**NASA DEVELOP National Program**



Maricopa County Department of Public Health and Arizona State University

*Fall 2016*

Maricopa County Health & Air Quality II

Modeling a Decade of PM10 Concentrations in Maricopa County, Arizona for Air Quality and Epidemiological Analysis

 **Technical Report**

Rough Draft – October 6, 2016

Tamara Dunbarr

Michael Pechacek

Jayanth Serla

Parul Singh

Lance Watkins, NASA Develop (Science Advisor)

David Hondula, School of Geographical Sciences and Urban Planning, Arizona State University (Science Advisor)

Previous Contributors:

Leslie Araujo

Jason Hodgson

# 1. Abstract

Exposure to air pollution is associated with respiratory and cardiovascular morbidity and mortality. Particulate matter (PM) less than 10 microns in diameter is a key contributor to poor air quality, posing a significant public health hazard. In Maricopa County, Arizona the combination of a scarcely vegetated, semi-arid landscape and high anthropogenic activity (e.g. the prevalence of industrial and agricultural activity, wood burning, and vehicle emissions) results in atmospheric concentrations of particulate matter that can exceed federal air quality standards and threaten human health. Despite the presence of one of the Southwest’s most innovative pollutant monitoring networks, which benefits from the latest technologies in advanced real-time detection and prevention, the vast size and complex geography of Maricopa County make it difficult to monitor particulate matter in all communities of interest. Moderate Resolution Imaging Spectroradiometer (MODIS) Level 2 Aerosol Optical Depth data were utilized in combination with Landsat 7 ETM+ thermal infrared data, environmental variables, elevation, and transportation to predict particulate matter concentrations throughout Maricopa County. The predicted particulate matter concentrations were further examined in the context of the US Environmental Protection Agency’s (EPA) Air Quality Index, which relates particulate concentration to health impact, to categorize the health risk around schools throughout the county. The end products produced from this project were given to the Maricopa County Department of Public Health and the Maricopa County Air Quality Department for future epidemiological research and public outreach efforts.

**Keywords**

Remote sensing, air quality, air pollution, particulate matter, health, EPA, NAAQS, mixed effects modeling

# 2. Introduction

* 1. ***Background Information***

Particulate matter (also commonly referred to as aerosols) include any atmospheric particles derived from anthropogenic and natural sources (Kaufman et al., 2002). Airborne particulate matter less than 10 microns in aerodynamic diameter (PM10) is a human respiratory health hazard and a large contributor to air pollution in urbanized areas. Traffic, industrial activity, and biomass burning (i.e., home heating and cooking) are prime examples of non-natural sources of PM10 (Han, 2006). Natural sources of the pollutant include volcanic ash, suspended dust from concentrated agricultural activity, and dust storms in arid and semi-arid areas, as in Maricopa County, Arizona.

Multiple studies confirm that excessive human exposure to PM10 has been largely demonstrated to have negative consequences on the human respiratory system and overall health. Many of these studies have successfully established an association between elevated and concentrated PM10 levels with respiratory, cardiovascular, and mutagenic diseases (Schwartz, 1993). The microscopic size of the particles allows them to penetrate deeper into the respiratory system causing a more severe level of exposure. Effects of long term exposure to particulate matter include hospitalization for asthma and chronic respiratory diseases.

Currently, across the United States, air quality and health departments sample small areas using ground-based air quality monitoring stations to help reduce PM10 emissions and subsequent epidemiological problems associated with PM10 inhalation. However, this leaves large swaths of land without PM10 monitoring, potentially exposing thousands or millions of people to environmental hazard (Prud’homme, 2013).

Many current methods used to estimate PM10 concentrations using ground-monitor point data involve interpolation (such as kriging and inverse distance weighted regression) and land-use regression models. These methods, although useful to an extent, do not provide the level of spatial and temporal granularity that is obtained from satellite-based measurements such as aerosol optical depth (AOD) which is a measure of extinction of sunlight in the atmosphere by aerosols. Satellite aerosol optical depth data is not always employed in these studies given its complex nature and because it requires a higher level of understanding. Still, there are some studies that use satellite data, though they typically involve large regional study areas (Kloog, 2011; Nordio, et al., 2013). Kriging is a favored interpolation method for *in situ* particulate matter point observations, however this does not give enough information to identify local sources of PM10 (and other pollutants), especially in smaller, county-level study areas (Pope et al., 2014).

Utilizing methodology outlined by Kloog (2011) and Nordio (2013), this study uses monthly MODIS aerosol optical depth and ground-based sensor data, along with meteorological, open space/land cover, transportation, and elevation data to create a model of predicted PM10 concentrations for areas not covered by ground-based monitoring sensors. This study focuses on the Maricopa county in Arizona. The study observes PM10 in these counties from January 1, 2006 to December 31, 2015. This area is unique climatically because it is a semi-arid desert area situated at approximately 33º in latitude, in the midst of a subtropical high-pressure band. This high-pressure area of the planet is characterized by stagnant air flow and little cloud cover. The surrounding mountains further restrict air circulation, creating a settling effect. This effect prevents air pollution from circulating out of the metropolitan area.

* 1. ***Project Partners & Objectives***

The objective of this project was to assist the Maricopa County Department of Public Health (MCDPH) and the Maricopa County Air Quality Department (MCAQD) with decision-making related to PM10 both in regard to Federal regulations and epidemiological concerns and inhalation risk. The MCDPH is interested in cross-referencing areas of high PM10 concentrations with other factors, such as schools in the study area where students may be effected by higher concentrations of particulate matter. The MCAQD will use the data to assess areas of greater concern that are not covered by ground-based PM10 monitors in order to address potential hotspots for pollution activity, both spatially and temporally. National Ambient Air Quality Standards dictate that PM10 concentrations are measured and reported every 24 hours. Only one exceedance per year is allowable under the Clean Air Act provisions (NAAQS, 2016).

The objective was to produce the PM10 MODIS-enhanced mixed effect model and archive of modeled PM10 concentrations from 2006 – 2015 at a 3km spatial and daily temporal resolution for the Maricopa County study area. This timeframe was chosen because it represents the best long-term intersection of both satellite and ground-based monitoring systems. This project falls within NASA’s Health & Air Quality National Application Area. The goal outlined above encourages the use of Earth observations in: assessing exposure of health-related hazards, the implementation of air quality monitoring, and public health outreach and intervention.

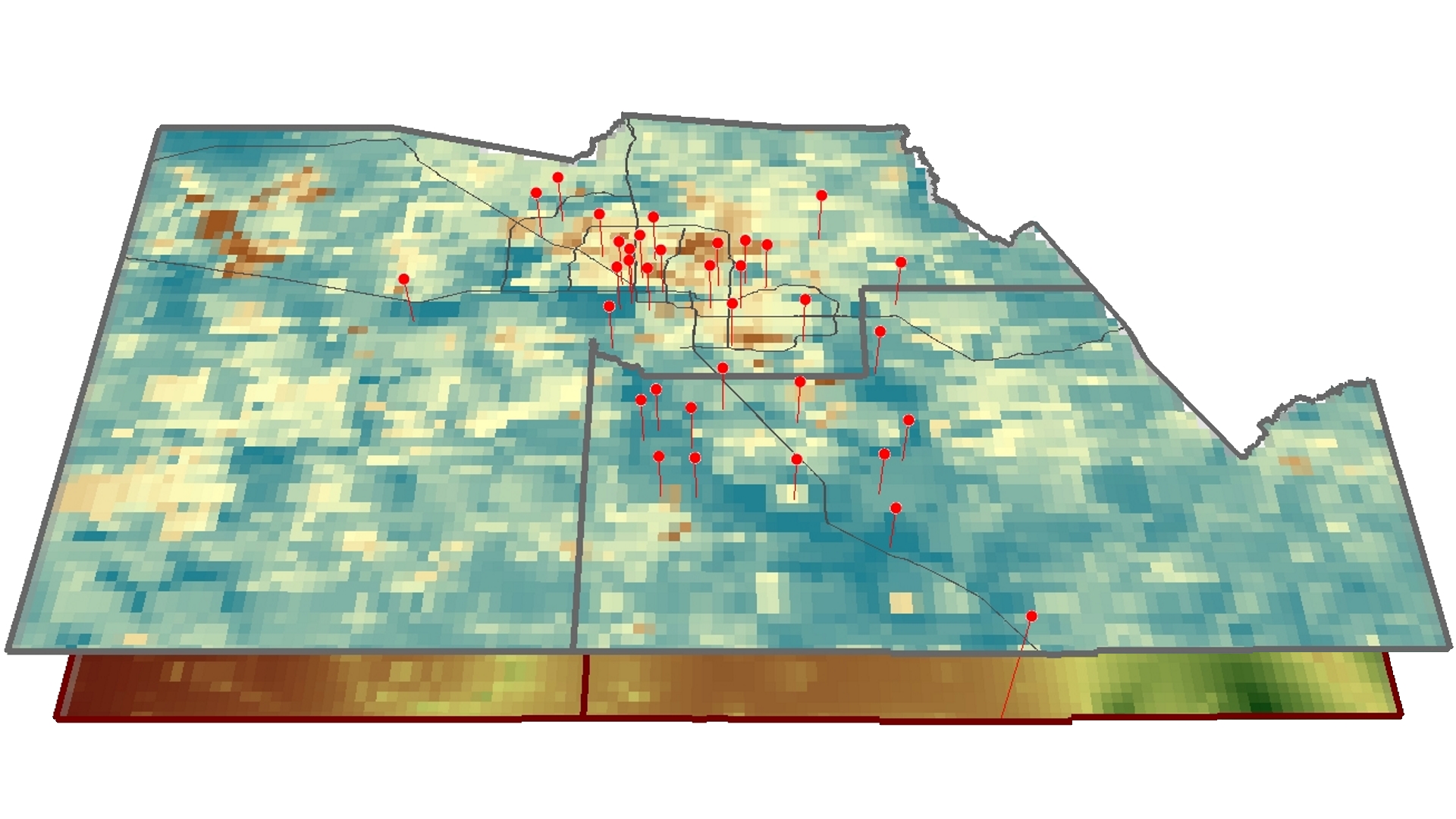


Figure 1. Location of all PM10 monitors (red vertical markers) in

Maricopa County and Pinal County, Arizona overlaid on an AOD raster surface.

# 3. Methodology

The MODerate Resolution Imaging Spectroradiometer (MODIS) is an instrument aboard the Terra satellite, the flagship satellite of the three satellites initially launched as a part of the NASA Earth Observing System (EOS) mission. MODIS is also operating on the second EOS satellite, Aqua, launched two years after the Terra satellite. For this study, the 3 km spatial resolution of MODIS data was best suited to analyze the Maricopa county study area when compared to other Earth observations such as Multi-angle Imaging SpectroRadiometer (MISR) and Cloud Aerosol Lidar Pathfinder Satellite Observations (CALIPSO).

All of the mentioned sensors, including MODIS, are satellites which are part of the “A-train”—a string of NASA earth-observing satellites closely following one another in a polar orbit.

Multiple potential predictors were examined for the required spatial resolution of a 3 km x 3 km grid: areas of open spaces, elevation, length of roads, and meteorological data.

To be able to create the model, the location of all the included sites were overlaid on the 3-km layers.

***3.1 Data Acquisition***

*MODIS Data (AOD)*

AOD data were obtained for the 2006-2015 period using Level 1 and Atmosphere Archive and Distribution System (LAADS Web), using Terra MODIS Level 2 Aerosol Product at 3km resolution (MOD04\_3K) and Aqua Level 2 Aerosol Product at 3km resolution (MYD04\_3K) from Collection 6 Aerosol Data from the Corrected\_Optical\_Depth\_Land parameter.

*Landsat 7 ETM+ Infrared Data*

The thermal infrared data were obtained for the 2006-2015 period from the USGS EROS archive using Glovis.

*County line data*

County border shapefiles were obtained from the TIGER/Line repository.

*Climate data*

Climate data were obtained from the Daymet 2-meter air temperature was retrieved from Oak Ridge Laboratory, Daily surface weather and climatological summaries. Data utilized was in the daily surface of minimum and maximum temperature from 2006 – 2015 was calculated into daily average temperature in degrees Celsius.

*Elevation*

Elevation data were obtained from the USGS National Elevation Dataset.

*Open space*

Land cover data were obtained from the United States Geological Survey (USGS) 2011 National Land Cover Database (NLCD), which was then clipped to the Maricopa and Pinal Counties. Land use data were acquired from the Maricopa Association of Governments’ 2012 Maricopa General Plan. These datasets were combined to determine the percentage of open space within the study area.

*Transportation*

Transportation data were obtained from the TIGER/Line repository for each year.

***3.2 Data Processing***

County border shapefiles were constrained to Maricopa county and all other layers were clipped to the extents of this layer for aesthetics regarding the area of interest.

The AOD variable comes directly from the MODIS data and is a measure of the aerosol concentration in any given area. Using this variable assists in accurately predicting the ground level PM amounts. MODIS data were geographically constrained to 34.5 N, 32 S, -110 E, and -114 W, temporally constrained day-to-day when all the variables were available for a given day. Only daytime files were used to increase accuracy because MODIS uses all of its available 36 sensors for daytime observations as opposed to 16 during nighttime observations (Toller, 2002). The files were processed into a geographic projection and output as a GeoTIFF raster with nearest neighbor processing via the LAADS post-processing system.

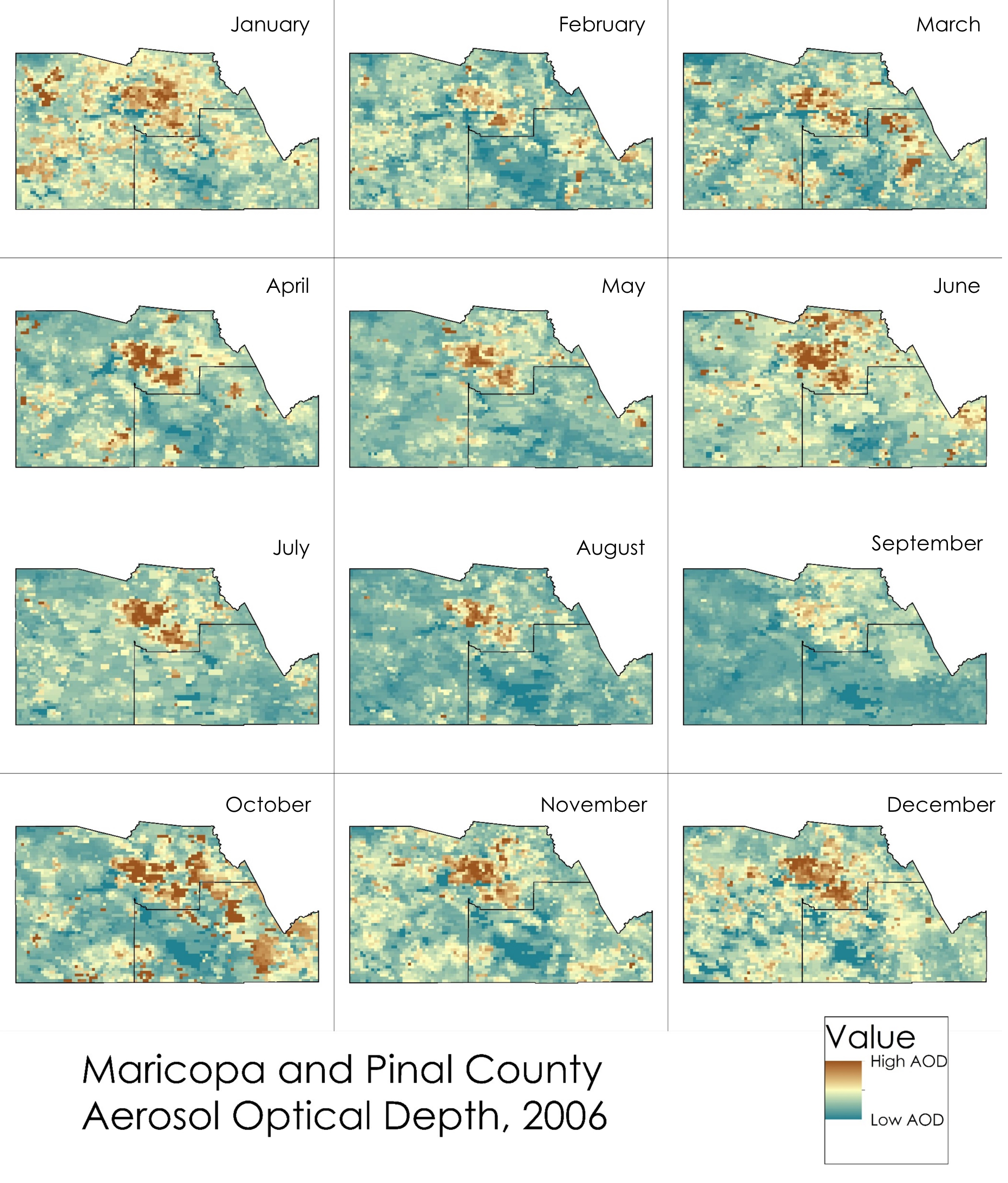


Figure 2. A compilation of all processed AOD maximum value monthly raster surfaces

from MODIS Level 2 data for 2006.

The thermal infrared data were obtained from Landsat 7 ETM+ SLC-off images held in the USGS EROS archive. DN values of the pixels were extracted and all the pixels with DN values of 0 and 255 were omitted in order to account for cloud cover, sensor saturation and the absence of spacecraft line correction.

The climate data was obtained from the Oak Ridge Laboratory; this data was in the format of tiles. ArcPy was used to write python script to convert netCDF format to raster, mosaic the tiles together, extract by mask to the study area, raster calculator to find the average temperature, and to find the temperature values of each pixels of the monitoring stations.

Elevation data were obtained from the USGS’s National Elevation Dataset. It was aggregated via bilinear interpolation to a 3-km cell grid and used as the primary reference for all other layers (the snap raster). This ensured that all 3-km raster cells were aligned properly for comparison.

Percent of open space was derived from the USGS’s 2011 National Land Cover Database as a raster, and from Landsat for 2006-2012 as a raster. Any area that was classified as non-open space within the context of PM10 (such as developed, urban area) was assigned a value of ‘0’, while all other open areas (such as shrub land, agricultural areas) were assigned a value of ‘1’. After reclassification, this raster dataset was aggregated and snapped to a 3km grid to match the elevation dataset. Each raster cell value indicated the percent of open space within that cell’s area (this value was obtained by calculating the 0/1 value average during aggregation). This process allowed for quick identification and a calculation that simplified the NLCD data into open-space and non-open space.

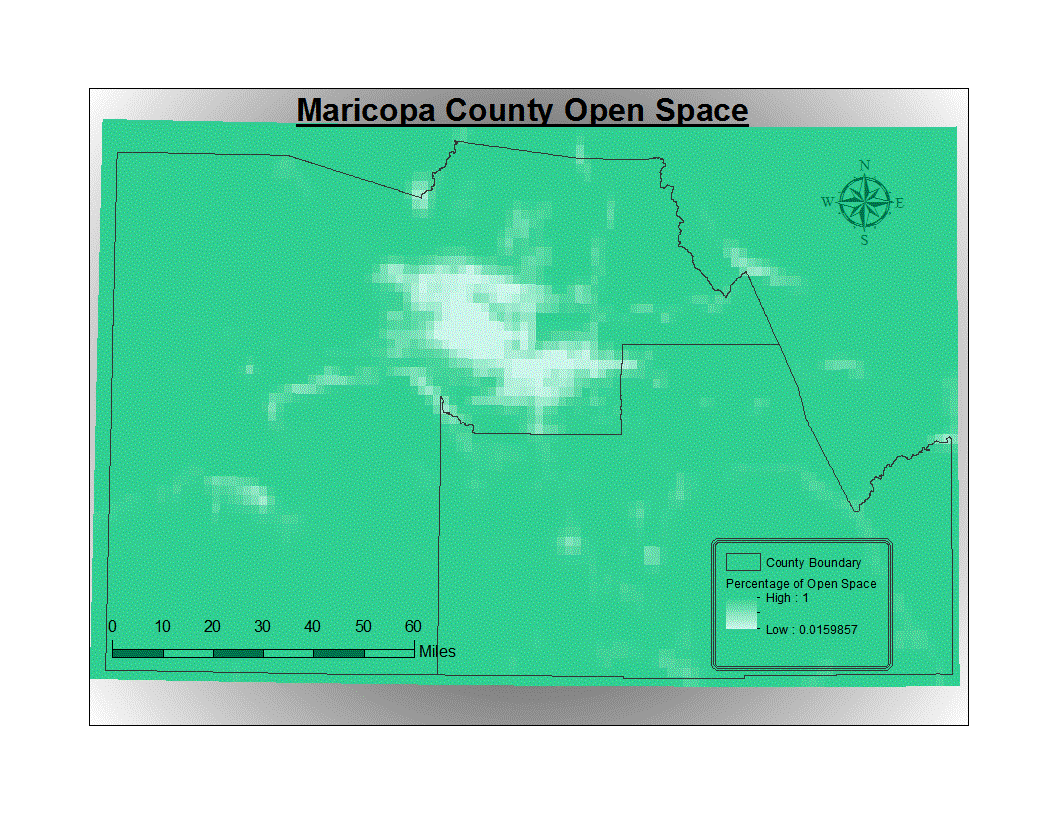


Figure 3. Percent of open space raster surface for Maricopa County and Pinal County derived from reclassification and aggregation of the USGS’ 2011 NLCD and snapped to a 3-km grid.

Transportation data were obtained from the TIGER/Line repository. These data were used to determine road densities for each 3-km grid square. The roads layer was clipped to the Maricopa County boundary lines and then a 3-km fishnet was created and overlaid the roads layer. A geo-processing intersect was performed on the roads with the fishnet as the feature which allowed the road segments to be sectioned off at the fishnet boundaries. Each 3-km fishnet cell was assigned an OBJECTID and the SUM of the road lengths were calculated for each unique cell. This allowed values to be assigned to each cell which provides a numerical value to input into the model.

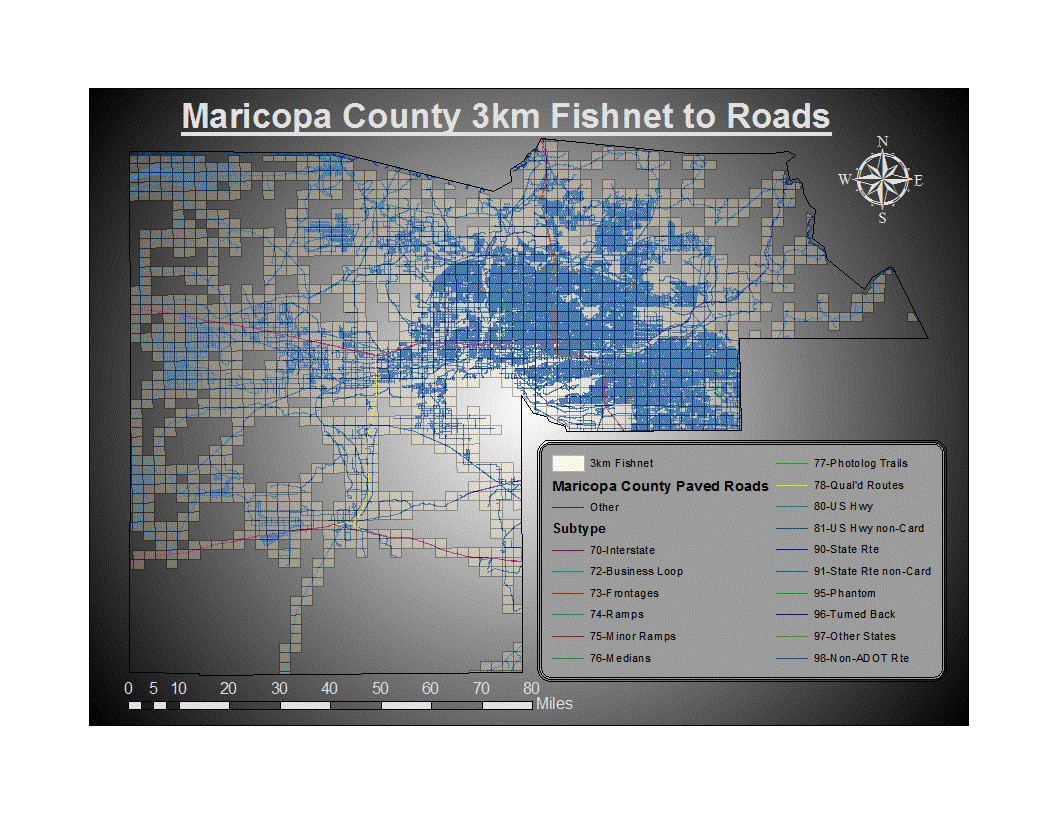


Figure 4. Map depicting the dataset used to produce

a road density rasterized surface via a 3-km fishnet grid.

***3.3 Data Analysis***

One of the key points to keep in consideration while analyzing these datasets in an effort to come up with a model to predict PM10 is that the raw monthly correlations between AOD and PM10 are not very strong. This is due to the fact that AOD observes and reports any particulates and pollutants in the air, so it is essentially a measure of how “dirty” the air is over a certain area. There are plenty of other effects that contribute to AOD levels, and PM10 is only one of those many particulates and pollutants. By incorporating other variables associated specifically with PM10 into the mixed effect model, we were able to distinguish that specific relationship and help improve the predictability of PM10 concentrations in areas currently unrepresented by PM10 monitors.

While many of the variables were unchanged, part of the continuation of this study is focused on refining the model used in the previous study, and overlaying that information with other elements to determine health or epidemiological implications of the study. Some of the similarities are the use of a generalized mixed model, the study area, and several of the data sets, such as the national elevation data set, and the ground-based PM station data set. Other data sets were enhanced with the use of additional or re-processed information. Satellite data at a finer spatial scale via the MAIAC algorithm of the MODIS data was used. Additionally, processing was done at a finer temporal scale by overlaying each day’s (the days where all the variables were available) satellite imagery on the study area, instead of aggregating to monthly data. Meteorological data was enhanced by the use of data from ground-based monitors via Daymet. Road density data was acquired for all study years instead of using a single year of road density for the entire representative study.

To set up the data in a format ready for model development, we first obtained all the PM10 monitoring station location information (latitude and longitude) and listed it on a table. Each monitoring station had its respective monthly maximum and average observations listed, and the monthly average temperature designated to each monitoring station was also included. To incorporate the raster datasets, the value of the pixel corresponding to each monitoring station location was identified and then added to the table. This was done for AOD, elevation, percent of open space, and transportation raster datasets. Therefore, for every monitoring station listed in the table, there was a corresponding monthly maximum PM10, average PM10, average temperature, the percent of open space, elevation value, and road density.

Before being imported into R, a refined dataset was also created to be compared to the complete dataset for model performance. The refined dataset included only Maricopa County monitoring sites that had complete data for all the corresponding variables. Ultimately, the entire statistical modeling process was carried out on both datasets to determine which dataset resulted in a better-performing model for predicting PM10.

Once imported into R, the lme4 package was used to create the mixed effect model. Given that the values for each of the variables vary greatly, all of the variables were also scaled beforehand to allow the model to run efficiently. Two base models relating AOD to PM10 average and PM10 max values were created as a starting point. More variables were added one by one, and AOD and temperature were added as random effects with respect to month and monitoring site.

Once all variables were included, the Akaike Information Criterion (AIC) was used as the main indicator for model performance, and the varying models were compared to one another based on their AIC score. The models were adjusted to optimize performance and PM10 predictability, and once those refined models were obtained, their corresponding R2 value was calculated.

# 4. Results & Discussion

A correlation between the DN values of the thermal infrared band images and the observed PM10 values could not be determined when the infrared data was used without taking any of the ancillary data into consideration. As a result, the thermal infrared data has not been used for further modeling.

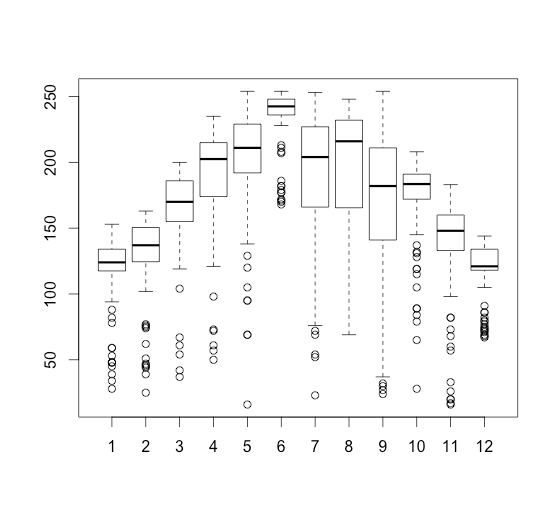


Figure 5. Variance of the thermal infrared DN values by month (DN value vs. Month)

# 5. Conclusions

There seems to be no plausible correlation between the thermal infrared data and observed PM­10 values when only the DN values of the thermal band images were used in the model. It could be that the thermal images would have to be processed further before any correlation can be observed. Aerosol optical depth, road density, temperature, elevation, and ground-based stations have a correlation just below 0.3 at the daily temporal scale, which is consistent with the results of the former study aggregating the data at a monthly scale. This is also consistent with correlations reached by Kloog, Nordio, et al. in semi-arid desert areas. Further research efforts suggest the incorporation of other variables or other modeling techniques to help refine the model further. Suggestions include barometric pressure readings and measurements of soil in the area.

# 6. Acknowledgments

The authors would like to thank Dr. Itai Kloog, Geography and Environmental Development Department, Ben-Gurion University of the Negev, Be'er Sheva, Israel.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL11AA00B and cooperative agreement NNX14AB60A.

# 7. References

Han, X., and Naeher, L. P. (2006). A Review of Traffic-related Air Pollution Exposure Assessment Studies in the Developing World. Environment International 32.1, 106-120. doi:10.1016/j.envint.2005.05.020

Kaufman, Y. J., Tanré, D., and Boucher, O. (2002). A Satellite View of Aerosols in the Climate System. Nature 419.6903, 215-223. doi:10.1038/nature01091

Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment, 45*(35), 6267-6275. doi:10.1016/j.atmosenv.2011.08.066

NAAQS Table. (2016, March 29). Retrieved June 30, 2016, from https://www.epa.gov/criteria-air- pollutants/naaqs-table

Nordio, F., Kloog, I., Coull, B. A., Chudnovsky, A., Grillo, P., Bertazzi, P. A., . . . Schwartz, J. (2013). Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite  
 data and land use regression over Lombardy, Italy. *Atmospheric Environment, 74*, 227-236.   
 doi:10.1016/j.atmosenv.2013.03.043

Pope, R., & Wu, J. (2014). Characterizing air pollution patterns on multiple time scales in urban areas: A landscape ecological approach. *Urban Ecosystems*, *17*(3), 855-874. doi:10.1007/s11252-014-0357-0

Prud'homme, G., Dobbin, N. A., Sun, L., Burnett, R. T., Martin, R. V., Davidson, A., . . . Johnson, M. (2013). Comparison of remote sensing and fixed-site monitoring approaches for examining air pollution and health in a national study population. *Atmospheric Environment, 80*, 161-171. doi:10.1016/j.atmosenv.2013.07.020

Schwartz, J., Slater, D., Larson, T. V., Pierson, W. E., & Koenig, J. Q. (1993). Particulate Air Pollution and Hospital Emergency Room Visits for Asthma in Seattle. Am Rev Respir Dis American Review of *Respiratory Disease, 147*(4), 826-831. doi:10.1164/ajrccm/147.4.826

Toller, G. N. "MODIS Level1B Product User's Guide." NASA/Goddard Space Flight Center (2002): 28.   
 Web. June 2016.

# 8. Content Innovation

**Content Innovation #1**

VPS

Emailed to [Lauren.M.Childs@nasa.gov](mailto:Lauren.M.Childs@nasa.gov) with filename AZVPS.mp4

**Content Innovation #2**

Glossary Viewer

* Aerosol Optical Depth (AOD) **-** The measure of aerosols in the atmosphere measured from the MODIS instrument on board the Terra and Aqua satellites. This measurement is not the same as a PM10 count, but with statistical modeling it can be highly correlated to PM10 in order to accurately predict PM10 concentrations over an area.
* AIC Score - The Akaike Information Criterion (AIC) is a measure of the relative quality of statistical models for a given set of data.
* Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) - CALIPSO is a joint NASA (USA) and CNES (France) environmental satellite, built in the Cannes Mandelieu Space Center, which was launched atop a Delta II rocket on April 28, 2006.
* Digital Elevation Model (DEM) - A Digital Elevation Model (DEM) is a digital cartographic/geographic dataset of elevations in xyz coordinates.
* Epidemiology - the branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health.
* In Situ - In *situ* is a Latin phrase that translates literally to "on site" or "in position".
* Kriging - In statistics, originally in geostatistics, Kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariance’s, as opposed to a piecewise-polynomial spline chosen to optimize smoothness of the fitted values.
* NASA Earth Observing System (EOS) - A program of NASA comprising a series of artificial satellite missions and scientific instruments in Earth orbit designed for long-term global observations of the land surface, biosphere, atmosphere, and oceans of the Earth.
* GeoTIFF - GeoTIFF is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file.
* Maricopa County Department of Air Quality (MCDAQ) - The Maricopa County Air Quality Department is a regulatory agency whose goal is to ensure federal clean air standards are achieved and maintained for the residents and visitors of Maricopa County, Arizona.
* Maricopa Department of Public Health (MCDPH) – The Maricopa County Department of Public Health is a regulatory agency whose goal is to protect and promote the health and well-being of all of our residents and visitors in Maricopa County, Arizona.
* Multi-angle Imaging SpectroRadiometer (MISR) - MISR is a scientific instrument on the Terra satellite launched by NASA on December 18, 1999. This device is designed to measure the intensity of solar radiation reflected by the Earth system (planetary surface and atmosphere) in various directions and spectral bands.
* MODerate-resolution Imaging Spectroradiometer (MODIS) - MODIS is a payload scientific instrument built by Santa Barbara Remote Sensing that was launched into Earth orbit by NASA in 1999 on board the Terra Satellite, and in 2002 on board the Aqua satellite.
* National Ambient Air Quality Standards (NAAQS) - The National Ambient Air Quality Standards (NAAQS) are standards established by the United States Environmental Protection Agency under authority of the Clean Air Act (42 U.S.C. 7401 et seq.) that apply for outdoor air throughout the country.
* National Centers for environmental Information (NCEI) - NOAA’s National Centers for Environmental Information (NCEI) are responsible for hosting and providing access to one of the most significant archives on earth, with comprehensive oceanic, atmospheric, and geophysical data.
* National Land Cover Database (NLCD) - NLCD 2006 is designed to provide the user both updated land cover data and additional information that can be used to identify the pattern, nature, and magnitude of changes occurring between 2001 and 2006 for the conterminous United States at medium spatial resolution.
* PM10 **–** Can either refer to a piece of particulate matter less than 10 microns in size, or the measure of PM10 concentration in the atmosphere. High levels of PM10 are of epidemiological concern because they are inhalable at that size.
* R - *R* is a language and environment for statistical computing and graphics.
* RStudio - RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.
* United States Geologic Survey (USGS) - The United States Geological Survey is a scientific agency of the United States government. The scientists of the USGS study the landscape of the United States, its natural resources, and the natural hazards that threaten it.
* DN – Digital Number is a value assigned to a pixel in a digital image. For an 8-bit image the DN value ranges from 0 to 255; 0 implying black and 255 meaning white.
* ETM+ - Enhanced Thermal Mapper Plus is an 8-band, multispectral sensor on the Landsat 7 satellite. It can detect electromagnetic radiation in the visible, near infrared and thermal infrared regions of the spectrum.

**Content Innovation #3**

Inline Supplementary Material

* Figure 1
* Figure 2
* Figure 3
* Figure 4
* Figure 5