**Associations between optical, physical and chemical properties of aerosols measured at ground-based networks**

Draft report: Meytar Sorek-Hamer, Meredith Franklin

July 13, 2019

**1. Introduction**

A large body of literature has shown that satellite-derived aerosol optical depth (AOD), typically retrieved at 550 nm wavelength, reliably correlates with mass-volume concentrations of fine mode particulate matter with aerodynamic diameter less than 2.5 um (PM2.5). Studies that have used satellite observations to generate PM2.5 have been instrumental for air pollution - health effects research.

The associations between AOD and different chemical components of PM2.5 are lesser known. A handful of studies using the Multiangle Imaging SpectroRadiometer (MISR), an instrument onboard the NASA Terra satellite that provides observations of optical properties by particle type (size, shape, absorption), have provided evidence that different optical properties relate to different physical and chemical properties of particulate matter. Results have been somewhat inconsistent, showing differences depending on geographic area of analysis, optical components used, and statistical tools applied.

The purpose of this analysis is to make a detailed examination of the statistical relationships between ground-level PM2.5 and PM2.5 chemical components (nitrate, sulfate, elemental carbon, organic carbon, dust) and optical measures of aerosols (e.g. aerosol optical depth, angstrom exponent). We use

**2. Methods**

The study encompasses the San Joaquin Valley region of central California (Figure 1). We leverage datasets available at four sites - Bakersfield, Fresno, Modesto, and Visalia. At these sites there are co-located instruments from EPA’s chemical speciation network (CSN), EPA’s air quality system (AQS) and NASA’s AERONET network.

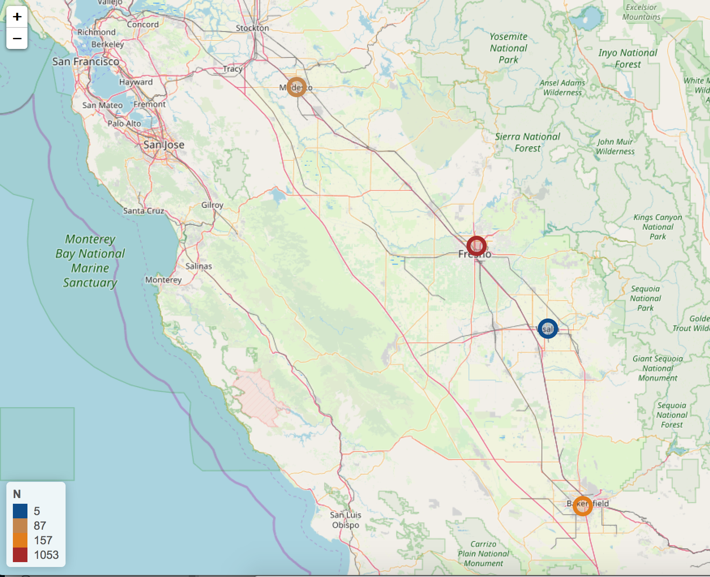


Figure 1. Map showing study area and data availability at four sites with co-located EPA PM2.5, PM2.5 speciation, and AERONET aerosol data.

Data

The CSN monitors are on a 1 in 3 or 1 in 6 day sampling schedule, providing PM2.5 mass and component PM2.5 concentrations of metals (e.g. Aluminium Al, Silicon Si, Calcium Ca, Titanium Ti, Iron Fe) obtained from X-ray fluorescence (XRF), ions (nitrate NO3- and sulfate SO42-) from ion chromatography, and carbons (organic OC and elemental EC) from thermal/optical analysis. To quantify dust we use the following equation:

dust = 2.2Al + 2.49Si + 1.63Ca + 1.94Ti + 2.42Fe

The AQS monitors provide daily concentrations of PM2.5 mass by the EPA’s Federal Reference Method, which is the highest quality gravimetric measurement method used for regulatory purposes.

AERONET sites are sunphotometers providing a “ground-up” measurement of aerosol optical properties at multiple wavelengths and have been used extensively to validate “top-down” satellite observations of related properties. Wavelength-specific AOD and angstrom exponents are the primary sunphotometer variables. Using quadratic log-log interpolation we calculated AOD 550 nm from AOD 440, 500, 675, 870 nm in log-log space. AOD at 550 nm is the most common wavelength retrieved from satellite instruments.

A retrieval-based AERONET product, called the inversion product, provides an additional suite of aerosol properties that help distinguish size (fine, coarse effective radius), shape (asymmetry), and absorption. We excluded sunphotometer and inversion variables that had a significant proportion of missing data (~90% missing). A list of the variables included in the analysis are shown in the Appendix.

In a separate test we examine data from the SPARTAN site in Rehovot, Israel. The SPARTAN network provides data for PM2.5 mass and speciation concentrations on an integrated 1 in 9 day sampling schedule, and is colocated with an AERONET site (We don’t have the speciation data for this site so we could only look at PM2.5 for now).

Statistical Methods

Prior to model building we examined a cluster-based correlation heat map (Figure 2), which provides the Pearson correlations between all pairs of AERONET variables grouped by a decision tree. To avoid collinearity in the regression models, we kept the most relevant of a group of variables that had a correlation coefficient > |0.9|. We then examined and picked a subset of variables connected at the mid-tier level of the tree to construct interactions.

We fit simple linear regression models separately for PM2.5 mass, sulfate, nitrate, EC, OC, and dust with AOD 550 nm as the sole predictor variable. Multiple linear regression models were again fit separately for PM2.5 mass, sulfate, nitrate, EC, OC, and dust, but with the combined AERONET sunphotometer and inversion product as predictor variables and model selection was conducted using the “all possible subset method”. This method constructs models based on all combinations from 1 to k variable models. We select the best model from the combinations based on highest R2, lowest RMSE, and Mallow’s Cp statistic that is close to k+1.

Model selection for the Fresno and Bakersfield sites were examined separately and in combination in a “total CA” analysis, which combined data from Fresno, Bakersfield, Modesto, Visalia, and a special DRAGON campaign in late 2012-early 2013 over the region (8 co-located sites with PM2.5 mass).

All models were cross validated (CV) with 10-fold CV, and we report the CV R2 and RMSE. Models were fit in R using the leaps() library.

**3. Results**

The merged data consists of N=608 AERONET-PM2.5-PM2.5\_speciation at the Fresno site, N=68 at Bakersfield, N=63 at Modesto, and N=5 at Visalia. The date ranges and summary statistics are shown in Table 1, along with summary statistics.

Table 1. Summary statistics by site. Medians and min-max are shown (most distributions lognormal), all PM mass and component concentrations are in units of ug/m3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Bakersfield | Fresno | All SJV | Rehovot, Israel |
| Matched Dates | 4/22/2002 – 6/27/2017 | 11/17/2012- 8/20/2018 | 4/22/2002- 8/20/2018 |  |
| AOD 550 nm | 0.1 (0.05-0.32) | 0.1 (0.3-2.0) | 0.1 (0.02-2.0) |  |
| N CSN | 68 | 608 | 739 |  |
| PM2.5 | 11.8 (2.0 -48.3) | 10.6 (0.6 -106.5) | 10.3 (0.6 -106.5) |  |
| Sulfate | 1.2 | 1.2 | 1.3 (0.0-3.5) |  |
| Nitrate | 0.9 | 1.7 | 1.6 (0.0-41.0) |  |
| OC | 2.8 | 3.8 | 3.6 (0.5-13.9) |  |
| EC | 0.5 | 0.6 | 0.6 (0.02-0.7) |  |
| Dust | 1.5 | 0.8 | 0.8 (0.0-8.7) |  |

Concentrations of PM2.5 had similar medians across SJV sites but given the larger amount of data in Fresno, there is a lot more variability (0.6 -106.5 ug/m3). Nitrate concentrations are generally higher in Fresno (median 1.7 ug/m3), as are OC (median 3.8 ug/m3), while dust concentrations are higher in Bakersfield (median 1.5 ug/m3). Sulfate concentrations are similar across the region, as are EC.

Interpolated AERONET AOD 550 also similar median across sites, but like PM2.5 the range is a lot greater in Fresno due to the large amount of data at that site.

The AERONET variables correlated in meaningful clusters (Figure 2). For example, effective radii and asymmetry factors show high positive correlations within size grouping (total, coarse, fine). AOD at different wavelengths are also highly positively correlated with each other. Angstrom exponents (AE) at all wavelengths have correlations >0.9, and as such we selected to use only the 440-675 nm AE in model selection.

After filtering variables based on proportion of missing observations, correlation, and constructing a manageable set of interactions, we ran the all subset method on a maximum of 20 variables, but choosing the models based on the 10 or fewer most important variables (Table 2).

Summary of PM2.5 mass

The simple linear regression between AOD 550 and PM2.5 has an overall R2 = 0.27 (0.28 in Fresno, 0.25 in Bakersfield). The best model with the inversion product is for Bakersfield (R2 =0.70 and RMSE = 5.03 ug/m3) but showed consistency with R2=0.65 in Fresno (RMSE = 10.12 ug/m3) and R2=0.62, RMSE = 9.17 ug/m3 using all available data in the SJV region. There is evidence of spatial heterogeneity in the strength of the association, which appears to be dependent upon the magnitude and variance in concentrations. Concentrations in Fresno are on average 3 ug/m3 greater than Bakersfield and 30% more variable.

There is consistency in the inversion products selected for PM2.5. The most important variables include: **Effective Radius (Fine), Angstrom Exponent 440-675, and AOD Extinction (fine, multiple wavelengths). Note that none of the AOD sunphotometer values appear in the list of best predictors of PM2.5.**

The simple linear regression between AOD 550 and PM2.5 at the Rehovot SPARTAN had R2 = 0.66 and RMSE of 25.03ug/m3

Summary Sulfate:

The simple linear regression between AOD 550 and sulfate has an overall R2 = 0.27 with RMSE = 0.51 ug/m3. Using the inversion products the R2 increased to 0.35 (even higher in Bakersfield 0.45) and decreased the RMSE to 0.48 ug/m3

The most important variables across sites are: **AOD 380, 440 ,1020 nm, Asymmetry Factor (fine), and the interaction between Asymmetry Factor Fine with Effective Radius Fine**. Some Asymmetry factors in the coarse mode appear in the best overall SJV model.

Summary Nitrate:

The simple linear regression between AOD 550 and nitrate is better than that of PM2.5 or sulfate R2=0.33 (ranging from 0.32 to 0.38). The inversion products greatly improve the model, increaing the R2 to 0.71 (RMSE = 3.81 ug/m3).

The most important variables are similar to that of PM2.5, likely because nitrates dominate the mass in this region. However, asymmetry factors and the total and coarse sizes appear rather than strictly fine: **Effective Radius (Fine and Total), Angstrom Exponent 440-675, AOD Extinction (fine, larger wavelengths), Asymmetry Factor Coarse 675,1020 nm.**

Summary of EC:

The simple linear regression between AOD 550 and EC is very poor in across the board R2 = 0.03 overall, RMSE 0.55 ug/m3. The inversion products help improve the associations greatly, bringing the R2 above 0.5.

The most important variables are: **Asymmetry factors fine and coarse, effective radius coarse and Angstrom exponent 440-870 (longer wl than others).**

Summary of OC:

Like EC, the simple linear regression between AOD 550 and OC is very poor in across the board R2 = 0.09 overall, RMSE 2.2 ug/m3. The inversion products help improve the associations,but not as much as EC, bringing the R2 = 0.34.

The most important variables are: **Asymmetry factors coarse at multiple wavelengths, Angstrom exponent 500-870 (longer wl than others), and Effective radius coarse interaction with asymmetry factor.**

* Note for EC and OC the selected variables are a bit more variable than the other components.

Summary of Dust:

The simple linear regression between AOD 550 and dust is the worst of all components R2= 0.006, RMSE 0.76 ug/m3. With the inversion products it improves to R2=0.44 and the RMSE reduces to 0.56 ug/m3.

The most important variables are: **AOD extinction Coarse at multiple wavelengths, Asymmetry factors Fine and Coarse at multiple wavelengths, and interestingly the Bakersfield and overall model AODs at 440, 500, 1020 nm appear in the list of best predictors. Recall dust is far higher in Bakersfield than Fresno.**

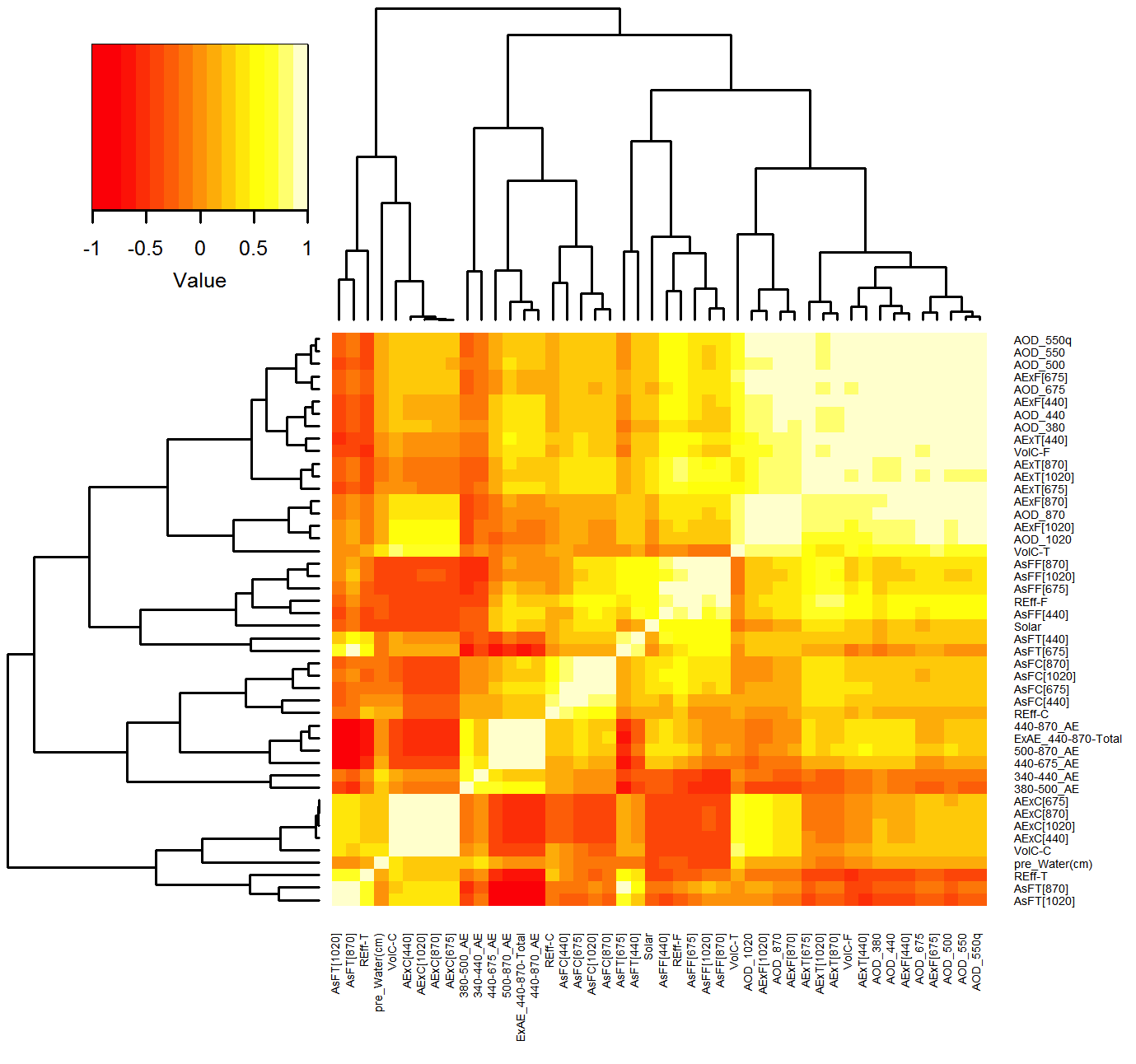


Figure 2. Correlation regression tree of AERONET sunphotometer and inversion products

Table 2. Summary statistics for the fitted regression models: univariate with AERONET log-log interpolated AOD 550 nm and multivariate with combined AERONET AOD (sunphotometer) and retrieved (inversion) products

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Site** | **No. obs** | **Response variable** | **Univariate Model (AOD550)**  **R2, RMSE (ug/m3)** | **Full Model**  **R2, RMSE (ug/m3)** | **Most contributing predictors (based on CV best fit)**  Best result out of 20, and not more than 10 (variables)  Interactions indicated by x between two variables |
| Fresno CSN (AQS) | 528  (3,107) | PM2.5 | R2=0.28  RMSE=13.8  (RMSE = 21.2) | R2= 0.65  RMSE=10.12 | `440-675\_Angstrom\_Exponent`  `AOD\_Extinction-Fine[675nm]`  `AOD\_Extinction-Fine[870nm]`  `AOD\_Extinction-Fine[1020nm]`  `Asymmetry\_Factor-Fine[440nm]`  `Asymmetry\_Factor-Coarse[870nm]`  `Asymmetry\_Factor-Coarse[1020nm]`  `REff-F`  `Asymmetry\_Factor-Coarse[675nm]`:`REff-C`  `440-675\_Angstrom\_Exponent`x`Asymmetry\_Factor-Fine[675nm] |
|  | 508 | Sulfate | R2=0.27  RMSE=0.51 | R2= 0.38  RMSE= 0.48 | AOD\_1020nm  AOD\_440nm  AOD\_380nm  `Asymmetry\_Factor-Fine[440nm]`  `Asymmetry\_Factor-Fine[675nm]`x`REff-F` |
|  | 508 | Nitrate | R2=0.37  RMSE=5.43 | R2=0.71  RMSE = 3.66 | `340-440\_Angstrom\_Exponent`  `AOD\_Extinction-Fine[870nm]`  `AOD\_Extinction-Fine[1020nm]`  `Asymmetry\_Factor-Coarse[870nm]` `Asymmetry\_Factor-Coarse[1020nm]`  `REff-T`  `REff-F` |
|  | 105 | OC | R2=0.09  RMSE=2.32 | R2=0.38  RMSE = 1.89 | `Asymmetry\_Factor-Coarse[440nm]`  `Asymmetry\_Factor-Coarse[675nm]`  `Asymmetry\_Factor-Coarse[870nm]`  `440-675\_Angstrom\_Exponent`x`Asymmetry\_Factor-Fine[675nm] |
|  | 105 | EC | R2=0.05  RMSE=0.58 | R2= 0.41  RMSE = 0.48 | AOD\_1020nm  AOD\_380nm  `440-870\_Angstrom\_Exponent`  `AOD\_Extinction-Coarse[675nm]`  `AOD\_Extinction-Coarse[870nm]`  `Asymmetry\_Factor-Coarse[440nm]`  `Asymmetry\_Factor-Coarse[675nm]`  `Asymmetry\_Factor-Coarse[870nm]`  `REff-F`  `Asymmetry\_Factor-Fine[675nm]`x`REff-F` |
|  | 520 | Dust | R2=0.003  RMSE=0.61 | R2=0.47  RMSE = 0.45 | `340-440\_Angstrom\_Exponent`  `AOD\_Extinction-Fine[1020nm]`  `AOD\_Extinction-Coarse[440nm]`  `AOD\_Extinction-Coarse[870nm]` `Asymmetry\_Factor-Fine[440nm]` `Asymmetry\_Factor-Fine[675nm]`  `Asymmetry\_Factor-Coarse[440nm]` `Asymmetry\_Factor-Coarse[870nm]`  `REff-C` |
|  |  |  |  |  |  |
| Bakersfield CSN (AQS) | 65  (250) | PM2.5 | R2=0.25  (R2 = 0.32)  RMSE=9.3  (RMSE = 23.2) | R2=0.70  RMSE = 5.03 | `440-870\_Angstrom\_Exponent`  `AOD\_Extinction-Coarse[440nm]`  `AOD\_Extinction-Coarse[675nm]`  `Asymmetry\_Factor-Fine[870nm]`  `Asymmetry\_Factor-Fine[1020nm]`  `REff-T`  `REff-F`  `440-675\_Angstrom\_Exponent`x`Asymmetry\_Factor-Coarse[675nm]`  `Asymmetry\_Factor-Fine[675nm]`x`REff-F` |
|  | 64 | Sulfate | R2=0.2  RMSE=0.58 | R2=0.45  RMSE = 0.54 | ‘AOD\_870nm’  `Asymmetry\_Factor-Coarse[440nm]` |
|  | 64 | Nitrate | R2=0.15  RMSE=4.94 | R2=0.54  RMSE = 1.39 | "`Asymmetry\_Factor-Fine[675nm]`"  "`REff-F`"  "`Asymmetry\_Factor-Fine[675nm]`x`REff-F`" |
|  | 49 | OC | R2=0.09  RMSE=1.60 | R2=0.36  RMSE = 1.3 | `Asymmetry\_Factor-Fine[440nm]`  `Asymmetry\_Factor-Coarse[675nm]`:`REff-C`  `440-675\_Angstrom\_Exponent`x`REff-F` |
|  | 49 | EC | R2=0.13  RMSE=0.46 | R2=0.50  RMSE = 0.29 | `Asymmetry\_Factor-Fine[1020nm]`  `440-675\_Angstrom\_Exponent`x`REff-C` |
|  | 62 | Dust | R2=0.03  RMSE=1.28 | R2=0.4  RMSE = 1.12 | AOD\_1020nm  AOD\_380nm  `AOD\_Extinction-Fine[870nm]`"  `AOD\_Extinction-Fine[1020nm]`"  `AOD\_Extinction-Coarse[1020nm]`" `Asymmetry\_Factor-Fine[440nm]`"  `Asymmetry\_Factor-Fine[675nm]`  `Asymmetry\_Factor-Fine[870nm]`  `Asymmetry\_Factor-Coarse[440nm]` |
| Total CA CSN (AQS) | 653 (3,892) | PM2.5 | R2= 0.27  RMSE=13.04  (RMSE = 20.68) | R2=0.62  RMSE = 9.17 | `440-675\_Angstrom\_Exponent`  `AOD\_Extinction-Fine[675nm]`  `AOD\_Extinction-Fine[870nm]`  `AOD\_Extinction-Fine[1020nm]`  `Asymmetry\_Factor-Fine[440nm]`  `Asymmetry\_Factor-Coarse[870nm]`  `Asymmetry\_Factor-Coarse[1020nm]`  `REff-T`  `REff-F`  `440-675\_Angstrom\_Exponent`x`Asymmetry\_Factor-Fine[675nm] |
|  | 622 | Sulfate | R2=0.27  RMSE=0.51 | R2= 0.35  RMSE=0.48 | AOD\_870nm  AOD\_440nm  AOD\_380nm  `340-440\_Angstrom\_Exponent`  `Asymmetry\_Factor-Fine[675nm]`  `Asymmetry\_Factor-Coarse[870nm]`  `Asymmetry\_Factor-Coarse[1020nm]`  `REff-F`  `Asymmetry\_Factor-Fine[675nm]`:`REff-F` |
|  | 620 | Nitrate | R2=0.33  RMSE=5.35 | R2= 0.71  RMSE=3.46 | `340-440\_Angstrom\_Exponent`  `AOD\_Extinction-Fine[870nm]`  `AOD\_Extinction-Fine[1020nm]`  `Asymmetry\_Factor-Coarse[675nm]`  `Asymmetry\_Factor-Coarse[1020nm]`  `REff-T`  `REff-F` |
|  | 154 | OC | R2=0.07  RMSE=2.2 | R2=0.34  RMSE=1.89 | `500-870\_Angstrom\_Exponent`  `Asymmetry\_Factor-Coarse[440nm]`  `Asymmetry\_Factor-Coarse[675nm]`:`REff-C` |
|  | 154 | EC | R2=0.03  RMSE=0.55 | R2=0.52  RMSE=0.4 | `440-870\_Angstrom\_Exponent`  `Asymmetry\_Factor-Fine[1020nm]`  `Asymmetry\_Factor-Coarse[440nm]` |
|  | 632 | Dust | R2=0.006  RMSE=0.76 | R2=0.44  RMSE = 0.56 | AOD\_1020nm  AOD\_500nm  AOD\_440nm  `340-440\_Angstrom\_Exponent`  `AOD\_Extinction-Coarse[440nm]`  `AOD\_Extinction-Coarse[870nm]`  `Asymmetry\_Factor-Fine[440nm]`  `Asymmetry\_Factor-Fine[870nm]`  `Asymmetry\_Factor-Fine[1020nm]`  `Asymmetry\_Factor-Coarse[440nm]` |
|  |  |  |  |  |  |
| Israel, Rehovot SPARTAN | 84 | PM2.5 | R2=0.66  RMSE=25.03 |  |  |
|  |  | Sulfate |  |  |  |
|  |  | Nitrate |  |  |  |
|  |  | OC |  |  |  |
|  |  | EC |  |  |  |
|  |  | Dust |  |  |  |

**Appendix**

List of Input variables (red from sunphotometer, black from inversion product):

AOD 550nm (quadratic log-log interpolation from AERONET AOD 440, 500, 675, 870nm)

AOD 675

AOD 500

AOD 440

AOD 380

AOD 800

CWV

Angstrom Exponent 440-675

Effective Radius - Total

Effective Radius - Coarse

Effective Radius - Fine

Asymmetry Factor Total 675nm

Asymmetry Factor Coarse 675nm

Asymmetry Factor Fine 675nm  
Asymmetry Factor Total 440nm

Asymmetry Factor Coarse 440nm

Asymmetry Factor Fine 440nm  
Asymmetry Factor Total 380nm

Asymmetry Factor Coarse 380nm

Asymmetry Factor Fine 380nm

Asymmetry Factor Total 870nm

Asymmetry Factor Coarse 870nm

Asymmetry Factor Fine 870nm

Asymmetry Factor Total 1020nm

Asymmetry Factor Coarse 1020nm

Asymmetry Factor Fine 1020nm  
Solar Zenith Angle

List of Interactions as Input variables:

Angstrom Exponent 440-675 : Effective Radius - Coarse

Effective Radius - Coarse : Asymmetry Factor Coarse 675nm

Angstrom Exponent 440-675 : Asymmetry Factor Coarse 675nm

Angstrom Exponent 440-675 : Effective Radius - Fine

Effective Radius - Fine : Asymmetry Factor Fine 675nm

Angstrom Exponent 440-675 : Asymmetry Factor Fine 675nm