Accurate ML Depth and Rough Particle Size Information Allow Sequential Daily Maps (“Near-Data”) of PM$_{2.5}$ Episodes

Robert B. Chatfield, Robert F. Esswein
NASA Ames Research Center
Robert.B_CHATFIELD@nasa.gov
Unusual composition is also a problem:

Varying mixtures of wood-burning smoke, ammonium nitrate, and occasional sea salt, little sulfate...

These are rarely included satisfactorily in regional or global models, and progress on mechanisms may be slow.

Thanks and acknowledgements to:
Lyapustin, Wang at GSFC, M. Shook, LaRC, Anderson, Beyersdorf, Zembia aerosol tea
the Ferrare lidar team at LaRC, the DISCOVER-AQ and Calif ARB teams,
the GEO-CAPE science definition team aerosol group,
NOAA Rapid Refresh (RAP) meteorological data stream
Thanks for research support by DISCOVER-AQ
Idea: Gain understanding, using the special profile information from PM monitors, aircraft and added ground instrumentation with MODIS;

Create algorithm that allows broader scope e.g., the MODIS Aqua data record and other thin wintertime ML's
Notice in this one wintertime period:

- Episodes last from 3 to 6 days
- Two major episodes are very different: Note MAIAC to PM2.5 to PM10 ratios
- Episodes of PM2.5 occur when MAIAC can make observations (little cloud cover) or sometimes one day after (over-riding high clouds)
A "Mixed Effects" Regression allowing variation from day to day helps greatly

\[ PM2.5_{is} = a \cdot \frac{(\tau_{Scat})_{is}}{(\Delta z_{Mixed \ Layer})_{is}} + b_i + c + \varepsilon_{is} \ldots (+ d_s) \quad \text{instance, Station} \]

i.e. ...

\[ (\tau_{Scat})_{is} \approx \frac{PM2.5_{is} - b_i - c}{a} (\Delta z_{Mixed \ Layer})_{is} \]

A mixed effects regression solves a larger linear algebra problem, and gives the same results as a possibly more familiar regression with "dummy variables" representing measurement days ...

\[ PM2.5_s = a \cdot \frac{(\tau_{Scat})_{is}}{(\Delta z_{Mixed \ Layer})_{is}} + b_1 x_{i=1} + b_2 x_{i=2} + \ldots + b_{n\text{day}} x_{i=n\text{day}} + \varepsilon_s , \]

\[ x_{i=1} \text{ takes the value 1 when } i \text{ is instance (day) 1, etc.} \]

... but with better statistical information and the possibility of broader formulation.

Note: Using the regression \( PM2.5 = \ldots \) gives information about the standard error of PM2.5 and some separation of the scattering and mixed-layer depth effects
More recent results are in fact better, but not completely graphed up:

We have expanded to include the entire state monitoring network, after preliminary data-quality review.

Sacramento and Davis may be in sufficiently different air-basin so as to deserve separate study. PM2.5 is mostly lower.

Expanding to these will give a better understanding of composition and other variations in the Valley

We are conducting a cross-validation experiment, looking at varying estimation data-sets and complete evaluation data-sets

Standard errors of estimates run up to \( \sim 8 - 9 \, \mu g\, m^{-3} \) and \( R > 0.87, R^2 = 0.75 \)
\[ \tau_{\text{Scat}} = [\rho_{\text{aer}} k_{\text{Scat}}](RH) C_{\text{StdCond}} \Delta z_{\text{Mixed Layer}}^{PM2.5} \]

We will assume that the quantity reported by MAIAC, based on much more complex optical paths and radiation theory, can fit this simple relationship. Similarly, PM2.5 is the dry mass of aerosol per standard cubic meter, and we assume that the EPA reference method PM2.5 reflects this appropriately.

**Simple Regression using ML depth**

\[ PM2.5_i = a \cdot \frac{(\tau_{\text{Scat}})_i}{(\Delta z_{\text{Mixed Layer}})_i} + c + \epsilon_i \quad \text{Instance, Station} \]

**Simplest regression, no compositional variations considered**

Simple multiple regression of PM2.5 vs the ratio (MAIAC aerosol optical depth) / (Mixed layer depth from RAP) gives a useful if scattered relationship.
Qian Tan is looking into differences of composition and size distribution from the DISCOVER-AQ region, and other sources.

Mixed-effects (a.k.a. random effects) regression allowing intercepts to vary with day, $i$, (instance) improves variance explained and standard error of PM2.5 estimate greatly. These are used for the maps shown on the right.

What does this mean?
- RH? ✗
- MAIAC difficulties? ✗?
- Dependence of relation on composition or size distribution?

**14 µg m$^{-3}$**
Colors: PM2.5 stations S (red) to N (blue)
Residual fitting error = 6.1 \mu g \text{ m}^{-3} \quad R^2 = 0.82
Random effects estimated by Day of Year but not Site

Ångström Exponent low this day

Ångström Exponent high this day
First SJq PM2.5 Episode, 2013!
Selected Days, Others are Available

Prior to Episode:
Generally low PM2.5 through SJq Valley; air evacuated through Delta and Bay

I-5 aerosol accentuated by flow to NNW, parallel to freeway

Bakersfield:
NNW flow weak; air recirculates
Prior to Episode:
Generally low PM2.5 through SJq Valley; air evacuated through Delta and Bay

I-5 aerosol accentuated by flow to NNW, parallel to freeway

Bakersfield: NNW flow weak; air recirculates
19th Episode builds: Weak flow towards NNW through SJq Valley, ... accentuating I-5 and South-SJq PM2.5 sources
21st Episode peaks Generally high PM2.5 through Western SJq Valley; low winds, slow spilling of PM2.5 towards Bay and towards South
27th, Strong Final Outflow:  
23rd – 26th, mixed patterns with clouds (no MODIS) as episode ends. On the 27th is the effect of a final cold frontal airmass (no MODIS, of course).
Feb 4th:
New episode begins in SE Valley

PM10
PM2.5
MAIAC (scaled)
Promising study areas

San Joaquin Valley
Eastern Los Angeles / Riverside
Salt Lake City
Doña Ana County (El Paso)
?? Iowa / Illinois hog farm area ??

Figure 11. Annual average and 24-hour (98th percentile of 24-hour concentrations) \( PM_{2.5} \) concentrations in \( \mu g/m^3 \), 2010.
Conclusions:

• The use of RAP mixed-layer heights along with MAIAC AOT provided detailed maps of the origins and transport of PM2.5 for two episodes in the SJq Valley, a region with specialized meteorology, aerosol composition, thin mixed layers, and severe aerosol-caused health effects. Earlier successes of mixed effects models without explicit ML depth information is understandable: these capture ML effects at the PM2.5 station (only).

• There are indications that size-distribution changes affect the AOT-to-PM2.5 relationship strongly, with large particles affecting AOT less (low Ångström exponent).

• Size distributions vary most at large scales, but N Valley and S Valley (or E Valley and W Valley near Fresno) may show important variations on certain days. This cannot be captured with current mixed-effects models. Four sunphotometers would likely capture this variation. Stations which report PM10 and PM2.5 concentrations may allow similar description, but current analysis is inconclusive. There should be ways to capture this size effect from ground measurements, or e.g., the forthcoming MAIA satellite instrument (D. Diner, PI), or from PARASOL-type measurements, or O₂ A-band / CO₂ band analyses (according to D. Crisp).

• The episodes range from 3 to 12 days with varied structure, and so these maps go far beyond other aerosol retrievals for the region in suggesting source-receptor relationships. SPARTAN photometer/filter networks should sample 3-day periods in the San Joaquin Valley.

• Insights from DISCOVER-AQ allows remote sensing studies to connect the varying chemistry and microphysics of few-day events. Further, careful, analysis of the DISCOVER-AQ data and also the permanent state-wide permanent PM2.5 network are indicated.

• For other regions, other meteorological models may be more appropriate. For the East Coast and the South, complicating effects of over-riding layers aloft will require careful synoptic typing. Above-ML layers remain a major problem.

• Inferences about sources from this “near-data” should provide source-driven 3-d modeling.
Conclusions:

- The use of RAP mixed-layer heights along with MAIAC AOT provided detailed daily maps of the origins and transport of PM2.5 for two episodes in the S Jq Valley, a region with specialized meteorology, aerosol composition, thin mixed layers, and severe aerosol-caused health effects. Successes of mixed effects models without explicit ML depth information is understandable: these capture ML effects at the PM2.5 station but extension to maps is more uncertain. These can guide forward models (with sources, transport).

- There are indications that size-distribution changes affect the AOT-to-PM2.5 relationship strongly, with large particles affecting AOT less (low Ångström exponent).

- Size distribution effects vary most at large scales, but N Valley and S Valley (or E Valley and W Valley near Fresno) may show important variations on certain days. This cannot be captured with current mixed-effects models. Four sunphotometers would likely capture this variation. Stations which report PM10 and PM2.5 concentrations may allow similar description, but current PM analysis is inconclusive. There should be ways to capture this size effect from ground measurements, or e.g., the forthcoming MAIA satellite instrument (D. Diner, PI), or from PARASOL-type measurements, or O₂ A-band / CO₂-band analyses (according to D. Crisp).

- The episodes range from 3 to 12 days with varied structure, and so these maps go far beyond other aerosol retrievals for the region in suggesting source-receptor relationships. Note: SPARTAN photometer/filter networks should sample 3-day periods in the San Joaquin Valley. MAIAC precision is not limiting?

- Insights from DISCOVER-AQ allow remote sensing studies to connect the varying chemistry and microphysics of few-day events. Further, careful, analysis of the DISCOVER-AQ data and also the permanent state-wide permanent PM2.5 network are indicated.

- For other regions, other meteorological models may be more appropriate. For the East Coast and the South, complicating effects of over-riding layers aloft will require careful synoptic typing. Above-ML layers remain a major problem. => SOREK-HAMER talk, Tuesday

- Inferences about sources from this “near-data” should provide source-driven 3-d modeling for subsident winter conditions.
DISCOVER-AQ data showed that scattering aerosol was almost always in the mixed layer in California, unlike many East Coast situations described by Alan Chu.

The P-3 aircraft made repeated profiles of $b_{scat}$ under standard-RH and ambient conditions. These indicated that light-affecting aerosol was nearly always confined to the ML and typically rather well-mixed. (Anderson group). R Ferare has reported that the lidar data from the King Air supports these ML depths and that their data gives good estimates related to PM2.5 under certain conditions.
If multiple AERONET sun-photometers may be used, retrieval of PM2.5 benefits from using the Fine Mode Fraction. The San Joaquin has had 1 permanent sun-photometer at Fresno.
The MAIAC dataset for the Aqua time period, since 2005, can provide an initial estimate of PM2.5 in Important wintertime situations to kick-start MAIA analyses.

Diner et al., the MAIA EV-I instrument