Mineral Dust Detect Algorithms Using Satellite Data

Janita Patwardhan[1], Peichang Shi[2], and Qianqian Song[3]

Advisors: Dr. Jianwu Wang[2] and Dr. Zhibo Zhang[3]

[1]Department of Mathematics & Statistics, UMBC
[2]Department of Information Systems, UMBC
[3]Department of Physics, UMBC

Acknowledgements: NSF, UMBC, HPCF

UMBC
AN HONORS UNIVERSITY IN MARYLAND

CyberTraining: Big Data + High-Performance Computing + Atmospheric Sciences
University of Maryland, Baltimore County, Spring 2018

Outline

- Introduction
- Data Preparation
- Our Algorithms
- Conclusion

Mineral Dust Aerosol

- Defined as soil particles in the air
- Adversely affects air quality and human health
- Changes temperature structure in the atmosphere

NASA Worldview, May 9 2007: https://go.nasa.gov/2IlNV7r

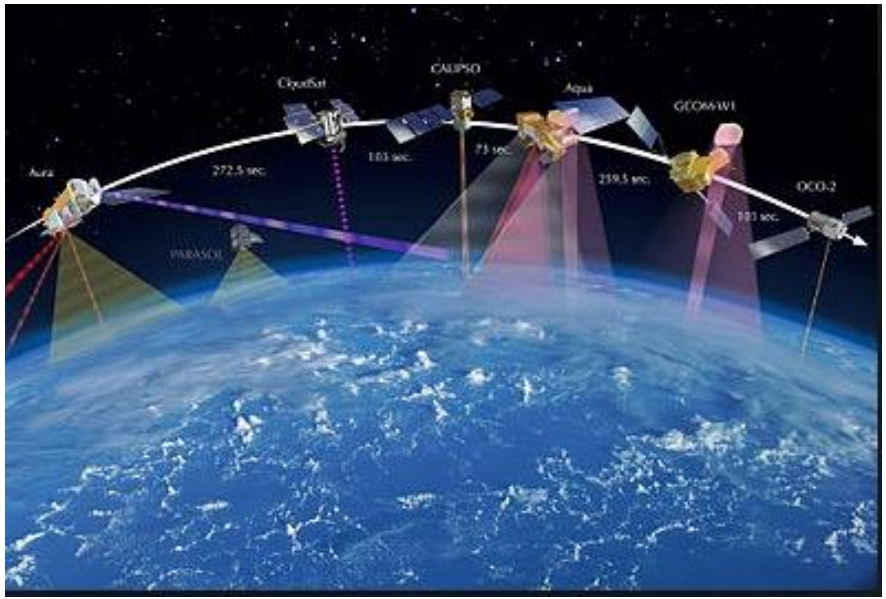
Satellite Data: MODIS and CALIPSO

MODIS:

- Passive sensor onboard Aqua and Terra satellite
- Images the entirety of the earth every 1-2 days
- 36 spectral bands

CALIPSO

- Satellite with active Lidar sensor
- Better able to detect dust by using depolarization



https://en.wikipedia.org/wiki/A-train_(satellite_constellation)

Project Objectives

1. Collocate the MODIS and CALIPSO data
2. Develop algorithms using MODIS data to detect dust over ocean
    a. Physically based algorithm
    b. Big data/machine learning algorithm
3. Design algorithms using MODIS data for detection over land
    a. Physically based algorithm
    b. Big data/machine learning algorithm
4. Determine accuracy rate of algorithms using CALIPSO data
5. Compare accuracy of our algorithms to published algorithms

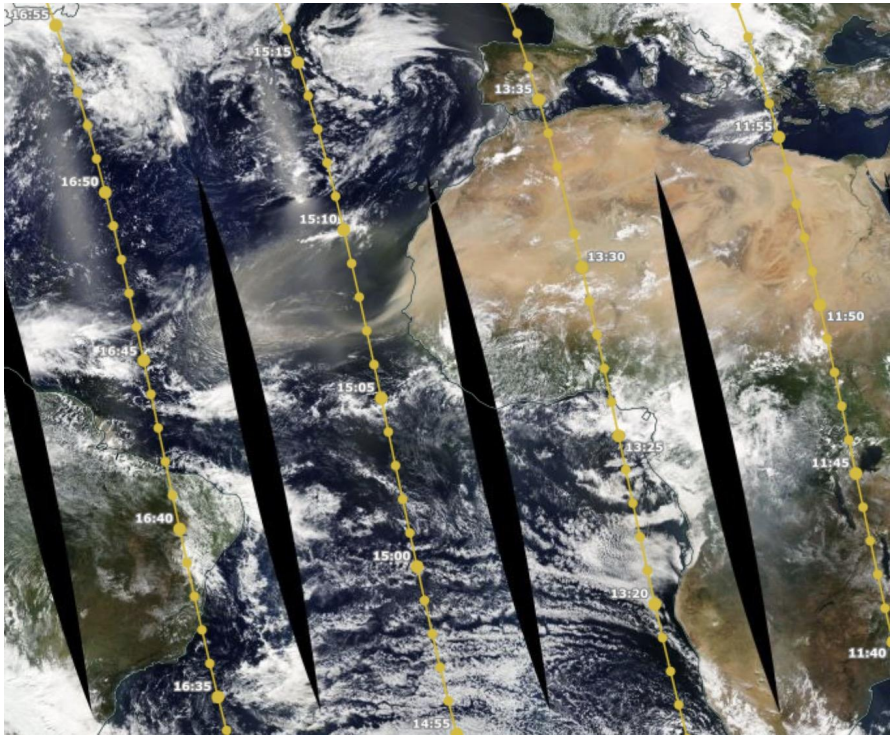
Project Objectives: Achieved

1. Collocate the MODIS and CALIPSO data
2. Develop algorithms using MODIS data to detect dust over ocean
   a. Physically based algorithm
   b. Big data/machine learning algorithm
3. Design algorithms using MODIS data for detection over land
   a. Physically based algorithm
   b. Big data/machine learning algorithm
4. Determine accuracy rate of algorithms using CALIPSO data
5. Compare accuracy of our algorithms to published algorithms

Outline

- Introduction
- Data Preparation
- Our Algorithms
- Conclusion

Collocating Data



NASA Worldview: https://go.nasa.gov/2KwnxV4

- Already have collocated data for CALIPSO and MODIS level-2 data
- CALIPSO data tells us dust for every 5 km
- MODIS data gives information for every 1 km
  - Took average over 5 pixels to get 5km resolution

Variables Utilized

- Existence of dust

Dust and clouds
Dust, no clouds

Dust

No dust or clouds
Clouds, no dust

No Dust

- Radiance/Emissivity from 36 spectral bands
- Geometry

Regions and Times Used

Northwest Coast of Africa (Atlantic Ocean)

- March 13, 2007
- May 9, 2007
- *July 15, 2007*
- March 31, 2008
- *June 22, 2009*
- April 22, 2010

Coast of Arabian Peninsula (Arabian Sea)

- November 5, 2009
- November 11, 2009

Outline

- Introduction
- Data Preparation
- Our Algorithms
- Conclusion

Simple Physical Algorithm

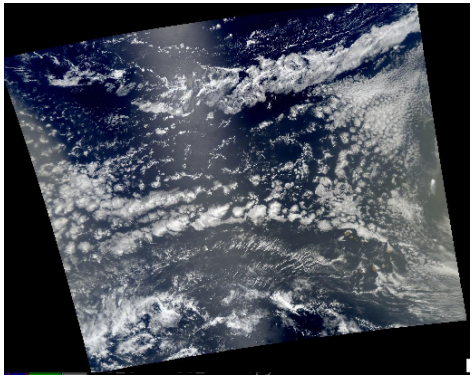
- Ackerman[1] showed that BTD(11-12um) of dust is smaller than that of clouds

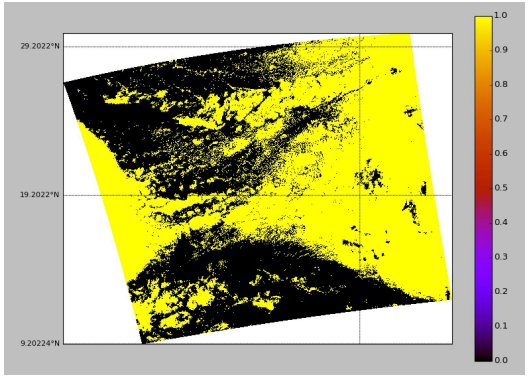
- Our physical algorithm:
  - BTD < threshold : dust
  - BTD > threshold : dust-free

- Threshold selection:
  - Applied different threshold for MODIS data along CALIPSO track
  - Calculated accuracy for different threshold with CALIPSO dust detection as reference
  - Found threshold ~ 0.8 gives highest accuracy around 60%~70%

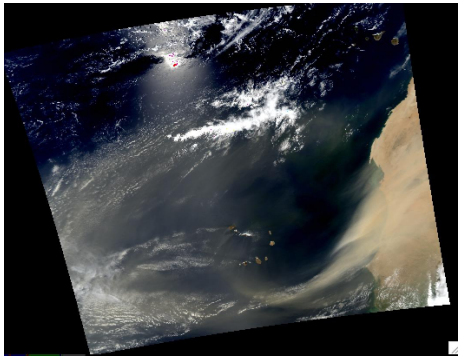
[1]Ackerman, Steven A. "Remote sensing aerosols using satellite infrared observations." *Journal of Geophysical Research: Atmospheres* 102.D14 (1997): 17069-17079.

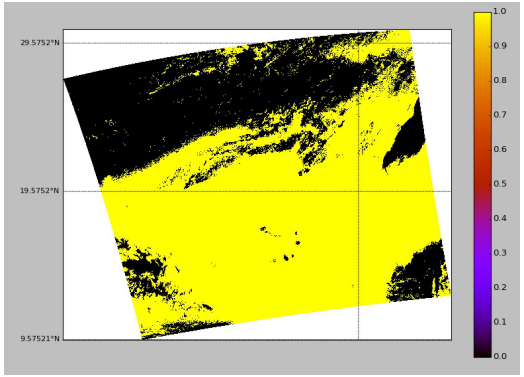
Results of Physical Algorithm





Use threshold = 0.8
July 15, 2007





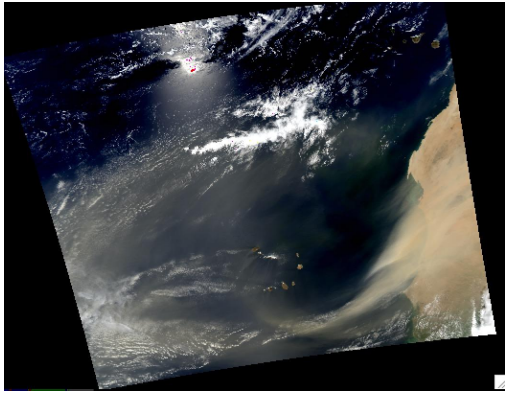
Use threshold = 0.8
June 22, 2009

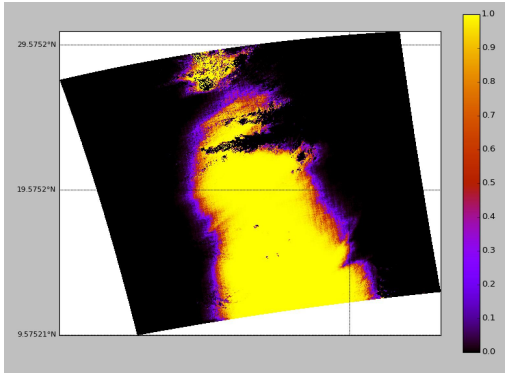
Machine Learning Method: Logistic Regression

- Summary of method
  - More intuitive and robust than other regressions
  - Not many assumptions needed
  - Goal is to select best set of predictor variables and create a model for outcome variable

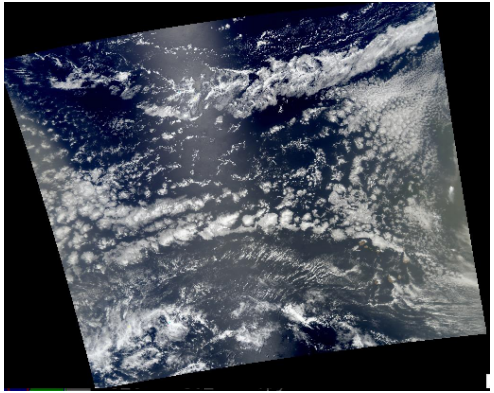
- Types of simulations run
  - Using all variables
  - Using 6 variables
    - R0.47
    - R0.47/R0.64
    - BTD(3.9um-11um)
    - BTD(11-12um)
    - R0.42um, R0.68
  - Using 6 variables with geometry

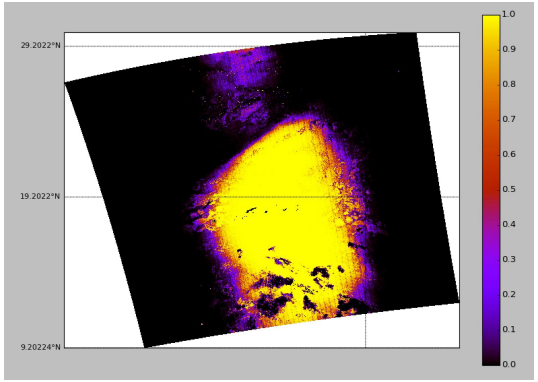
Logistic Regression Results: All Variables





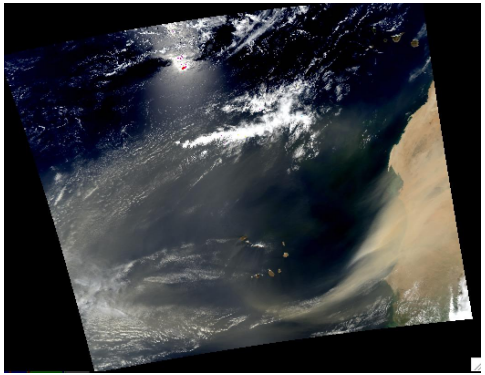
June 22, 2009

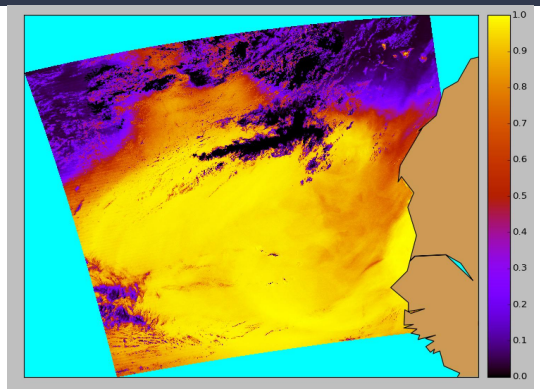


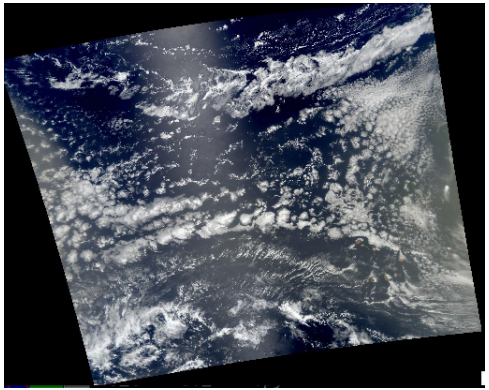


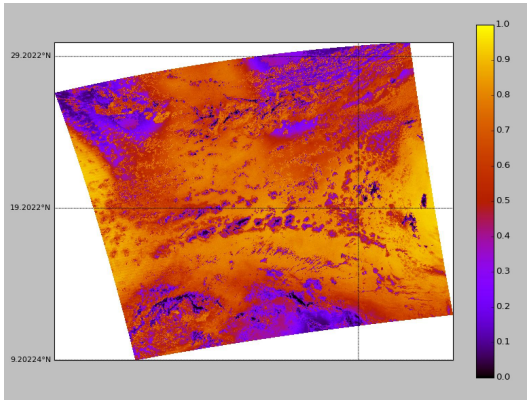
July 15, 2007

Logistic Regression Results: 6 Variables



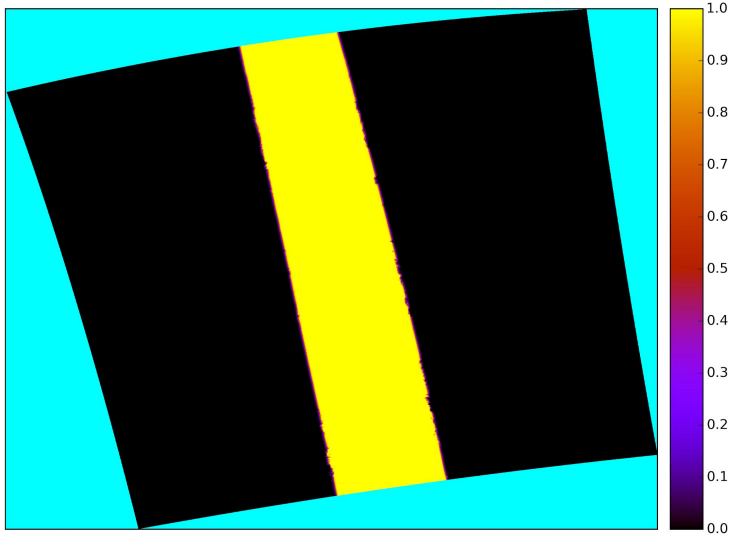






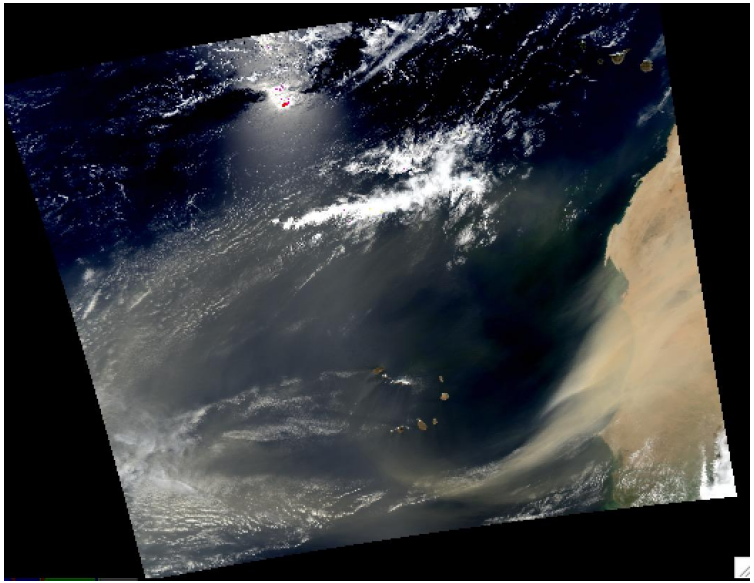
- R0.47
- R0.47/R0.64
- BTD(3.9um-11um)
- BTD(11-12um)
- R0.42um, R0.68

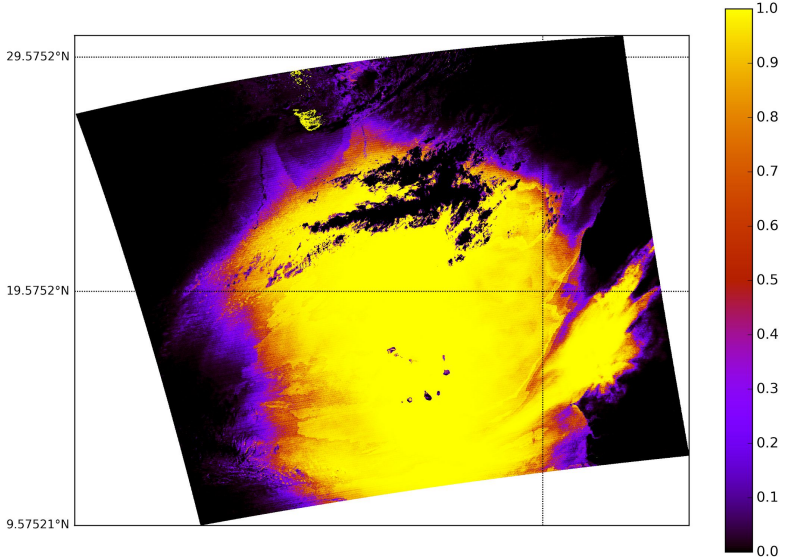
Logistic Regression Results: 6 Variables and Geometry



June 22, 2009

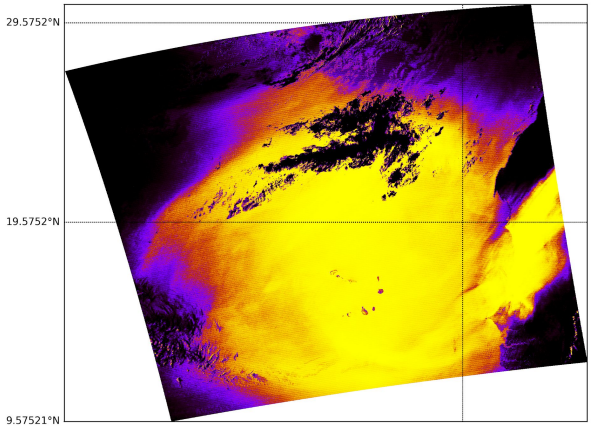
Logistic Regression Results: Multiple Data Sets with All Band variables



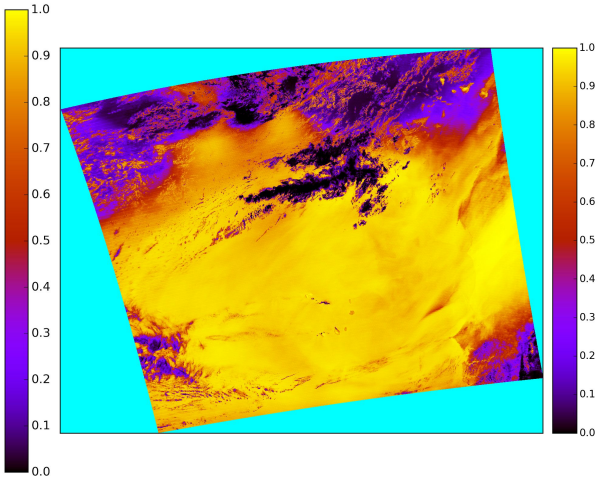


June 22, 2009

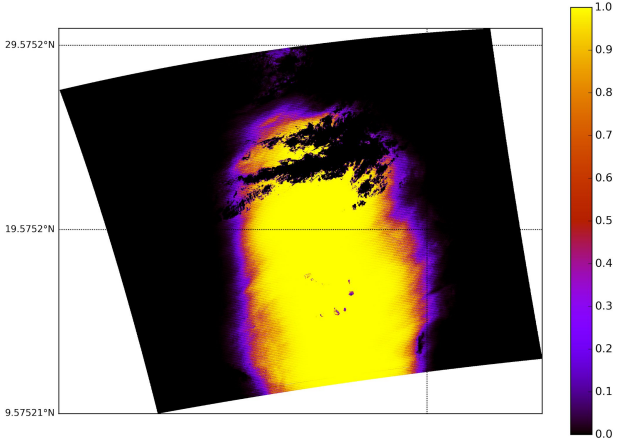
Multiple data vs Single data



Multiple dataset with parsimonious variables



Multiple data with 6 variables



Sing dataset with parsimonious variables

Outline

- Introduction
- Data Preparation
- Our Algorithms
- Conclusion

Accuracy Comparison: July 15, 2007

	Accuracy Rate	False Positive	False Negative
Physical Algorithm	50.8%	44.6%	4.6%
LR- All variables	90.5%	2.1%	7.4%
LR- 6 variables	78.3%	6.6%	15.1%
LR- 6 variables and geometry	81.7%	7.5%	10.8%
LR- Multiple data sets, all variables and geometry	92.4%	2%	5.6%
LR-Multiple datasets, parsimonious model	89.5%	2%	8.4%
LR single dataset , parsimonious model	93.2%	1.7%	5.1%

Future Work

- Try other machine learning techniques for the machine learning algorithm
- Create an algorithm for detecting dust over land
- Combine land and ocean algorithms
- Decrease false positive and negative rates
- Compare accuracy rates to those of published algorithms

Thank You!