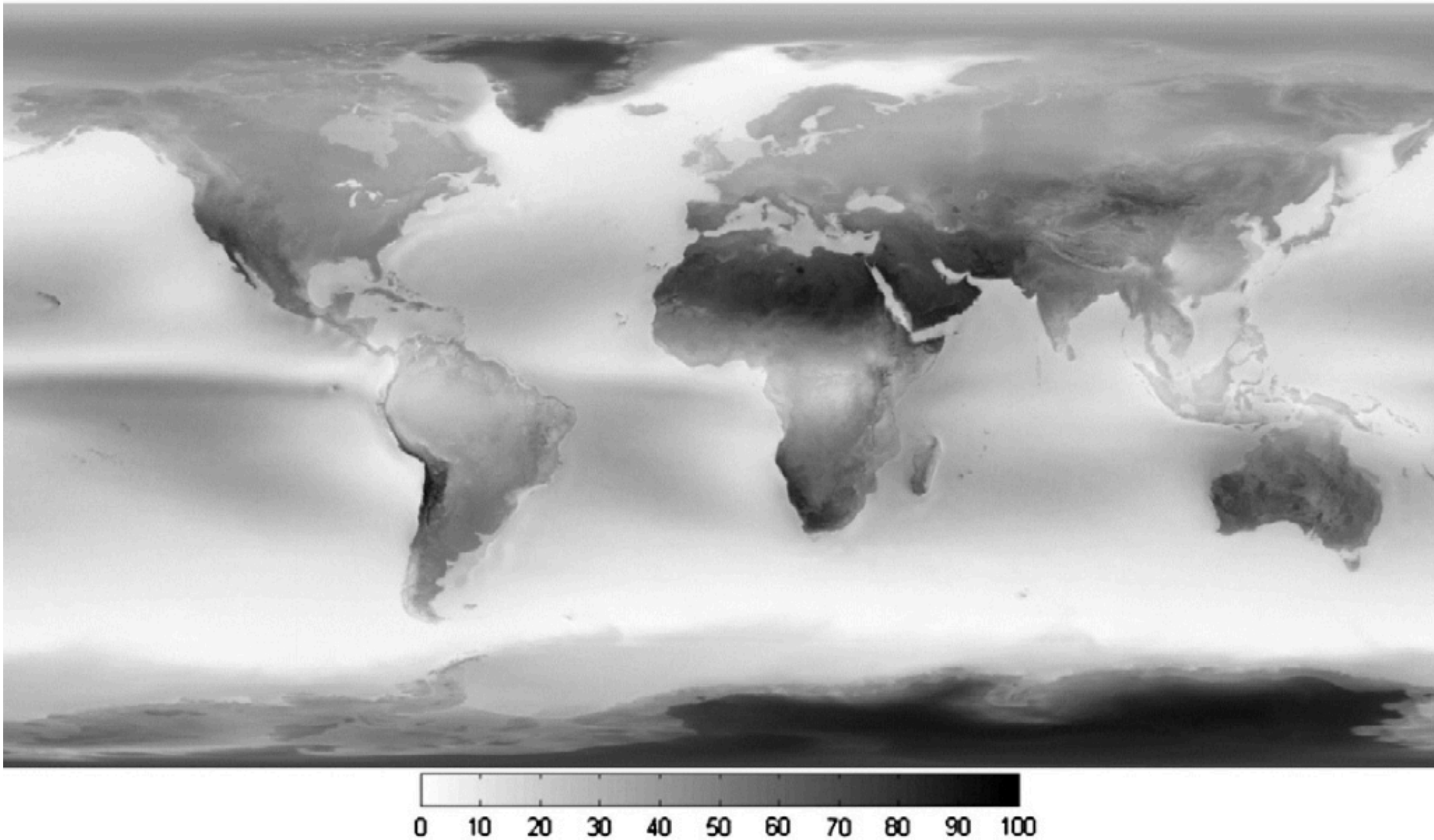
Jet Propulsion Laboratory
California Institute of Technology

Realtime Cloud Screening with AVIRIS-NG

David R. Thompson, Robert O. Green, Didier Keymeulen, Sarah K. Lundeen, Yasha Mouradi, Daniel Nunes, Rebecca Castaño, Steve A. Chien
Jet Propulsion Laboratory California Institute of Technology

The Earth is a cloudy place



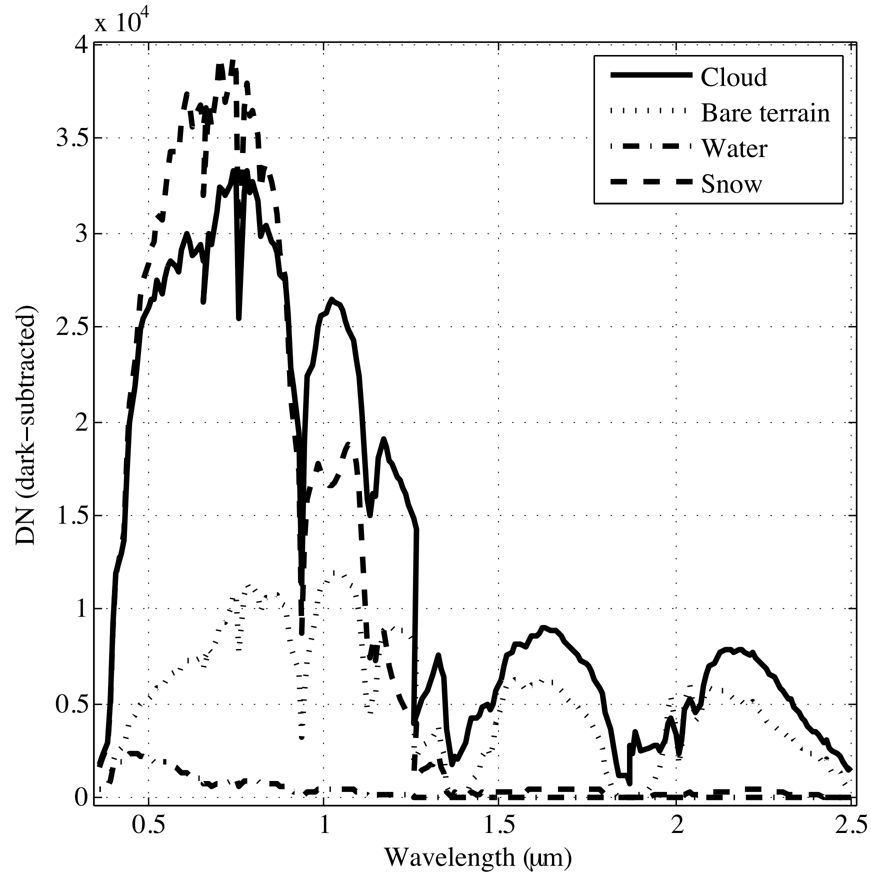
[Mercury et al. 2012]

Why onboard cloud screening?

- Spectrometer data rates approach 1Gb/s
- This imposes costs on buffering, transmission, analysis and curation
- For many analyses, 50-70% of scenes are unusable due to clouds [Eastman 2011]
- Excising bad data at the sensor benefits the entire processing and analysis pipeline
- Can be used to point gimbaled sensors

Previous cloud screening methods

- MODIS cloud mask
  [Ackerman 08]
- Physical Models
  [Gómez-Chova 07, Taylor 12]
- Thermal IR
  [Minnis 08]
- Pattern recognition
  [Lee 90]
- Onboard EO-1
  [Griffin 03]



Our unique requirements

- Use raw instrument DN values
- Low computational complexity
- Operate on-board in real time on existing computing hardware
- Excise opaque, non-cirrus clouds
- Reduce average data volume by 50% (based on a continuous year of ISS operations)
- Negligible false positives

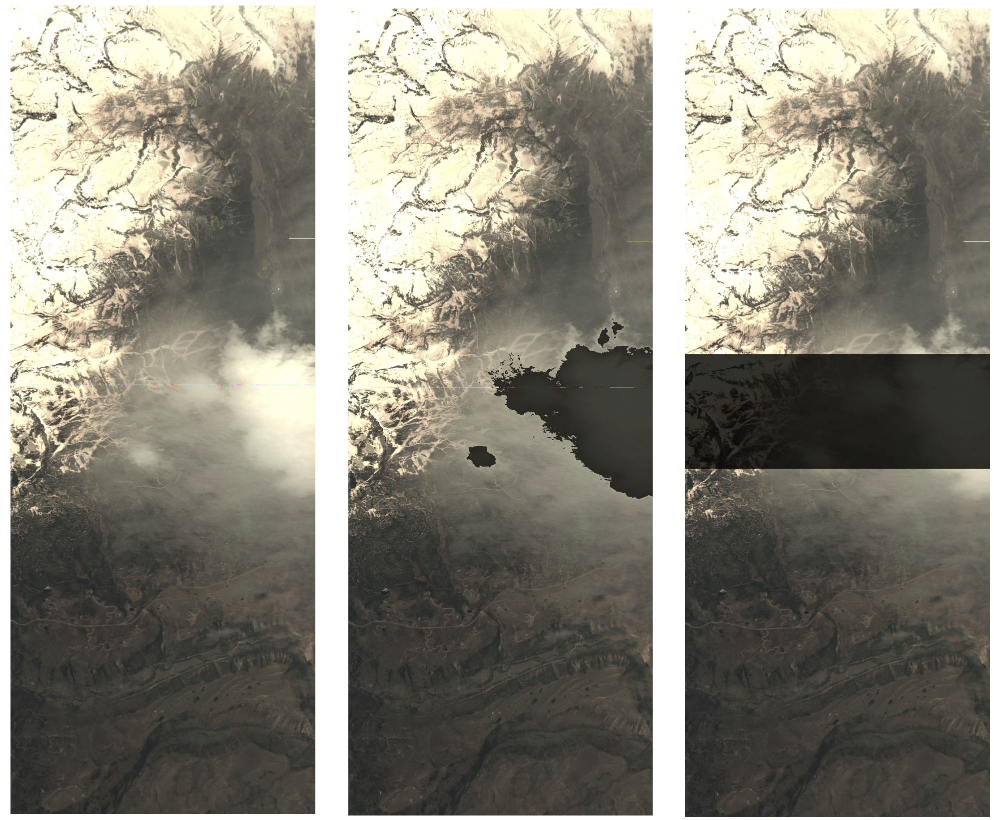
Concept of Operations

In advance

1. Model brightness of clouds and terrain
2. Use imaging geometry to predict optimal channel thresholds

Onboard

1. Apply thresholds to recognize cloud pixels
2. Excise image blocks with a large number of cloudy pixels



Simulation dataset

AVIRIS Classic high-altitude (ER-2) flights 2009-11

Labeled all cloud pixels by hand, excluding sunglint

Used half for training, half for testing

Land cover	Clear pixels	Cloudy pixels	Source
Water	6.7×10^8	2.2×10^7	UMD
Evergreen needleleaf forest	3.7×10^7	1.4×10^6	UMD
Evergreen broadleaf forest	5.8×10^6	5.8×10^5	UMD
Deciduous needleleaf forest	-	-	UMD
Deciduous broadleaf forest	2.1×10^8	4.4×10^6	UMD
Mixed forest	1.3×10^8	3.7×10^6	UMD
Closed shrublands	1.1×10^6	6.5×10^5	UMD
Open shrublands	1.3×10^8	1.8×10^6	UMD
Woody savannas	1.3×10^8	2.6×10^6	UMD
Savannas	-	-	UMD
Grasslands	8.9×10^7	3.1×10^6	UMD
Croplands	1.0×10^8	7.3×10^6	UMD
Urban and built-up	3.1×10^7	2.9×10^4	UMD
Snow and ice	1.3×10^8	4.4×10^6	
Barren	9.7×10^7	1.7×10^2	
Ocean glint	3.6×10^8	1.5×10^7	

Brightness model

Top of Atmosphere (TOA) reflectance

Instrument DN

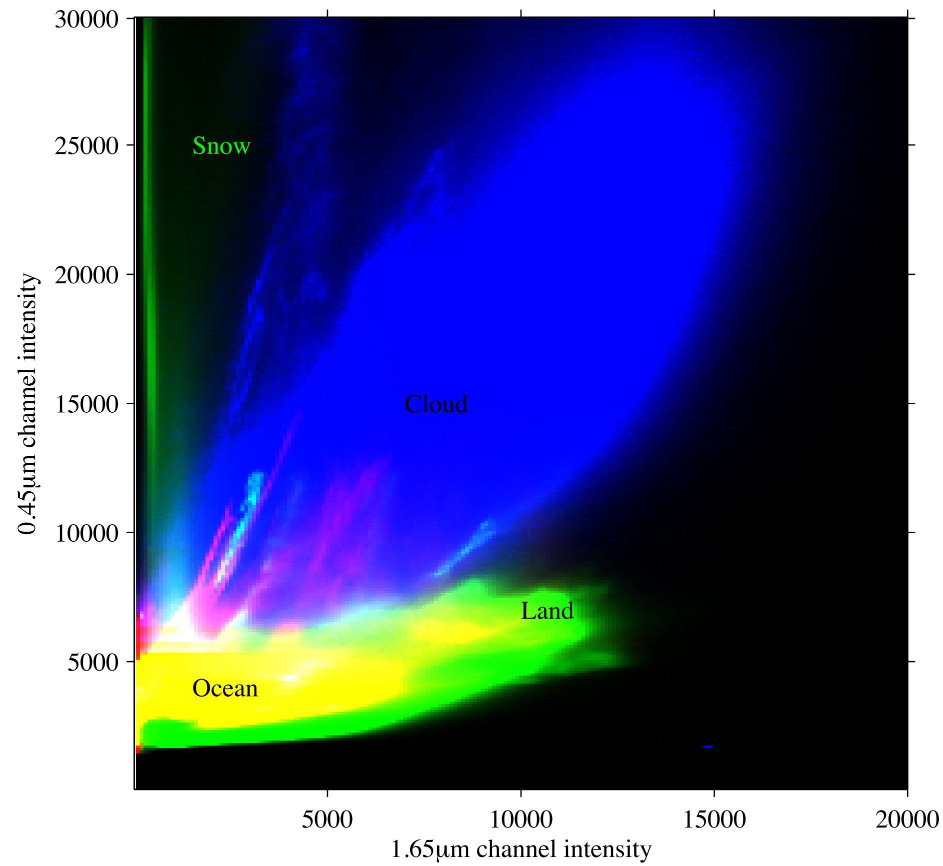
$$\mathbf{z} = \frac{\pi d^2}{\cos(\theta)\mathbf{s}}\mathbf{g}(\mathbf{y} - \mathbf{b})$$

Solar illumination term

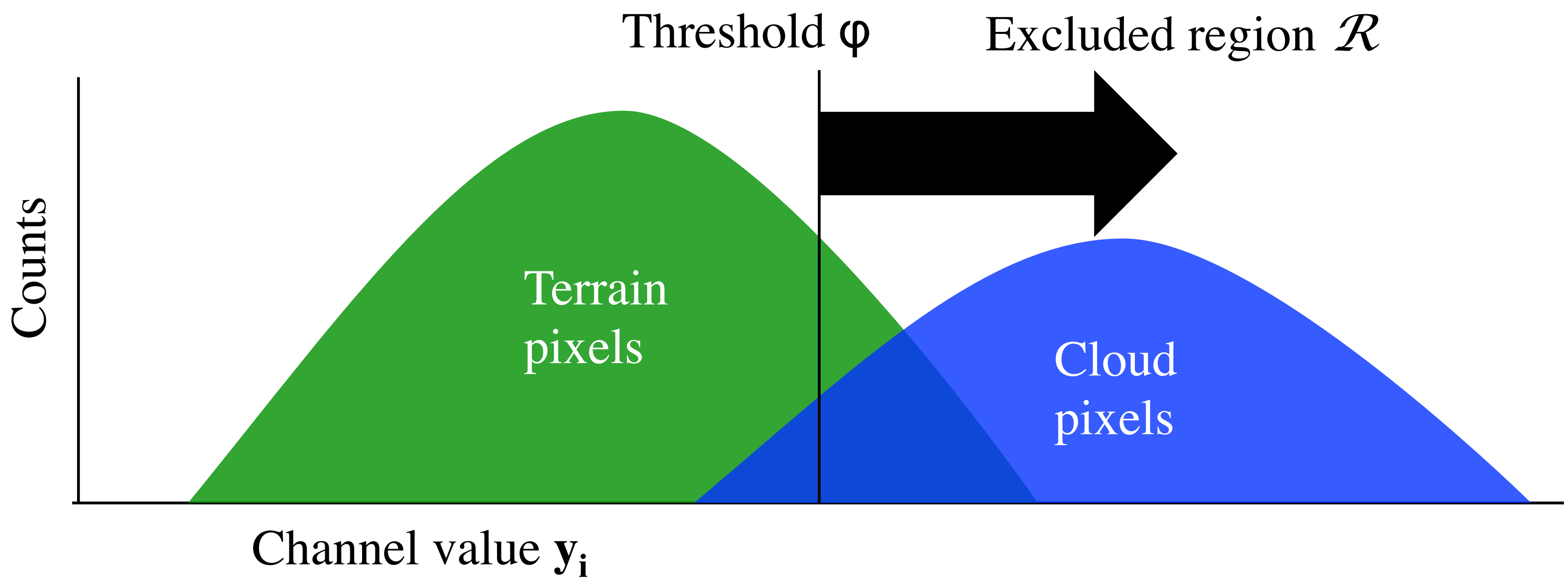
Gain

Dark current

Store **z** values in a histogram



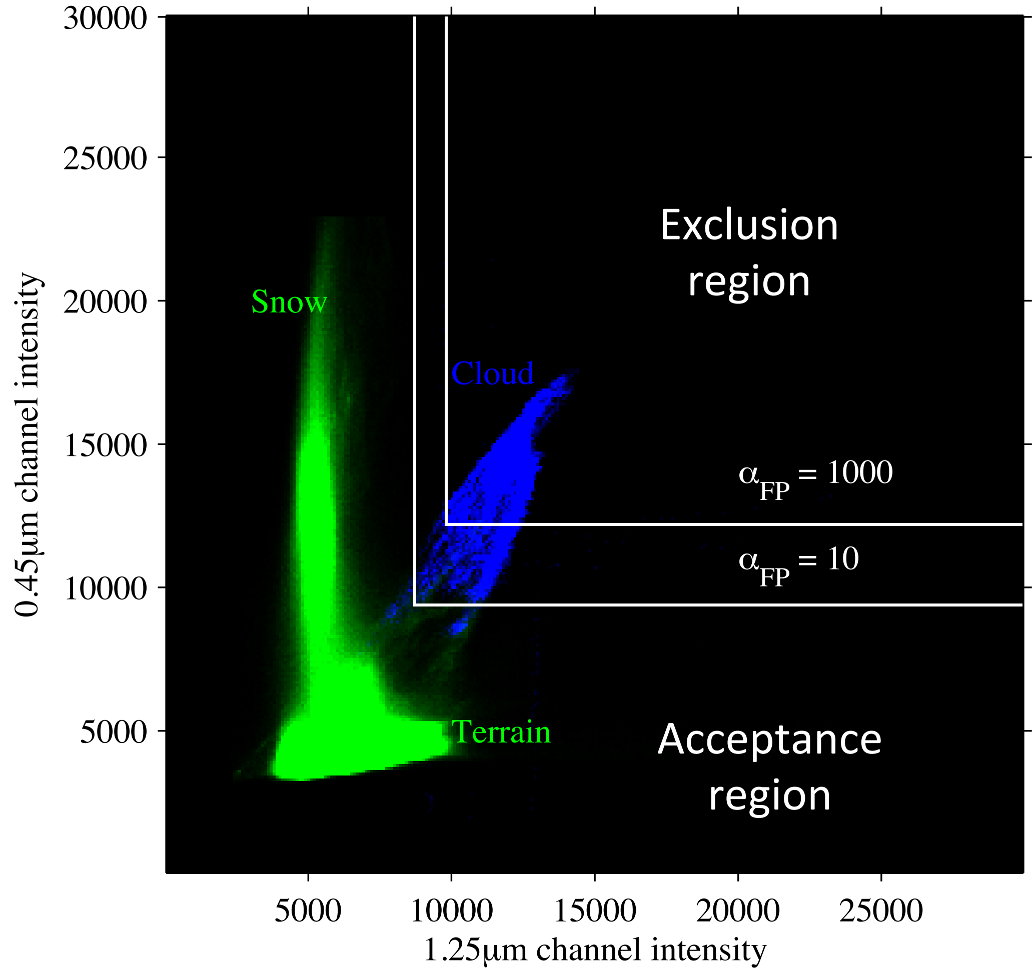
Calculating optimal thresholds



Calculating optimal thresholds

Must set thresholds on multiple channels at once

Right: Lenient and strict false positive costs



Calculating optimal thresholds

Use Bayesian decision theory to balance false negative and false positive risks

$$\underbrace{E[\mathcal{L}]}_{\text{Expected loss}} = \int_{\mathcal{R}} \underbrace{\alpha_{FP} P(\mathbf{y} \mid \mathbf{x}, c_1)\; P(c_1) d\mathbf{y}}_{\text{False positive cost}} + \int_{\mathbb{R}^d \setminus \mathcal{R}} \underbrace{\alpha_{FN} P(\mathbf{y} \mid \mathbf{x}, c_2)\; P(c_2) d\mathbf{y}}_{\text{False negative cost}}$$

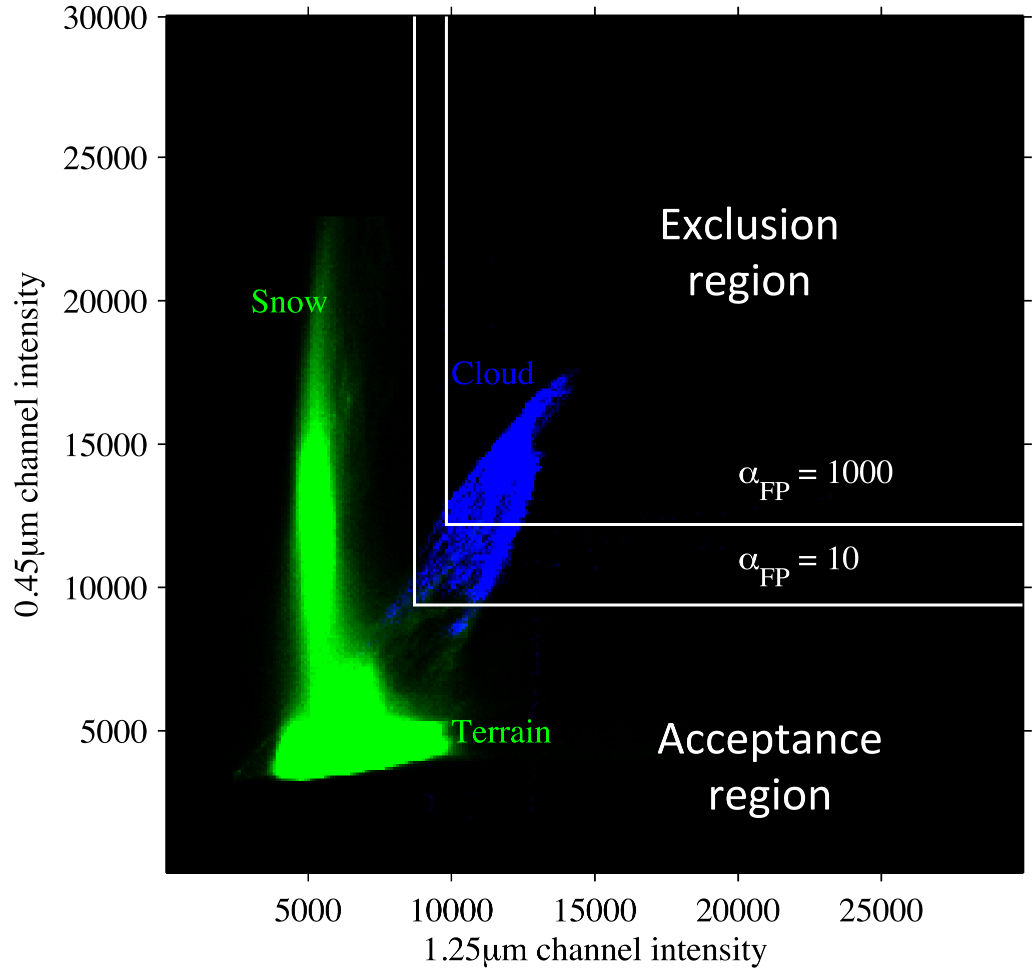
Acceptance region is free parameter

C_1 = clear, C_2 = cloudy
x = state vector, y = observation

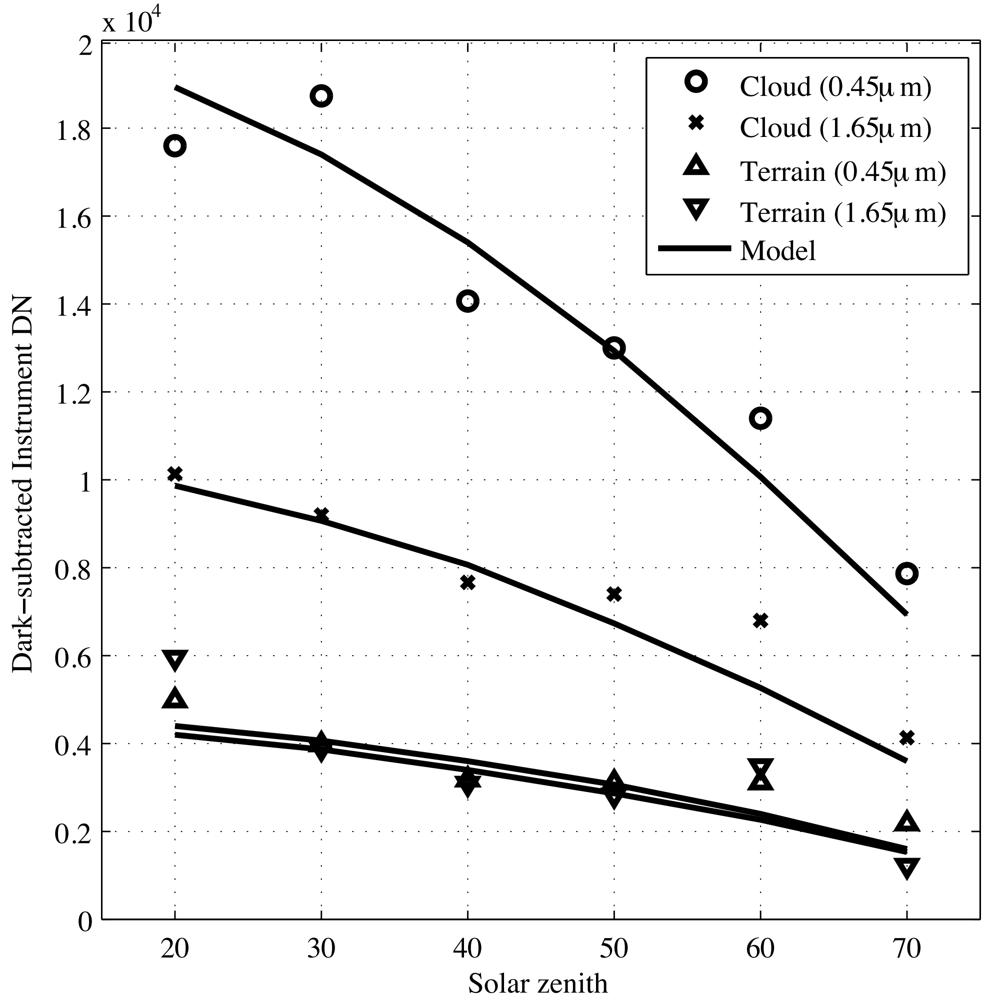
Calculating optimal thresholds

α_{FP} parameter represents our tolerance for false positives

A grid search identifies optimal threshold pairs

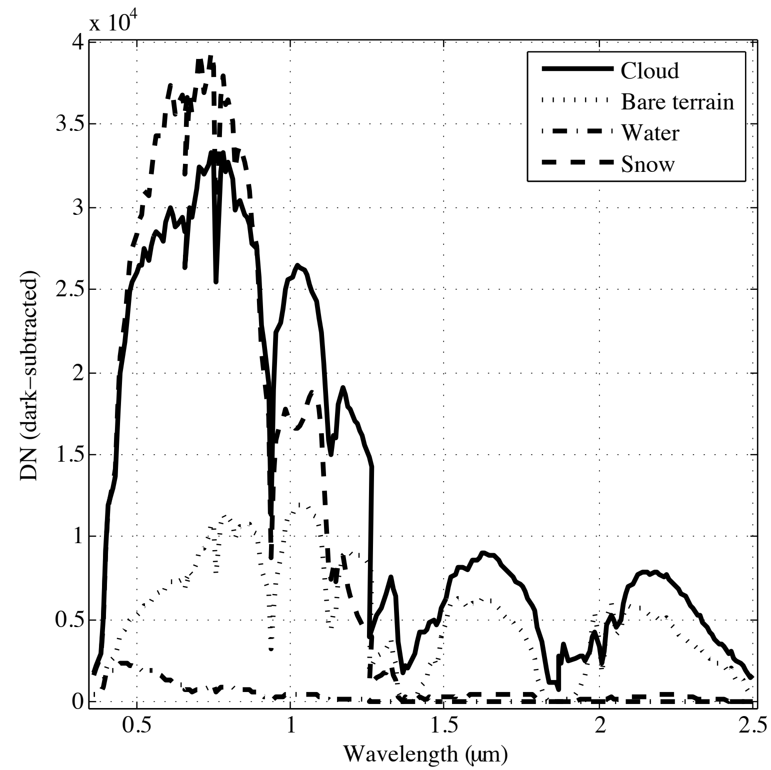


Brightness vs. SZA



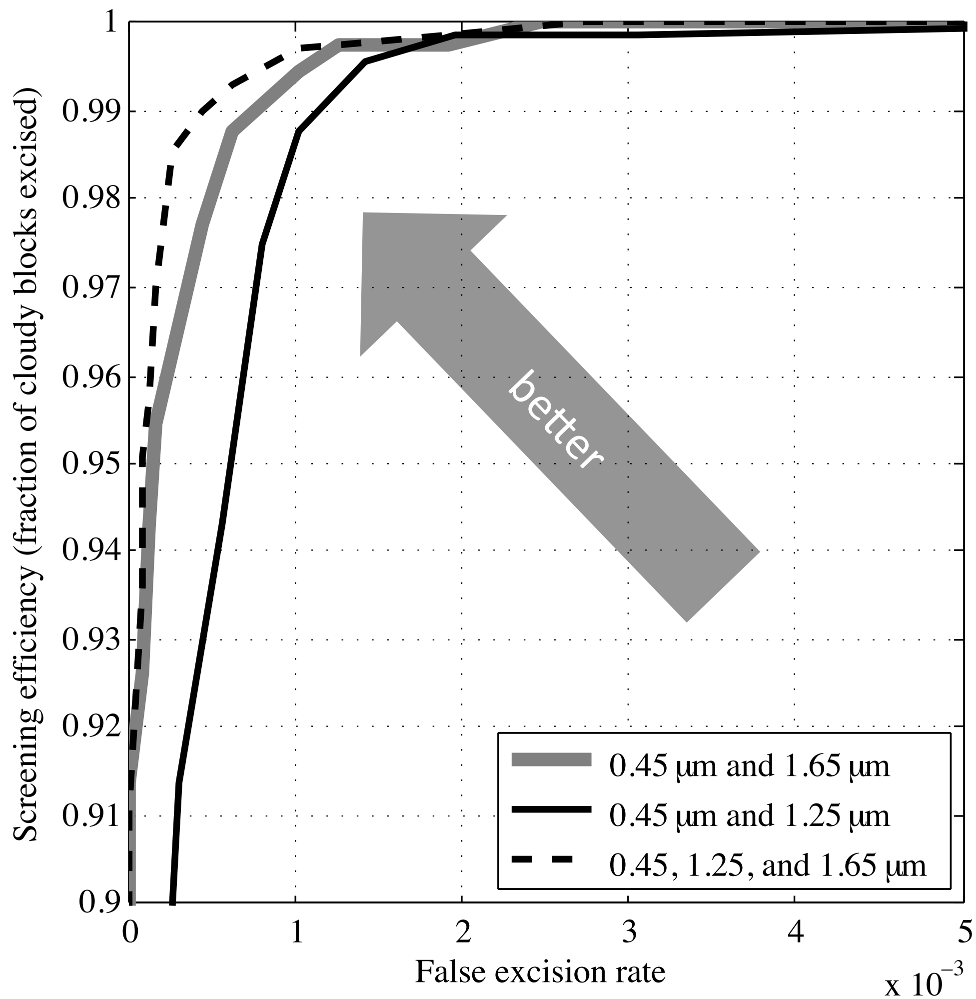
Which channels?

- Mutual Information (MI) scores channels' utility with respect to cloud/ non-cloud classification
- The combination of 0.45 µm and SWIR channels performs best

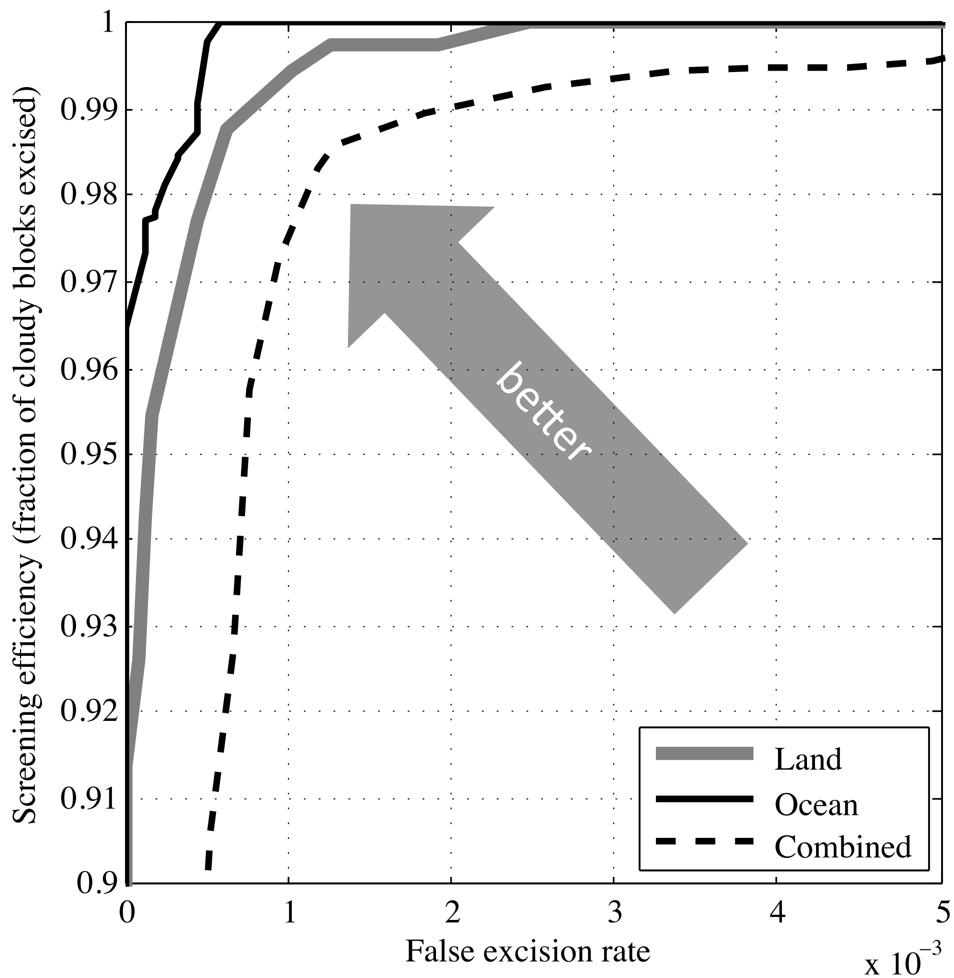


		MI when combined with				
Channel	MI	0.66 μm	0.86 μm	1.25 μm	1.38 μm	1.65 μm
0.45 μm	0.45	0.50	0.50	$\boxed{0.63}$	0.59	$\boxed{0.63}$
0.66 μm	0.44		0.51	0.62	0.59	0.62
0.86 μm	0.43			0.61	0.58	0.61
1.25 μm	0.54				0.60	0.58
1.38 μm	0.54					0.61
1.65 μm	0.40					

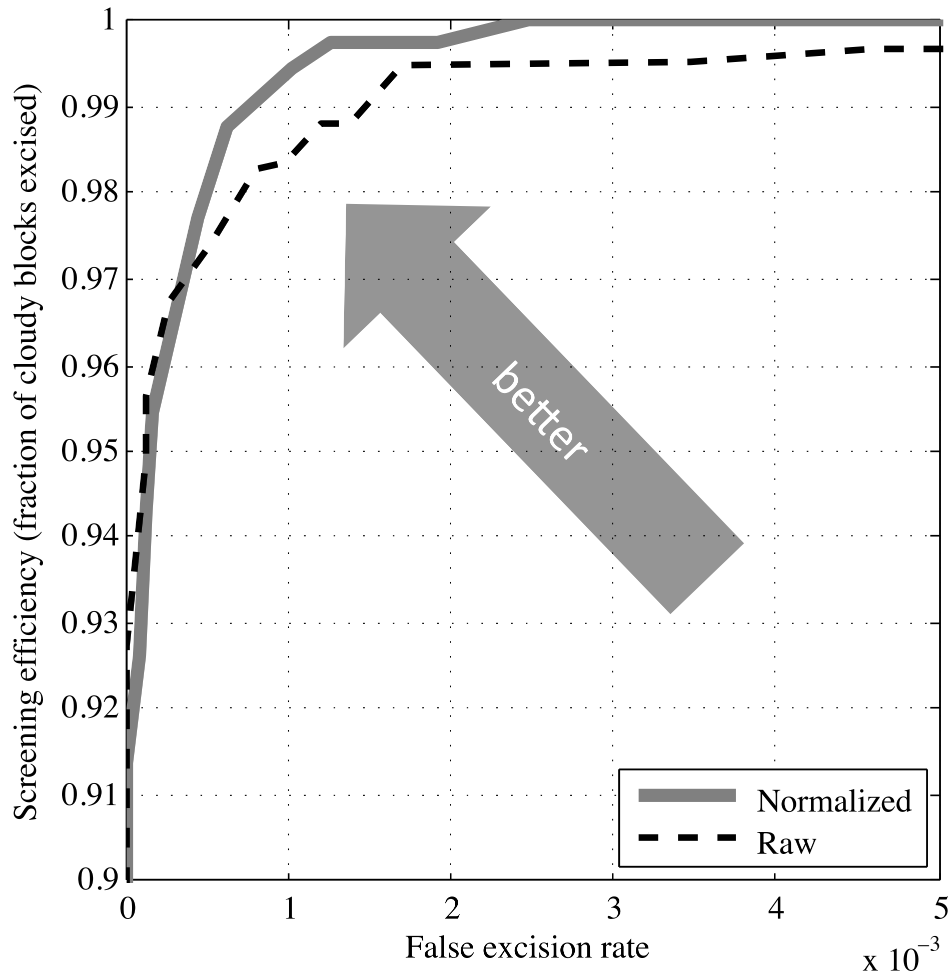
Performance: channel selection



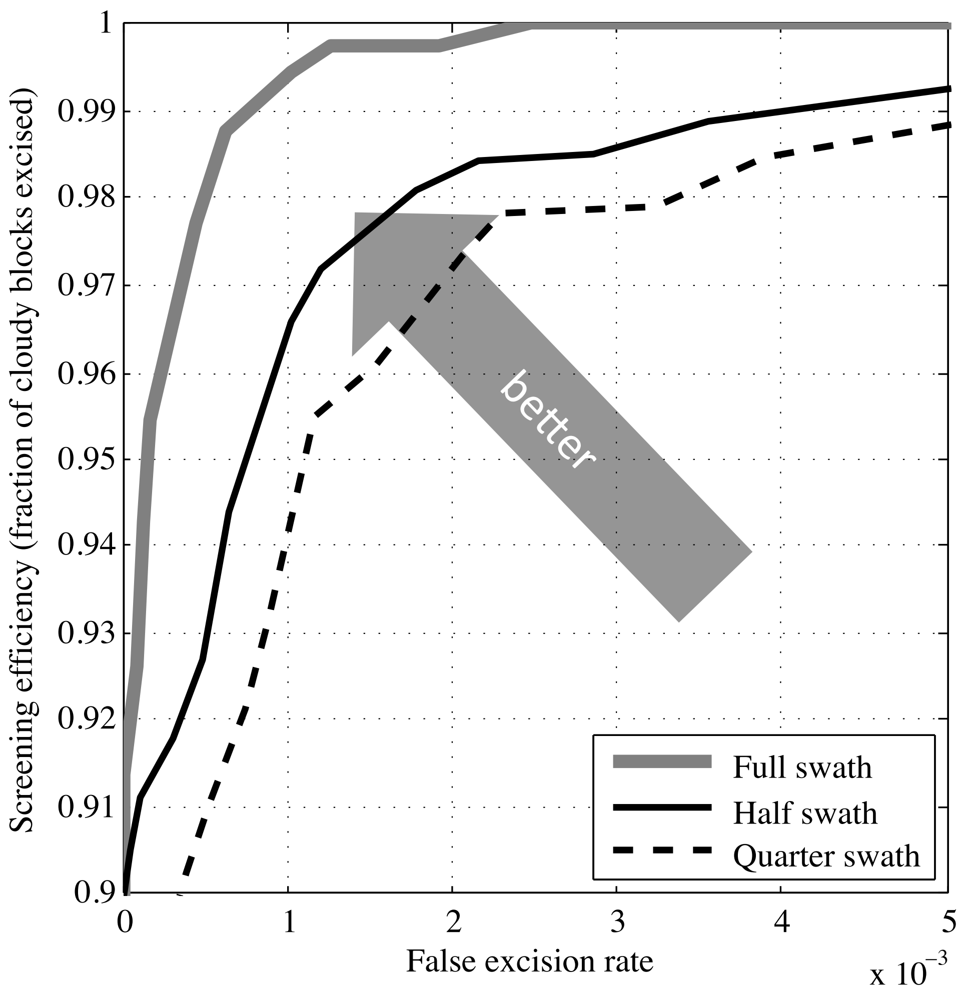
Performance: terrain types



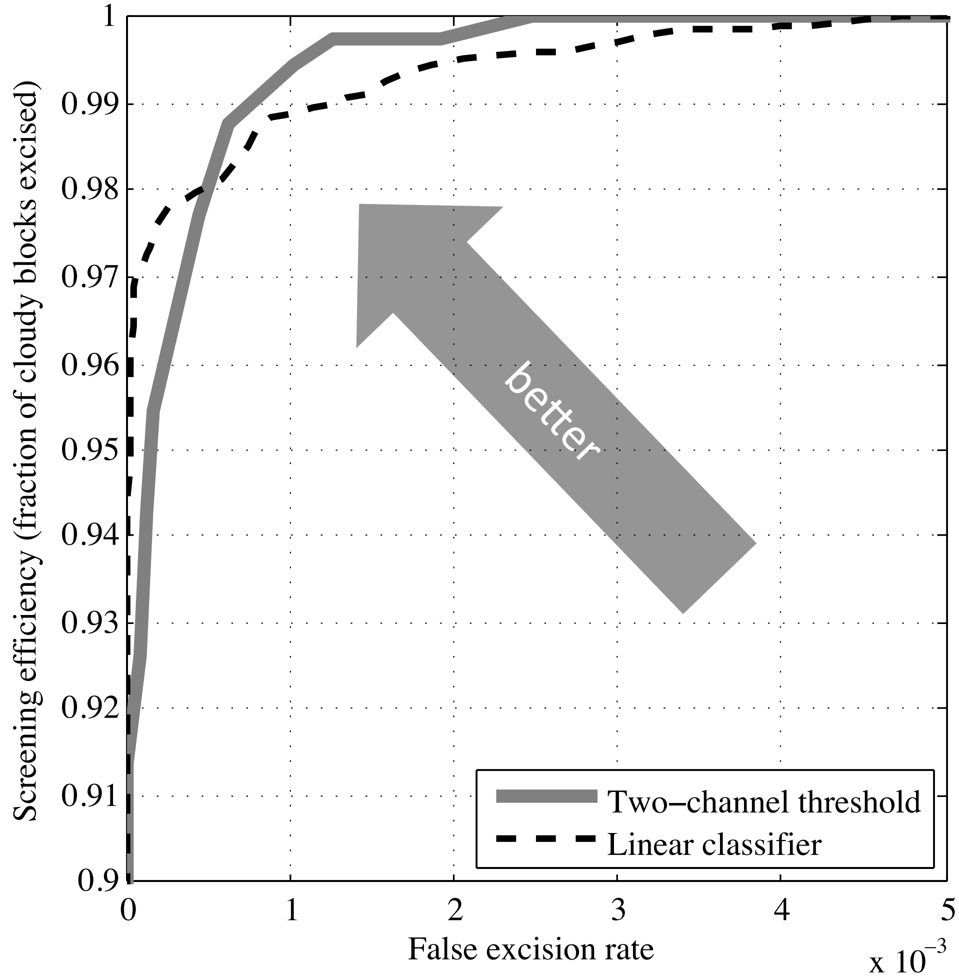
Performance: solar normalization



Performance: spatial aggregation



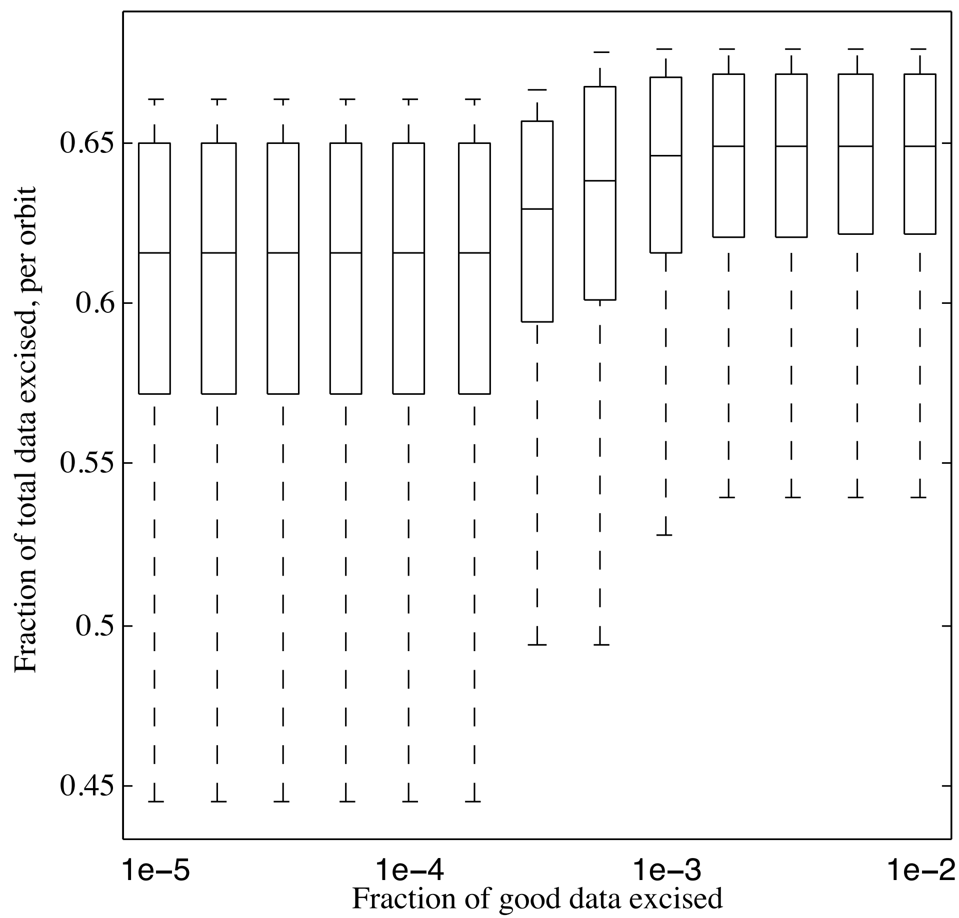
Performance vs. linear classifier



ISS Orbit simulation

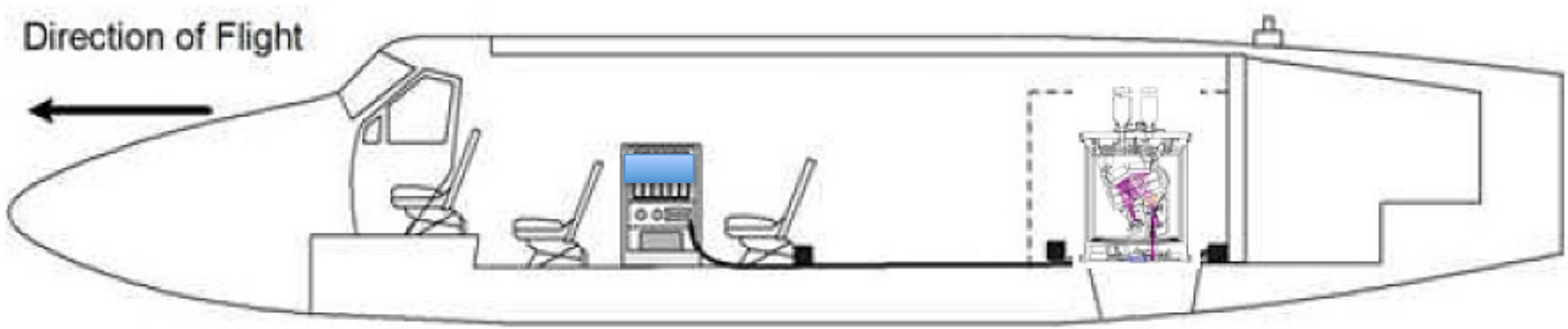
Simulated a continuous year of operations on the ISS, ignoring sun glint

50% data volume reductions are achievable



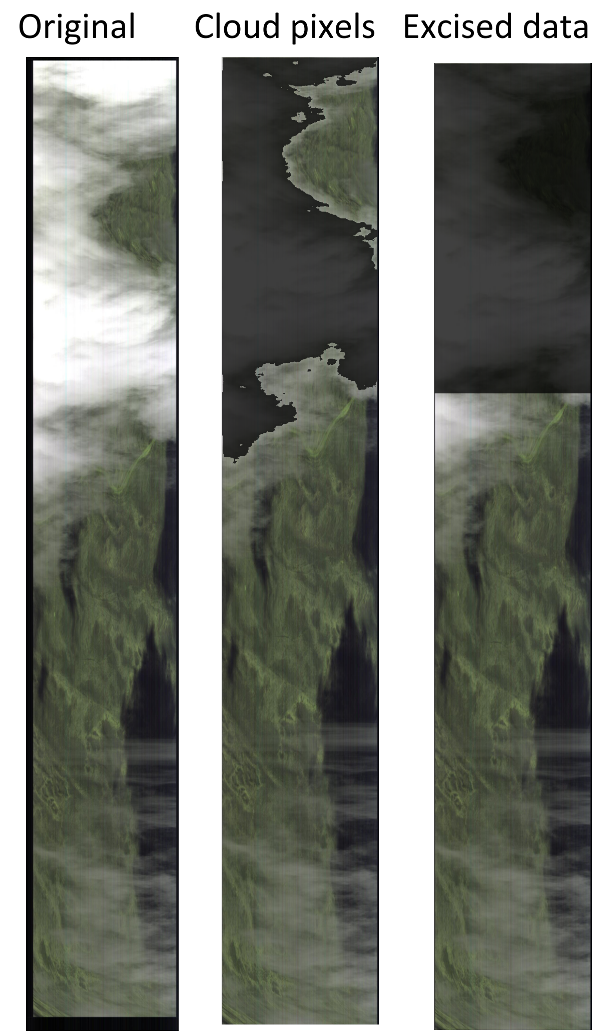
AVIRIS-NG demonstration

- A second computer in parallel with the main recording path
  - National instruments PXIe (4 CPU cores)
  - Matlab on Windows OS
- Computes thresholds from GPS data at the start of each flightline
- Performs cloud screening online in real time at 0.5 Gb/s



AVIRIS-NG demonstration results

- Evaluated performance during a campaign near Casper, WY
- Real time on-board operation with live monitoring
- Autonomous operation istransparent to Operator
- 142 flightlines total over 1TB
- Committed **no false positives during the one week campaign**
- Excised clouds from a handful of cloudy images (perfect performance)



Conclusions

- For typical Earth orbiting missions, realtime cloud screening could provide >50% data reduction with negligible false alarms
- Can turn the system off over snow for additional confidence
- Bayesian decision theory is a principled way to set channel thresholds
- Algorithm has very low computational complexity and integrates easily with real-time instrument data processing

Thanks!

- **Charles Sarture and the AVIRIS team**
- **Dr. Pantazis Mouralis, Jason Hyon** and the JPL Earth Science Directorate
- MODIS land cover data were obtained through the online Data Pool at the **NASA Land Processes Distributed Active Archive Center (LP DAAC)**, USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota
- This project made use of Vowpal Wabbit, a fast out-of-core learning system sponsored by **Microsoft Research, Yahoo! Research, and a large open source development community**