

Version 1.0 – CO₂ Processing

Soil CO₂ concentration, $\chi_{c,soil}$, was measured from gas collected in a hollow, porous disk buried in the soil, pulled through between 40' to 100' of 1/4" I.D. Dekabon tubing at approximately 6 LPM for 30 seconds, and measured using a LI-800/LI-820 gas analyzer. Every half-hour each disk was sampled, one per minute. The analyzer was recorded every second during the 30 seconds of pumping.

Note, all analyzers were with an averaging length of 0 s.

Note, there were no flow controllers on any of the pumps, thus flow rates were approximate and variable over time.

In theory, the 30 second waveform of each CO₂ sampling had three parts: initial seconds with random values (gas being purged from the tubing), the $\chi_{c,soil}$ measurement (gas in the collector disk), and final seconds with decreasing concentrations (gas pulled from the soil surrounding the disk). Not every sample behaved this way.

Processing Algorithm

A computer algorithm was written to help interpret the 30 second waveforms and select the most likely measurement corresponding to $\chi_{c,soil}$.

Some assumptions were made for the sampling. First, the minimum, initial characteristic, and maximum arrival times that corresponded to $\chi_{c,soil}$ were known for each tube:

Rep2ControlSubsample1 = 6.1, 8.8, and 16 s
Rep2ControlSubsample2 = 8.1, 11.7, and 20 s
Rep2Slash = 8.4, 12.2, and 20 s
Rep2Broadcast = 8.7, 12.6, and 20 s
Rep3Control = 4.5, 6, and 12 s
Rep3_4InchChips = 8.5, 10 and 16 s
Rep3_2InchChips = 7.5, 9 and 15s

An arrival time was predicted from the characteristic time adjusted for battery voltage (an empirical relationship was determined where flow rate was related proportionally to battery voltage, and arrival time was inversely proportional to battery voltage).

A linear fit from the last 10 seconds of each waveform was calculated. Residuals were calculated and standardized between the entire waveform the linear fit. Significant residuals were flagged when their absolute value was greater than 2.306 (t-test with $t=2.306$, $df = 8$, $p = 0.025$ one sided, note this applies specifically to the last 10 standardized residuals, but was extended to all of the other residuals). The significant residuals were weighted by their separation in time from the predicted arrival time such that residuals occurring closer to the predicted arrival time were given more weight. The weighting function was

$$\varepsilon_{i,test} = \varepsilon_i \bullet \left[\frac{1}{1 + |t_{i,adj} - t_{predicted}|} \right]^2$$

where each weighted standardized residual, $\varepsilon_{i,test}$, was calculated from each standardized residual, ε_i , the adjusted measurement time corresponding to each residual, $t_{i,adj}$ (adjusted from battery voltage) and the predicted arrival time, $t_{predicted}$. Only those residuals occurring within ± 2.5 seconds of the predicted arrival time were considered.

The slope of the linear fit was tested for downward or upward (t-test with $t=2.306$, $df = 8$, $p = 0.025$ one sided). For downward slopes, a peak was chosen if there existed a maximum positive weighted residual. A step function transition was chosen if there was a last (in time) negative residual followed by a non-significant residual. For upward slopes, a dip was chosen if there existed a maximum negative weighted residual. A step function transition was chosen if there was a last (in time) positive residual followed by a non-significant residual. For neutral (or flat) slopes, a peak or a dip was chosen if there was a maximum absolute weighted residual. For all other cases the shape was considered questionable.

The type of waveform was recorded (i.e., downward-peak, flat-questionable, etc.) and a tally of the occurrence of each type of waveform was updated (time constant = 24 hours). The mode of the type of waveform was selected (i.e., the most common type).

If the shape was not questionable, then $\chi_{c,soil}$ was selected as the CO_2 measurement corresponding to the time of the peak, dip, or step transition. The adjusted measurement time was used to update the prediction of the arrival time (time constant = 24 hours).

If the shape was in question, the mode of previous waveforms was used to decide if this current waveform was acceptable. For example, if the waveform was upward-questionable and the mode was downward-peak, then the waveform was considered unacceptable. If the waveform was downward-questionable and the mode was downward-peak, then the waveform was considered acceptable. For acceptable waveforms, $\chi_{c,soil}$ was selected as the CO_2 measurement corresponding to the predicted arrival time. This time was not used to update further predictions. For unacceptable waveforms $\chi_{c,soil}$ was set as a missing value.

	Down-Peak	Down-Step	Down-Quest	Flat-Peak	Flat-Quest	Flat-Dip	Up-Quest	Up-Step	Up-Dip
Down-Peak	1	1	1	1	1	0	0	0	0
Down-Step	1	1	1	1	1	0	0	0	0
Down-Quest	1	1	1	1	1	1	0	0	0
Flat-Peak	1	1	1	1	1	0	1	0	0
Flat-Quest	1	1	1	1	1	1	1	1	1
Flat-Dip	0	0	1	0	1	1	1	1	1
Up-Quest	0	0	0	1	1	1	1	1	1
Up-Step	0	0	0	0	1	1	1	1	1
Up-Dip	0	0	0	0	1	1	1	1	1

Table of allowable transitions between waveform types.

Despiking of CO₂ Data

After the CO₂ was processed to yield $\chi_{c,soil}$ values, bad values were removed using the median despiker. Half-hour values were despiked using $\chi_{c,soil}$ values between 0 and 20,001 ppm, a 5th order median filter (five-and-a-half hour window), with a threshold of ± 250 ppm. Data was removed and the despiker was repeated until all spikes were identified. After that, the data were gap filled using a 5th order median-filter (five-and-a-half hour window). Gaps larger than two-and-a-half hours were not filled.

Matching Frozen Tubes to Observed Notes

During most site visits, every tube was noted as being open or blocked/frozen (from audible clues or low analyzer pressure). The $\chi_{c,soil}$ data was removed when the tubes were noted as frozen. By inspection, the data was retained that was similar to known open times (i.e., during the spring/summer/fall), and data was removed that was similar to known blocked/frozen times (i.e., noted during site visits). Usually there was a discontinuity when the tube switched between frozen and unfrozen. Data was removed manually.

Finally, any other bad data or anomalies were removed manually.