10. Descriptive Data Analysis and Interpretation

Goals

A large quantity of data will be collected in support of Project MOHAVE. The descriptive data analysis and interpretation component of the study is intended to summarize the main features of the data as well as especially interesting cases, and offer physical explanations whenever possible. In contrast to the attribution analyses described in Section 11, this section will organize the data in a manner that will allow inference of effects from different sources, but will not generally be quantitative sufficiently to permit source apportionment.

Descriptive Statistics

Descriptive statistics will include calculation of means, standard deviations, skewness, and extreme values of the variables. In addition, time series of the data will be presented. These will include time series of the extinction coefficient (b_{ext}), tracer, sulfate, nitrate, organics, light-absorbing carbon, fine soil, and various trace elements and meteorological variables, for example. Correlations between variables will also be calculated.

Extinction Budget

Light extinction is caused by scattering and absorption by particles and gases. In general, particle scattering is the primary component of extinction, although in remote areas of the southwest, scattering by gases that compose the atmosphere (Rayleigh scattering) is a significant fraction on the clearest days. Black carbon (from diesel engines, forest fires, etc.) is the principal agent of particle absorption, and is occasionally an important contributor to haze in the study region. NO_2 is the only common gaseous pollutant that absorbs in the visible portion of the spectrum and is not likely to be a significant contributor to haze in GCNP.

The extinction budget analysis involves determining the contribution to extinction by all the major aerosol components. There are two fundamentally different approaches to estimate the extinction budget. A statistical approach uses multivariate analysis to explain the optical parameter (b_{ext} or b_{scat}) by a linear combination of the components. These components are the concentrations of the pollutant species (e.g., crustal, sulfate, nitrates, elemental and organic carbon, etc.) multiplied by best-fit determined coefficients interpreted as extinction efficiencies. The hygroscopic particle species (e.g., sulfate and nitrate) include a function of relative humidity to incorporate the effects of water upon the extinction efficiencies of these species.

An externally mixed aerosol (i.e. separate aerosol components are not contained within the same particles; for example, sulfate-coated crustal particles would not constitute an external aerosol mixture) is implicitly assumed by the statistical approach for extinction budget analysis. The extent to which this assumption is true, and the implication of it being violated, are hard to estimate in any individual situation. In general, the greatest impact of non-external mixtures is thought to be associated with an interpretation of how changes in aerosol composition would affect atmospheric optics. In other words, there is increased uncertainty associated with the prediction of how visibility will respond to changes in emission caused by violation of this assumption.

In addition to the concern about implied assumptions, any use of multivariate statistics carries with it the concerns caused by use of possibly highly covariant independent parameters, and the use of measured parameters with large differences in relative measurement uncertainty. Both of these concerns can result in biased results. However, there are standard approaches to detect and minimize the impacts of these concerns.

The other approach to estimating extinction efficiencies for the various aerosol components is by first principle calculations (Mie Theory). These calculations require as input, certain particulate characteristics such as the distribution of particle size, shape, and indices of refraction. Generally, the size distribution can be estimated from size segregating sampler measurements. For Project MOHAVE, this will be done with the DRUM impactors for some components, such as sulfur and crustal components, but not for others, such as organic carbon and nitrate species. A functional relationship between water and the hygroscopic particles must be assumed to estimate its effects on particle size. Particle shape is generally assumed to be spherical, and the refractive indices are assumed to be the same as the bulk indices for the various measured particle chemical components. Assumptions must also be made concerning the nature of the aerosol mixture (i.e., external, internal, or some combination) in order to calculate the extinction efficiencies of the components. The extent to which these deficiencies and assumptions affect the calculated extinction is unknown.

In spite of the uncertainties discussed above, extinction budget analysis done by the two approaches generally results in similar extinction efficiencies. Since Project MOHAVE will use both Mie Theory and statistical approaches, the results can be intercompared for consistency, and reconciled with extinction efficiency values from the literature to arrive at best estimates of the extinction budget.

Empirical Orthogonal Function Analysis

Empirical orthogonal function (EOF) and possibly other types of eigenvector analysis will be done to help summarize the data and gain insights into possible physical mechanisms at work. When working with large amounts of data, EOF analysis is especially useful by effectively reducing the dimensionality of the data set. A large number of observations at many locations can be reduced into a reasonable number of spatial patterns (eigenvectors), with a time series associated with each eigenvector showing the time variability of each pattern.

The EOF analysis is purely a statistical technique that attempts to account for most of the variability in a data set by a few eigenvectors. Although no physics is explicitly included in the analysis, the data set represented by the eigenvectors is certainly affected by physical processes. EOF analysis in conjunction with sound physical reasoning, including knowledge of meteorological conditions, location of emission sources, etc. can help in the formation of hypotheses and provide a qualitative check of receptor and deterministic modeling results.

The variables for which EOF analysis is likely to be done are sulfate, SO_2, tracer, elemental carbon, organic carbon, fine soil, and certain trace elements. EOF analysis of the vector wind field may be done as well. EOF analyses of the modeled output wind and concentration fields will be investigated as a method to help organize the large amount of model output. Additional EOF analysis using two (or more) parameters such as sulfate and tracer can be used to identify common jointly occurring patterns of more than one parameter.

Meteorological Classification

A meteorological classification scheme will be developed and applied to the study years and several previous years. The scheme will classify days into types on the basis of similarity of meteorological parameters. There are several reasons for doing a classification. One reason is to compare the frequency of each weather pattern during the study year with other years to determine how representative the study year is. Each pattern is likely to have transport of visibility affecting pollutants from different areas; the relative frequency of patterns for the study year compared to long-term averages can help put the impacts during the study year into perspective. It also provides a logical method of stratifying the data of the study year into a manageable number of patterns. Averaged spatial patterns of sulfate, etc. along with the variation within each pattern can reveal the main pathways for transport of both hazy and clear air into the study area. Contributions from individual source areas may also be inferred from the concentration fields associated with each pattern.

The meteorological classification scheme can aid in the interpretation of the EOF analyses. The time series of the EOF analyses indicate the times a particular eigenvector is significant. By determining the corresponding meteorological pattern most commonly associated with each eigenvector, it is easier to interpret the physical factors associated with the eigenvectors.

The classification scheme will also be used to study the MPP outage of June-December 1985. Sulfate concentration levels and spatial patterns associated with each weather pattern will be compared for the outage year and other years with SCENES data. This will help put bounds on the contribution of MPP to regional sulfate levels.

Of critical importance in the classification scheme are surface and upper air wind speed and direction, atmospheric moisture and thermal stratification. The wind data are necessary to account for the transport and dispersion properties of the flow. Moisture is necessary to determine the potential for aqueous phase oxidation of SO_2 to sulfate and washout. Thermal stratification is needed to know if pollutant emissions are likely to remain trapped in basins or are mixed through a deeper layer of the atmosphere.