Award Number: W81XWH-11-2-0165

TITLE: Linking NASA Data with Environmental Exposures and Health Outcomes in Theater of War

PRINCIPAL INVESTIGATOR: David Lary

CONTRACTING ORGANIZATION: University of Texas at Dallas
Richardson TX 75080

REPORT DATE: October 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland  21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
14. ABSTRACT
The hypothesis of this study is that a suite of remote sensing data products on atmospheric aerosols used in their meteorological context and processed by machine learning can provide a daily estimate of the global PM2.5 abundance. This information is of considerable value to Global Health Surveillance (GHS), providing a capability to routinely estimate troop deployment exposure to elevated levels of particulate matter (PM) globally, significantly contributing to DoD-wide force health protection initiatives. We have exceeded our promised goal and have provided a daily global estimate of the PM2.5 distribution from February 2000 up through the present (we promised from 2006-present).

15. SUBJECT TERMS- none provided

16. SECURITY CLASSIFICATION OF:

<table>
<thead>
<tr>
<th>a. REPORT</th>
<th>b. ABSTRACT</th>
<th>c. THIS PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17. LIMITATION OF ABSTRACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>UU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>18. NUMBER OF PAGES</th>
<th>13</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>19a. NAME OF RESPONSIBLE PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>USAMRMC</td>
</tr>
</tbody>
</table>

| 19b. TELEPHONE NUMBER (include area code) | 000 |
Table of Contents
Introduction..................................................................................................................................... 4
Hypothesis....................................................................................................................................... 4
Technical Objectives....................................................................................................................... 4
Method ............................................................................................................................................ 5
Relating Aerosol Extinction to PM$_{2.5}$ Abundance ................................................................. 5
Key Research Accomplishments and Reportable Outcome ......................................................... 9
References..................................................................................................................................... 10
**Introduction**

Numerous studies have associated acute and chronic exposures to high levels of particulate matter (PM$_{10/2.5}$) with health outcomes such as increased hospital admissions, increased respiratory/circulatory symptoms, and decreased lung functions. These exposures, which come from a variety of sources such as blowing sand and dust, smoke, vapors, and aerosols, are common in many areas throughout the globe where U.S. military personnel are deployed in support of our national defense.

Addressing such health concerns requires an accurate assessment of the small-aerosol-particle concentration near the ground—for example, the issue of air pollution and its effects on health. Various networks of ground-based sensors provide routine measurements of PM$_{2.5}$, but their spatial coverage is rather sparse, especially in Third World countries. On the other hand, multiple space-borne sensors measure the total aerosol optical depth (AOD), often with nearly global daily coverage. However, while we have good global coverage for AOD, it is often not the best proxy for the near-surface aerosol concentration. This can be for a variety of reasons, for example, aerosols can be transported at high altitudes without being present near the land surface, while still contributing to a high total AOD. While there is a growing interest in using satellite data, there is the issue that the currently available satellite data products do not provide accurate data on the near-surface PM$_{2.5}$ abundance that we need for health studies (Hoff and Christopher 2009).

We propose to break through these limitations by bringing together data from multiple sensors and new machine learning methodology. We will also be using new methodology that has recently won recognition as a NASA Aura mission science highlight and the 2010 IEEE Geoscience and Remote Sensing Society Letters Prize Paper Award. This methodology takes into account the cardinally nonlinear relationship between the near-surface abundance of PM$_{2.5}$ and AOD, which is a function of the boundary-layer height, humidity, surface pressure, surface wind speed, and surface type. This is a major achievement, as, although we have numerous observations, we do not yet have a complete theoretical understanding of this cardinally nonlinear relationship between the near surface abundance of PM$_{2.5}$ and AOD.

**Hypothesis**

The hypothesis of this proposal is that a suite of remote sensing data products on atmospheric aerosols used in their meteorological context and processed by machine learning can provide a daily estimate of the global PM$_{2.5}$ abundance. This information is of considerable value to Global Health Surveillance (GHS), providing a capability to routinely estimate troop deployment exposure to elevated levels of particulate matter (PM) globally, significantly contributing to DoD-wide force health protection initiatives.

**Technical Objectives**

The goal of this study is to provide a quantitative understanding of the intrinsically nonlinear, multivariate relationship between the abundance of PM$_{2.5}$ in the atmospheric boundary layer and Remotely Sensed Aerosol Optical Depth (AOD) and extinction products. This is encapsulated in a software system that is capable of routinely providing a global data product for DoD health applications. As this project nears completion we see that we the basis for an operational system to serve the DoD health system. This can provide global coverage and therefore has the potential to provide Global Health Surveillance (GHS) with a capability to routinely estimate troop deployment exposure to elevated levels of PM globally, significantly contributing to DoD-wide force health protection initiatives.

Realizing our goal required two components. The first is to use the appropriate temporally and spatially varying meteorological context of the latest version of each satellite product, as well as in-situ ground truth observations of PM$_{2.5}$ abundance. The precise context of observations is critically important, as there is significant temporal and spatial variability in the abundance of PM$_{2.5}$, so careful attention must be paid to ingesting/fusing the satellite observations at both the appropriate time and place.

The second required component uses nonlinear, nonparametric, multivariate machine learning to address the issues for which we do not yet have a complete theoretical description encapsulated in our Numerical Weather Prediction (NWP) models. It would obviously be ideal if we had a complete theoretical understanding of the multivariate, nonlinear relationship between PM$_{2.5}$ and AOD, in which case we would gladly dispense with the machine learning.
However, as this most desirable state currently eludes us, the array of tools we have for multivariate, nonlinear, nonparametric machine learning has proved invaluable to a wide variety of applications and has already won significant recognition within NASA. In this study the nonlinear, multivariate issue that we dealt with is the multivariate, nonlinear dependence of the abundance of PM$_{2.5}$ in the atmospheric boundary layer on AOD, humidity, temperature, boundary-layer height, surface pressure, wind speed, and surface type. As mentioned earlier, our previous work in this area has won wide recognition as ground breaking.

**Method**

NASA has a constellation of satellites flying in close formation called the “A-Train” (Figure 1). Several of these satellites host instruments that make a variety of aerosol observations. These instruments include Terra MODIS (Remer et al. 2005) and MISR (Kahn et al. 2005), launched in December, 1999; Aqua MODIS, launched in May, 2002; Aura OMI (Torres et al. 2007), launched in July, 2004; and CALIPSO CALIOP (Mcgill et al. 2007; Winker et al. 2007), launched in April, 2006. We also have aerosol observations from SeaWIFS (Hooker and McClain 2000), launched in August, 1997, on GeoEye's OrbView-2 satellite.

The aerosol optical depth (AOD), $\tau$, is a measure of the light extinction at a given wavelength by atmospheric aerosols, in a vertical column from the earth’s surface up to the top of the atmosphere. Several of the A-Train instruments provide a daily global picture of the total aerosol optical depth. For example, MODIS provides the total AOD across its swath at a resolution of 10 km; the SeaWIFS resolution is 1.1 km. A new MODIS product at 3 km resolution should soon be available. The 3 km product introduces more noise but does capture fine (more urban scale) aerosol structure that is missed by the 10 km product. MODIS, OMI, and SeaWIFS provide the total global aerosol burden but not how it is distributed vertically, whereas other instruments provide detailed vertical aerosol structure but do not provide the contiguous global coverage of MODIS, OMI, and SeaWIFS. For instance, while CALIPSO provides corrected backscatter and extinction profiles at a 120 m vertical resolution, at altitudes below 20 km it does not provide contiguous horizontal coverage. MISR also provides some vertical information for cases with higher optical depths and distinct plume boundaries but at a coarser resolution than CALIPSO. The CALIPSO observations provide a set of high vertical resolution “curtains” underneath the satellite flight path. The CALIPSO curtains span the globe daily; however, there are substantial gaps between these curtains. Since CALIPSO completes 14.55 orbits per day, at the equator there is a separation of 24.7° in longitude between each successive curtain.

**Relating Aerosol Extinction to PM$_{2.5}$ Abundance**

The relationship between the PM$_{2.5}$ abundance at the earth’s surface and the boundary layer optical depth or aerosol extinction depends on a variety of factors that change both seasonally and geographically. These factors include the humidity, temperature, boundary-layer height, surface pressure, wind speed, and surface type (Liu et al. 2004a; Liu et al. 2004b; Hutchison et al. 2005; Gupta et al. 2006; Koelemeijer et al. 2006; Liu et al. 2007a; Liu et al. 2007b; Liu et al. 2007c; Pelletier et al. 2007; Gupta and Christopher 2008; Hutchison et al. 2008; Zhang et al. 2009).

When using a multi-linear analysis of the relationship between the AOD observed by MODIS and PM$_{2.5}$ it is found that better correlations are observed principally over the eastern United States in summer and fall (Zhang et al. 2009). The southeastern United States has the highest correlation coefficients, at more than 0.6. The southwestern United States has the lowest correlation coefficient, at approximately 0.2. Several factors are at work here. One is that the entire aerosol loading does not usually reside in the boundary layer; hence, using AOD alone as a proxy for PM$_{2.5}$ will invariably result in significant error. For example, on the West Coast, a significant fraction of the AOD is due to smoke events where substantial amounts of aerosol are above the boundary layer. Additional reasons for the poor correlation in the southwest may be associated with the humidity and land surface type. In addition, the correlation depends on the version of the satellite retrieval. For example, MODIS v5.2.6 AOD retrievals demonstrate better correlation with PM$_{2.5}$ than v4.0.1 retrievals, but they have much less coverage because of the differences in the cloud-screening algorithm (Zhang et al. 2009). We address these issues by using a fully non-linear, multivariate, non-parametric machine learning approach.

(Gupta et al. 2006) found that correlation between AOD and PM$_{2.5}$ increases as the mixing-layer height decreases. Larger wind speed can induce high mixing-layer height, which can change the correlation between AOD and PM$_{2.5}$. The relative humidity (RH) can affect the AOD-PM$_{2.5}$ by altering the optical properties of the aerosols. The higher the relative humidity, the larger the portion of light that is scattered, hence the larger AOD (Hoff and Christopher
We address this issue in this study by using the humidity and boundary-layer height contemporaneously with each observation used. The humidity comes from the meteorological analyses. The boundary-layer height also derives from the meteorological analyses and can be verified with the available LIDAR data. The meteorological analyses we use are the NASA Modern Era Retrospective Analysis for Research and Applications (MERRA) analyses produced by the Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (GMAO). The correlation between AOD and PM$_{2.5}$ is also related to the surface pressure and wind speed (Smirnov et al. 1995; Lyamani et al. 2006; Choi et al. 2008; Rajeev et al. 2008). We address this issue by using the surface pressure and wind speed contemporaneously with each observation. The surface pressure and wind speed also come from the meteorological analyses.

Table 1. Training dataset statistics and global 2000-2012 correlation coefficients.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>R</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqua Deep Blue</td>
<td>8,233</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Aqua Standard</td>
<td>30,298</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Terra Deep Blue</td>
<td>4,011</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Terra Standard</td>
<td>19,718</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Figure 1 Validation scatter diagrams showing the performance of the machine-learning algorithm for the two MODIS sensors using the standard and Deep Blue algorithms. In each case the x-axis shows the observed abundance of PM$_{2.5}$ ($\mu$g/cm$^3$) as observed by in-situ instruments. The y-axis shows the abundance of PM$_{2.5}$ ($\mu$g/cm$^3$) estimated by the machine learning based on the satellite and meteorological data products.
Several studies have sought to overcome this limitation by using satellite-derived Aerosol Optical Depth (AOD) with regression and/or numerical models to estimate ground-level PM$_{2.5}$ within the Earth’s boundary layer. Zhang et al. (2009) presented a comprehensive study for the 10 EPA regions across the United States using multi-linear regression between the PM$_{2.5}$ abundance observed by the EPA and the Moderate Resolution Imaging Spectroradiometer (MODIS) AOD and a set of meteorological parameters. The best correlations of PM$_{2.5}$ with AOD were observed for the eastern states in summer and fall, with EPA region 4 having a correlation coefficient of more than 0.6. The poorest correlations were observed for the southwestern states, with EPA region 9 having a correlation coefficient of approximately 0.2. Weber et al. (2010) extended the study of Zhang et al. (2009) for five EPA monitoring sites in the Baltimore/Washington DC Metro area by considering AOD from MODIS, the Multi-Angle Imaging Spectroradiometer (MISR), and the Geostationary Operational Environmental Satellite (GOES). The PM2.5 estimates of Zhang et al. (2011) and Weber et al. (2010) are made available through the Infusing satellite Data into Environmental Applications (IDEA) website (http://www.star.nesdis.noaa.gov/smcd/spb/aq/).

In an elegant study Van Donkelaar et al. (2006) presented a global estimate of the long-term average PM$_{2.5}$ concentrations between 2001-2006 using both satellite observations of AOD from MODIS and a global chemical transport model to estimate $\eta=$PM$_{2.5}$/AOD. The 3D chemical transport model used was GEOS-Chem. Van Donkelaar et al. (2006) found significant spatial agreement with North American PM$_{2.5}$ measurements (correlation coefficient of 0.77) and with non-coincident measurements elsewhere (correlation coefficient of 0.83).

In this study we have used a proprietary machine learning approach to estimate $\eta=$PM$_{2.5}$/AOD entirely from observations. We used PM$_{2.5}$ observations from the United States, Europe, Africa, Australia and Asia to create a comprehensive training dataset spanning more than a decade. We then used this training dataset to estimate $\eta$ as a function of the satellite AOD at multiple wavelengths and all the associated parameters that are available with the AOD (such as the angstrom exponent, scattering angle, cloud masks, surface reflectivity, and viewing geometry) and the meteorological analyses. Fifty independent trainings were performed using this training dataset, for each of these fifty trainings there was a random selection of 66% of the data for use in the training, with 34% of the data left out. The statistics shown in Table 1 and Figure 1 is the mean solution for these fifty independent trainings. Very careful attention is paid to ensure that the PM$_{2.5}$ observations and satellite observations are coincident in space and time to within a great circle separation of 0.02˚ (approximately 2 km) and a time window of 30 minutes. This is done for the standard and Deep Blue retrieval algorithms of MODIS Terra and Aqua. This can be thought of as the global fully non-linear multivariate extension to the pioneering work of (Zhang et al., 2009).

The results of this comprehensive training are shown in the table below. The performance of the approach we have used here is substantially better than that of the previous studies. Our worst performance has a correlation coefficient of 0.85, which is better than the best performance of the previous studies 0.83 for the non-coincident measurements of Van Donkelaar et al. (2006). It should also be noted that our values are global, so include the west coast of the United States which, as mentioned above, is typically more challenging to reproduce (Zhang et al., 2009).

As can be seen from Table 1 and Figure 1, we successfully used machine learning to describe the multivariate relationship between PM2.5 and a suite of parameters including AOD. Example PM$_{2.5}$ distributions are shown in Figure 2. These daily distributions can be used to provide the time evolution of PM$_{2.5}$ exposure for individual personnel (e.g. Figure 3).
Figure 2. Example distributions of PM$_{2.5}$ (µg/m$^3$) for March 1, 2012.
Figure 3. The Time evolution of PM$_{2.5}$ can be provided for more than a decade. The example above is for Taiyuan, Shanxi province, China.

**Key Research Accomplishments and Reportable Outcome**

The key accomplishment of this study has been successfully using machine learning to provide daily global analyses of PM$_{2.5}$ from March 2000 up until the present. This is twice the length of the period we promised in the proposal. The fidelity of this analysis (as can be seen from Figure 1 and table 1) is significantly better than that of previous studies (Van Donkelaar et al., 2006, Zhang et al., 2009, Zhang et al., 2011). These PM2.5 analyses are of considerable value to Global Health Surveillance (GHS), providing a capability to routinely estimate troop deployment exposure to elevated levels of particulate matter (PM) globally, significantly contributing to DoD-wide force health protection initiatives.
References


