MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886
February 21, 2017

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

To: VSRT Group
From: Alan E.E. Rogers
Subject: Ozone spectrometer data file format

The ozone spectrometers under construction for the CCLI project will use "compressed" file format to reduce the volume of data which will be sent to Haystack day. For easy exchange between machines of different endian an ASCII format is preferred but the Spectra will be encoded using 2-bytes for each spectral point. These 2-bytes use the 6-bit email attachment ASCII codes:

0→25	A→Z
26→51	*a→z*
52→61	0→9
62	+
63	/

The format is space delimited as follows:

yyyy:ddd:hh:mm:ss decimal_hours fstart

fstep fcal fcalamp total_pwr_db added fields staname

spect_vsrt_number peak s 256_point_spectrum

where

yyyy   year i.e. 2009

ddd   day of year i.e. 1-365 or 1-366 in leap yr

hh   hour Universal time

mm   minute

ss   second

decimal_hours hours in 9.5f (see note for multichannel data)

fstart   frequency of first spectral point MHz 9.4f (see note for multichannel data)

fstep   frequency spacing MHz   9.7f

fcal   frequency of calibration signal MHz 9.4f

fcalamp   amplitude of calibration signal   9.5f

total_pwr_dB total power in dB   9.5f

6 added fields for multichannel data

1

staname   station name in up to 12 characters

spectrometer number spectxxx (where xxx = spectrometer number)

peak   peak magnitude of spectrum in K (used to normalize 12-bit values 9.5 f

s   fixed ASCII lower case s (used as marker)

256_point_spectrum   512 characters

Each pair of characters represents a spectral point which can be reconstructed by

$$\left((64a + b) - 2000\right) \times peak/2000 \; K$$

where a is the 6-bit code of the 1st character

and b is the 6-bit code for the 2nd character

example

2009:018:14:25:59   14.43306   1322.1420   0.0024414

      1320.5347   0.7357 23.54290   bridgewater

      spect002   1.09244   s   YHTB…..

files are named

      yydddhh.sn

where yy   yr   i.e. 2009=09

      ddd   day of year

      hh   UT hour

      s   fixed character

      n   spectrometer number 03d

A single "record" with the encoder spectrum is written every 90 seconds as a line with 627 characters and a new file is started each day. The data transferred from 6 spectrometers will be about 3.6 MB per day.

Note: Multichannel spectrometer data

Fields changed

Decimal-hours   channel saturation flag

fstart   indicates channel number for calibration frequency and frequency calibration signal strength

Fields added   6 fields added which indicate power in each channel in dB and noise cal Y-factor in each channel. These alternate as follows. On an even channel number they indicate power and on an odd channel number they indicate Y-factor

MOSAIC-2 data format

Type ASCII

Format is indicated in Column 19 of each line.

Blank- original format

Field number	Format
1	YYYY:ddd:HH:MM:SS
2	This field is blank
3	Frequency start (MHz)
4	Frequency step (MHz)
5	Frequency of calibration (MHz)
6	Amplitude of calibration
7	Total power
8	Station name
9	Spec%03d station number
10	Spectrum normalization
11	S – spectrum marker
12	Spectrum in base64 character set

"0" or "1" multichannel format

Field number	Format
1	YYYY:ddd:HH:MM:SS
2	"0" or "1"
3	Saturation indicator
4	Channel number
5	Frequency step (MHz)
6	Frequency of calibration (MHz)
7	Amplitude of calibration
8	Total power

"2" 2-channel format

Field number	Format
1	YYYY:ddd:HH:MM:SS
2	"2"
3	Channel number (alternates 0 and 1)
4	Saturation indicator
5	Frequency step (MHz)
6	Frequency of calibration (MHz)
7	Amplitude of calibration (MHz)
8	Total power
9	Station name
10	Spect%3d station number
11	Spectrum normalization
12	s- spectrum marker
13	Spectrum in base64 character set

"a" Beagle Bone 2-channel format
The total number of fields on a line depends on the number of channels used by the spectrometer. The fields are defined as follows:

Field number	Format	Description
1	YYYY:ddd:HH:MM:SS	Time: YYYY=year, ddd=day of year, HH=hour, MM=minute, SS=second
2	Literal letter "a"	Format version indicator
3	%d	Number of channels
4	%9.4f	Frequency of first point in spectrum (MHz)
5	%9.7f	Spacing of points in spectrum (MHz)
6	%d	Channel saturation flag
7	%9.4f	Calibrator frequency (MHz)
8	%9.5f	Calibrator "amplitude"
9	%.5f	Channel power (dB)
10	%.2f	Channel Y-factor (not available, reported as "nan")
6+5*C	%9.5f	Total power (dB)
6+5*C+1	%s	Station name
6+5*C+2	spect%03d	Station number (prefixed by "spect")
6+5*C+3	%9.5f	Maximum absolute value of ozone line spectrum before normalisation
6+5*C+4	Literal letter "s"	Spectrum marker
6+5*C+5	512-character string	256-point spectrum encoded using base64 character set

Fields 6-10 are repeated C times where C is the number of channels.

The normalisation and encoding of the ozone spectrum is described in VSRT Memo #51.