Testing Land Coverage Classification Algorithms for Optimizing Flood Detection in Hyperspectral Image Data

Benjamin Huynh, Maria T. Patterson (Advisor)

Center for Data Intensive Science, University of Chicago

Summary

- We want to improve on existing algorithms for detecting water from satellite imagery in order to effectively monitor floods.
- Current methods are limited to using few spectral bands due to onboard computational constraints.
- Our methods utilizing all available spectral bands significantly outperform algorithms currently onboard the EO-1 satellite.
- We thus leverage existing cloud-based infrastructure to quickly process and classify large hyperspectral images with high accuracy.

Background

- NASA's Hyperion instrument on its Earth Observing-1 (EO-1) Satellite provides hyperspectral imagery, covering 242 spectral bands (from .4 to .25 μm).
- Standard scenes are 37 km $\times$ 42 km, amounting to 1.5-2.5 GB of data per scene.
- High spectral resolution creates a high-dimensional feature space, potentially allowing for high predictive performance.
- Reflectance values from hyperspectral images can thus be used to detect water.

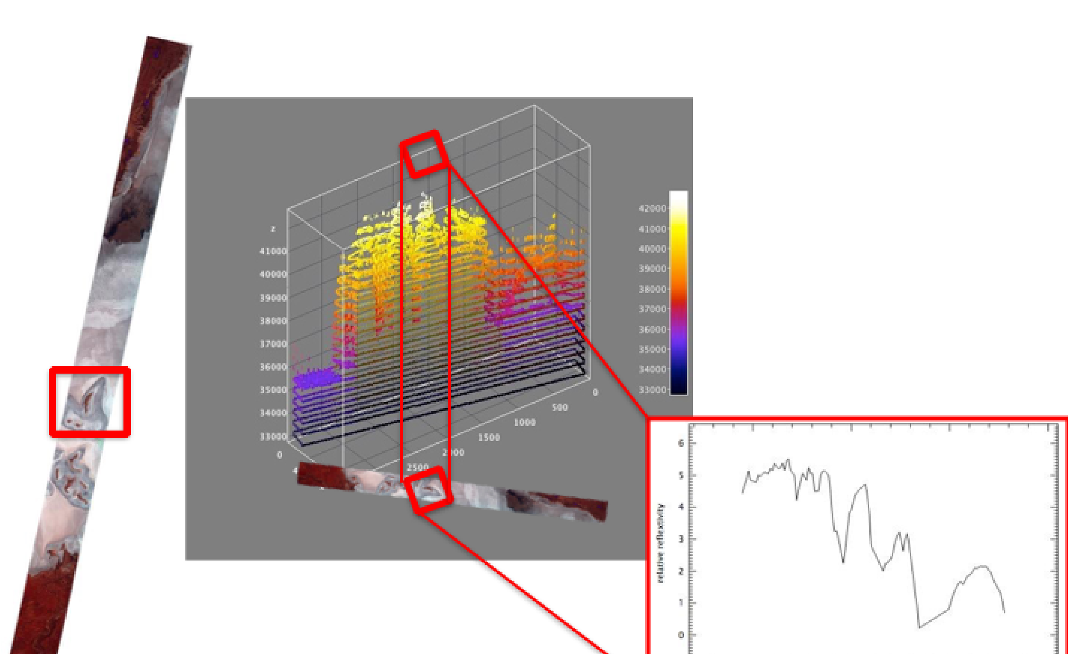

Figure 1: Left: An RGB interpretation of a Hyperion scene. Right: A representation of relative reflexivity vs wavelength from the Hyperion scene.

Problem: Data Access and Limited Computational Ability

- Current onboard sensors are limited to only using 3 of the 242 spectral bands for water detection.
- Hyperion scenes are difficult to access in large volumes for scientific processing.
- How do we quickly classify hyperspectral images using all 242 spectral bands?

Data Acquisition

- Project Matsu, a cloud-based collaboration with NASA, provides a framework for fast access to hyperspectral EO-1 images.
- We leverage Project Matsu's framework to access and construct a training set of hyperspectral data, based on diverse types of surface materials (clouds, dry land, vegetation, water).

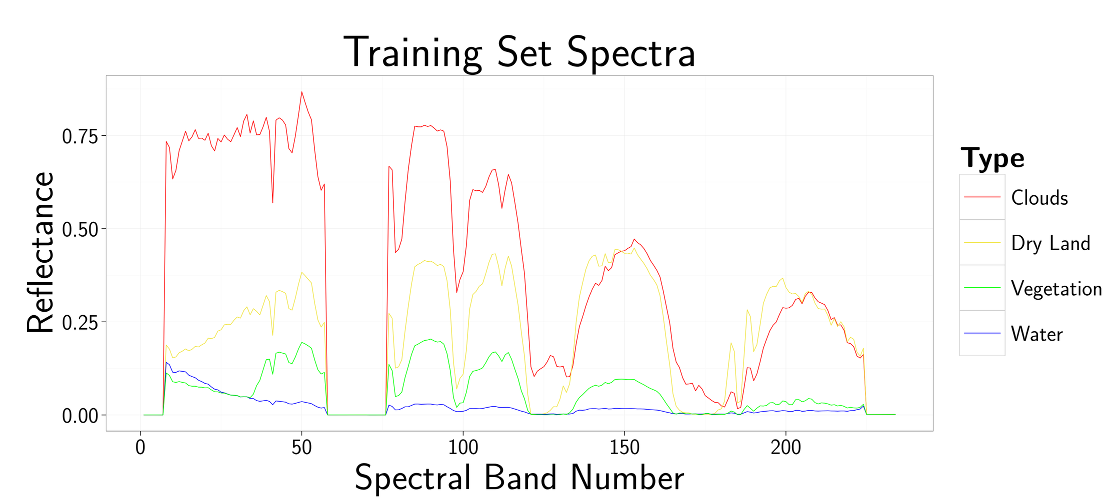

Figure 2: Spectra of the training set, separated by type of material.

Image Pre-Processing

- Input data: raw radiance values per band per pixel.
- The radiance for each band is divided by its respective solar irradiance, then geometrically corrected for solar elevation/distance at observation time.
- Output data: at-sensor reflectance values per band per pixel.

Results: Model Comparison

- 60/20/20 training/test/validation set used to train and compare models.
- Compare models trained on all bands vs. models used onboard EO-1, resulting in 5% - 20% accuracy increase.

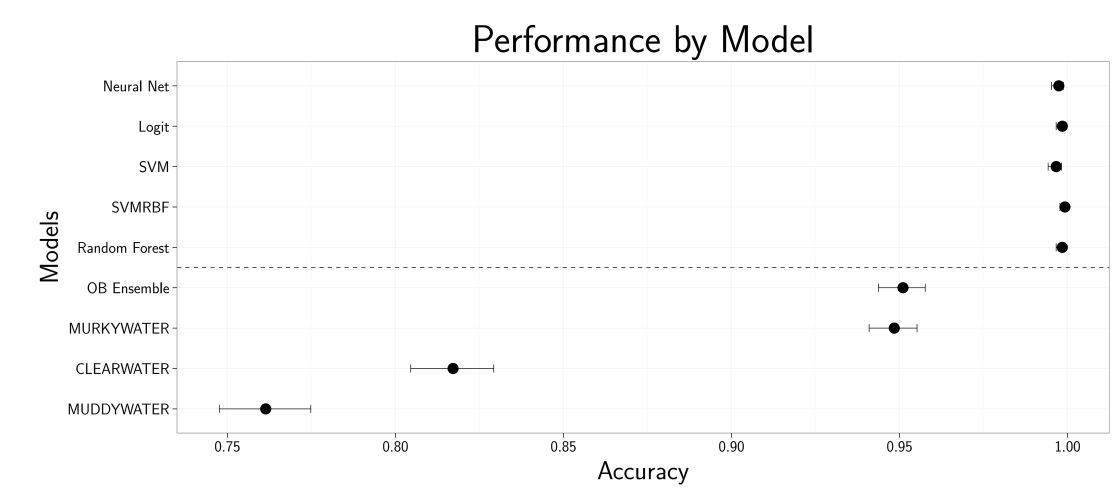

Figure 3: Comparison of model accuracy in terms of water detection. Models above the dotted line are modern machine learning algorithms trained on all 242 spectral bands. Models below the dotted line are reproduced versions of the onboard algorithms, trained only on two to three bands.

Results: Spectral Band Performance

- Feature ranking/selection can be used to determine optimal band selection.
- Many bands unused by onboard algorithms have high predictive performance.
- Bands used by onboard algorithms are suboptimal for flood detection; better band combinations can be chosen.

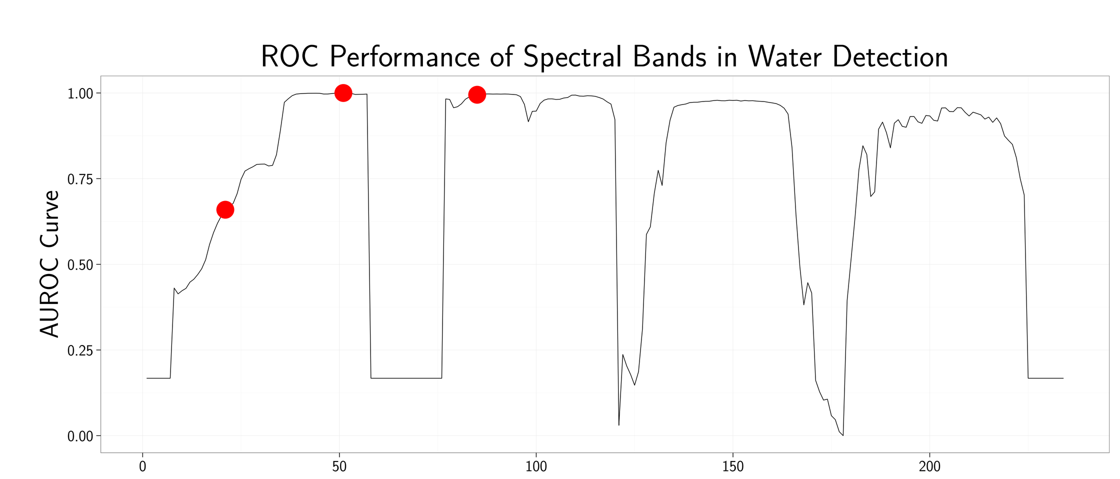

Figure 4: The area under ROC curve for each individual spectral band, in terms of ability to detect water pixels. Bands used by the onboard classifiers are shown as red dots.

Conclusion

- Because of the efficient framework provided by Project Matsu, we get quick overnight characterization of possible flood scenes with greater accuracy (5% - 20% increase) than existing methods.
- Feature selection methods can be used to select optimal spectral band combinations for water detection without additional computational expense.
- Classification of hyperspectral images can potentially be used to accurately detect various types and mixtures of surface materials.

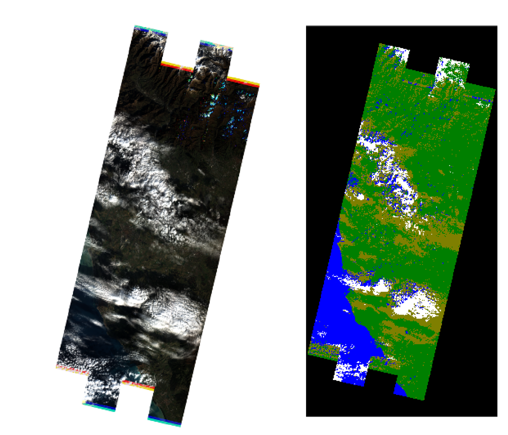

Figure 5: Left: An RGB reproduction of a satellite scene. Right: A land-coverage classified version of the same scene.

Acknowledgments

This work made use of the Open Science Data Cloud (OSDC) which is an Open Cloud Consortium (OCC)-sponsored project. The OSDC is supported in part by grants from Gordon and Betty Moore Foundation and the National Science Foundation and major contributions from OCC members like the University of Chicago. This work was supported by the National Science Foundation Partnerships for Research and Education (PIRE) Award Number 1129076. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

THE UNIVERSITY OF CHICAGO

OSDC OPEN SCIENCE DATA CLOUD

Additional Info

To see the datasets used and a reproducible code tutorial, follow the QR code below, or visit benhuynh.github.io/waterdetection.html. The author can be reached at benhuynh@uchicago.edu.