VSRT MEMO #062

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886

June 1, 2009

Telephone: 781-981-5407
Fax: 781-981-0590

To: VSRT Group
From: Alan E.E. Rogers and Vincent Fish
Subject: Suggested Ozone spectrometer projects for students

The data from 11 GHz Ozone spectrometers at

Chelmsford High School
Haystack Observatory Moran Building
Bridgewater State College, Bridgewater MA
Union College, Schenectady, NY
Univ of North Carolina At Greensboro

Is available at http://www.haystack.mit.edu/ozone and will shortly be available from

Southeast Missouri State University
Kent State University, Ohio

The Graphical User Interface (GUI) allows students to retrieve selected spectra from the radiometers. The CGI script allows the user to select sites UT date range, local time range and Sun elevation range. The student can make their own analysis of the spectra using a spreadsheet like Excel, a computer program like C, a scripting language like Python or an analysis package like Octave or Matlab.

The GUI also allows the user to fit a theoretical ozone spectral profile to the data and hence provide an estimate for the mesospheric ozone concentration as a function of date or local time. ASCII summary files and plots can be obtained. It is suggested that a student interested in gaining a detailed understanding of the instrument and the data could work directly with the spectra perhaps only using the GUI to provide suitably filtered sets of spectra. If the student prefers to start by examining the diurnal or seasonal variation of the ozone the student can use the online processor. The online processor may also be used to look for correlations of the mesospheric ozone with environmental conditions like cloud cover.

Suggested projects:

1] Statistical analysis of the spectra. Averaging spectra.

Procedure:

Select nighttime data by setting Sun elevation filter. Download at least 10 days of data from one or more sites. Average the spectra over different time spans to see how the noise in the spectra decreases with number of spectra averaged. Each spectrum from the download is itself a 10 minute average from each ozone spectrometer. See VSRT memo #056.

2] Statistical analysis of the spectra. Estimation of mesospheric ozone.

Procedure:

Select nighttime data by setting Sun elevation filter. Download 10 days of data from one or more sites. Average the spectra to obtain a single spectrum. Ideally, find a theoretical ozone spectrum that fits the measured spectrum "best" by minimizing, Q, the sum of the residuals squared.

$$Q = \sum_{i=0}^{N-1} \left(d_i - t_i \right)^2$$

where d_i is the i^{th} point of the measured spectrum

and t_i is the i^{th} point of the theoretical spectrum

The theoretical spectra is dependent on several parameters like the ozone concentration, temperature etc. These parameters can be estimated, along with their error with the aid of matrix algebra. Starting with a simple analysis the theoretical spectrum might be approximated by a fixed shape multiplied by a constant, c, representing the ozone concentration.

In this case

$$Q = \Sigma \left(d_i - cf_i \right)^2$$

where f_i is a fixed spectrum which could be simple as "pulse" of 3 points centered on the peak index

f_i = 1 for $i = p-1, p, p+1$ = 0 elsewhere.

The value of c which minimizes Q is given by

$$c = \sum d_i f_i \Big/ \sum f_i$$

For the simplified theoretical spectrum

$$c = \left(d_{p-1} + d_p + d_{p+1} \right) \Big/ 3$$

That is a rough approximation of the ozone concentration can be obtained by averaging the center point with one point on either side.

A better method is to fit a theoretical spectrum with a least 2 parameters: one to represent the ozone concentration and another to account for a constant instrumental bias in the measured spectra. The GUI analysis will fit a 3 parameter model of a constant plus separate ozone concentrations below and above 80 km.

Once you have "programmed" a method of estimating the ozone concentration using Excel or another analysis tool you can compare your estimate for the ozone concentration at night with the estimate of the data taken during the day. See VSRT memos #042 and 059.

3] Study the diurnal variation of mesospheric ozone.

Procedure:

Obtain the ozone concentration estimate for a range of local time using your own analysis or using the online analysis. The online tools will allow you to compare the results from each site separately or to average the data from all sites. In addition, the online tools can separately estimate the ozone concentration for a region below 80 km and a region above 80 km. Study the chemistry which affects the ozone in the mesosphere and at least qualitatively explain the day/night variation and the rates of change at sunrise and sunset.

4] Study the seasonal variation of mesospheric ozone.

Procedure:

Obtain the nighttime ozone concentration estimate for each day or several day average over a large time span using your own analysis or the online tools. The online tools will allow you to obtain the seasonal variation from each site or groups of sites. Study the chemistry of the creation and destruction of ozone in the mesosphere. Explain how the presence of water vapor in the mesosphere would affect the ozone. Search the web to find data on the seasonal variations in the environment of the mesosphere to compare with the seasonal variations seen by the ozone spectrometers.

5] Study the evidence for variations of the ozone estimates from the spectrometers.

Comments of motivation

While the diurnal variation of the mesospheric ozone is clear and there is no question that the mesospheric ozone increases dramatically at night, other variations and their cause are less clear. Seasonal variations look convincing but we don’t yet know how well these variations will repeat from year to year. Other potential causes of variation, like Earth tides which effect the temperature of the mesosphere, water vapor circulation in the mesosphere and the injection of protons from “proton events” are less clear. The ozone spectrometers are not perfectly calibrated or perfectly stable so there are significant instrumental errors.

Procedure:

Obtain ozone estimates from your own analysis or use the online tools. Start with 4 day averages from each site. Estimate the error in each 4 day average. Compare the results from separate sites. Are the difference in results between sites consistent within the errors? The ozone spectrometer sensitivity is reduced by water vapor in the beam which is mainly within a few km of the spectrometer. A more severe loss will occur when it is raining at the site of the spectrometer. The presence of rain will also produce a more significant increase in the “total power.” Compare the ozone estimates with total power to determine the significance of the instrumental error due to rain. Study the methods of determining the statistical significance of a hypothesis. For example, in order to determine if the ozone in the mesosphere is correlated with the cloud cover at night one might start by plotting the estimated ozone from the spectrometers as a function of time along with a measure of cloudiness below the point that the antenna beam intersects the mesopause. If the plot shows in an absolutely clear manner that the ozone follows the cloudiness by either increasing every time it is cloudy or decreasing every time it is cloudy then it is likely that the ozone in the mesosphere does depend in some way on the cloudiness. However, it is more likely that the plot may just show a possible correlation whose significance will need to be tested and even if a statistical test indicates a significant correlation it may be the result of instrumental errors. So testing a hypothesis based on "real world" data is not a simple task. How often have you heard that according to trials something has an effect on your health only to have the conclusion reversed at a later date because of a new trial or the discovery of some other influential factor? Also be aware that hypothesis testing is based on statistics so that nothing is completely certain. For example there is an absolutely negligible, but non zero, probability that a heavy object will levitate as a result of the thermal motion of its molecules because all its molecules move in the same direction. In practice this will never happen. There is a small, but distinct, possibility that the object may briefly levitate as a result of an earthquake, however. To look for a correlation we start by calculating the normalized correlation, ρ,

$$\rho = \sum_{i=0}^{i=N-1} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) \Big/ \left[\left(x_i - \overline{x}\right)^2 \sum \left(y_i - \overline{y}\right)^2\right]^{1/2}$$

where x_i are the values of ozone

y_i are the values of cloudiness

N is the number of data products

$\overline{x}$ and $\overline{y}$ are the averages of x_i and y_i respectively

and $-1 \le \rho \le 1$

The next step is to estimate the significance of any correlation from the probability that the value of ρ could have resulted from a null hypothesis. For reasonably large number of data values (> 20) the quantity:

$$t = |\rho| \Big/ \left[\left(1-\rho^2\right)\big/\left(N-2\right)\right]^{1/2}$$

follows "student's t" probability distribution. For N>20 there is less than 5% probability that t has a value which exceeds about 1.7. For example if we have an ozone and cloudiness value for each day of the year for an entire year, a 10% correlation corresponds to t=1.9 and a probability of less than 5%, so the correlation is not "real." That is, the apparent correlation may just be the result of noise in data.

Even if the correlation is large enough to be significant it may still be "indirect" or instrumental in origin. Another check on the significance can be made by generating values of x and y using a random number generator. Also one could shift the values of y in time and see if the correlation peaks at zero lag i.e., at no shift. More information on statistical significance tests can be found on the web.

Cloudiness was chosen just as an environmental factor that might influence the ozone in the mesosphere via a change in the Earth's albedo and corresponding change in the temperature of the mesosphere. Other environmental factors may have an influence are "Rossby Gravity" waves, Earth tides and space weather. Weather patterns show intra-seasonal variations and oscillations. Variations with a periodicity in the 30-60 day range are known as Madden-Julian Oscillations (MJO). MJO are an ongoing area of research. The effects of the MJO have been seen in the temperature and winds of the mesosphere and thus might be expected to influence the ozone in the mesosphere. It also suggested that the ozone data be correlated with sine and cosine functions of varying period to see if there is any statistically significant periodicity in the ozone concentration.

Notes on weather data:

There is limited free weather data available from the National Weather Service. The site http://data.nssl.noaa.gov allows one to obtain a list of various surface parameters. For example, selecting list data and entering @WAL and RELH, TEMPC gets the relative humidity and temperature for the meteorological station at Wallops Island VA. However, the site only allows one to get data for one day at a time and there is no "cloudiness index." A possible proxy for cloudiness is the relative humidity or the inverse of the temperature difference between day and night.

Weather data can be obtained for free from http://www.wunderground.com in comma-delimited data files. For any location (e.g., Westford Massachusetts), choose a Weather Station at the bottom of the page. Click on the station name (in the Station Location column), and a Daily Summary page will load. For a day-by-day summary of weather conditions, click on the Yearly link at the top of the Daily Summary panel, or select a custom time period by clicking on Custom. The data are presented in tabular form in the Tabular Data panel and can be obtained in a machine-readable format by clicking on the Comma Delimited File link. These data can be copied to a local file using cut and paste or saved via the web browser (but beware of possible html formatting in the latter case).