ATMS 451 Winter 2013   Chemistry Data Analysis Worksheet

The purpose of this analysis is to allow you to apply concepts we have been discussing throughout the course to the study of atmospheric chemistry and to introduce you to some programming methods often used in the analysis of atmospheric observations. **You may want to use some of these methods in the analysis of your weather station data.**

The data set contains several chemical and physical parameters measured *in situ* from June – July, 1999 outside of Nashville, TN as part of a multi-investigator and multi-platform air quality and chemistry study. Specifically, these include observations of nitric oxide (NO), nitrogen dioxide (NO_2), ozone (O_3), NO_2 photolysis frequencies (J_{NO2}), and air temperature obtained at a ground-based site 8 km from downtown Nashville. Most observations are 1-minute averages of higher time resolution data. Gaps in the data set were caused by power outages or malfunctioning of one or more instruments. **You are only required to examine the NO_2 observations.**

You are encouraged to use Matlab throughout this process because of its flexibility and ability to handle large data sets relative to Excel. The data set is available both as a .mat (MATLAB) file and an ASCII text file. The latter can be loaded into other spreadsheet programs (such as Excel). MATLAB is accessible from computers in the facility on the 6th floor in the ATG building. You should plan to work there with your group members during the scheduled lab periods.

Download the data set from the Share Space

1. Distinguishing Instrumental Imprecision from Atmospheric Variability

The goal of this exercise is to assess whether minute-to-minute variability in the NO_2 data can be assigned to "natural" variations of the system (i.e., atmospheric variability) or to instrumental "noise".

1.1 Determine the frequency distribution of fractional point-to-point deviation in the NO_2 data, also known as the normalized adjacent difference. That is, calculate a vector such as

$$\frac{2\left[NO_2(i)-NO_2(i+1)\right]}{NO_2(i)+NO_2(i+1)} \approx \frac{NO_2(i)-NO_2(i+1)}{NO_2(i)} \tag{1.1}$$

Create a frequency distribution of this point-to-point variation with at least 500 bins. HINTS: In Matlab "diff(x)" produces an adjacent differences vector from x which is N-1 in length, where N is the number of x values. To achieve eqn. (1.1), divide the result by no2_ppb(1:end-1). [N, X] = hist(x,bins) produces a frequency distribution of x among the number of bins specified by "bins". "N" is the count in bin located at X. Show a well-labeled figure of the histogram.

1.2 Focusing on adjacent differences located within +/- 100%, produce a normal distribution that reasonably reproduces the observed distribution between fractional point-to-point differences of – 1 and 1. You might be able to use the "normfit" function in Matlab. Type "help normfit" for details. If not, start by using "std" and "mean" functions for the observed adjacent difference distribution to create test Gaussians via equation (1.2)

$$f = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \qquad (1.2)$$

where μ is the mean and σ is the standard deviation. The Gaussian approximation will be used below. Note Eqn. (1.2) is the normalized Gaussian such that the integral of *f* from negative to positive infinity is 1. You may want to normalize your distributions simply so that the maximum of the frequency distribution is at 1. This allows rapid comparison of the widths (assuming the tails are controlled by the same native function).

1.3 The operators of the laser-induced fluorescence instrument that was used to detect NO_2 claim that the instrumental precision is dominated by the random nature associated with counting rare (fluorescence) events. The statistics of such measurements are governed by the Poisson distribution for which the standard deviation, σ_x, about the mean number of fluorescent events, μ_x, in some time period, is equal to the square root of the mean:

$$\sigma_x = \sqrt{\mu_x} \qquad (1.3)$$

As the number of fluorescent events scales linearly with NO_2 mixing ratio, C_{NO2}, the relative uncertainty due to precision error scales as $1/\sqrt{C_{NO2}}$.

The operators state that for a constant NO_2 mixing ratio of 1 ppb, the 1-minute averages of measurements would have a standard deviation of ${}^{p}\sigma_{NO2}$ = 0.01ppb. The 'p' superscript is to distinguish this precision quantity from the *total* uncertainty discussed later.

Using this stated precision at 1 ppb NO_2, calculate a mean precision uncertainty for the NO_2 mixing ratios observed over the campaign.

1.4 Create another Gaussian that represents the point-to-point variation *expected solely from instrument imprecision.* Show a well-labeled plot of the two distributions on the same x-axis: 1) Gaussian approximation to the observed point-to-point variation and Gaussian approximation to variation expected by instrumental noise. Comment on the assumption that the observed point-to-point variation can be fit by a normal (Gaussian) distribution.

1.5 Given that each adjacent NO_2 observation is typically separated by 1 minute, what fraction of the observed minute-to-minute variation in NO_2 can be explained by instrument noise? What does this tell you about the importance of natural variability on these short timescales?