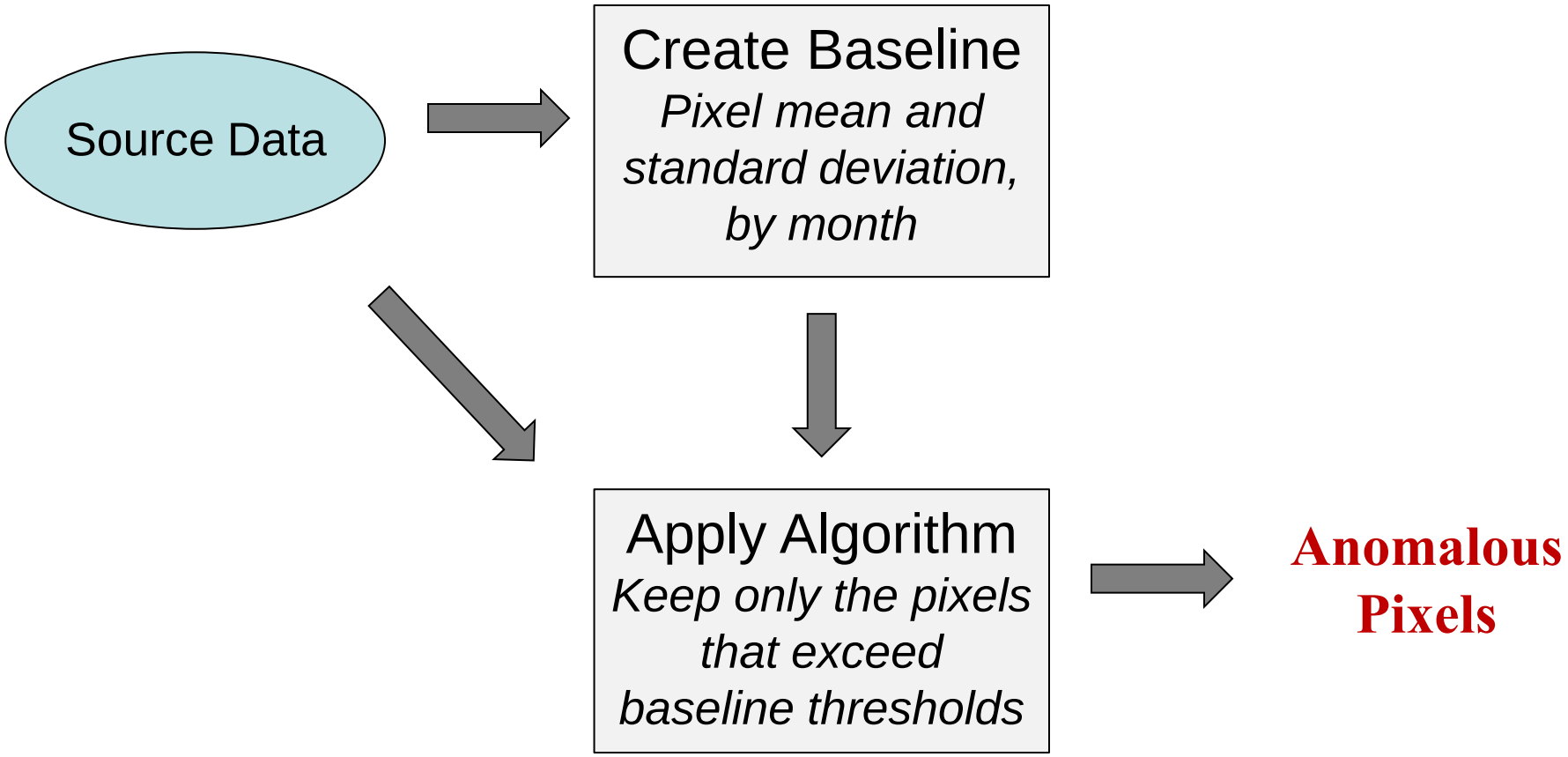
~25 years of daily imagery

5 km pixel resolution

3 IR channels, 2 visible channels

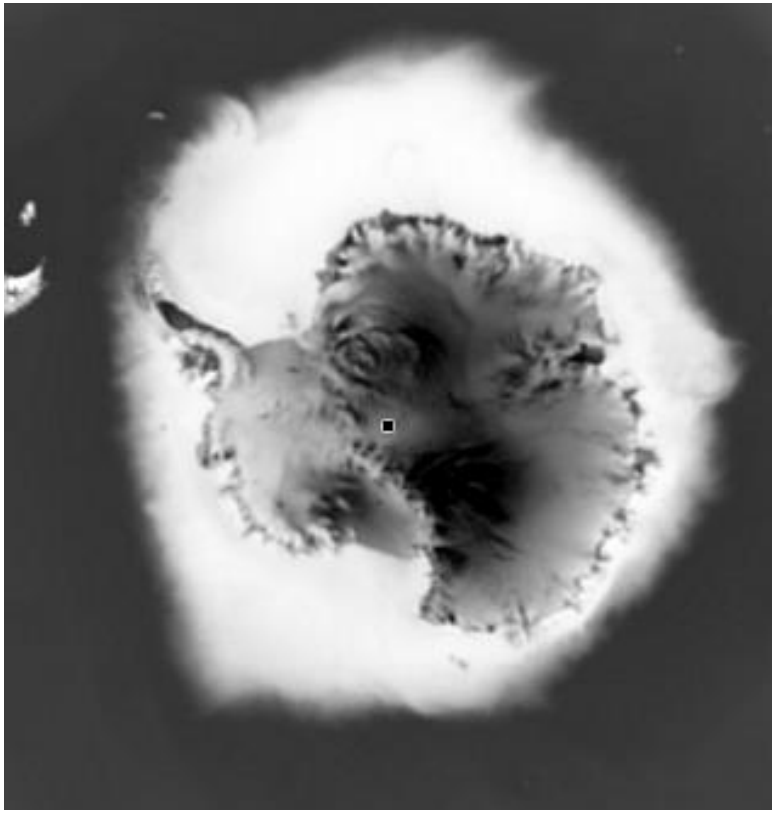
Albedo and surface temperature

Condensation Algorithm Development

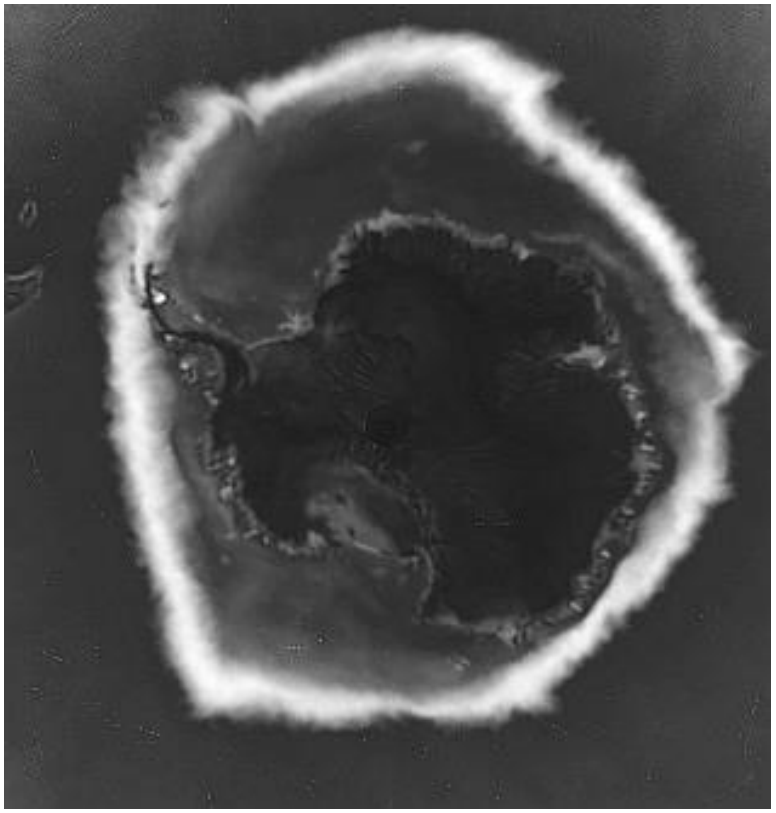


Condensation Process – Baseline

DMSP SSM/I (passive microwave, 25 km resolution)
19 GHz Vertical Polarization
Antarctica, September 1990 – 2014



Mean Brightness Temperature (Tb)

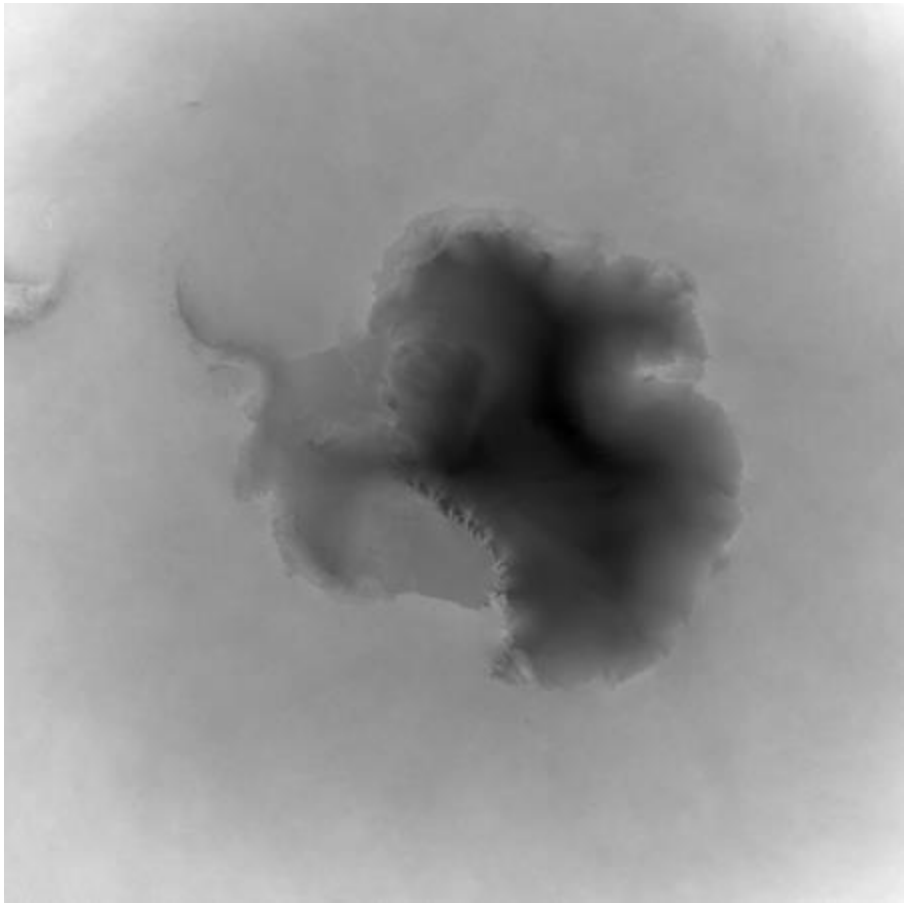


Standard Deviation

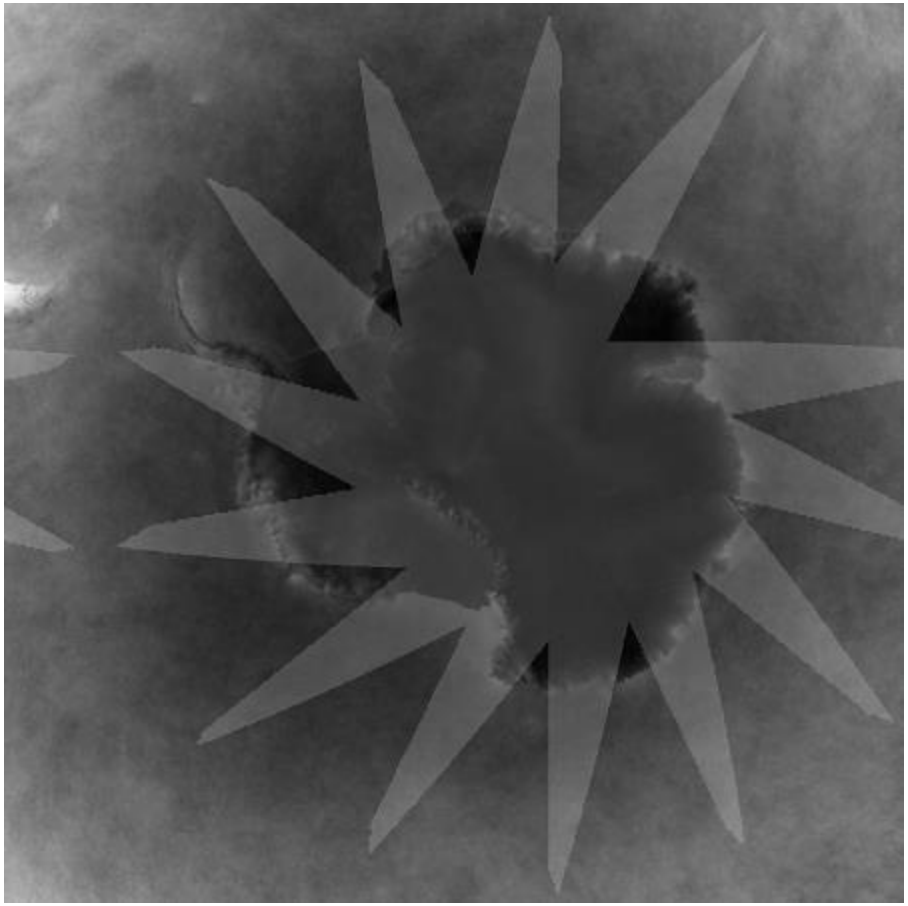
Condensation Process – Revealing Data Quality

AVHRR Polar Pathfinder, Surface Temperature

Antarctica, February, 1981-2005

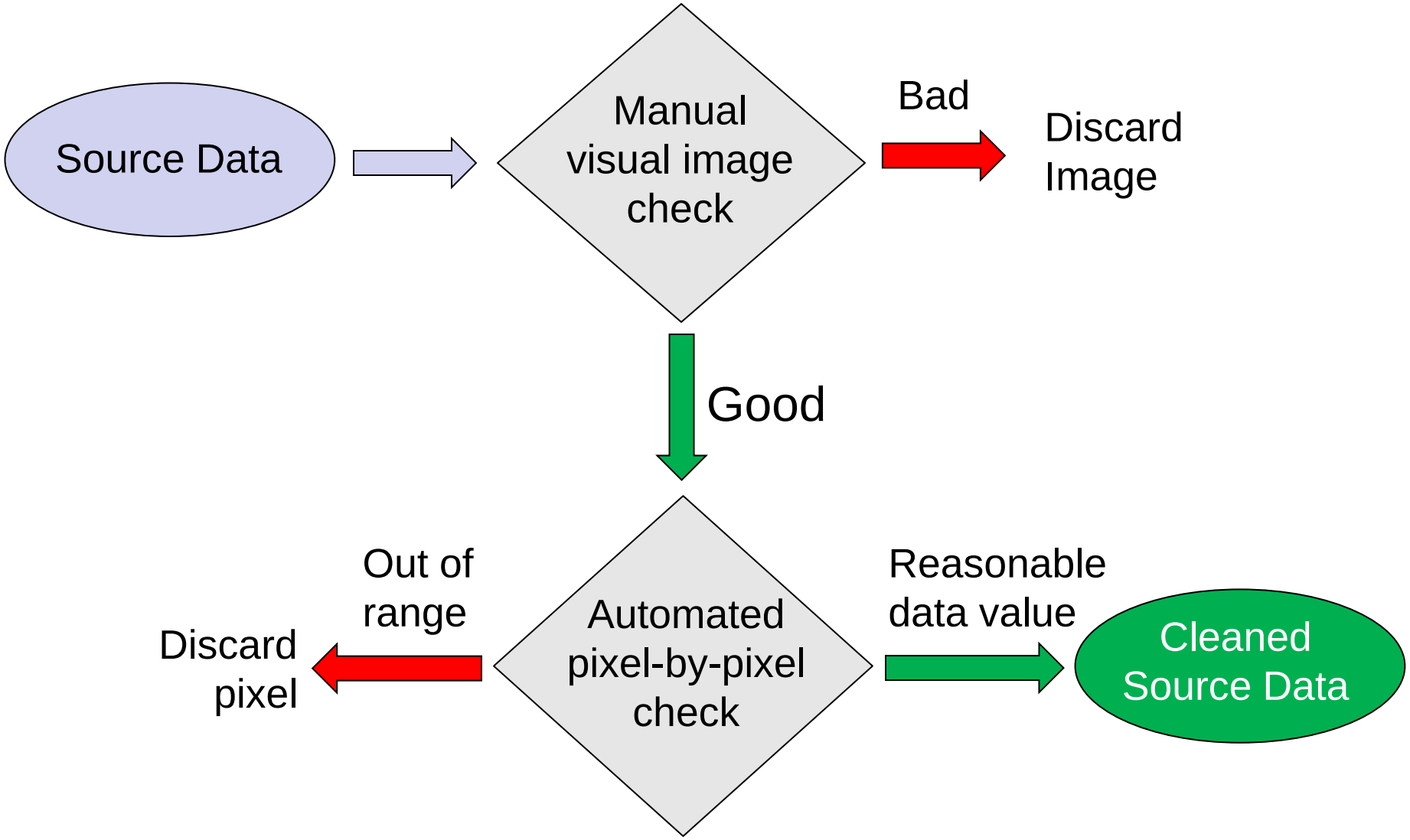


Mean Surface Temperature

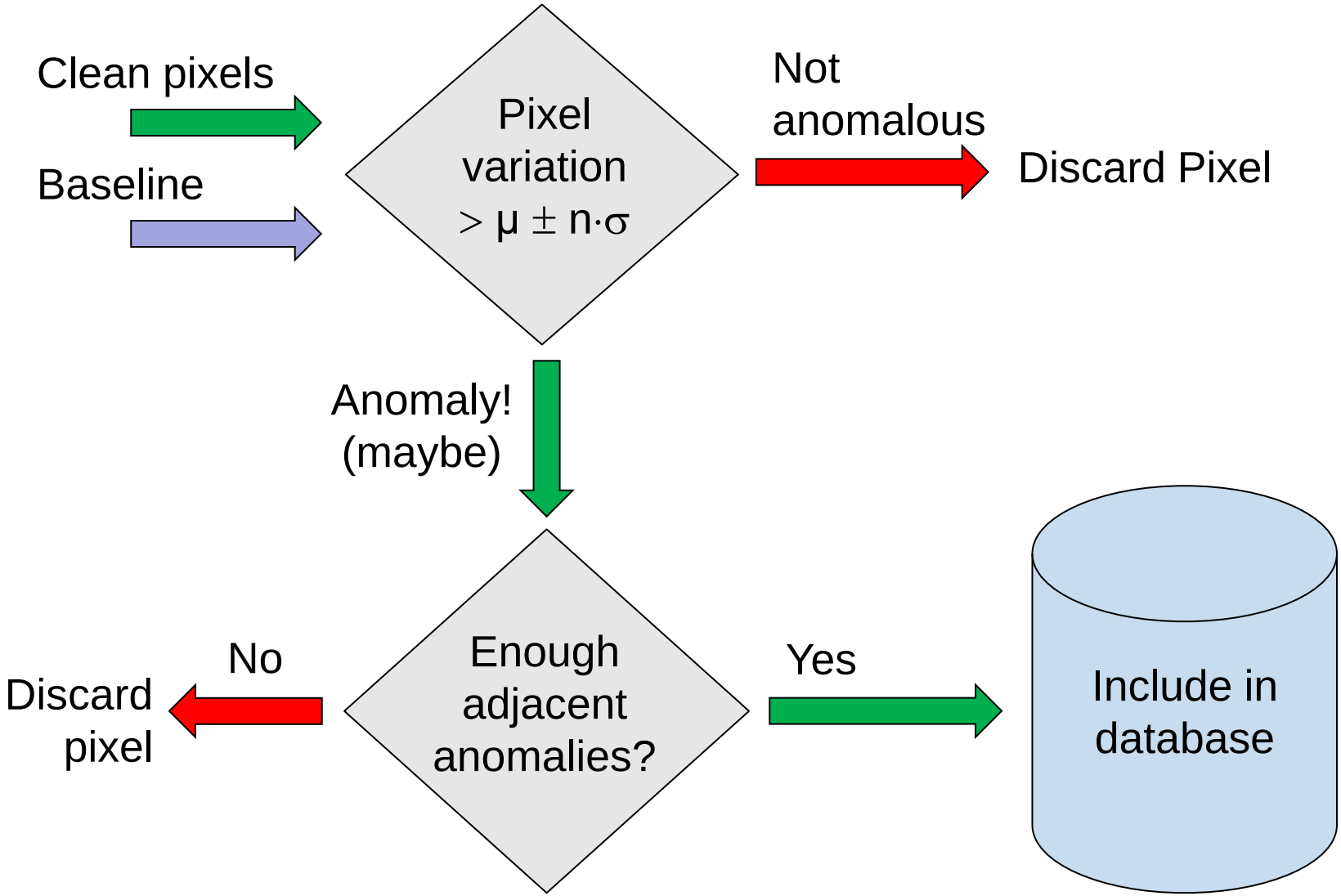


Standard Deviation

Condensation Process – Data Cleaning

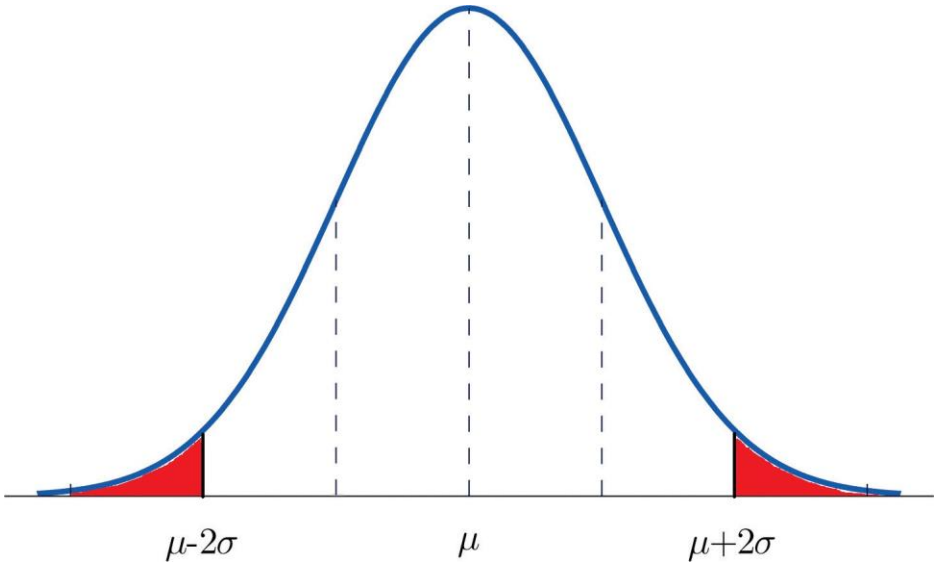


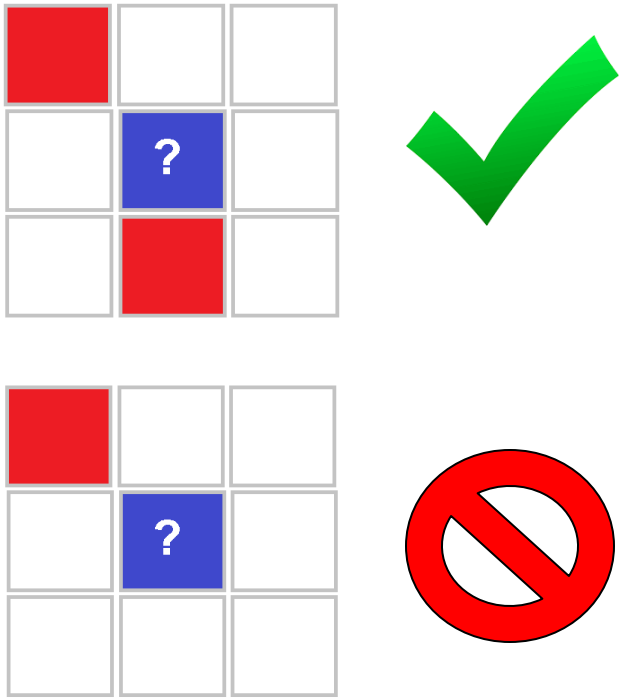
Condensation Process – Algorithm



Condensed Database - Explained

"The anomaly database contains pixels that are more than 2 standard deviations from the monthly mean, and adjacent to at least 2 other anomalous pixels."





Condensation Efficiency: 25 to 1

SSM/I Imagery, Southern Hemisphere, 1990-2014		
Freq. (GHz)	Anomalies	Percent Anomalies
19H	35444218	3.71
19V	34242309	3.59
22V	36236336	3.79
37H	35734214	3.74
37V	35786088	3.75
85H (1992-2009)	89497257	3.38
85V (1992-2009)	92563801	3.49
Total:	359504223	**3.57**

SSM/I:

Source data ~20 GB
Condensed to 0.71 GB

Reduced to 3.57% of the original.

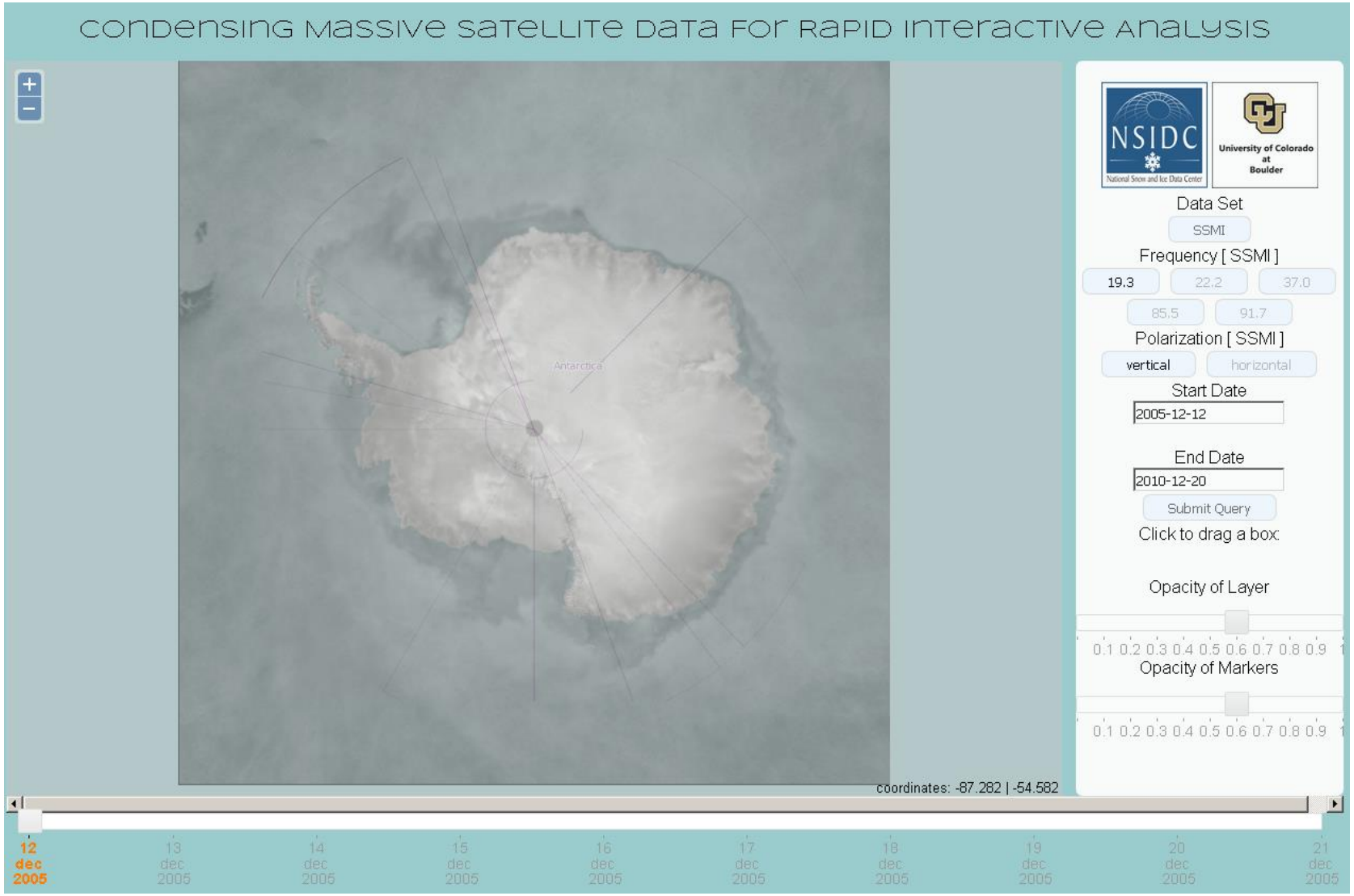
AVHRR, Southern Hemisphere, 1981-2005		
Channel	Anomalies	Percent Anomalies
Albedo	217977324	1.51
Channel 1	510534229	3.55
Channel 2	500451422	3.48
Channel 3	601637261	4.18
Channel 4	641882775	4.46
Channel 5	634990619	4.41
Surface Temp	641947045	4.46
Total:	3107473630	**3.60**

AVHRR:

Source data: 235 GB
Condensed to 8.5 GB

Reduced to 3.6%

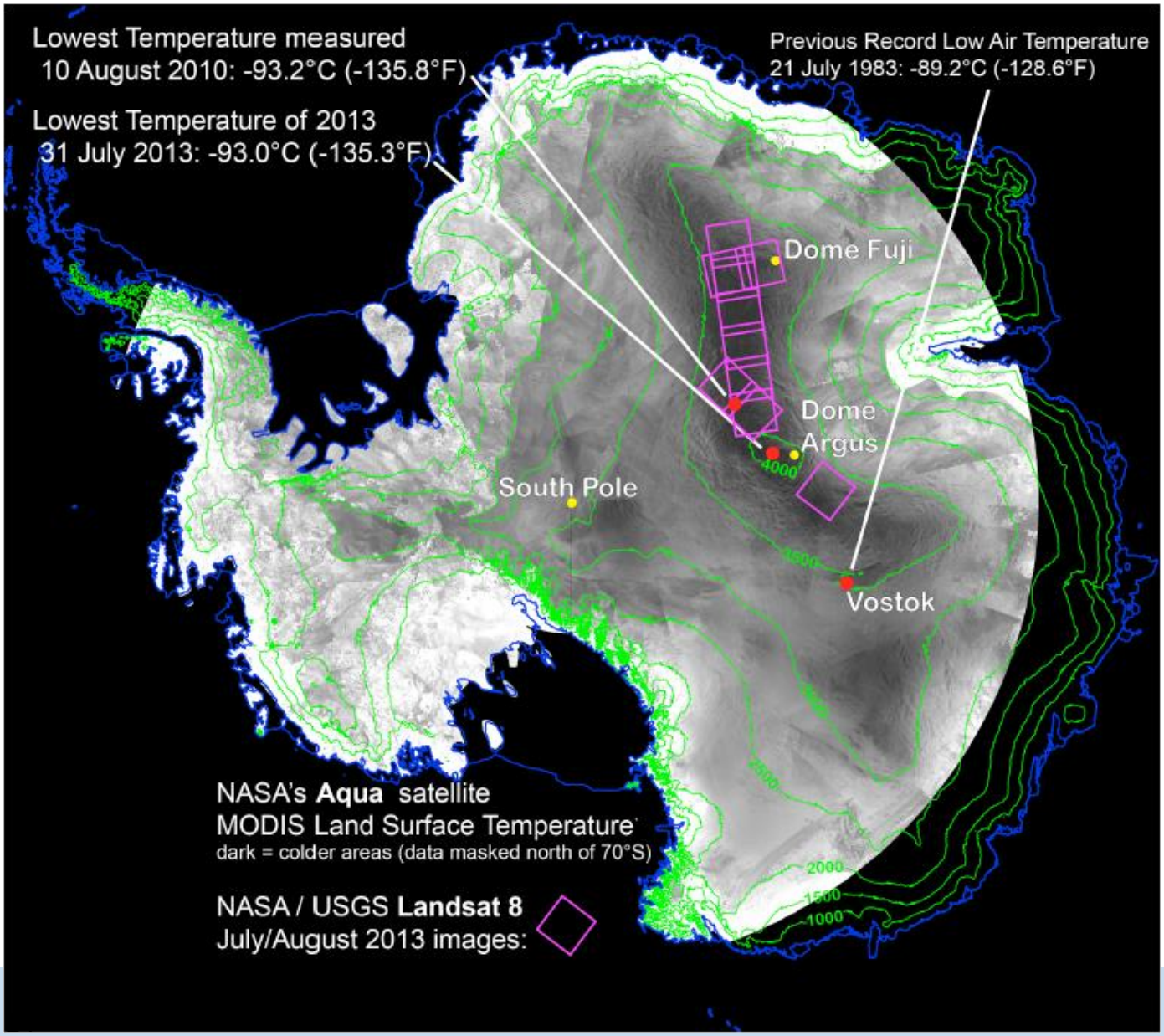
User Interface



What can you do with it?

Examples

Example Analysis – Coldest Temperatures



Scambos et al. (2013), "The Coldest Place on Earth: -90C and below from Landsat 8 and other satellite thermal sensors."

AGU Fall Conference, 2013

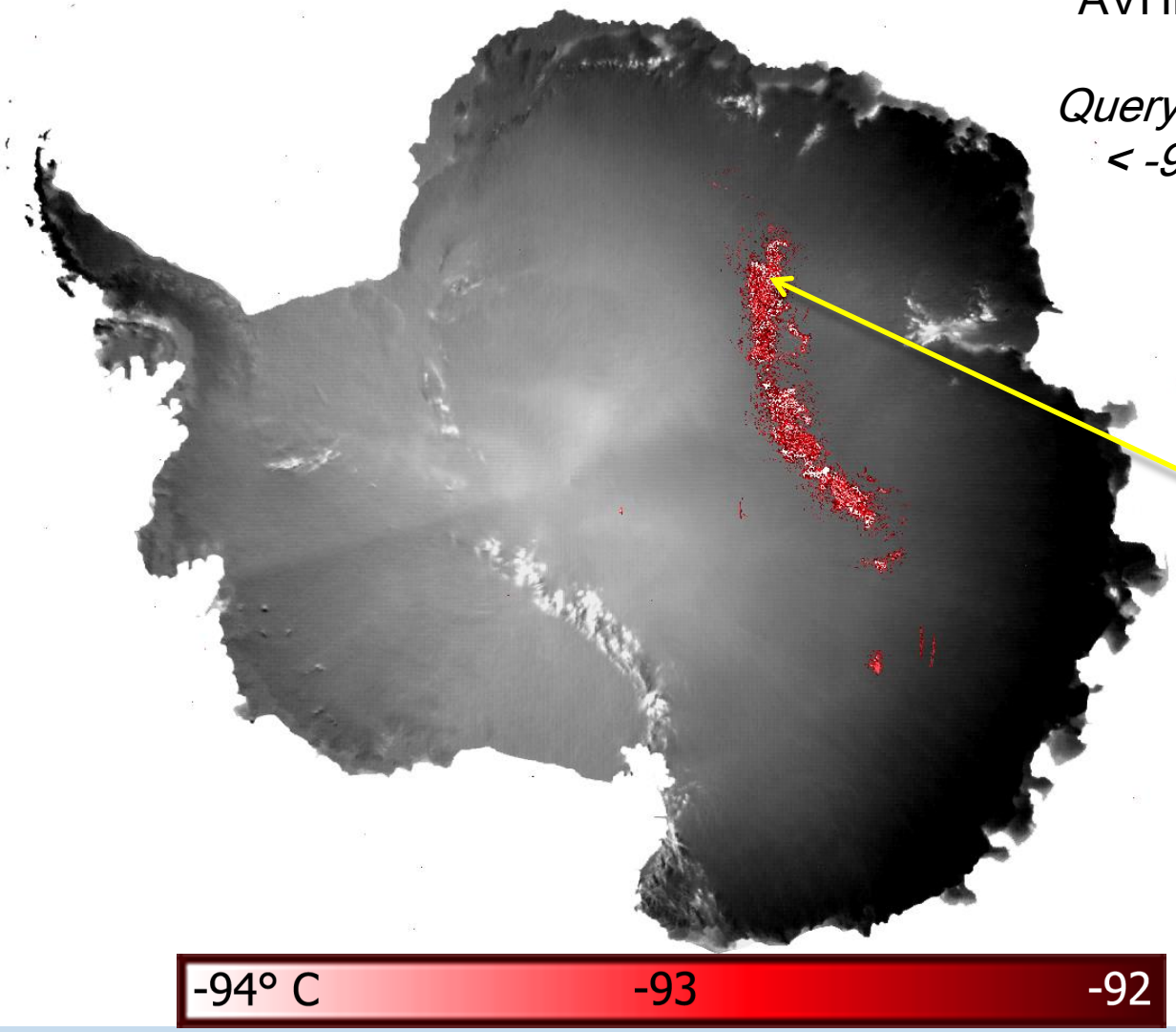
NSIDC

Example Analysis – Coldest Temperatures

10-minutes!

AVHRR condensate database:

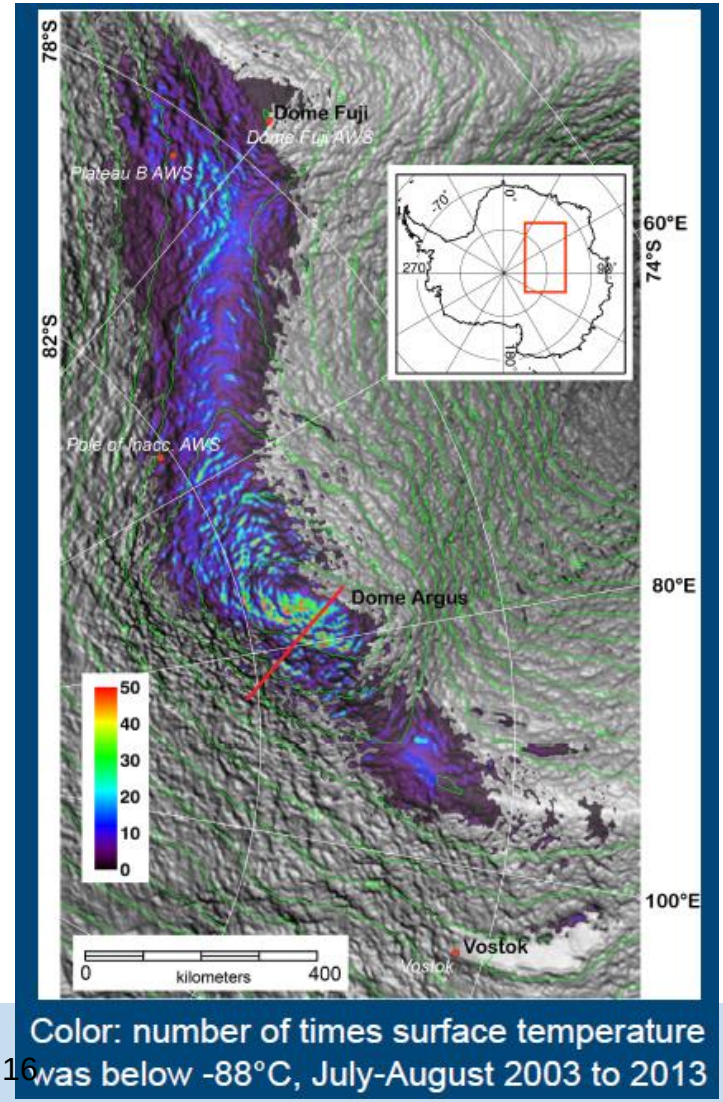
Query results for surface temp, < -92 C for July and August, 1981 - 2005



Coldest recorded temperature: -96.2°C August 20, 1997 (preliminary)

Example Analysis – Coldest Temperatures

From: Scambos et al. (2013), "The Coldest Place on Earth" *AGU Fall Conference, 2013*



Color: number of times surface temperature was below -88°C, July-August 2003 to 2013

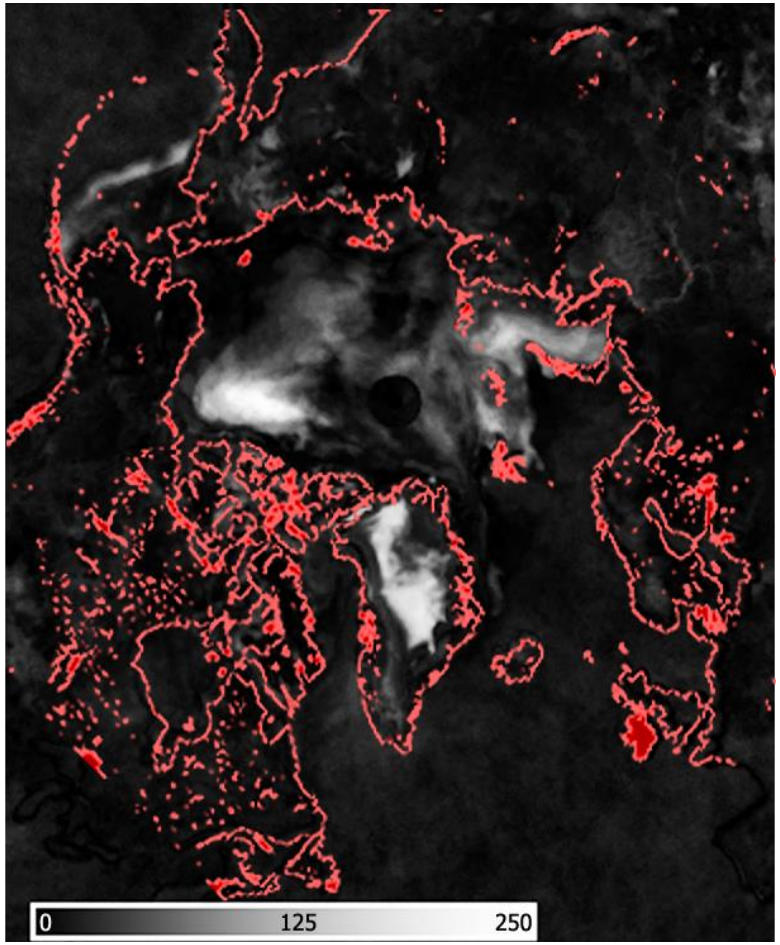
From: the AVHRR surface temperature condensate database: 1981 - 2005

Number of times below -88°C, 1981-2005

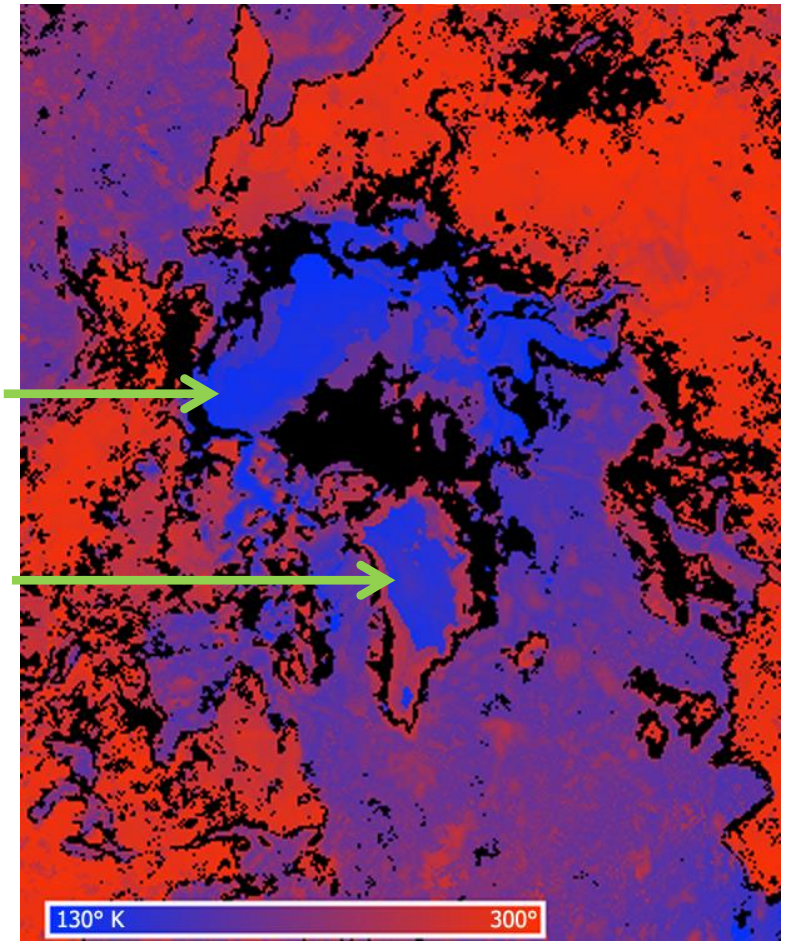
Example Analysis – Arctic Anomalies

Passive Microwave (SSM/I) 37 GHz horizontal polarization

Number of anomalies, 2012



What kind of anomaly? Tb, May-Aug



Anomaly Detection -- Limitations and Caveats

- Only looking at anomalies. “Normal” data is ignored.
- Slow trends are not visible.
- However, you *can* see temporal trends in anomalies.
- Noisy datasets and clouds increase climatology variability.
- Areas of high variability will hide anomalies. Conversely, anomalies will stand out areas of low-variability.

Conclusion

- We can condense large datasets by 95% or better.
- Rapid analysis and data exploration.
- Conceptually easy to grasp.
- Surprising results, reveals unexpected phenomena.
- Not intended for data archival.
- Future plans:
  - Database storage still needs work.
  - User Interface improvements
  - Use a much bigger dataset: MODIS

Thank you!

Made possible by the *National Science Foundation*, award #1251257

And the National Snow and Ice Data Center