Quality Assurance and Quality Control of Soil CO_2 Flux Datasets

Rick Garcia

September 22, 2011

TRANSCRIPT

INTRODUCTION:

Moderator:

Before we get started, I would like to pop up our poll question for this webinar. The question is and should be showing on your screen now, "What is the most critical factor in data quality." The research question, experimental design, research tools and instruments or data quality and screening. Go ahead and click the radio button next to your answer choice and we'll give everyone a few moments to select an answer.

SLIDE 1:

Speaker:

Thanks Ashlee. Obviously there is no right or wrong answer to our polling question. We have all heard the terms QA and QC. What do these terms mean?

SLIDE 2:

In the context of research, a simple definition of Quality Assurance is a process implemented before data are collected to ensure quality.

SLIDE 3:

Quality control, on the other hand, is a process implemented after data are collected to ensure quality.

SLIDE 4:

In scientific research, some of the elements of quality assurance should include the research question, an appropriate experimental design, adequate measurement tools, proper installation and usage of the tools and thorough preliminary testing.

SLIDE 5:

The questions to be answered and experimental design are beyond the scope of this webinar, but obviously, fundamental to any research.

SLIDE 6:

However, we at LI-COR do have quite a lot of experience at Soil CO_2 flux measurements and you may want to review some of our application notes and poster papers as you consider your experimental design. One note, which may be of interest in this regard, is titled, Soil CO_2 flux measurements: addressing spatial variability and determining the number of readings required. This note is available on our website.

SLIDE 7:

As I just mentioned, we at LI-COR have quite a lot of experience in Soil CO_2 flux measurements. This experience is built into the design of the LI-8100 system. See our previous webinars on Soil CO_2 flux if you would like to hear details on why we believe the LI-8100 system is a great tool for measuring Soil CO_2 flux.

SLIDE 8:

Our manuals give detailed information on installation and usage of the LI-8100. For example, it is important that the soil collar provide a mechanically stable measurement interface between the chamber and the soil. Soil collars are covered in Chapter 2 of the manual.

SLIDE 9:

Although weather tight, the instrument connectors and cases were not designed for submersion.

SLIDE 10:

In long term, unattended measurements at sites which may be exposed to intense precipitation, it is important to keep the instruments up off the soil using bricks, blocks or even a pallet as in this example.

SLIDE 11:

Direct sunlight can cause overheating of the instruments. Some sort of shading is appropriate under these conditions. The shading could be as simple as foam insulation nailed to the top of four stakes pounded into the ground

SLIDE 12:

When measuring multiple chambers, the potential for air leaks increases. Therefore, it is important to perform a leak test after installing all the chambers in the field to ensure that there are no leaks. The leak test is covered in Chapter 10 of the manual.

SLIDE 13:

Preliminary testing of the system is critical. Test to determine the optimum observation length, observation delay and purge times for your specific conditions. Be sure that the data from the system are reasonable. Do your first quality screening runs on these preliminary data. For unattended measurements, be sure to examine the starting CO_2 concentrations carefully. Depending on the site and chamber positioning, a pre-purge time may be needed to be added. Also, if you have attached any ancillary sensors to the systems such as soil temperature or moisture probes, this is the right time to confirm that the data from these probes in reasonable.

SLIDE 14:

Quality control is the process we implement after the data are collected to ensure quality.

SLIDE 15:

Elements of a quality control program for data from the LI-8100 should include data archiving, checking system setup and constants, checking instrument diagnostics, checking key variables used in the flux calculations and checking the computed flux values.

SLIDE 16:

It should be emphasized at the outset, that before any quality control process of data occurs, one should always archive the original data. It would be best to have two copies, which are archived on separate systems.

SLIDE 18:

After archiving, we load a copy of the data into the file viewer software for analysis. The LI-8100 file viewer software is a powerful tool. It makes it easy to get an overall summary of an entire season's data from a multi chamber experiment and, at the same time, drill down into the quality and characteristics of an individual observation. In the next few slides, we will walk through the steps you might follow using the file viewer software to do quality screening of a data set. I have chosen to break down the process into steps, which make sense for me. Your process may look different, but it is important to have a strategy before you start the experiment. With experience, the strategy will probably evolve. In my case, I chose to evaluate the information by categories.

The system constants and setup information, instrument diagnostics, variables used in the flux calculations and finally the flux values themselves. It should be emphasized here that just because the final flux values are different from what we expected or what we have experienced in the past or the variances are large, does not mean we have poor data that should be suspect or, worse yet, thrown out.

In Step 1, we load or open a file or multiple files. This is covered on page 24 of the file viewer manual.

SLIDE 19:

First, click the open file button on the toolbar or select the File Open from the main menu.

SLIDE 20:

Then use the Open File dialogue box to navigate to the desired category, desired directory and select a file to be viewed.

SLIDE 21:

Use control plus click to select an individual file or, by using shift plus click, you can select multiple files.

SLIDE 22:

If multiple files are selected when you click open, you will be given a choice of combining the files or keeping them separate.

SLIDE 23:

In Step 2, set up the displayed variables. This is covered on page 11 in the manual.

SLIDE 24:

First click on the change variable icon in the toolbar.

SLIDE 25:

Choose whether the variable you want is a measured or a header variable. We will be doing some of each. If the variable is a measured variable, click the appropriate radio button for the initial value or mean or range. In some cases, you may want to display the mean and the range or the initial value in the range for the same variable.

SLIDE 26:

Click the variables and add them to the included list. When you're done, click 'OK'.

SLIDE 27:

Some important diagnostic variables are the bench or IRGA temperature, the electronics and multiplexer temperatures, system voltage, the multiplexer flow and time to complete chamber closure.

SLIDE 28:

The bench or IRGA temperature is important because control of the IRGA temperature minimizes drift and to some degree, noise in the CO_2 signal.

SLIDE 29:

The electronics and MUX temperatures are important for system reliability and integrity.

SLIDE 30:

System voltage can impact system reliability and interruption in operation.

SLIDE 31:

Multiplexer flow is critical to the assumptions of good mixing and representative sampling in the flux computations.

SLIDE 32:

And finally, complete chamber closure is critical to measurements. The time to close the chamber can be a good indicator of complete closure.

SLIDE 33:

In order to pinpoint the sources of problems in the data, we need to have a fundamental grasp of the flux equation and the key variables in the flux calculations. This understanding will help us in investigating of the possible failure of a sensor in the system or a weakness in the experimental protocol or implementation of the protocol.

SLIDE 34:

In the equation, FC is the Soil CO_2 flux,

SLIDE 35:

V is the system volume including the chamber, tubing and IRGA volumes,

SLIDE 36:

P sub 0 is the initial pressure in the chamber,

SLIDE 37:

W sub 0 is the initial water vapor mole fraction,

SLIDE 38:

R is the gas constant,

SLIDE 39:

S is soil surface area and

SLIDE 40:

T sub 0 is the initial air temperature in the measurement chamber.

SLIDE 41:

Finally, DC Prime DT is the initial rate of change in water corrected CO_2 mole fraction.

SLIDE 42:

Key variables that I would suggest screening for from the flux computation include the initial values of atmospheric pressure, air temperature and dilution corrected CO_2 concentration. Although RH is not directly in the equation, water vapor mole fraction is. But relative humidity may be more useful to screen for because, when the relative humidity approaches condensing conditions, there could be concerns about the gas analyzer operation and measurements. Also, remember that RH is a function of water vapor concentration and air temperature so you should always look at chamber temperature at the same time that you are looking at RH.

SLIDE 43:

Most of the time, careful screening of the underlying variables will result in a data set with few, if any, outliers in the calculated flux values. However, it can be instructive to sort the data by the flux values calculated from the linear model and compare these values to the flux computed with the exponential model. The iterations required for convergence in a solution of the exponential model can also be used for quality flags.

SLIDE 44:

In Step 3, we set up charts so you can observe the results of screening outlier data.

SLIDE 45:

Briefly, click the add chart button,

SLIDE 46:

Choose the variable to be charted and the x-axis

SLIDE 47:

And finally create a chart group so you can screen multiple variables in one pass. For more details in setting up charts and chart groups, check out page 13 in the manual.

SLIDE 48:

After setting up the variables and charts, you can sort observations by a specific field or column of the data. Click the header of the column once to sort maximum to minimum.

SLIDE 49:

Click again to sort minimum to maximum. This procedure will allow us to quickly find observations which have variables values which are outliers or just don't make sense.

SLIDE 50:

As an illustration, I chose some data from and experiment that we conducted in 2007 using 16 chambers.

SLIDE 51:

The initial data set contained over 40,000 observations, including what appeared to be a number of outliers.

Moderator:

Rick, I'm going to interrupt you with a question. The audience asks, "Why does the file viewer software compute two flux values, one with an exponential regression and one with a linear regression?"

Speaker:

Thanks Ashlee. Let me look for a slide that can address that question. Ah, here we go. To answer that question, let's take a closer look at a single observation specifically zooming in on the aggression analysis.

SLIDE 52:

Here we see the water vapor corrected CO_2 concentration inside the chamber plotted against time. At time zero, the chamber closes and CO_2 begins to increase. It's clear from the data and makes sense physically that the rate of increase in CO_2 is greatest when the chamber first closes. If we fit the data with an exponential model as depicted by the red line through the data in this slide, and then solve for the initial slope, we can see that a linear fit of the data depicted by the black line through the data underestimates the true flux. In most cases, the exponential fit will be the best estimate of the flux. However, if the data are extremely noisy or if the there are anomalies in the data, it can be difficult to fit it.

SLIDE 53:

In fact, here is an example where the data are noisy and the alcarhythm never converged to fit the exponential model. In this case, we reached 10 iterations without convergence as you can see in the slide where iterations are circled in red and the system only reported the linear flux values. While I'm on this slide, I might point out another observation. As you can see, the CO_2 concentration before the chamber closed during the negative times was much higher than the CO_2 concentration at closure. This occurs occasionally depending upon where the open chamber is set, whether it is over vegetation or some respiring material, you might get very high CO_2 concentrations. As you set up your preliminary testing, you should look for this and be sure to add a pre-purge time so that you can get CO_2 concentrations down to ambient before the chamber closes.

SLIDE 54:

During the screening for instrument diagnostics, we found several records with little or no flow. Without flow, the measured change in CO_2 concentration is typically small and may be upward or downward. The regression analysis shown here was from a record with no flow. The stepwise decrease in CO_2 concentration leads me to believe that there may have been a flow blockage which caused the low flow.

SLIDE 55:

After screening for problems in the fundamental variables, we screened for outliers in the flux values from the linear model. We removed 16 records which had flux values indicating CO_2 uptake and 12 records where values were extremely high and may have been due to a rodent inside the chamber.

SLIDE 56:

After screening a total of 32 records from the original 40,000 records, I was left with this view.

SLIDE 57:

To dig a little deeper into the data and look at more detail, I chose to look at a one week segment of the data. There appears to be a discontinuity in the data between June 13th and June 14th. However, since I have carefully screened the data using my predetermined process, I can be more confident that the changes I see are real.

SLIDE 58:

In fact, there was a 29 millimeter rain on the afternoon of June 13th. This is in conjunction with the fact that some of the soil chambers had significantly more decomposing crop residue probably explains the variation we observed during this period in the data.

One take home message from this illustration is that, if we had thrown out the data where there were unexpected on its face and on its face difficult to explain changes in flux, we would have lost an opportunity to learn more about the system we were characterizing. Anytime data have more variance than we expected, change in unexpected directions or have greater magnitudes than we expected we should look for the underlying cause and not just assume that the data are suspect.

SLIDE 59:

Thank you very much for your time. This is the end of my presentation. Ashlee, do we have any questions from the audience?

Moderator:

Thanks Rick. I do have one question for you and it's I have relative humidity values that don't make sense. What can I do?

Speaker:

Ashlee, that's a tough question, but let me try to address it. Remember I mentioned earlier that relative humidity is a function of water vapor concentration and air temperature. It's possible that the relative humidity data don't make sense because your air temperature sensor is broken. That's the first thing to check. If, in fact, that is the case, what you could possibly do is use air temperature from a different chamber and substitute that into your data set to recompute a relative humidity based on that air temperature. If the relative humidity doesn't make sense because you have liquid water in the system, then unfortunately there is a good possibility that there will also be problems with your CO_2 data as well. In that case, you'll probably want to screen the data carefully to ensure that you don't have CO_2 numbers that are also unrealistic. So, are there any more questions Ashlee?

Moderator:

I actually don't have any more questions, but it looks like our panelists are still working on some of the questions that were submitted throughout the talk. I'd like to remind our audience that, if they would like to continue their conversation with the panelists or with Rick, they can do so offline or via email. That information should be on the screen for you. The last thing I'd like to do is to let everyone know that there's an on demand replay of this presentation. Just watch your email or visit www.licor.com/env/webinars for information on how to access that. So thank you for joining us today and we hope to see you at our next webinar.