Machine learning approach to retrieving physical variables from remotely sensed data

Fazlul Shahriar

July 16, 2015

Introduction

- There is a growing wealth of remote sensing data from hundreds of space-based sensors.
- However, not all data of interest may be available.
  - Not in the sensor specification
  - Hardware failure
- Statistical analysis and machine learning techniques can be used to
  - Produce virtual integrated sensors
  - Solve problems important to scientists (e.g. for ocean study)

Remote Sensing Instruments

- Onboard Polar Orbiting Satellites:
  - Moderate Resolution Imaging Spectroradiometer (MODIS)
  - Visible Infrared Imaging Radiometer Suite (VIIRS)
  - Hyperion
- Onboard Geostationary Satellites:
  - GOES-R Advanced Baseline Imager (ABI)
  - Advanced Himawari Imager (AHI)

Overview

- Virtual Sensing
  - Restoration of the Aqua MODIS 1.6 µm Band
  - Estimating a 13.3 µm VIIRS Band
  - Estimating True Color Imagery for GOES-R
- Machine Learning Approach to Clear-sky Classification

Restoration of the Aqua MODIS 1.6 µm Band

- Fifteen out of 20 detectors in this band are broken.
- NASA publishes band 6 with the missing scanlines filled in using columnwise linear interpolation.
- Wang et al. (2006): Missing data can be estimated by fitting, at good band 6 detectors, a cubic polynomial expressing the band 6 pixel values as a function of band 7.
  - Better than basic interpolation.
  - Assumes the existence of a global relationship between bands 6 and 7.

Restoration of the Aqua MODIS 1.6 µm Band

- Rakwatin et al. (2007): a local cubic regression approach.
  - Applied histogram matching destriping to the image as a pre-processing step to reduce detector-to-detector striping artifact.
- Shen et al. (2011): a within-class local fitting algorithm
  - An unsupervised classification is used to separate various scene types.
  - Band 7 is used to perform the prediction within each class.

Restoration of the Aqua MODIS 1.6 µm Band

- QIR (2012): restore band 6 using bands 1-5 and band 7, and a local window as input.
- Band 5 contains information about band 6 not captured by band 7:

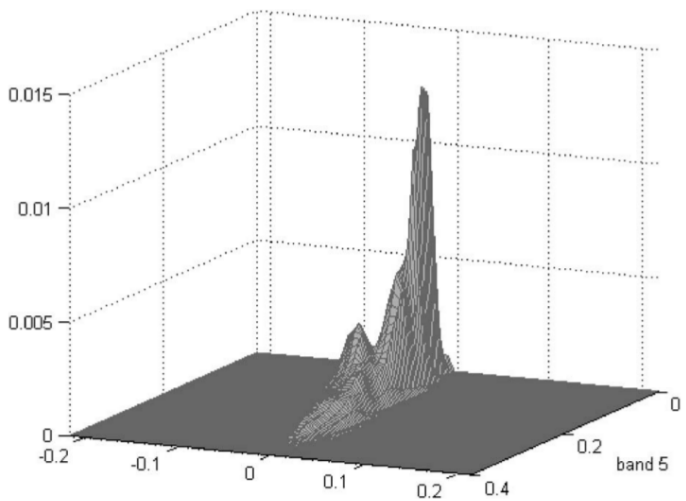


Fig. 1. Joint PDF of band 5 radiance and the residual of band 6 radiance with the portion predicted by band 7 subtracted.

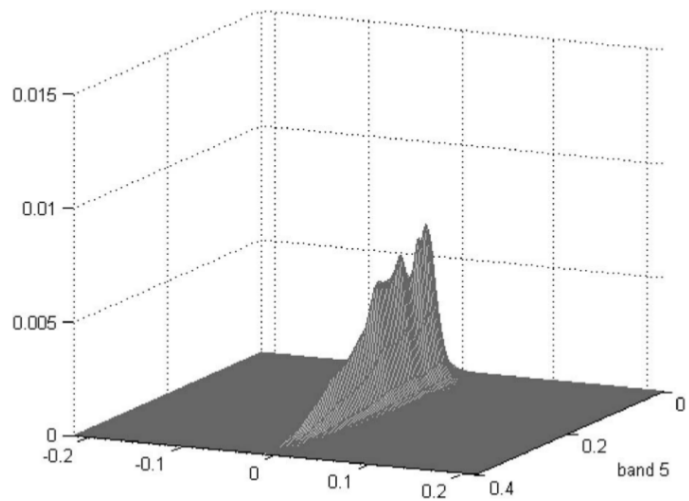


Fig. 2. Product of band 5 probability and residual of band 6 radiance with the portion predicted by band 7 subtracted.

Restoration of the Aqua MODIS 1.6 µm Band

- Restoration function F_T for a tile is found by minimizing the training error

$$\text{Error}(F_T) = \sum_{p \in V_T} |F_T(\mathbf{x}(p)) - z(p)|^2 \tag{1}$$

where $z(p)$ is the true band 6 value for pixel p, $\mathbf{x}(p)$ are values in a spatial window centered at pixel p for bands 1-5 and 7, and V_T is the set of pixels in working detectors in band 6.

Restoration of the Aqua MODIS 1.6 µm Band

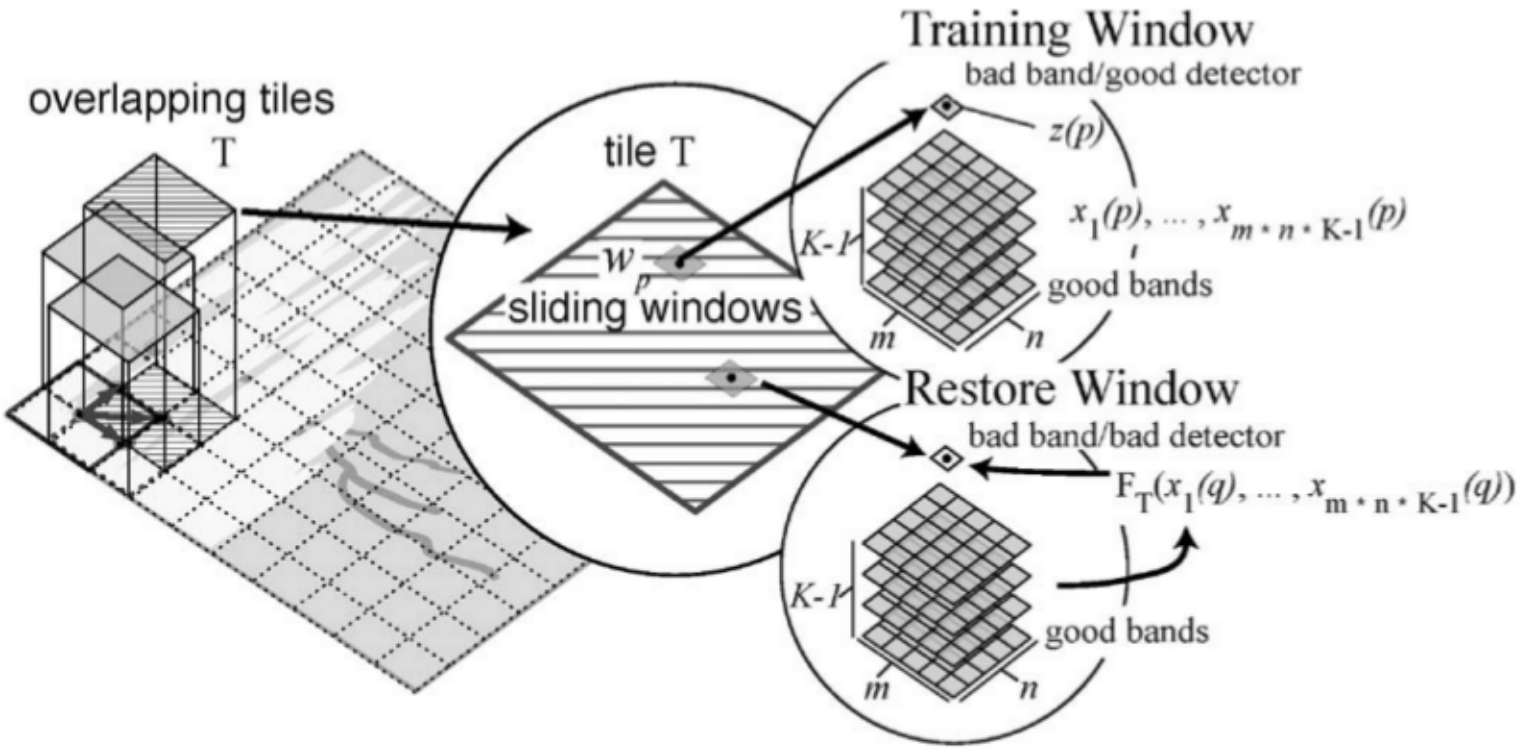


Fig. 4. Diagram showing the process of determining the restoration function for a tile.

Restoration of the Aqua MODIS 1.6 µm Band

- Li et al. (2014): Robust M-Estimator Multiregression (RMEMR)
- Regression is done locally in tiles of size 20×20 which are slided by a fixed step size (usually ten).
- For each tile p, it minimizes the error

$$\sum_{i,j} w_{i,j}(R_{i,j,6} - \widehat{R}_{i,j,6})^2, \qquad (i,j) \in \Omega_p \tag{2}$$

  where Ω_p consists of all the values originating from non-broken detectors, $\widehat{R}_{i,j,6}$ is the estimated value at band 6 at a pixel, and $R_{i,j,6}$ is the true value at the same pixel.
- Weights $w_{i,j}$ are calculated by a Huber M-estimator based on the residue.

Restoration of the Aqua MODIS 1.6 µm Band

- Local cubic regression in does consistently and significantly better the global cubic regression.
- The QIR algorithm, which uses more bands but linear regression instead cubic regression, performs better than band 7 based methods.
- RMEMR algorithm produces better restoration than QIR, and it's also has faster running time than QIR.

Estimating a 13.3 µm VIIRS Band

- VIIRS does not have a band at 13.3 µm.
- The state of the art cloud-top pressure (CTP) algorithm developed for the GOES-R ABI instrument requires 13.3 µm values.
- Cross et al. 2013: A 13.3 µm band can be estimated from existing VIIRS bands and the Cross-track Infrared Sounder (CrIS).
- MODIS and Atmospheric Infrared Sounder (AIRS) is used for evalutation.

Estimating a 13.3 µm VIIRS Band

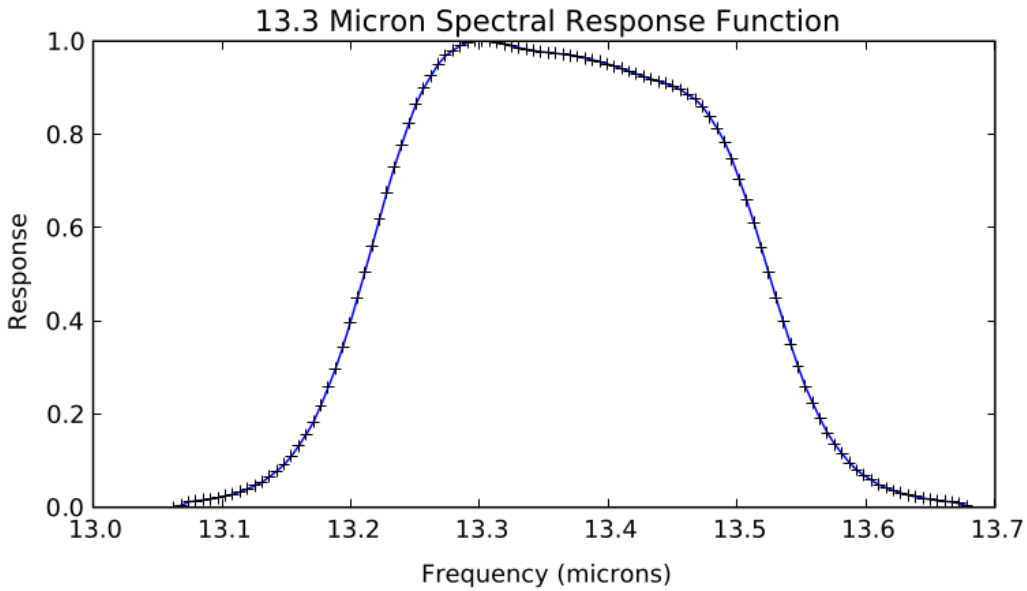


Fig. 3 The "+" markers indicate the spectral response values at 109 central wavelengths where AIRS measurements are available.

Estimating a 13.3 µm VIIRS Band

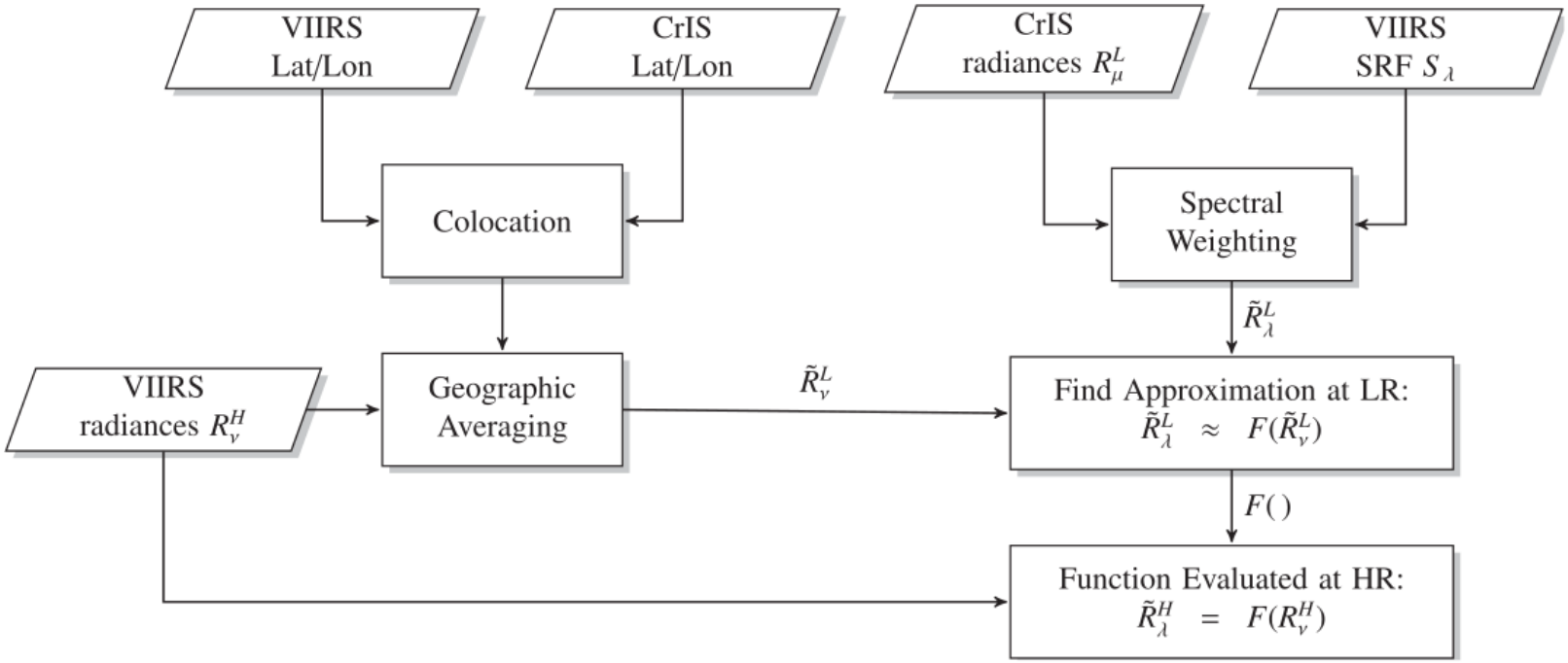


Fig. 2 Block diagram of statistical estimation algorithm.

- F is implemented using the k-d tree data search algorithm.
- F was extended to take latitude and longitude as input arguments in addition to source radiance values.

Estimating True Color Imagery for GOES-R

- GOES-R will not have green band.
- Since GOES-R ABI data is not yet available, training data typically consists of simulated ABI constructed from similar MODIS bands as proxy.
  - MODIS also contains a green band.
- Look-up table (LUT) based methods: Hillger et al. (2011), Miller et al. (2012).
  - Both LUT use values from 470 nm (blue), 640 nm (red), and 860 nm (near infra-red (NIR)) bands as inputs and predicts 550 nm green band value.

Estimating True Color Imagery for GOES-R

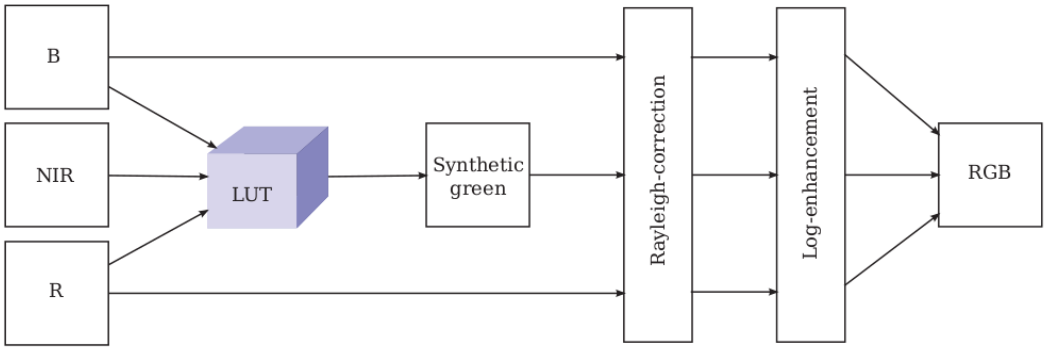


Figure 1. Block diagram for generating true-color/RGB images, with application of Green-LUT, followed by Rayleigh-correction, followed by log-enhancement.

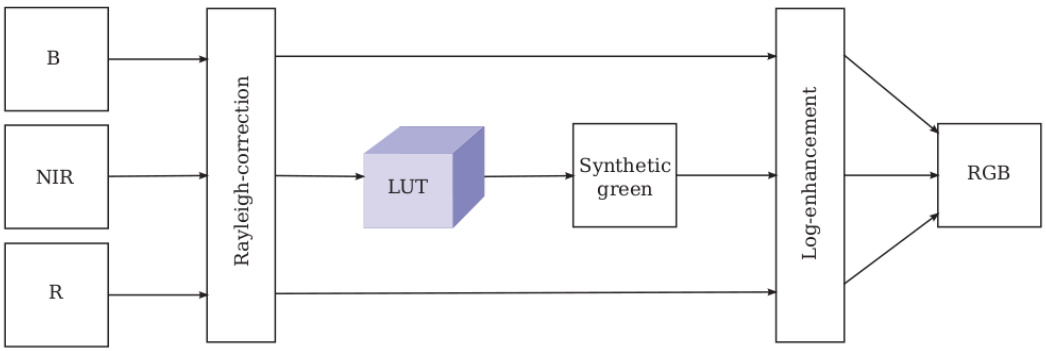


Figure 2. Block diagram for generating true-color/RGB images using Rayleigh-corrected reflectances prior consulting the table

Estimating True Color Imagery for GOES-R

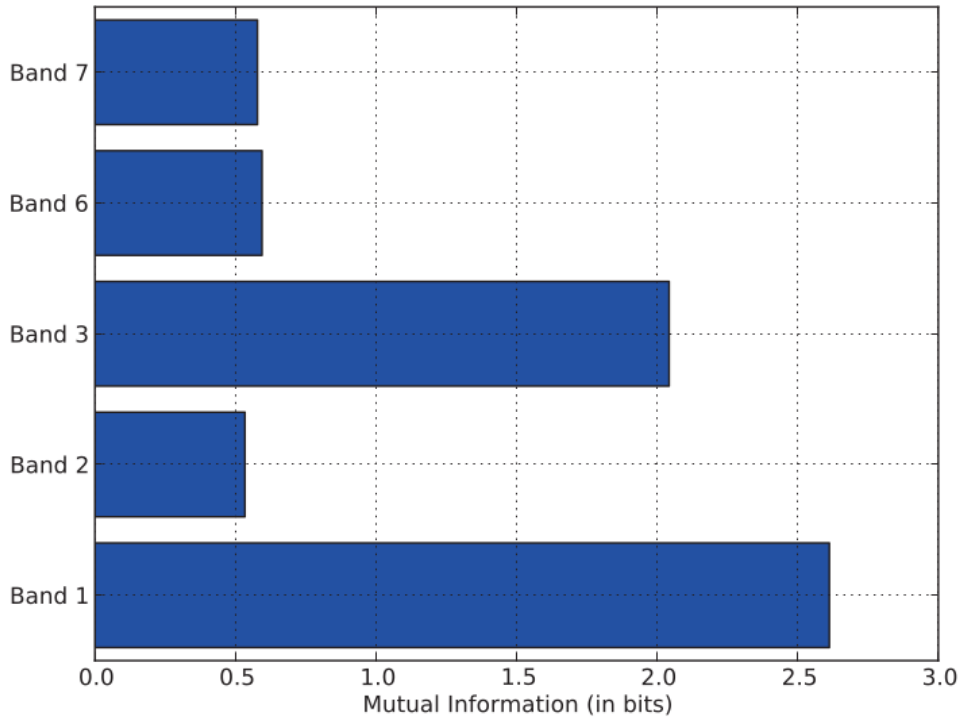


Figure 4. Mutual information of bands 1, 2, 3, 6 and 7 with green band 4 in bits.

- Bands: (1) Red, (2) NIR, (3) Blue, (6) NIR, (7) NIR.

Estimating True Color Imagery for GOES-R

- Gladkova et al. (2011):
  - The approximation function for green will depend on five spectral parameters as opposed to the three.
  - The piecewise constant green approximation is replaced by a piecewise linear approximation over the Voronoi cells associated with the sampling points.
    - Curse of dimensionality
  - A variably spaced sampling points is employed, as opposed to the lattice-spaced sampling points.
    - K-means clustering

Estimating True Color Imagery for GOES-R

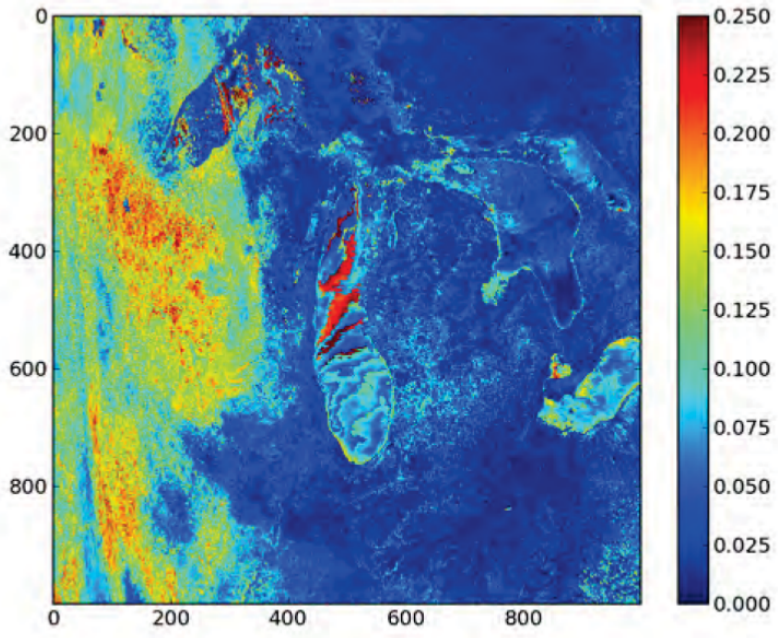


Figure 10. Relative error using LUT as in^2 .

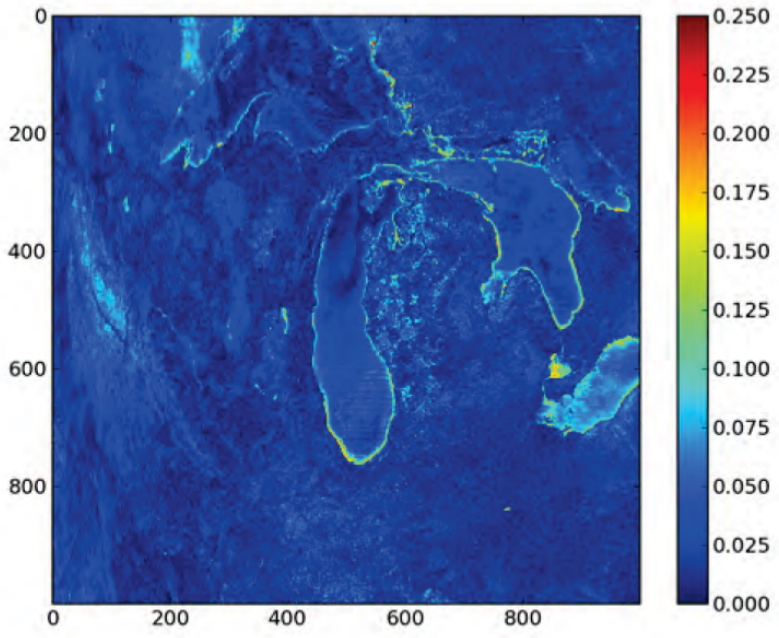
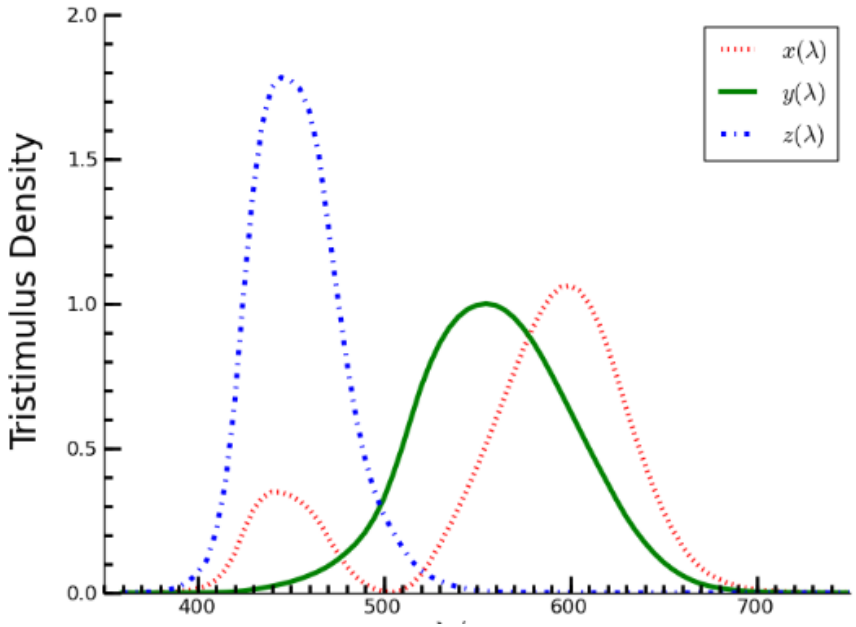


Figure 11. Relative error using multi-linear piecewise approximation.

Estimating True Color Imagery for GOES-R

- Grossberg et al. (2011): estimate true color imagery using CIE 1931 XYZ color space.
- The tristimulus values X, Y, Z:



$$X = \int_0^\infty I(\lambda)x(\lambda)d\lambda,$$
$$Y = \int_0^\infty I(\lambda)y(\lambda)d\lambda,$$
$$Z = \int_0^\infty I(\lambda)z(\lambda)d\lambda,$$

- It's impossible to distinguish two colors with spectral radiance I_1 and I_2 if and only if $X(I_1) = X(I_2)$, $Y(I_1) = Y(I_2)$, and $Z(I_1) = Z(I_2)$.

Estimating True Color Imagery for GOES-R

- Training and verification data sets are obtained from the hyper-spectral imager Hyperion: 6 visible and near visible bands of ABI, and coincident values for X, Y, and Z.
- A non-linear regression method produces a predictor of the values of X, Y, and Z.
  - Fits a separate multi-linear model per K-means cluster

Machine Learning Approach to Clear-sky Classification

- The NOAA Advanced Clear Sky Processor for Oceans (ACSPO) sea surface temperatures (SSTs) product retrieved from VIIRS data.
- The ACSPO clear-sky mask (ACSM) which identify the clear-sky pixels.
- ACSM is created from comparisons of retrieved SST with a first guess (reference) SST (daily Canadian Meteorological Centre CMC product), reflectance thresholds, and spatial uniformity tests. (Petrenko et al. (2010))
- The ACSM generally performs well on a global scale, but there exists
  - *cloud leakages* (cloud miss-classified as clear-sky), and
  - even more *false alarms* (clear sky miss-classified as cloud)

Machine Learning Approach to Clear-sky Classification

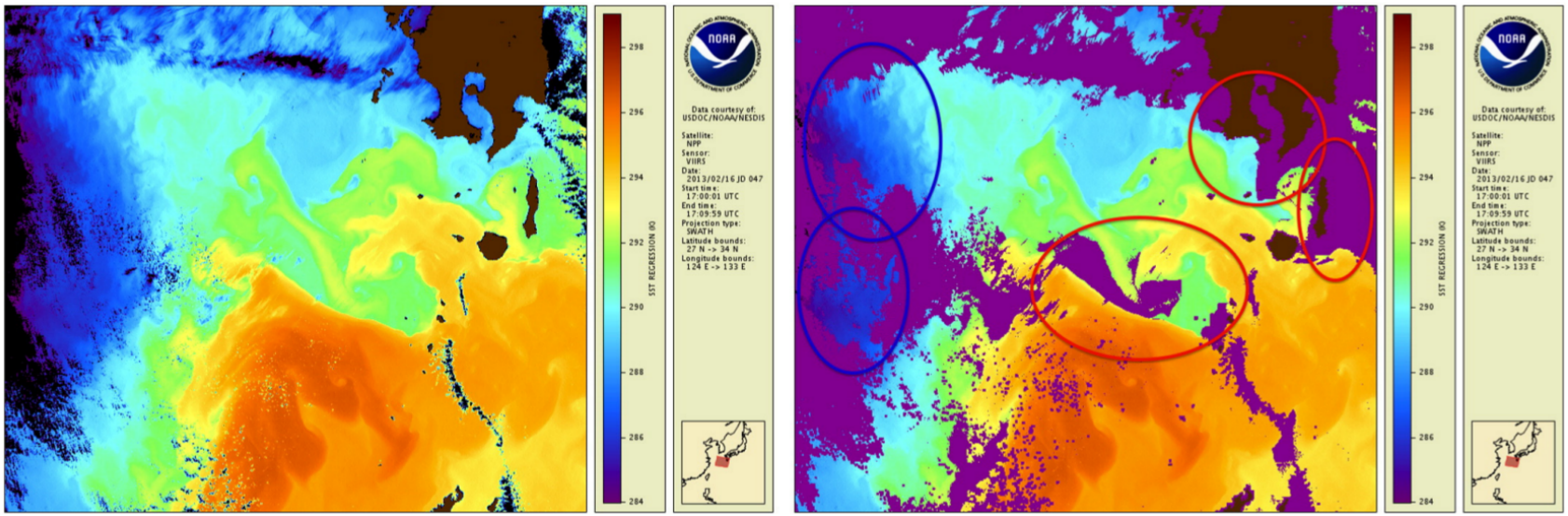


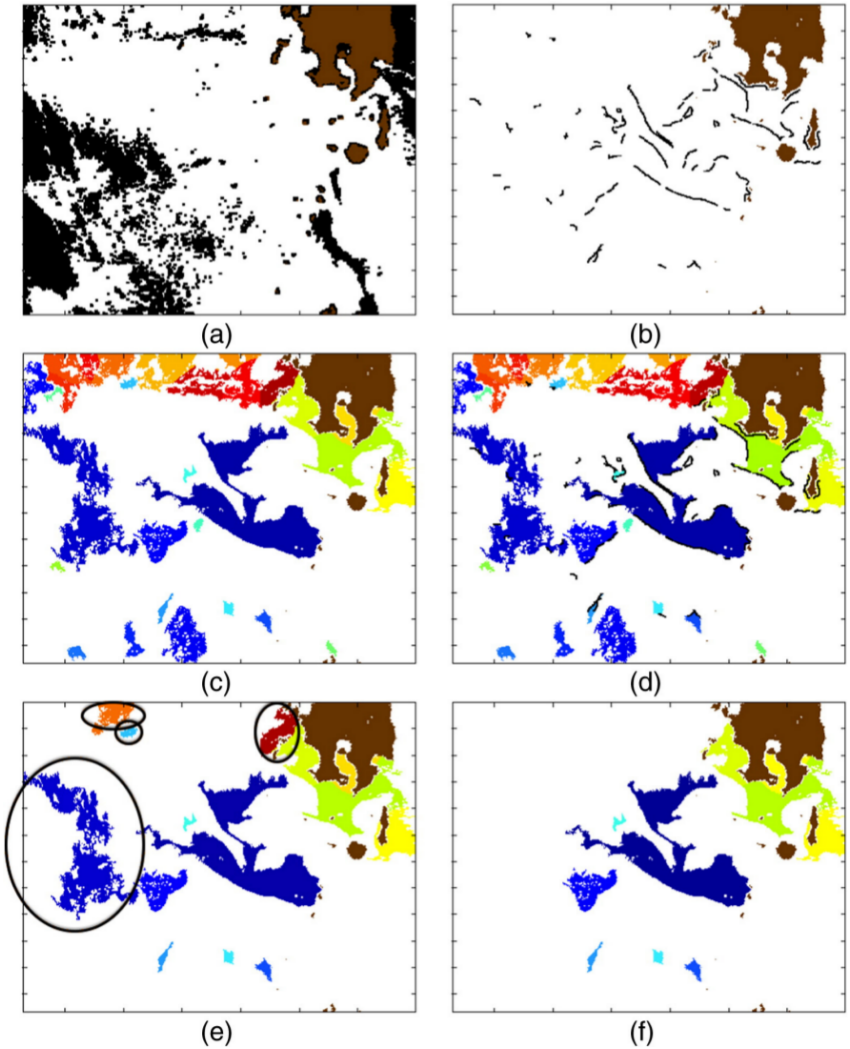
Fig. 1. East China Sea, 16 February 2013, 22:00–22:10 UTC (night). Land rendered in brown, ACSM cloud pixels in magenta and pixels with out-of-scale cold SST values in black; (a) VIIRS SST (without mask); and (b) VIIRS SST with ACSM overlaid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Red is false alarm, and blue is cloud leakage.

Machine Learning Approach to Clear-sky Classification

- Steps to the algorithm proposed in Gladkova et al. (2015):
  1. Narrow down the search domain
  2. Determine SST gradient ridges (SGR). Found in the SST gradient magnitude domain using image processing tools such as morphological dilation, erosion, thinning, and connected components.
  3. Determine spatially connected regions with retrieved SST smaller than the reference SST, using the watershed algorithm on ΔSST.
  4. Ridge Adjacency Test
  5. Corner Cases Test

Machine Learning Approach to Clear-sky Classification



Conclusions

- Missing data can be estimated by a function learned from the training data
  - multilinear regression, polynomial fitting, and nearest neighbor search, etc.
- Sometimes we need use simulated data from a proxy instrument.
- Local features can be taken into account by creating the predictor on a local patch (e.g. QIR) or on segmentation classes (e.g. GOES-R RGB prediction).
- In the clear-sky classification problem, unsupervised learning and image processing techniques are used instead of a training data set.