# NERT DADS: A Near-Real-Time Dust Aerosol Detection System

Pablo Rivas-Perea
School of Computer Science and Mathematics
Marist College
Poughkeepsie, New York, USA
Pablo.Rivas@Marist.edu

Juan Cota-Ruiz
Dpt. of Electrical and Computer Engineering
Autonomous University of Ciudad Juarez
Ciudad Juarez, Chihuahua, Mexico
jcota@uacj.mx

## ABSTRACT

Global climate is in part affected by airborne particles known as dust aerosols, which are abundant in dry deserted areas such as the northwestern region of Africa. Researchers have found that dust aerosols can travel long distances, even across continents, to participate in the life cycle of other ecosystems. However, in spite of dust aerosols being good for nature, elevated concentrations and prolonged exposure to dust aerosols may deteriorate the quality of life for human beings. Motivated by the potential benefit to our communities we propose a system that detects dust aerosols in near-real time, allowing people to study their own geographical region and prepare for overcoming adverse scenarios caused by major dust events. The proposed system makes use of data science algorithms to estimate the probability of dust aerosols based on carefully chosen multispectral data. During training, our system performs to a 92% of accuracy, and produces a probabilistic view that enables researchers to observe and study the behavior of dust aerosols at a global scale, facilitating the modeling of dust behavior as it relates to global climate.

## Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*Parameter learning*; I.4.10 [**Image Processing and Computer Vision**]: Image Representation—*Statistical*; I.5.1 [**Pattern Recognition**]: Models—*Statistical*

## General Terms

Algorithms

## Keywords

Support vector machines, Supervised learning by regression, Geographic visualization, Dust aerosols, Dust storms, Multispectral analysis

## 1. INTRODUCTION

Current efforts in understanding climate change rely on the understanding of global ecosystems. The number of variables to consider is significantly high, motivating researchers to breakdown the problem into parts that describe specific global climate tasks. One of such is the understanding of the role of dust aerosols transport, i.e., the mobility of airborne dust particles, in global climate. Dust aerosols can be very beneficial, but also detrimental to humans in severe dust events, such as dust storms or haboobs.

Research suggests that dust storms that originate in the Saharan desert positively contribute to one of the largest known eco-systems: the Central Amazon Basin [13]. Dust aerosols, in this case, provide the means for several smaller natural eco-systems to continue their life cycle. Furthermore, recent findings suggest that Saharan dust storms hinder the formation and strength of hurricanes over the Atlantic Ocean [8].

However, dust aerosols are also associated with respiratory illness, particularly under long exposure or high concentrations of dust [9, 12]. Severe dust events close to inhabited geographical regions can cause more harm than good if the awareness and preparedness level of the population is low. The most commonly known negative effects of major dust events include an increase in cases of respiratory illness and allergic reactions, and near zero visibility in nearby airports and roads, which are associated with deadly accidents.

Therefore, understanding how dust aerosols are transported is important to reduce the risk of illness and potential damage in communities. Our research addresses the problem of detecting dust aerosols using data science tools over multispectral data, providing researchers with global near-real time (NRT) data for the study of dust transport and global climate. In particular, we use a support vector regression (SVR) model that produces sparser solutions in terms of the number of support vectors [10], and as input data we use NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) spectral bands in the near infra-red range. MODIS data has proven to be successful in the analysis of dust aerosols [5, 4]. However, state-of-the-art research on dust aerosols focuses mostly in analyzing specific dust events, and only a few provide models that can be applicable globally, still fewer process data daily and make it available to the public. To the best of our knowledge, our research is the only methodologies that produce high-resolution NRT data that is available to the public.

This paper is organized as follows: Section 2 briefly addresses the spectral data that is used as features and it explains the SVR model. Section 3 discusses the process production and publication of the dust detection. Details

Bloomberg Data for Good Exchange Conference.
28-Sep-2015, New York City, NY, USA.

on how to search and retrieve the data are discussed in Section 4. Finally, in Section 5 we discuss our conclusion and briefly mention our future plans of development.

## 2. SVR ANALYSIS OF DUST AEROSOLS

NASA's MODIS instrument is currently aboard two satellites, Terra and Aqua, that scan the Earth in opposite directions and capture data across different spectral frequencies. Such frequencies are usually known as “spectral bands”. The information captured by the instrument is sent directly to the Earth and processed to produce “aerosol products” at different levels of abstraction [6]. The more specific the data product, the deeper the level is, and the longer it takes to be produced. The Land, Atmosphere Near real-time Capability for EOS (LANCE) project from NASA, currently provides the fastest way of accessing aerosol products that have been produced to less strict scientific standards allowing near-real time (NRT) availability.

The spectral bands that have proven to be more effective in finding dust aerosol signatures are those in the near-infrared spectrum [11]. Specifically, we use the following four bands and their corresponding spectral frequency range: B20 ($3.66 - 3.84\mu$m), B29 ($8.40 - 8.70\mu$m), B31 ($10.78 - 11.28\mu$m), and B32 ($11.77 - 12.27\mu$m). We use more than 30 known dust events as ground truth and extract four-dimensional feature vectors for each geographical coordinate that was known to contain dust aerosols. One single dust event may contain data from hundreds of thousands geographical coordinates, depending on the size and extension of the dust event, especially, since one single geographical coordinate covers a 1km area. The highest resolution of the bands we use is 1km per geographical coordinate, which can also be interpreted as a point on Earth, or as a pixel in a spectral image. The complete training data set contained more than 42 million four-dimensional samples, which required a large-scale learning algorithm to model the probability of dust aerosols.

SVRs are an alternative for data scientists looking to output probability estimates to the highest degrees of performance. However, SVRs, in their native formulation, require significant computational resources to solve the optimization problem that produces a model based on support vectors capable of predicting a specific pattern. Specifically, we used a learning algorithm based on a linear programming formulation that was aimed to produce the sparsest solution, resulting in the smallest number of support vectors necessary to solve the problem [10]. The algorithm is specialized in solving large-scale problems using decomposition algorithms that solve smaller linear programming problems efficiently with Interior Point Methods. Our results on the training dataset produced a 92% of accuracy and the visual results are remarkable. Since, we trained the SVR to learn the probability of a single data point being dust, the output is an estimate of the probability of the input being dust, which allows direct visualization of the output as an image.

Figure 1 shows the case of a massive dust event that was captured by NASA Terra's MODIS instrument in October 18th of 2012 in two overpasses: 16:55 UTC and 18:35 UTC.

The figure depicts a dust storm affecting the American states

**Figure 1: Dust event of October 18th of 2012. Satellite: Terra. Overpasses: 16:55 and 18:35 UTC.**

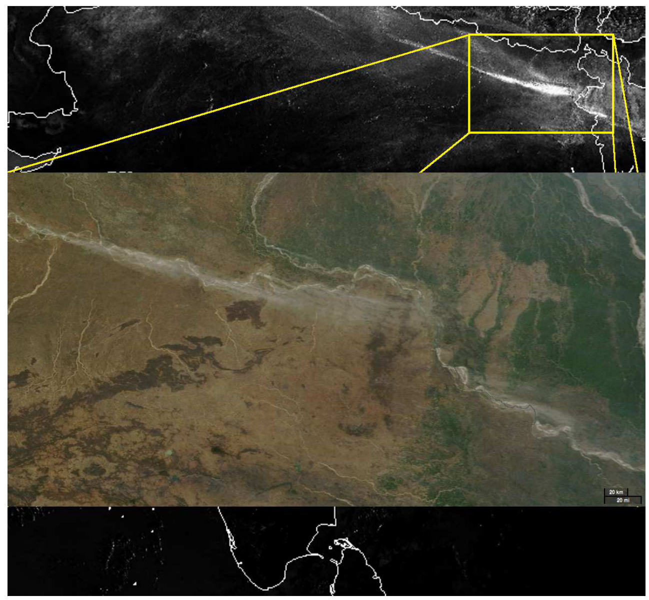

**Figure 2: Dust event of April 5th of 2013. Satellite: Terra. Overpass: 05:25 UTC.**

of Nebraska, Colorado, Kansas, and Oklahoma. Strong winds originated from the Dakotas and extended to Oklahoma. The area monitored by Terra-MODIS suggests that the dust storm size could cover the state of Oklahoma almost in its entirety. The left picture corresponds to the true color image and the right one is the result of our algorithm. From black to white, our visualization indicates low to high probability, respectively. Clearly, our system shows high probabilities of dust aerosols present in the exact regions of the dust storm, providing details that are not visible in the true color image. The massive storm of Figure 1 reportedly caused fatal traffic accidents and respiratory illnesses were associated with the dust event. The effects of the storm were observed even three days after the event, and the dust was tracked all the way up to the East Coast.

Our system also detected a large plume of dust near Bihar, India. Figure 2 shows the dust plume that extends over 600 miles. The data was acquired with NASA's Terra satellite on April 5 of 2013 at 05:25 UTC.

The figure indicates a correct detection of the dust plume and provision of details not trivially observed in the true color representation, which is magnified for visual validation. It is fortunate that there were no reports of accidents, injuries, or other major adversities related to this event.

An archive of these and other successful detections is available on-line at `http://www.lsmmalab.com`; however, our actual live detection system is running and available to the public in the following web address `http://dust.reev.us`. In the next section we briefly describe how the data is produced and published into such website.

## 3. DATA PRODUCTION & PUBLICATION

We have a server computer dedicated to run our NRT Dust Aerosol Detection System (DADS). The server has an Intel i7 64-bit chip with 8 cores with 2 threads each, has 16GiB of RAM, and runs Ubuntu 14.04.3 LTS on a 300GiB disk with a 1TeB disk for archiving the data. The server automatically produces the data following the next steps:

1. Access NASA's LANCE servers to retrieve the data at the moment it becomes available.
2. Process the data from NASA using MATLAB and extract the features.
3. Load the trained SVR and process the features and save the output at a 1km resolution, which is the highest.
4. Reproject the output using the Mercator approach. Reproject at 1, 2, 4, 8, 16, 32, and 64km resolution with coastal and country lines.

The raw outputs from Step 3 are saved as a JPEG files for speed and space convenience, while the reprojected images are saved as PNG image files for better quality. The publication of the data to a web server follows the next steps:

1. Move the newly produced data to the archives and overwrite previously existing data.
2. Produce an `index.html` file to display the latest 11-days in the archive.
3. Upload to the web server only those non-existing or newly modified files
4. Remove from the web server all data older than the latest 11 days.

The publishing steps are triggered after the processing steps are finished. The most computationally expensive operations in the production stage are the SVR processing of features and the reprojection of the results. In the stage of publication, the most expensive operation is the production of the `index.html` file. The slowest operation in the stage of production is the retrieval of the data from NASA, while in the stage of publication is the upload to the web server. It is important to point out that the `index.html` produced is purposely generated in a manner that facilitates web-scraping. Nonetheless, the access to the data should not require scraping techniques since it is easily accessible, as explained in the next section.

## 4. SEARCH AND RETRIEVAL OF DATA

The data that we are making available at the website `http://dust.reev.us` is organized by Julian dates. A Julian date is the consecutive number of the current year with the smallest being 001 and the largest 365 or 366 in a leap year. E.g., January 1st corresponds to 001, February 1st corresponds to 032, and December 31st on a leap year corresponds to 366. In our website, the `index.html` file contains a menu with the latest 11 Julian days to facilitate navigation and browsing. The index file also has a menu to choose the resolution of the reprojected results, which by default is 32km because it fits well with most screens and loads faster. The raw output is available at the lower part of our website, which contains thumbnails of the actual output that facilitates the visual search for dust events.

Besides searching the data using the `index.html` file, data robots can directly download the data by knowing four things: the year, the Julian date, the satellite granular time, i.e., the overpass time, and the resolution wanted. However, if only the reprojected data is wanted, the Julian day and resolution should suffice.

To retrieve the reprojected data, one can simply access the following:

```
http://dust.reev.us/DDD/earth-RES.png
```

where `DDD` is the three-digit Julian, and `RES` is the resolution wanted, e.g., 4km, 8km, or 16km. As an example, consider the following:

```
http://dust.reev.us/221/earth-8km.png
```

which would download August 9th's reprojected data at an 8km resolution. On the other hand, to download the raw output, the exact granular time is needed. The granular time is separated in 5-minute intervals and it is expressed in UTC time. To retrieve the raw output, one can access the following:

```
dust.reev.us/DDD/AYYYYDDDHHMM/AYYYYDDDHHMM-RES.jpg
```

where `YYYY` denotes the four-digit year, `HH` and `MM` denote the two-digit hour and two-digit minute of the granular time, respectively. Consider the following example:

```
dust.reev.us/220/A20152201025/A20152201025-1km.jpg
```

which would download August 8th's raw data for the year 2015 using granular time 10:25 UTC. If the user does not have the granular times available beforehand, they can easily be scraped from the `index.html` file, or from NASA's rapid response team website.

## 5. CONCLUSION AND FUTURE WORK

Using a data science methodology we were able to develop a near-real time dust aerosol detection system that we call

"NERT DADS". It uses multispectral features from satellite scans to detect the signatures of dust aerosol while neglecting the spectral signatures of clouds, fires, land cover, vegetation, sea cover, and snow. To date, researchers have found our methodology relevant for their studies of dust aerosols [1, 2, 7, 15, 14, 3]. We consider this a success story of sharing data for the well-being of our community, of our planet, as it enables the general public to understand the geographical region surrounding them and increase preparedness as global climate changes. And not only the general public but also researchers who are responsible for finding the underlying causes of major dust events and their role in global eco-systems, and perhaps more importantly, their role in climate change itself.

The next logical steps in our research are to make it more available to the general public and much faster. With respect to increasing availability of the data we are considering a better structure of our website, perhaps with the same functionalities that the "Worldview" tool from NASA's Earth Observing System Data and Information System (EOSDIS) [1], specifically, a larger data archive available and the possibility of overlaying the dust detection over true color images. We also think that a free smartphone app can increase availability. The app would show regional and global imagery and will give warnings related to poor air quality due to dust events nearby or otherwise. With respect to speed we will consider options to reduce the latency between NASA server and our server, which would significantly increase the speed of the whole process.

## 6. ACKNOWLEDGMENTS

PRP performed part of this work while at NASA Goddard Space Flight Center under the supervision of Dr. James C. Tilton. The authors acknowledge the support of the Large-Scale Multispectral Multidimensional Analysis (LSMMA) Laboratory (`http://www.lsmmalab.com`), and also credit and acknowledge that all MODIS data used to create the training dataset was acquired via NASA's EOSDIS, Earth Observing System Clearing House (ECHO), and Warehouse Inventory Search Tool (WIST).

## 7. REFERENCES

[1] M. Akhlaq, T. R. Sheltami, and H. T. Mouftah. A review of techniques and technologies for sand and dust storm detection. *Reviews in Environmental Science and Bio/Technology*, 11(3):305–322, 2012.

[2] E. El-ossta, R. Qahwaji, and S. S. Ipson. Detection of dust storms using modis reflective and emissive bands. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(6):2480–2485, 2013.

[3] E. E. A. El-Ossta. *Automated dust storm detection using satellite images. Development of a computer system for the detection of dust storms from MODIS satellite images and the creation of a new dust storm database.* PhD thesis, University of Bradford, 2013.

[4] P. Ginoux, J. M. Prospero, T. E. Gill, N. C. Hsu, and M. Zhao. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on modis deep blue aerosol products. *Reviews of Geophysics*, 50(3), 2012.

[5] Y. Kaufman, I. Koren, L. Remer, D. Tanré, P. Ginoux, and S. Fan. Dust transport and deposition observed from the terra-moderate resolution imaging spectroradiometer (modis) spacecraft over the atlantic ocean. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 110(D10), 2005.

[6] R. Levy, L. Remer, Y. Kaufman, D. Tanré, S. Mattoo, E. Vermote, and O. Dubovik. Revised algorithm theoretical basis document: Modis aerosol products mod/myd04. *Greenbelt, MD: NASA Goddard Space Flight Center*, 2006.

[7] C. Liu, P. Fieguth, and C. Garbe. Background subtraction and dust storm detection. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 2179–2181. IEEE, 2012.

[8] C. Mungai. Ten years after hurricane katrina, it's going to be quiet in the americas this year - thank the sahara for that. *Mail and Guardian Africa*, 2015.

[9] A. D. Oxman, D. C. Muir, H. S. Shannon, S. R. Stock, E. Hnizdo, and H. Lange. Occupational dust exposure and chronic obstructive pulmonary disease: a systematic overview of the evidence. *American Review of Respiratory Disease*, 148(1):38–48, 1993.

[10] P. Rivas-Perea and J. Cota-Ruiz. An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods. *Pattern Recognition Letters*, 34(4):439–451, 2013.

[11] P. Rivas-Perea, J. Rosiles, and J. Cota-Ruiz. Statistical and neural pattern recognition methods for dust aerosol detection. *International journal of remote sensing*, 34(21):7648–7670, 2013.

[12] D. A. Schwartz, P. S. Thorne, S. J. Yagla, L. F. Burmeister, S. A. Olenchock, J. L. Watt, and T. J. Quinn. The role of endotoxin in grain dust-induced lung disease. *American journal of respiratory and critical care medicine*, 152(2):603–608, 1995.

[13] R. Swap, M. Garstang, S. Greco, R. Talbot, and P. Kållberg. Saharan dust in the amazon basin. *Tellus B*, 44(2):133–149, 1992.

[14] M. S. Wong, F. Xiao, J. Nichol, J. Fung, J. Kim, J. Campbell, and P. Chan. A multi-scale hybrid neural network retrieval model for dust storm detection, a study in asia. *Atmospheric Research*, 158:89–106, 2015.

[15] F. Xiao, M. S. Wong, K. H. Lee, J. R. Campbell, and Y.-k. Shea. Retrieval of dust storm aerosols using an integrated neural network model. *Computers & Geosciences*, 2015.

[1] `https://worldview.earthdata.nasa.gov`