Near Real-Time Dust Aerosol Detection with Support Vector Machines for Regression

Pablo Rivas-Perea, Pedro Rivas-Perea, Juan Cota-Ruiz, and Raul Aragon Franco

Pablo.Rivas@Marist.edu

MARIST

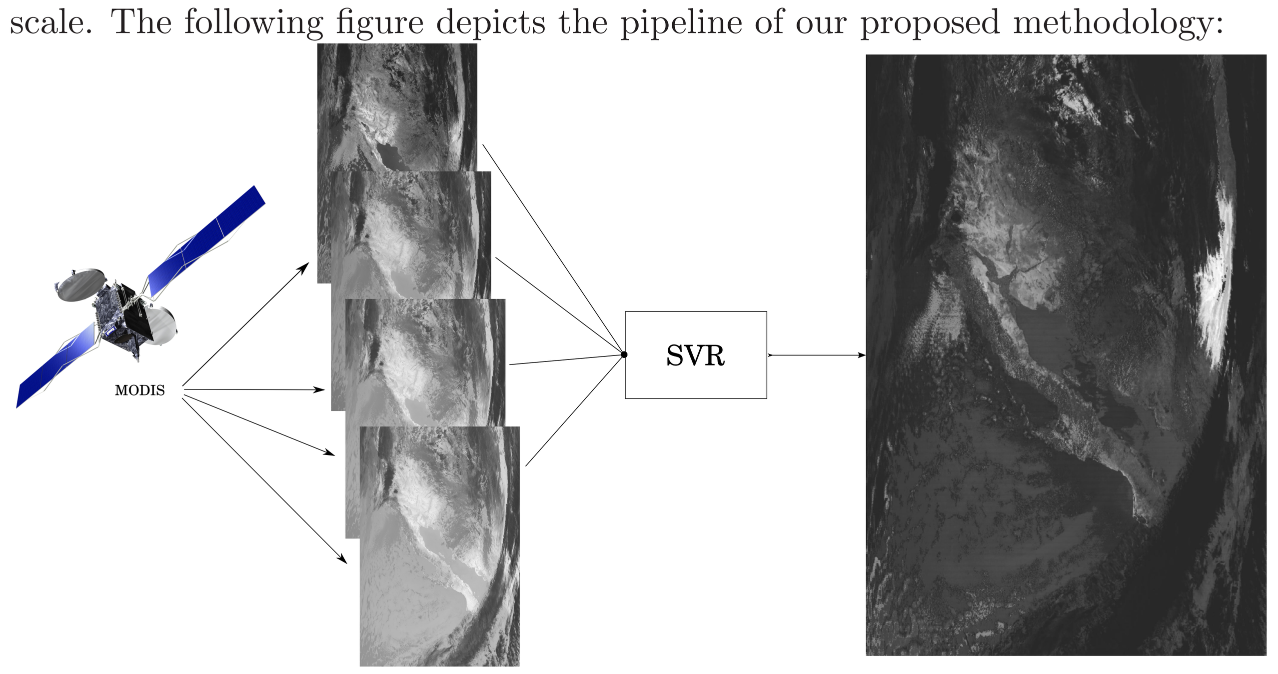
Contribution: Fast Machine Learning-Based Dust Aerosol Detection

Fewer support vectors allow fast computation of kernel distances in SVR for faster estimation of dust aerosol probabilities. The proposed methodology allows for near real-time processing of MODIS data.

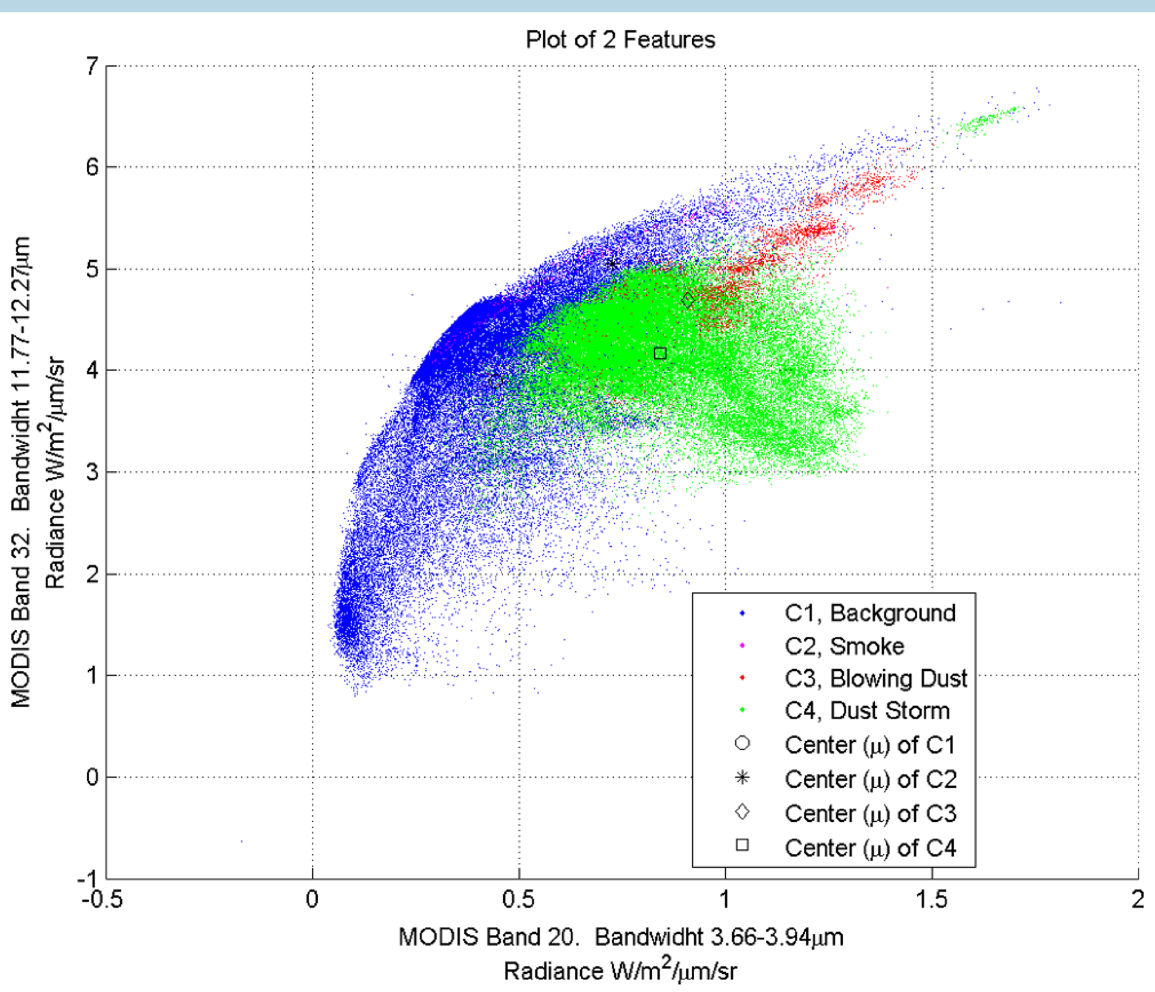
Dust Storms: A Machine Learning Problem

Dust aerosol detection and tracking is typically possible through satellite observations in the visible and near-mid infrared spectrum [1]. Usually, dust aerosol analysis begins with a case study and its fundamental geophysical science principles and moves forward to build robust models. However, given the advances in machine learning algorithms operating over massive datasets, we can now reconsider using large satellite data repositories for aerosol analysis with machine-learning algorithms [2].

In [3] we proposed an algorithm to learn from massive datasets based on Support Vector Machines for Regression and here use it to detect dust aerosols on a global scale. The following figure depicts the pipeline of our proposed methodology:

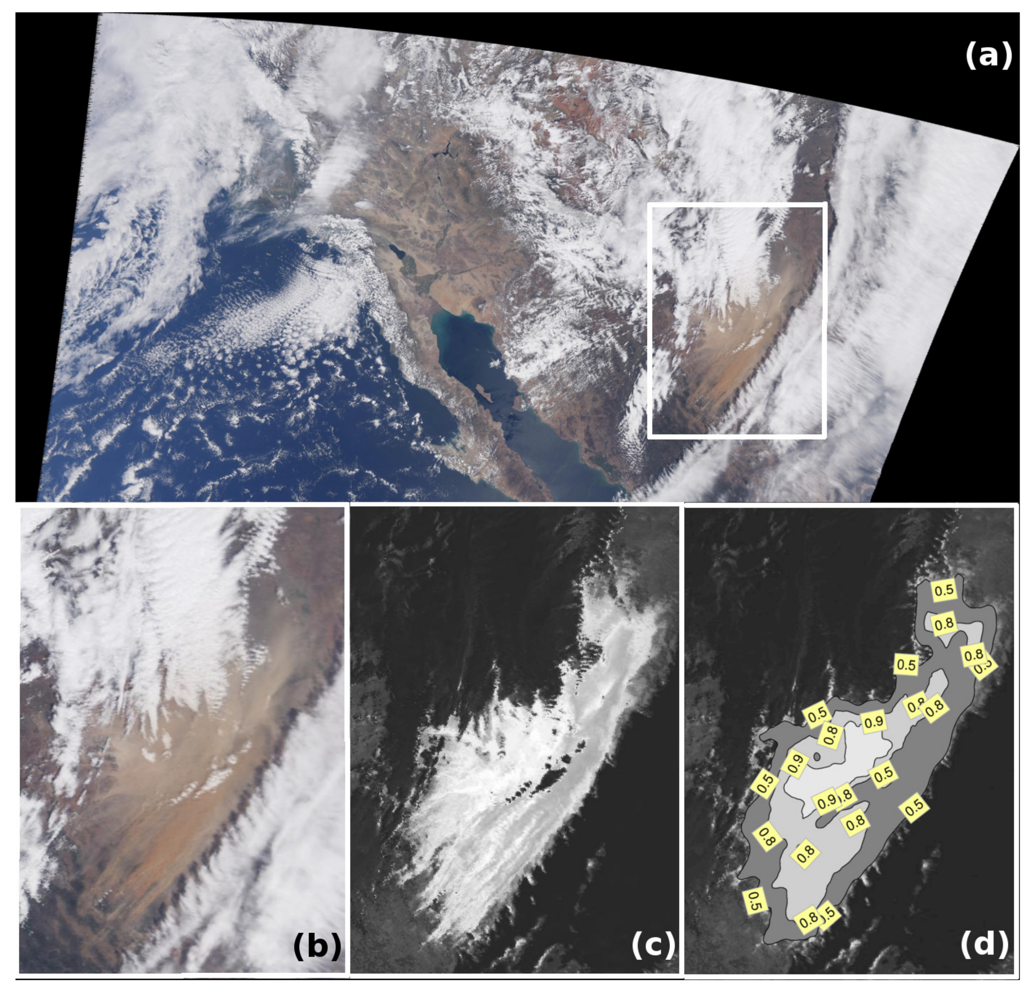


Analisis of Dust Events and Ground Truth



The model was trained with images from 38 dust events from 2001 to 2015. The dataset accounts for 97 million feature vectors with four elements. The analysis of the data is based on the manual tagging of four major classes: Background, C_1, which includes sea cover, clouds, vegetation, etc.; smoke, C_2; blowing dust, C_3, which are minor dust events; and dust storm, C_4, which are major dust events. The figure depicts the distribution of the different classes along two spectral bands $B_{20}^{3.66-3.84\mu m}$ and $B_{32}^{11.77-12.27\mu m}$, out of a total of four spectral bands which include also $B_{29}^{8.40-8.70\mu m}$ and $B_{31}^{10.78-11.28\mu m}$. From the figure we see that some classes are difficult to separate with traditional machine learning methods. We decided to simply make an SVR estimate the probability of dust aerosols. Note, however, that our ground truth also is confirmed using several other methodologies including those by Ackerman, Hao, and also the well known Aerosol Optical Density (AOD).

SVR Estimates Probability of Dust Aerosols



We trained an SVR with our ground truth to estimate the probability of dust aerosols [4]. The output of the SVR is then saved as a gray-scale image where black corresponds to zero probability and white to high probability.

The SVR hyper-parameters were estimated using gradient-descent methodology and the optimal parameters were $\sigma = 0.125$, $\eta = 0.5$, and $\epsilon = 0.1$. The algorithm used in our research aims to minimize the number of support vectors which leads to faster and efficient implementations of the algorithm. It enables its usage in NRT scenarios. The figure above shows that using a simple threshold on the estimate of the probability may lead to image segmentation based on specific confidence intervals. For example, if one wants to display only major dust events and leave out minor dust activity.

Processing Algorithm

Establish connection with NASA's Land, Atmosphere Near real-time Capability for EOS (LANCE) servers and begin downloading satellite data

Load SVR with $\sigma = 0.125$

for *each data granule* **do**

  Extract $B_{20}^{3.66-3.84\mu m}, B_{29}^{8.40-8.70\mu m}, B_{31}^{10.78-11.28\mu m}, B_{32}^{11.77-12.27\mu m}$

  Estimate probability of dust, $\mathbf{y} \leftarrow \text{SVR}(B_{20}, B_{29}, B_{31}, B_{32}, \sigma)$

  for *each desired resolution* **do**

    Reproject $\mathbf{y}$ at resolution

  Publish $\mathbf{y}$ and reprojected imagery

Close connection to NASA

This algorithm is currently running in near real-time [5] and is freely accessible at `dust.reev.us`

References

[1] Mishra, Manoj K., Prakash Chauhan, and Arvind Sahay. "Detection of Asian dust storms from geostationary satellite observations of the INSAT-3D imager." *International Journal of Remote Sensing* (2015).

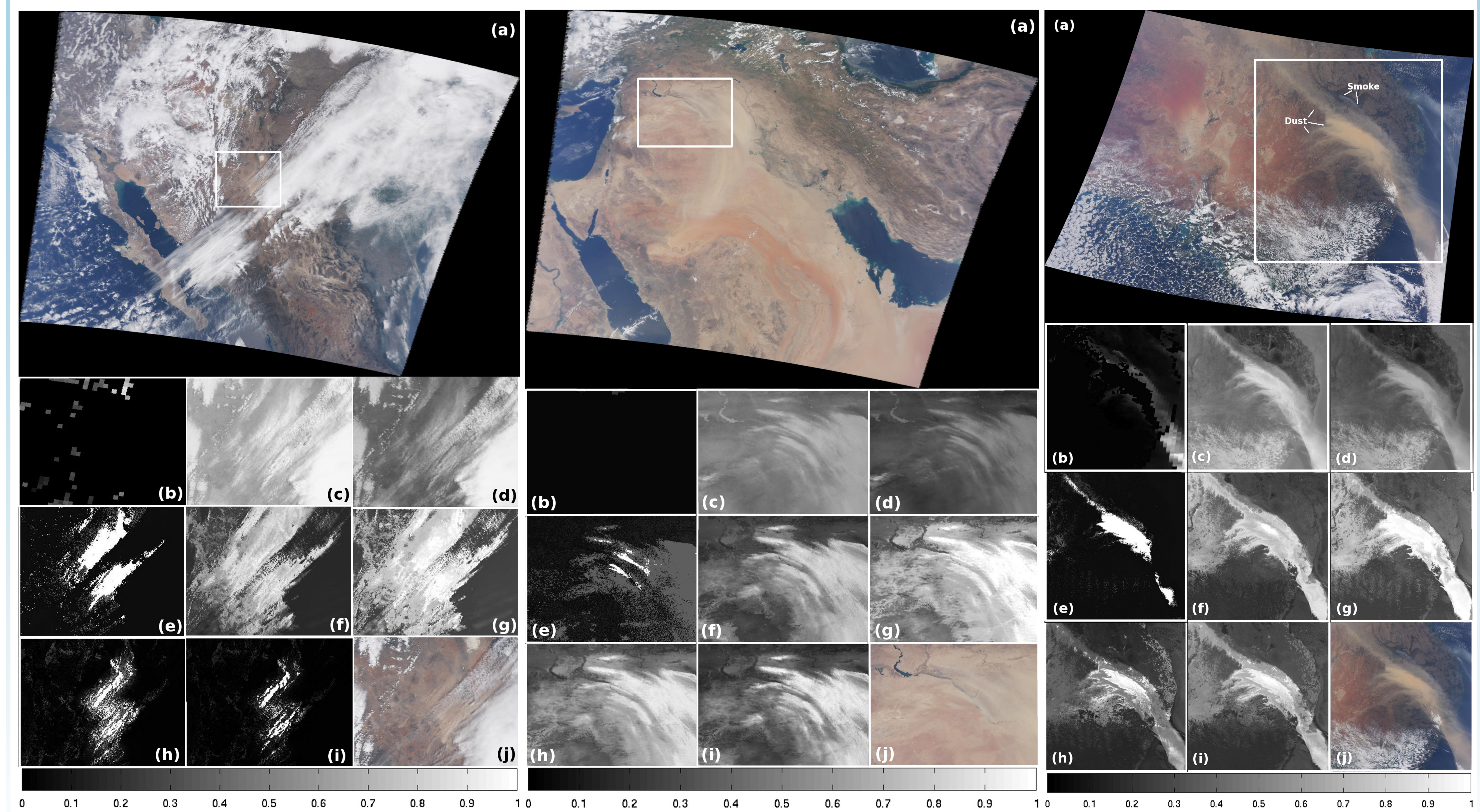
[2] Lary, David J., et al. "Machine learning in geosciences and remote sensing." *Geoscience Frontiers* (2015).

[3] Rivas-Perea, Pablo, and Juan Cota-Ruiz. "An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods." *Pattern Recognition Letters* (2013).

[4] Rivas-Perea, P., J. G. Rosiles, and J. Cota-Ruiz. "Statistical and neural pattern recognition methods for dust aerosol detection." *International journal of remote sensing* (2013).

[5] Rivas-Perea, Pablo, and Juan Cota-Ruiz. "NERT DADS: A Near-Real-Time Dust Aerosol Detection System." *Data for good exchange* (2015).

Results and Comparison with Other Methods



Left. Southwestern US, 10 April 2001, 18:05 UTC. (a) True color. (b) AOD. (c) Ackerman's method. (d) Hao's TDI. (e)-(i) Dust probability using ML, PNN, FFNN, LP-SVR, and proposed LP-SVR. (j) ROI. **Center**. Middle East, 5 June 2009, 07:50 UTC. **Right**. Australia, 26 September 2009, 00:35 UTC.

NRT Global Coverage: Dust Event Near The East Coast of The US. 19 October 2012