

MOSAIC Data Output Guide

<http://www.haystack.mit.edu/ozone>

Background information

Format of avout.txt files

When the data are downloaded in the form of an avout.txt file, the column headings are located in the column preceding the data.

2 sets of units

Depending on the data requested, the quantity used to characterize mesospheric ozone intensity will vary. In spectral plots, the data plotted are antennae temperature, measured in milliKelvin (mK). This is a unit used by radio astronomers to characterize signal strength, and can be thought of as the temperature of a blackbody emitting the equivalent signal at the frequency being studied. In ozone vs. time plots, the data are in terms of a volume mixing ratio, measured in parts per million. Clearly the two methods of measuring are related, since both are a measure of the strength of the signal. The specifics of the conversion can be found in an upcoming memo.

Multiple Kinds of Time

When the data are collected at each of the MOSAIC sites (approximately every 75 seconds), they are marked with the Universal Time (UT) of the measurement. This means the times recorded are EST plus 5 hours or EDT plus 4 hours. Every day, the data are averaged into 10 minute chunks in order to reduce the file sizes and provide a manageable amount of data. However, the times that are usually of scientific interest are local times, and, because so much of the mesospheric ozone's behavior is diurnal (changing at sunrise and sunset), the local time used is an "equinox" time, where sunrise is always set to be 6:00 am and sunset is always set to be 6:00 pm (or 18:00). This allows us to average together data over a whole year to study the diurnal trends without any need to account for the differences in sunrise time. These times are listed in the data in decimal form, ranging from 0.0 to 23.99.

Fitting Above/Below Together or Separately

The MOSAIC system uses the fact that mesospheric ozone is not pressure broadened (due to the relative scarcity of molecules in the mesosphere) to look "through" the higher density ozone located in troposphere and stratosphere (below 80 km) to the low density ozone located in the mesosphere (above 80 km). Of course, there is some signal collected from the lower ozone, but the signal is so broad that it can be modeled separately. When "above and below 80 km fit together" is selected on the interface main page, only one spectrum is modeled and fit. When "above and below 80 km fit separately" is selected, you will receive two spectra or two sets of ozone vs. time graphs.

Ozone spectrum

What you see (graphical output)

The x-axis is frequency, with 0.0 = 11.0725 GHz (the center of the ozone line). The y-axis is "antenna temperature," which just is a measure of the strength of the signal being received. The data are plotted in staircase form to make clear the number of channels (64) and the frequency width (9.75 MHz) of each. The data are automatically fit to the theoretical curve for the shape of

the line. If you have selected above and below 80 km fit separately, you will see two curves fitting the data, one corresponding to the ozone above 80 km (from the mesosphere), and one corresponding to the ozone below 80 km (from the troposphere and stratosphere).

What you get (avout.txt file)

You will receive a lot of data. The columns are separated by tabs, and if you save the file, you can import it into excel, which knows how to turn the text file into a spreadsheet. The name for each column is in the preceding column.

The time/date columns (“year” “day” “hour” and “minute”) are probably self explanatory. The only important thing to note is that these times are in Universal Time (UT). These data will be listed for every 10 minutes in the time you requested.

“num_rec” displays the number of spectra averaged into that 10 minute period. (The MOSAIC spectrometer actually reads out every 75 seconds or so, so usually there are either 8 or 9 spectra in any given 10 minute block.)

“sun_el” is the sun’s elevation at the location of the spectrometer at the time of the spectrum. This is at ground level, and since the mesosphere is much higher (80 km above Earth’s surface), the sun sets later and rises earlier than it does on the surface. “d The early columns are the date and time of the spectrum.

“tpwr” is the total power measured by the system during that 10 minute period, measured in dB.

“rms” is the standard deviation in the spectrum, in Kelvin, after subtracting the average value.

“ltm” is the “equinox” local time at the location of the spectrometer. This is an artificial local time that makes 6.0 sunrise and 18.0 sunset every day, regardless of the season. (This is so that you can average data from multiple days in different parts of the year without introducing error due to the differences in the start of the day.)

“spect__” is the designation for the spectrometer that collected the data. Each of the 5 operational MOSAIC units has a number, so when you see spect000, you know that spectrum was collected at Chelmsford High School. What follows are 64 columns of numbers, corresponding to the (10 minute average) spectral readout from the spectrometer at the time indicated.

What you get (sum.txt file)

This contains the parameters for the curve automatically fit to the data displayed in the graphical output. As in the data from the avout.txt file, the name of each element is in the column preceding it.

“peak” is the maximum value of the curve fit to the spectrum, in milliKelvin (mK). For students trying to characterize the “value” of the signal, this is probably the best number to use (if you want the “answer” in mK).

“rmsresid” is the root-mean-square (like an average) of the residuals of the spectrum relative to the curve fit. The larger this number is, the worse the fit of the data to the curve. If the curve were a perfect fit with no noise, the value would be zero. The value is in mK.

“theory” is the theoretical value of the rmsresid based on the system noise temperature. This would be the value of the rmsresid if all the error in curve fit came from random noise. In practice, the actual rmsresid is normally larger than theory due to systematic changes in the system or interference in the measured spectrum, which results in deviation from the expected profile.

“fit_high” This is the peak value of the curve fit to the data, in ozone vmr measured in parts per million (ppm). This is another good way to characterize the “value” of the spectrum.

“fith_err” is a measure of the uncertainty in the fit_high. It is the standard deviation of the residuals to the fitted curve.

“fit_low” will be equal to zero if you have selected to fit above and below 80 km together. If you have fit them separately, this will give the value of the peak of the curve fit to the below 80 km data (the curve seen in the dotted line in this type of output.)

“fitl_err” is a measure of the uncertainty in the fit_low peak. Will also be zero if you have selected to fit above and below 80 km together.

“avtpwr” is the average total power for the spectrum, in dB.

Following that, there will be an indication of the dates selected for the spectrum.

Ozone vs. local time

About the averaging

The data are averaged according to “equinox” local time, with a center value of the time listed and a time interval equal to what is selected on the interface. The system combines the spectra from all the selected spectrometers within the time ranges indicated, using the “ltm” value (the equinox local time). For example, for 20 minute averaging, the data point listed at 1.00 hour will contain the average of data from equinox local times 0:50 to 1:10, or 0.833 to 1.167.

What you see (graphical output)

Time is on the x-axis, in hours, taken from the “equinox” local time. The time will always go from 0 hours to 24 hours, with day located in the middle of the graph and night located on either side. On the y-axis is ozone volume mixing ratio, in parts per million. This is an indication of the strength of the mesospheric ozone signal. The data are automatically fit to a curve that assumes a constant value for ozone at night and during the day, with an exponential function to connect them. If you have selected above and below 80 km fit separately, you will see two graphs like this, one corresponding to the ozone above 80 km (from the mesosphere), and one corresponding to the ozone below 80 km (from the troposphere and stratosphere).

What you get (avout.txt file)

“time” is probably self-explanatory. You will see values from 0 to 24, with the number of times included depending on what time interval you averaged for. For example, if you average over 30 minutes of data, you will see 0.0, 0.5, 1.0, 1.5, etc.

“fit_h” gives, for each time interval, the peak of the curve fit to the spectral data, in ozone vmr in ppm. This is the same as the fit_high parameter provided in the sum.txt output from the spectral graph. This is just a measure of the relative strength of ozone at each time. If you have selected to fit above and below 80 km together, this is the parameter corresponding to both parts of the spectrum. If you have opted to fit them separately, it corresponds to above 80 km, or the mesospheric contribution.

“err” gives the uncertainty associated with the peak value in fit_h. This is the same as fith_err in the sum.txt file from the spectral output, and is the standard deviation of the residuals to the fitted curve.

“fit_l” will be zero if you have opted to fit data from above and below 80 km together. It will be non-zero if you have fit them separately, and equal to the peak of the curve fit to the averaged

data (in ozone vmr in ppm) from below 80 km. It is the same as the “fit_low” parameter in the sum.txt file from the spectral curve fit.

“err” is the uncertainty associated with the peak value in fit_l. It will be zero if you have opted to fit above and below 80 km together.

Ozone vs. Date

About the averaging

The system combines the spectral data from throughout each day and from each of the spectrometers selected, averaging them to produce a spectrum. The date plotted is the center of the averaging. If you select to average over 10 days, it will take 5 days before and 5 days after the chosen start date for the first data point. The curve that would fit this spectrum is determined, and its peak value is used to characterize the strength of the ozone on this day. This graph is best seen when only nighttime data is used, since the daytime levels of ozone are so low.

What you see (graphical output)

Time is on the x-axis, in days. The days are counted from January 1, 2008, so a date of 367 corresponds to January 1, 2009 (2008 was a leap year), and a date of 732 corresponds to January 1, 2010. Ozone volume mixing ratio (in parts per million, or ppm) is plotted on the y-axis, which is a measure of ozone concentration.

What you get (avout.txt file)

“date” This is the date being plotted, with the first number being the chosen start date from the GUI.

“fit_h” This is the peak value of curve fit to the averaged spectrum, in ozone vmr in ppm. IF you have opted to fit above and below 80 km together, this is the parameter corresponding to all parts of the atmosphere observed. If you fit them separately, it corresponds to above 80 km, or the mesosphere.

“err” is the uncertainty associated with the fit_h peak value.

“fit_l” will be zero if you have opted to fit above and below 80 km together. If you have fit them separately, this will correspond to the peak value of the curve corresponding to the data from below 80 km.

“err” is the uncertainty associated with the fit_l peak value.