

NISTIR 8070

Greenhouse Gas Emissions and Dispersion

4. Quality Control Process of in-situ INFLUX Tower-based Observation Data

Heming Hu
Kuldeep Prasad
Israel Lopez-Coto
Subhomoy Ghosh
James R. Whetstone

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8070>

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

NISTIR 8070

Greenhouse Gas Emissions and Dispersion

4. Quality Control Process of in-situ INFLUX Tower-based Observation Data

Heming Hu
Kuldeep Prasad
Israel Lopez-Coto
Subhomoy Ghosh
NIST Engineering Laboratory

James R. Whetstone
NIST Office of Special Programs

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8070>

September 2015



U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Under Secretary of Commerce for Standards and Technology and Director

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

**National Institute of Standards and Technology Internal Report 8070
Natl. Inst. Stand. Technol. Internal. Publ. 8070, 35 pages (September 2015)**

**This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8070>**

Greenhouse Gas Emissions and Dispersion

4. Quality control process of in-situ INFLUX tower-based observation data

Heming Hu, Kuldeep Prasad, Israel Lopez-Coto,
Subhomoy Ghosh, James R. Whetstone
National Institute of Standards and Technology, Gaithersburg, MD

Abstract

The Indianapolis Flux Experiment (INFLUX) was designed to develop top-down methods for measurement and modeling of urban fluxes of greenhouse gases (GHG) to the atmosphere with spatial and temporal scale of one km. and one hour respectively. A small bias in the observation data relative to the background can introduce significantly large bias in the computed fluxes if it is relatively high to the emission enhancement. Hence, it is important to maintain accuracy of *in situ* GHG data for unbiased estimation of urban fluxes. Accuracy of GHG observation data can be influenced by quality control (QC) process used for error detection, instrument calibration and error correction. Currently different organizations employ different QC processes on the same raw data resulting in significantly different QC'ed data. This report introduces a standardized approach integrating all present techniques. It consists of four major sequential steps: data screening, gas source based data separation, calibration and time-lag correction, and data averaging. QC'ed data for a set of twelve towers are compared with that from other organization for a period of two months. The influences of different methodologies in the QC process were analyzed, and the important parameters were optimized to reconcile the differences. Finally, recommendations in different phases of QC process are presented to provide guidelines for tower observation.

1 Introduction

Long-lived greenhouse gases (GHG), including CO₂, CH₄, CO, contribute greatly to global warming. An accurate understanding of their sources and sinks is fundamental for developing mitigation and adaptation policies [1]. Recent data shows cities contribute more than 70% of the fossil fuel emissions to the Earth's land surface. However, uncertainties in GHG emission inventories at individual city level are either very large or unknown. Sometimes uncertainties range from 50% to as high as 100%.

The Indianapolis Flux Experiment (INFLUX) was designed to develop top-down methods for measurement and modeling of urban fluxes of GHG to the atmosphere with 10% uncertainty at spatial and temporal scale of one km. and one hour respectively. Top-down modeling is usually based on in-situ tower data. Cavity ring down spectrometers (CRDS) instruments have been mounted on various towers (measurement locations) in the city of Indianapolis to quantify urban emission flux of CO₂ and CH₄, with data being collected at approximately every second. This data can be utilized as a central component of an inverse modeling to calculate GHG sources temporally and spatially [2, 3].

The concentration enhancement due to the city emissions is usually very small relative to the background concentration, so even a tiny bias in measurement can be large with respect to the enhancement. Figure 1 shows a bar plot of time-averaged CO₂ concentration (day time scale: 17:00 – 22:00 UTC, and monthly scale: Sep. – Oct. 2013) from INFLUX towers. Figure 1 clearly shows site 1 as the tower recording least average concentration and site 3 as the tower with the largest. If we consider site 1 as the background, then the largest enhancement is only 4.2 μmole/mole compared to the background (396.3 μmole/mole). A small bias in the observation data can result in a large bias in the computed fluxes because of the low enhancement. Observation bias of only 0.4 μmole/mole will mean a flux bias of 10% for CO₂ at site 3, and it would be worse for other sites. Observation bias is just one of the contributions for flux uncertainty. To achieve a flux uncertainty of 10% or better for INFLUX, the accuracy of the observation data should be very high.

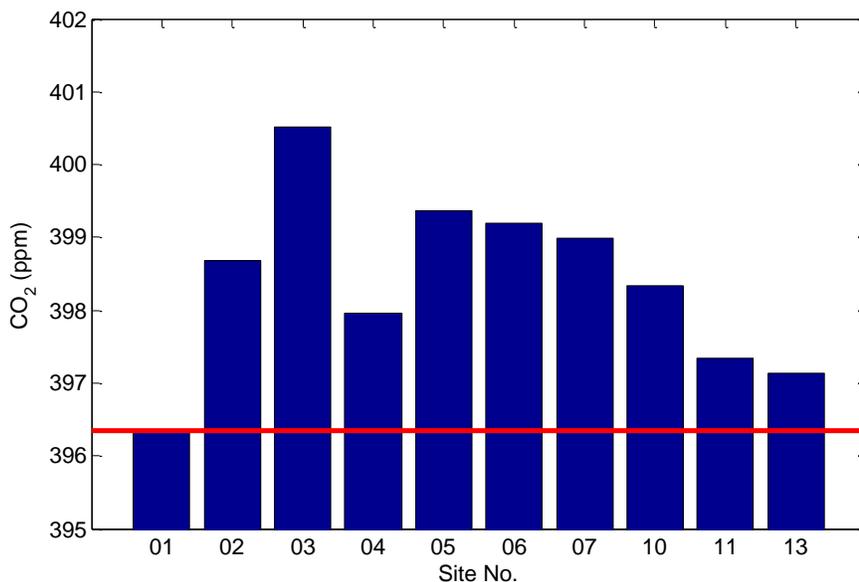


Figure 1: Observed daytime CO₂ concentration in Indianapolis (Sep. – Oct. 2013).

Due to increased observation frequency, a large amount of data can be obtained from the instruments mounted on various towers. The raw data contains many unreasonable data points due

to mechanical problems, electrical problems, and software problems of the sampling system, which should be flagged and removed before the data can be used in the inversion system. There are also some necessary correction that should be performed, such as calibration and time lag correction. Quality control (QC) is defined as the operational techniques and activities that are used to fulfill requirements for data quality, including missing data detection, error detection and possible error corrections [4]. Accurate measurement data depend not only on an accurate instrument, but also on a careful QC process of the raw data. To ensure the highest possible reasonable standard of accuracy for optimum use by all possible users, a well-designed quality control system is vital, and effort shall be made to correct all erroneous data and validate suspicious data detected by QC procedures.

Every database has its own demands and features, and has corresponding QC process. The Global Atmosphere Watch (GAW) Programme of the World Meteorological Organization (WMO) provides a framework for the development and implementation of integrated greenhouse gas (GHG) observations. Report No. 206 [5] of GAW reviewed WMO data quality objectives in greenhouse gas measurements, calibration, quality control, data management and archiving, and made several recommendations on the WMO data quality objectives. The Global Greenhouse Gas Reference Network of the U. S. National Oceanic and Atmospheric Administration (NOAA) constitutes a major part of the GAW network. Andrew et al. [6] described the tower measurements of this reference network, including instrumentation, uncertainty analysis, and recommendations for future high-accuracy GHG monitoring efforts. The Environmental Protection Agency (EPA) has released a quality assurance handbook [7] for Air Pollution Measurement Systems to standardize the data quality related objectives, methods and management.

The accuracy of GHG observation data can be influenced by the QC process. Currently different QC processes are used by different organizations, which can result in some non-ignorable differences for the same raw data. In an effort to standardize the QC process for greenhouse gas measurement data, a comprehensive four-step approach is presented in this report and the QC'ed data of INFLUX towers are compared with that from another organization for a period of two months. Existence of a large amount of real-time data, heterogeneous data stream and high requirement of data accuracy, stability and consistency were some of the major challenges.

The main objective of this report is to develop more effective data checking methods, to create recommendations for different phases of quality control, and to formulate general guidelines for flagging data. Section 2 provides background of INFLUX tower based sampling systems, data management and data contents. Section 3 describes the QC process including data screening, data separation, calibration, and data averaging, and also discusses impact of different methodologies in the QC process. Section 4 shows detailed comparison between our QC'ed data and that from

other organization. Finally, Section 5 summarizes our major findings of this report and provides detailed recommendations.

2 Data collection

2.1 INFLUX tower based sampling system

There are twelve observation towers located in and around Indianapolis, where the measurements of CO₂, CH₄ and CO are being taken. A map of the tower locations [8] is shown in Figure 2. Most of them measure the enhanced GHG concentration for urban area, and the others for suburban area. For example, SITE01 can measure the background concentration from upwind direction of Indianapolis. INFLUX project uses Picarro¹ instruments with the Cavity Ring-Down Spectroscopy (CRDS) technology to measure concentration of GHG in and around Indianapolis. They can achieve an overall inter-laboratory comparability as stipulated by the WMO standards [5] (100 ppb for CO₂, 2 ppb for CH₄ and 5 ppb for CO) without drying the samples to very low levels [9]. A sampling system at each tower also includes a drying system to control air humidity level with the ability of reducing the mixing ratio of water vapor below 1%. In fact, the measured water ratios of most sites are less than 0.2% except for Site 2 at which the water ratio was less than 0.6% due to a shorter Nafion tube.

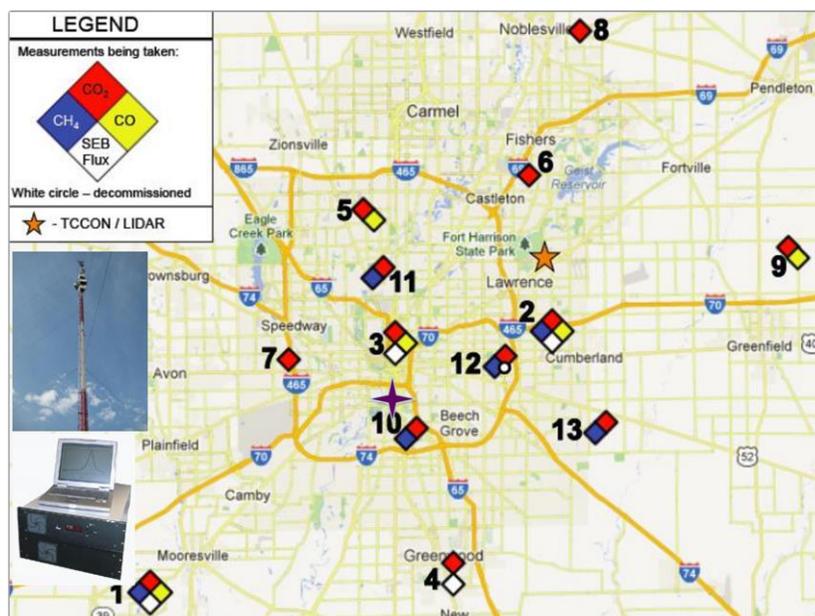


Figure 2: Distribution of INFLUX towers (Site 12 has been decommissioned)

¹ Certain commercial equipment, instruments, or materials are identified in this report in order to adequately specify the materials used and the experimental procedure. Such identification does not imply recommendation of endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Table 1 shows detailed information on the data source including installation date, site elevation, sampling heights and instrument. The sampling systems were installed on 13 towers during the period of 2010 to 2013. Site 12 has been decommissioned since Apr. 2013. Four of the 12 towers have more than one sampling heights (two towers with three heights and two towers with four heights). Four different models of CRDS instrument with different capacity of species sampling are used. The serial number of instrument has two components, first part indicates model no. and the latter part indicates corresponding instrument no. Site 4, 6 and 7 with older-version instruments can only measure concentration of CO₂, and the other sites can also measure CH₄, CO or both in addition to CO₂.

Table 1: Observation tower demographics

Site No.	Installation Date	Sampling Heights (m AGL)	Species measured	Serial number of instrument
SITE01	09/22/2010	10/40/121	CO ₂ /CO/CH ₄	CFKADS2025
SITE02	09/24/2010	10/40/136	CO ₂ /CO/CH ₄	CFKADS2005
SITE03	06/01/2012	10/20/40/54	CO ₂ /CO	CKADS 2023
SITE04	08/18/2012	60	CO ₂	CADS06
SITE05	03/30/2012	125	CO ₂ /CO	CKADS2024
SITE06	07/15/2013	39	CO ₂	CADS10
SITE07	03/29/2012	58	CO ₂	CADS05
SITE08	05/28/2013	41	CO ₂ /CH ₄ /CO	CFKADS2079
SITE09	03/30/2012	10/40/70/130	CO ₂ /CO	CKADS2025
SITE10	03/28/2012	40	CO ₂ /CH ₄	CFADS2155
SITE11	04/17/2013	130	CO ₂ /CH ₄	CFADS2160
SITE12	Aug 2012 – Apr 2013	40	CO ₂	CADS10
SITE13	04/17/2013	87	CO ₂ /CH ₄	CFADS2159

2.2 Data management system and data contents

A data center receives data on a daily basis in ASCII format. On a given day, more than 1 file from any instrument indicates interruption(s) in the system. Source control of the data is encoded in first 3 parts of every file names. Old-version instrument at Site 4, 6 and 7 uses the name format “serialNo-date(yyyymmdd)-startTime(hhmm)-Mailbox-col.dat.gz” (for example, “CADS05-20131201-0000-Mailbox-col.dat”). Other sites use “serialNo-date(yyyymmdd)-startTime(hhmmss)Z-DataLog_User-col.dat.gz” (for example, “CFADS2160-20131201-000019Z-DataLog_User-col.dat.gz”)formatting. The mapping of serial numbers of the instruments to the site numbers can be found in Table 1.

Table 2 gives a short summary of the fields in a typical data file. Each file has 25 columns. Columns containing time stamp information, concentration, and instrument conditions are the most important subjects at discussion in this paper. Section 3.2.2 discusses QC process over the mentioned columns. There are some columns needed to be mentioned. The time stamp (column 1 & column 2) is in UTC format and the “FRAC_DAYS_SINCE_JAN1” (column 3) is the number of days from Jan 1st every year.

Table 2: Data file variables

Column	Name	Unit	Explanation
1	DATE	/	The sampling date in “yyyy-mm-dd”
2	TIME	/	The sampling time in “HH:MM:SS.000” where “SS.000” is the second with 3 digits
3	FRAC_DAYS_SINCE_JAN1	/	The serial date which returns 0 on Jan 1 st of every year.
4	ALARM_STAT US	/	The alarm for instrument status. 0 is for normal, and 1 is for abnormal. Sometimes it happens to be 65536 which can also be regarded as 1.
5	CH4	μmole /mole	The CH ₄ concentration without vapor correction.
6	CH4_dry	μmole /mole	The dry concentration of CH ₄ .
7	CO	μmole /mole	The dry concentration of CO.
8	CO2	μmole /mole	The CO ₂ concentration without vapor correction.
9	CO2_dry	μmole /mole	The dry concentration of CO ₂ .
10	CavityPressure	Pa	Pressure in the cavity of instrument, ~140 Pa.
11	CavityTemp	°F	Temperature in the cavity of instrument, ~45 °F.
12	DasTemp	°F	Room temperature
13	H2O	%	The raw concentration of water vapor
14	InletValve	/	The opening of inlet valve of cavity which is always kept open at a fixed position.
15	OutletValve	/	The opening of outlet valve of cavity which varies between fully closed and valve fully open to ensure the stability of cavity pressure.
16	galpeak14_final	/	
17	h2o_pct	%	The pct concentration of water vapor.
18	h2o_reported	%	The reported concentration of water vapor. Used for the vapor correction.
19	peak84_raw	/	Raw signal for CO
20	solenoid_valves	/	At most 6 inlets have been used for sampling of different gas sources, at most 2 out of them represent reference gas source.
21	species	/	An indicator function of the gas species update: 1 if CO ₂ , 2 if CH ₄ , 3 if

			H2O and 4 if CO is updated respectively.
22	b_h2o_pct	%	Absorption line of H ₂ O that are adjacent to the line of CO
23	peak_14	/	
24	vpeak_14	/	
25	wlm1_offset	/	

The concentrations of different species are recorded with the unit of ppm ($\mu\text{mol/mol}$) which is the mole fraction of substances in dry air (dry air includes all gaseous species except water vapor). The CRDS instrument uses a built-in algorithm to convert measured wet concentration into dry concentration. The CO₂ dry concentration (column 9) can be calculated from the actual wet air concentration (column 8) and the concentration of water vapor (column 18) using a quadratic polynomial fitting [10],

$$(CO_2)_{dry} = (CO_2)_{wet} / (1 + aH_{rep} + bH_{rep}^2) \quad (1)$$

A similar approach can be used for obtaining dry CH₄ concentration,

$$(CH_4)_{dry} = (CH_4)_{wet} / (1 + cH_{rep} + dH_{rep}^2) \quad (2)$$

where H_{rep} is the water vapor concentration reported by the instrument. The fitted values are shown in Table 3. The experimental fitting introduces uncertainty. The drier the air the lower is the uncertainty.

Table 3: Constants in equation (1) and (2)

Parameter	a	b	c	d
Value	-0.01200 ± 0.0009	$-(2.674 \pm 0.18) \times 10^{-4}$	-0.00982 ± 0.0006	$-(2.393 \pm 0.1) \times 10^{-4}$

For SITE04/06/07, the CADS instrument updates water vapor data (column 13 / 17 / 18) with a frequency of 1/10th of Co₂ update frequency. . We can calculate the time cycle based on the time stamps of raw data. Table 4 shows that the mean time cycle for each species is within 1~4 s in the normal working condition. An increase in time cycle indicates instrument malfunction.

Table 4: Mean time cycle at normal condition (unit: s)

Site	SITE01	SITE02	SITE03	SITE04	SITE05	SITE06	SITE07	SITE08	SITE09	SITE10	SITE11	SITE13
CO ₂	2.3	2.2	2.7	2.2	2.8	1.4	1.6	2.3	2.7	3.3	3	3.4
CH ₄	2.3	2.2	/	/	/	/	/	2.3	/	3.3	3	3.4
CO	2.3	2.2	2.7	/	2.8	/	/	2.3	2.7	/	/	/

3 Flow diagram of the QC process

3.1 Overview

A comprehensive QC process has been developed and tested for the in-situ tower data collected as part of the INFLUX experiments. The purpose of the QC process is three fold:

- a) Identifying suspicious values, certain errors etc. and flagging them;
- b) Plausible correction of errors
- c) Averaging of data at different frequencies

The QC process is developed using a four-step procedure as shown in Figure 3. The first step is reading raw data from the received files and data screening. The second step is separating the raw data based on gas sources and species with an index matrix. QC'ed data can be calculated in the third step with the index matrix. Calibration correction and time lag correction can also be appended in this step. Finally, averaged data at different frequency can be obtained based on the QC'ed data. Intermediate matrices can be stored in '.MAT' format instantaneously. Although almost every part of the above process can be automated, but it is always instructive to have a manual inspection on the data problems.

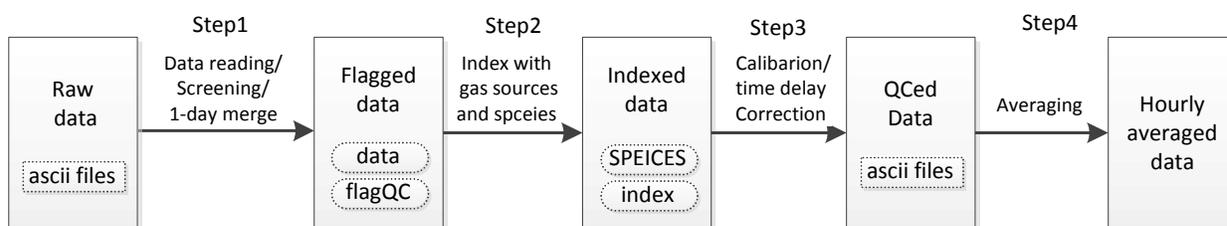


Figure 3: Flow diagram of the QC process

3.2 Step1: Data reading and data screening

3.2.1 Data reading

The first step of the QC process is reading the raw data from various files, and recognizing site numbers by file name. The QC process can directly read the compressed files and store data in a Matlab matrix. Lines with missing fields are deleted in this step. If there are more than one file for a day, these files are combined into one matrix. Absent data with the value -9999.99 are replaced by NaN. The time stamp consists of the “DATE” (column 1) and “TIME” (column 2). A MatLab function *datenum* is used to convert the time stamp (in string format) to serial date (floating number).

3.2.2 Data screening

The data screening of raw data is used to identify unreasonable values which are outside the normal range defined in the Table 5. The normal concentration ranges for CO₂ / CH₄ / CO are shown, and the data is erroneous if the values are out of this range. The ranges for “CavityPressure” and “CavityTemp” are the normal operating conditions for instrument which mean the data are not

very reliable if operating conditions are not met. There are yet other rules of data consistency that need to be checked. The date should be consistent with the date in file name and the “FRAC_DAYS_SINCE_JAN1” (column 3). The “CO2” (column 8) and “CO2_dry” (column 9) should be consistent as described in vapor correction of equation (1), and the “CH4” (column 5) and the “CH4_dry” (column 6) should be consistent as described in equation (2).

Table 5: Items of data screening

No.	Items	Column	Normal range	Notes
1	DATE, TIME	1, 2		Date and time should be in a correct format. The date should be consistent with the date in file name, and also the “FRAC_DAYS_SINCE_JAN1” (column 3).
2	ALARM_STATUS	4	0/65536	65536 is also acceptable.
3	CO2_dry	9	(10~10000) μ mole/mole	Vapor correction using equation (1).
4	CH4_dry	6	(0.1~1000) μ mole/mole	Vapor correction using equation (2).
5	CO	7	(0.01~100) μ mole/mole	
6	CavityPressure	10	(140 \pm 4) Pa	
7	CavityTemp	11	(45 \pm 0.02) °F	
8	solenoid_valves	20	0, 1, 2, 4, 8, 16	
9	Species	21	1, 2, 3, 4	
10	h2o_reported	18	(0~0.6)%	The vapor mixing ratio after Nafion drying system.
11	InletValve	14	50000	Inlet valve should be kept at a fixed position
12	OutletValve	15	15000~65000	Standard deviation in a running window of 15 data is acceptable in range of 0.1~1000.

* Items with NaN value will be ignored in the data screening procedure.

The data screening produces a matrix *flagQC* with a same size as *data* to flag if the certain value passes the data screening. The element of *flagQC* is 0 or 1, and 1 means the corresponding item fails the check. If there is a failed item, this line of data will be abandoned in the following steps and will not be distributed to the potential users. All the data including those failing the data screening are still stored in the INFLUX data servers for necessary inspection or rewinding.

There are also several different conditions for data correction. The first condition is that all the data from SITE01 failed the vapor correction check of “CO2_dry”. All “CO2” values are in the wrong order of magnitude (software setup problem), whereas the CO2_dry values are right. So “CO2” values are converted to NaN and vapor correction check is skipped over.

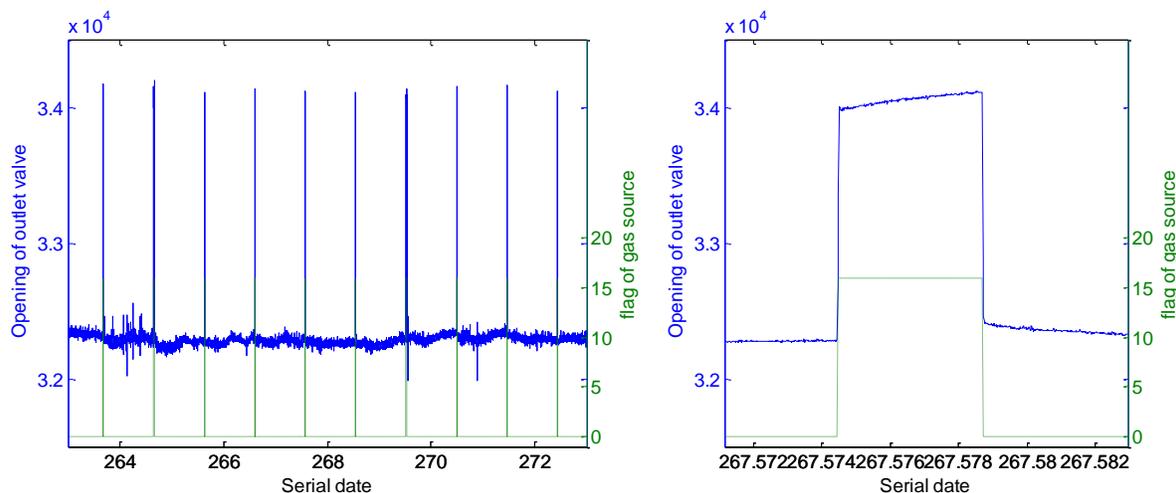
The “Solenoid_valves” (column 20) is an important parameter flagging the gas source which the instrument is measuring. There are some “solenoid_valves” with decimal values during the

transition period for most of sites as shown in Table 6 with a yellow color, and there are also some unreasonable integral transition values which are not for that sample height, like zeros in (b) and (c) of Table 6. Those are just software errors which don't influence the validity of the concentration of species. Those transition values are amended to be their former values to avoid losing those points.

Table 6: Software problems for the “solenoid_valves”

(a) only a decimal value				(b) two zeros before decimal value				(c) zero between decimal values			
DATE	TIME	valves	species	DATE	TIME	valves	species	DATE	TIME	valves	species
12/5/2013	3:32:57	2	1	12/6/2013	21:51:53	2.00	1	12/6/2013	4:32:43	2	1
12/5/2013	3:32:57	2	2	12/6/2013	21:51:54	2.00	2	12/6/2013	4:32:43	2	2
12/5/2013	3:32:58	2	4	12/6/2013	21:51:54	2.00	4	12/6/2013	4:32:44	2	4
12/5/2013	3:32:59	2	3	12/6/2013	21:51:55	2.00	3	12/6/2013	4:32:45	2	3
12/5/2013	3:32:59	2	1	12/6/2013	21:51:55	0.00	1	12/6/2013	4:32:45	2	1
12/5/2013	3:32:59	2	2	12/6/2013	21:51:56	0.00	2	12/6/2013	4:32:45	2	2
12/5/2013	3:33:00	2.53	4	12/6/2013	21:51:57	2.53	4	12/6/2013	4:32:46	2	4
12/5/2013	3:33:01	4	3	12/6/2013	21:51:57	4.00	3	12/6/2013	4:32:47	2	3
12/5/2013	3:33:01	4	1	12/6/2013	21:51:58	4.00	1	12/6/2013	4:32:47	2	1
12/5/2013	3:33:02	4	2	12/6/2013	21:51:58	4.00	2	12/6/2013	4:32:48	2	2
12/5/2013	3:33:03	4	4	12/6/2013	21:51:59	4.00	4	12/6/2013	4:32:49	1.53	4
12/5/2013	3:33:03	4	3	12/6/2013	21:52:00	4.00	3	12/6/2013	4:32:49	0.00	3
12/5/2013	3:33:05	4	2	12/6/2013	21:52:00	4.00	1	12/6/2013	4:32:49	3.19	1
12/5/2013	3:33:06	4	4	12/6/2013	21:52:00	4.00	2	12/6/2013	4:32:50	4	2
12/5/2013	3:33:06	4	3	12/6/2013	21:52:01	4.00	4	12/6/2013	4:32:51	4	4
12/5/2013	3:33:07	4	2	12/6/2013	21:52:02	4.00	3	12/6/2013	4:32:52	4	3

There are a few data from SITE11 without “solenoid_valves” values due to some reasons, and the reference gas and sampling gas values are in-correct. “solenoid_valves” has to be rebuilt using the relationship to other parameters. With correlation analysis, the “OutletValve” (column 15) can be regarded as the proper parameter and can be used to rebuild the flag of gas sources, as shown in Figure 4. The opening values larger than a certain value can be flagged as reference gas, and the rebuilt flag of reference gas are moved forward by 5 points to ensure the calibration are all valid.



(a) Relation for a 10-day period
 (b) Local zoom of a calibration period
 Figure 4: Relation between opening of outlet valve and flag of gas sources (SITE10)

3.2.3 Review of the failed points

After the automatic data screening, the failed points should be inspected manually to ensure the data screening is effective. For every site, there is a log file to record the information of the failed points, as shown in Figure 5, including the column flag, the line number, and the corresponding raw data of the failed lines.

```

SITE01-CFRADS2025-20131201-000035Z-DataLog_User-col.dat.gz
timeConsistency,ALARM_STATUS, CO2_dry,CH4_dry,CO, cellPres,cellTemp, gasSourceFlag,SpeciesFlag
0 0 0 0 0 1 0 0 0 :78441 735569.515556 0 409.161504 2024.002275 150.704773 143.738541 44.998009 8.000000 3
0 0 0 0 0 1 0 0 0 :79699 735569.523877 0 417.124301 2462.268807 236.621579 137.609565 45.001823 0.000000 2
0 0 0 0 0 1 0 0 0 :79700 735569.523888 0 417.124301 2462.268807 231.326804 137.609565 45.001823 0.000000 4
0 0 0 0 0 3 0 0 0 :NaN
SITE01-CFRADS2025-20131201-161118Z-DataLog_User-col.dat.gz
timeConsistency,ALARM_STATUS, CO2_dry,CH4_dry,CO, cellPres,cellTemp, gasSourceFlag,SpeciesFlag
0 0 1 1 1 0 1 0 0 :1 735569.674463 0 0.000000 0.000000 -1.295900 139.993130 44.900871 0.000000 1
0 0 1 1 1 0 1 0 0 :2 735569.674470 0 0.000000 0.000000 -1.295900 140.017665 44.900871 0.000000 2
0 0 1 1 0 0 1 0 0 :3 735569.674497 0 0.000000 0.000000 158.278346 140.017665 44.900871 0.000000 4
0 0 1 1 0 0 1 0 0 :4 735569.674508 0 0.000000 0.000000 158.278346 139.983109 44.902550 0.000000 3
    
```

Figure 5: Appearance of log file

If the log file is not clear enough to inspect the failed points, a scatter figure as shown in Figure 6 can also be used where the red points show the failed points. Such plots may be easier to find and confirm the existing problems in the datafiles.

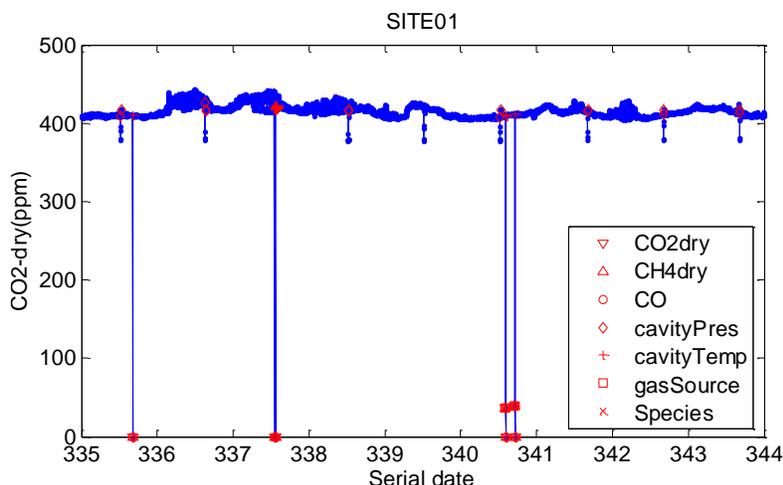


Figure 6: Scatter figure for inspecting the unreasonable raw data

3.3 Step2: Separation and calibration

There are usually several sampling heights using only one instrument for some sites. Figure 7 shows five gas sources including three sampling heights and two reference gases at SITE01. Different gas sources use different “solenoid_valve” value, which is 0, 2, and 4 for sampling gas from 121 m, 40 m and 10 m respectively, and 8, 16 for the two reference gases. In addition, the CRDS instrument can measure several species in turn and update them in the data file one by one. All the data are stored in the same file, and it is very important to separate them by the different gas sources and species.

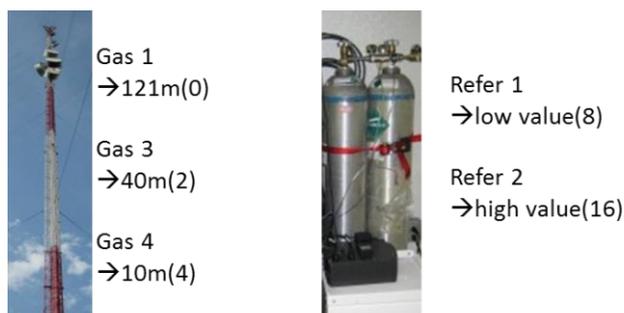


Figure 7: Sketch of sampling gas and reference gas (SITE01)

First of all, the data of different species can be easily separated (column 21) to get the corresponding dry concentration. The separation of different gas sources is complicated. It takes several seconds to make the sampling gas flow to instrument cavity and stabilize the measurement value, after “solenoid_valve” changes. The data in first several seconds should be abandoned, so it is critical to find the point at which the “solenoid_valve” value changes.

In Figure 8, the dotted lines with different color are the last point before “solenoid_valve” changes, and the dashed lines are that for reference gas. The data in the first several seconds of every

sampling height are abandoned (flagged with red points), and only the points in the last several minutes are kept for the reference gas. The criteria of exact time for each sites will be discussed in section 4.1. It is also necessary to inspect the plot manually for the point-wise errors such as spikes and jumps in the middle of sampling period of each height.

The points in each group of sampling height are acceptable only if they are more than 20. For the first group of data in a file, it needs to consider the last group of data in the previous file to decide the total number of points acceptable or not.

There are some time gaps in or between files. If the time gaps are larger than 1 min, data in first several seconds are also abandoned to avoid problems occurring frequently after the instrument recovers from malfunction.

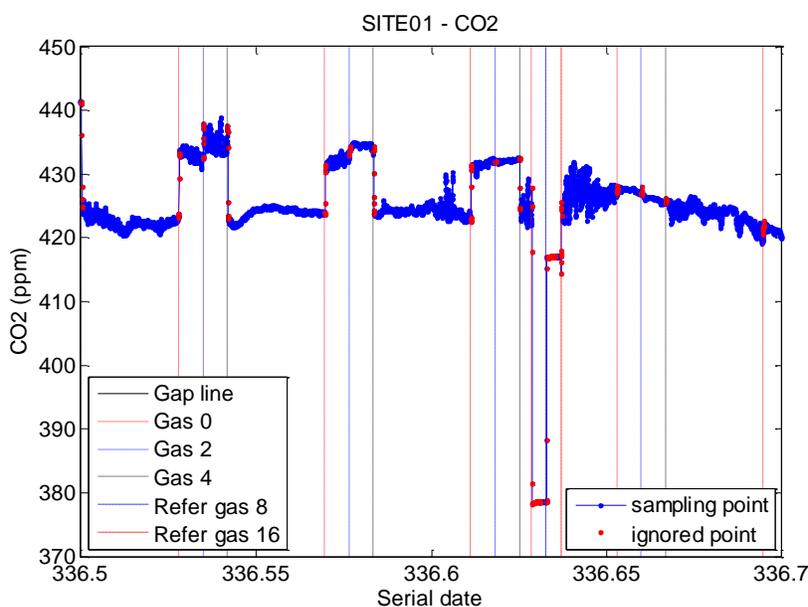


Figure 8: Separation based on the gas sources

The outputs of separation are two matrixes for each species, *SPECIES* and *index*. The four columns of *SPECIES* are serial date, concentration, “solenoid_valve” and validness flag. The first column of *index* is the line number of last point before “solenoid_valve” changes which is the dot lines and dash lines in Figure 8, and the second column is “solenoid_valve”. With the two matrixes, data for every species and gas sources can be separated clearly.

If “solenoid_valve” of *index* is 16 or 8, the corresponding group of data are reference gas, and the averaged value C_m in the last several minutes are also calculated and stored as the third column of *index*. For the following lines of *index* for sampling gas, the third column will be the measurement value of previous “solenoid_valve” of 16, and the fourth column will be the value of previous “solenoid_valve” of 16. If there is just one reference gas, the forth column will be NaN. If no reference gas before, the third or fourth column will also be NaN.

	1	2	3	4
2284	7.3557e+05	411.7361	0	0
2285	7.3557e+05	411.7626	0	0
2286	7.3557e+05	411.7163	0	0
2287	7.3557e+05	411.7790	0	0
2288	7.3557e+05	411.7603	0	0
2289	7.3557e+05	411.7966	0	0
2290	7.3557e+05	411.8055	0	0
2291	7.3557e+05	411.8207	0	0
2292	7.3557e+05	411.7983	2	1
2293	7.3557e+05	411.5172	2	1

	1	2	3	4
31	14308	4	NaN	NaN
32	14472	8	378.9944	NaN
33	14636	16	411.0951	NaN
34	15404	0	411.0951	378.9944
35	15677	2	411.0951	378.9944
36	15952	4	411.0951	378.9944
37	17048	0	411.0951	378.9944
38	17322	2	411.0951	378.9944
39	17596	4	411.0951	378.9944
40	18693	0	411.0951	378.9944
41	18967	2	411.0951	378.9944

Figure 9: Matlab matrix *SPECIES* and *index* for the indexed data

3.4 Step3: Output of QC’ed data after calibration and time delay correction

Based on the indexed data, the QC’ed data can be calculated with the valid data of *SPECIES* (the lines with 4th column of 0) and the line numbers (1st column) in *index* for each sampling height. The calibration correction is applied in this step. The difference $\Delta C_{calib} = C_m - C_r$ between measured value C_m and reference value C_r are used to correct the sampling gas till the next valid measurement of reference gas. The reference value C_r is from the laboratory precise calibration of the reference gas. If there are two reference gases at one site, their averaged value will be used. For some days without valid calibration, the previous calibration will be used.

There is also a time lag correction which is related to the flow time from the sampling point to the cavity of instrument. The time lag Δt will be subtracted from the time stamp of raw data before output of the QC’ed data. It is calculated as following,

$$\Delta t = \pi r^2(h + L)/q \tag{3}$$

Where, r is the radius of tubing which is 5/32 in for those towers, h and L are the sampling heights and extra length from the bottom of tower to the instrument respectively, and q is the flow rate of the gas which is 140 mL/min for CADS system (SITE04/06/07) and 240 mL/min for others sites. The calculated time lags for different sites and different sampling heights are shown in Table 7.

Table 7: Time lag for different sites and sampling heights

Site No.	Extra length (m)	SolenoidValves = 0		SolenoidValves = 1		SolenoidValves = 2		SolenoidValves = 4	
		height(m)	timeLag(s)	height(m)	timeLag(s)	height(m)	timeLag(s)	height(m)	timeLag(s)
SITE01	10	121	6.75			40	2.58	10	1.03
SITE02	10	136	7.53			40	2.58	10	1.03
SITE03	6	54	3.09	40	2.37	20	1.34	10	0.82
SITE04	10	60	6.19						
SITE05	10	125	6.96						
SITE06	10	39	4.33						

SITE07	5	58	5.57						
SITE08	10	41	2.63						
SITE09	10	130	7.22						
SITE10	8	40	2.47						
SITE11	12	130	7.32						
SITE13	9	87	4.95						

After the two corrections, the QC'ed data can be outputted into ASCII files for different sites, different heights and species. Appearance of output of QC'ed data at raw frequency is shown in Figure 10, including site name, sampling height, date and time, serial date, and the specie concentration.

Site	z (mAGL)	Date	Time	SerialDate	CO2 (ppm)
SITE01	121	2013-12-01	00:11:08	335.007731	408.11
SITE01	121	2013-12-01	00:11:10	335.007757	408.11
SITE01	121	2013-12-01	00:11:12	335.007783	408.08
SITE01	121	2013-12-01	00:11:15	335.007810	408.05
SITE01	121	2013-12-01	00:11:17	335.007836	408.04
SITE01	121	2013-12-01	00:11:19	335.007862	408.07
SITE01	121	2013-12-01	00:11:22	335.007888	408.11
SITE01	121	2013-12-01	00:11:24	335.007914	408.07
SITE01	121	2013-12-01	00:11:26	335.007940	408.07
SITE01	121	2013-12-01	00:11:28	335.007967	408.06
SITE01	121	2013-12-01	00:11:31	335.007993	408.09
SITE01	121	2013-12-01	00:11:33	335.008019	408.08

Figure 10: Appearance of output of QC'ed data at raw frequency

3.5 Step4: Averaging at different time intervals

The QC'ed data at raw frequency are too huge for potential users. It is necessary to average them at different time intervals. The QC process can also output the averaged data based on the QC'ed data obtained in STEP3. Figure 11 gives an example of the format for hourly averaged data, including mean concentration, standard deviation and count in this hour period. The Time (column 4) is the start time of the hour for the corresponding line. The count is the number of valid points in this hour.

Site	z (mAGL)	Date	Time	serialDate	CO2 (ppm)	std (ppm)	count	CH4 (ppb)	std (ppb)	count	CO (ppb)	std (ppb)	count
SITE01	121	2013-12-01	00:00:00	335.000000	408.19	0.12	1050	1975.1	1.5	1050	155.5	4.8	1050
SITE01	121	2013-12-01	01:00:00	335.041667	407.72	0.23	1047	1965.5	3.3	1047	153.5	5.3	1047
SITE01	121	2013-12-01	02:00:00	335.083333	407.83	0.22	1048	1962.2	1.3	1047	153.6	5.2	1047
SITE01	121	2013-12-01	03:00:00	335.125000	409.10	0.21	1046	1961.7	1.7	1047	159.9	5.6	1047
SITE01	121	2013-12-01	04:00:00	335.166667	410.00	0.21	1046	1958.4	1.7	1046	154.0	5.5	1046
SITE01	121	2013-12-01	05:00:00	335.208333	409.86	0.41	1046	1962.1	0.9	1047	153.4	4.9	1046
SITE01	121	2013-12-01	06:00:00	335.250000	408.87	0.21	1045	1966.1	2.9	1045	145.4	5.2	1046
SITE01	121	2013-12-01	07:00:00	335.291667	408.19	0.16	1044	1980.4	1.7	1043	145.9	4.9	1043
SITE01	121	2013-12-01	08:00:00	335.333333	408.54	0.14	1047	1989.0	2.0	1046	145.7	4.7	1046
SITE01	121	2013-12-01	09:00:00	335.375000	408.99	0.09	1045	1996.5	1.1	1044	147.0	4.7	1045
SITE01	121	2013-12-01	10:00:00	335.416667	409.26	0.11	1044	2003.7	2.5	1044	148.5	4.9	1044
SITE01	121	2013-12-01	11:00:00	335.458333	409.41	0.09	1044	2014.3	2.2	1044	149.7	4.7	1045
SITE01	121	2013-12-01	12:00:00	335.500000	409.18	0.13	719	2026.8	1.8	718	150.4	5.1	718

Figure 11: Appearance of output of QC'ed data at 1h frequency

3.6 Discussion on the QC parameters

3.6.1 Valid calibration time

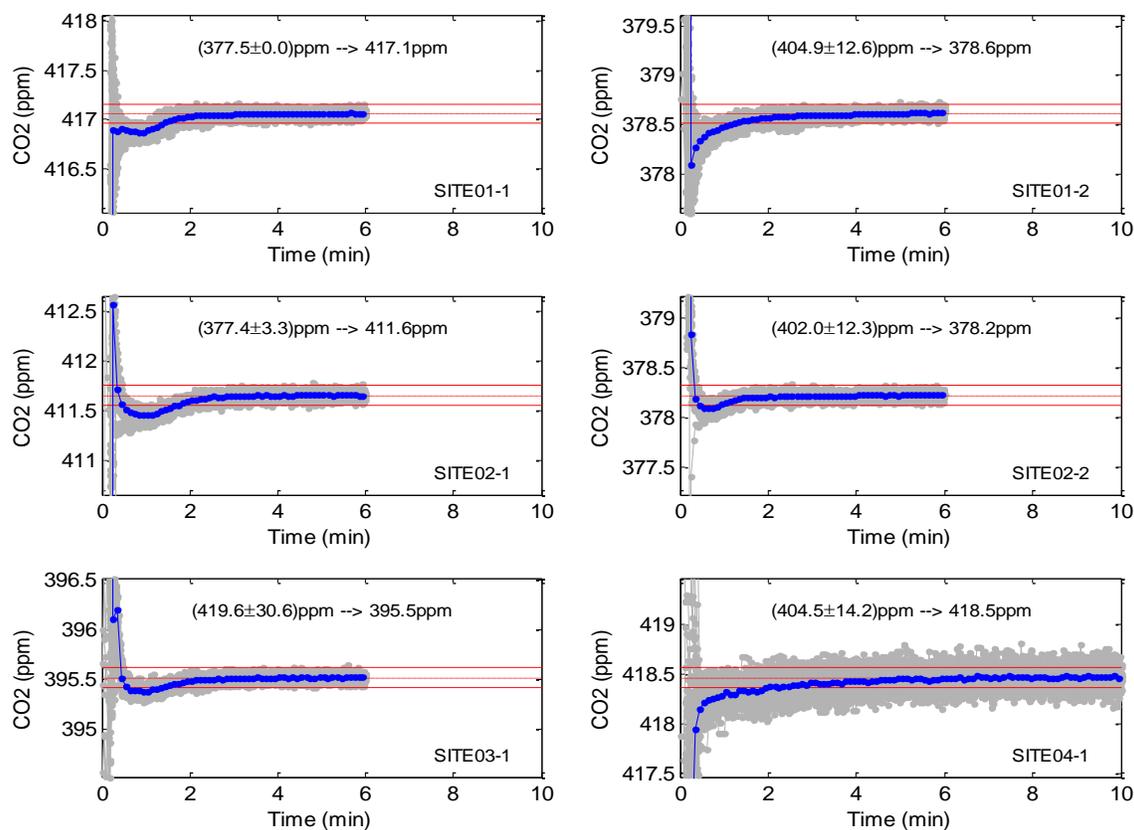
The data of the reference gas for the first several minutes are not valid. It takes several minutes to make sure the reference gas has arrived in the instrument cavity and the measurement value has been stabilized. The Figure 12 shows the time series of concentration of different species after “solenoid_valve” changes to the reference gas for different sites. To wipe off the change between different calibrations, the average measurement value C_{mi} has been converted to the reference value C_r , so the time series of concentration are

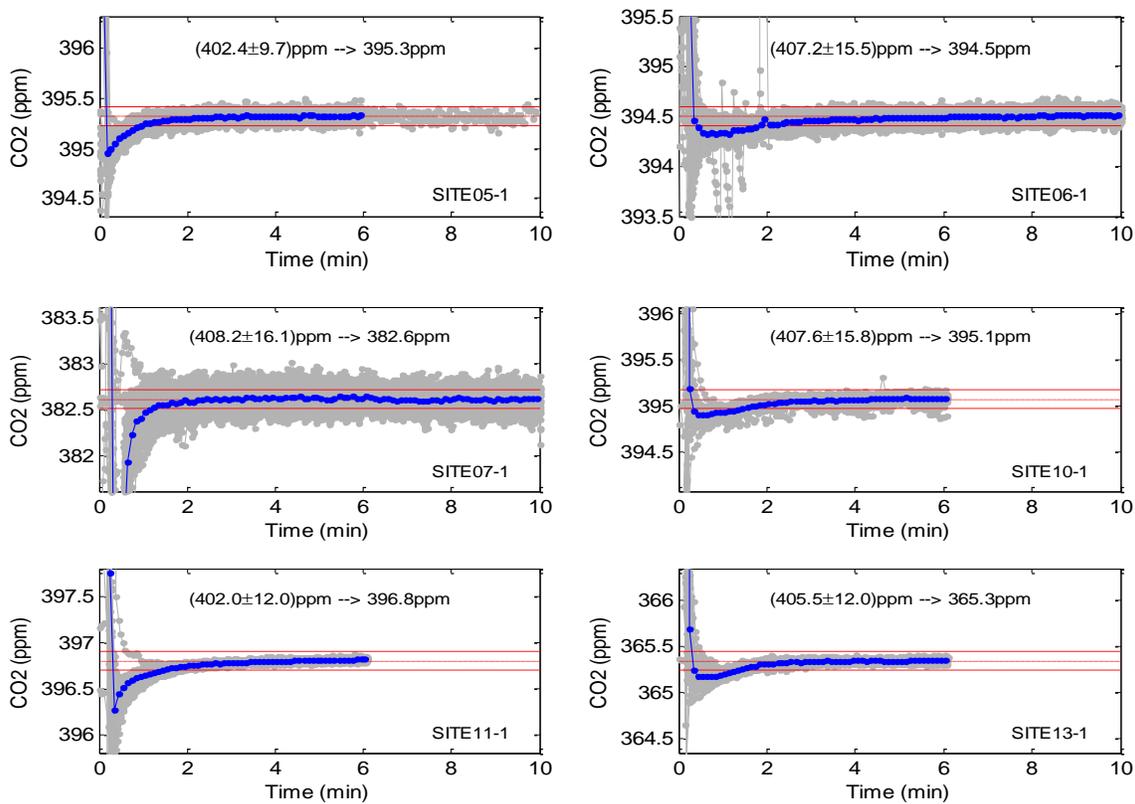
$$C'_i = C_i \times \frac{C_r}{C_{mi}} \quad (4)$$

where C'_i , C_i are the converted and original concentration respectively.

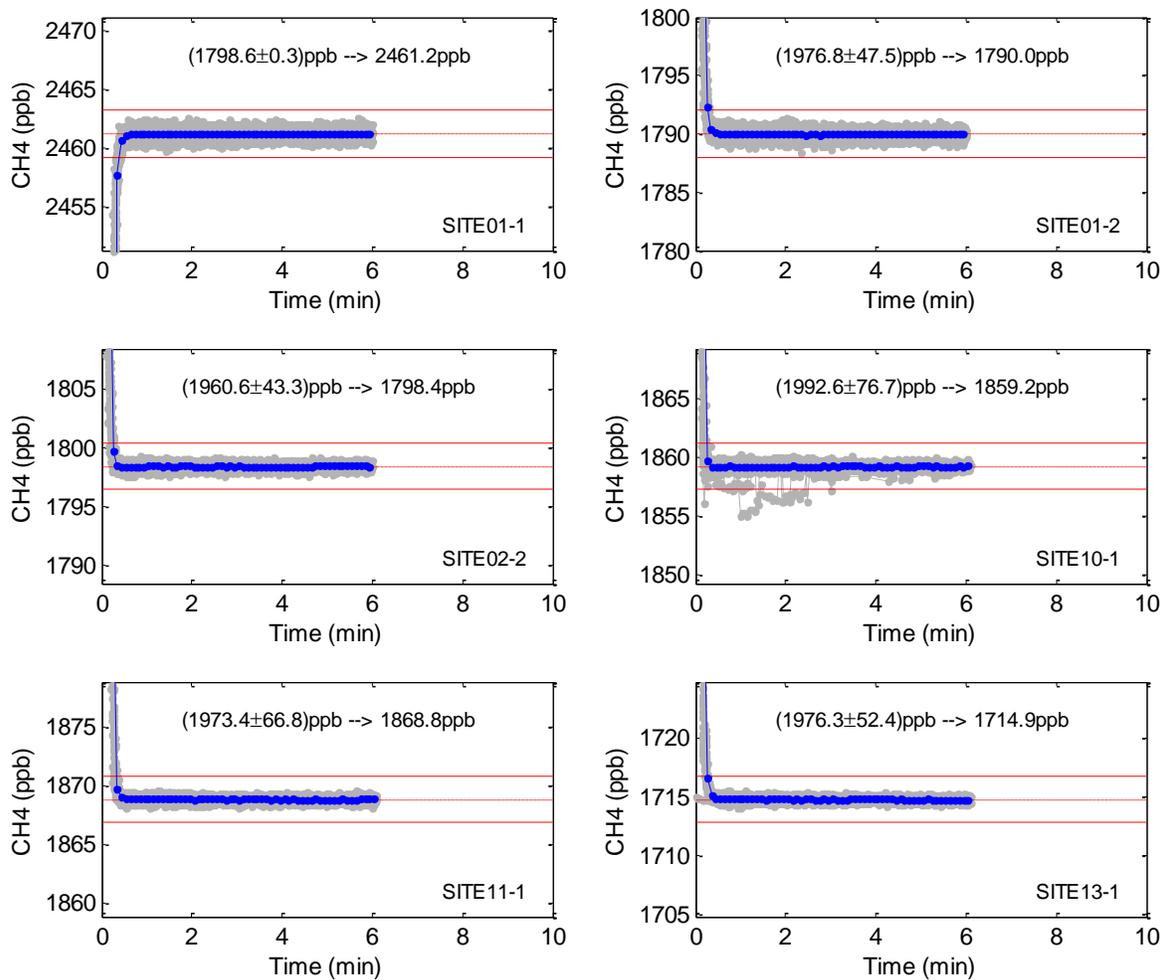
For the trend of CO₂ in the calibration, the red dash line is the mean value of the reference gas, and the two red solid lines are the limit of $\pm 0.1 \mu\text{mole/mole}$. The stabilized value of CO₂ can fall in the range of $\pm 0.1 \mu\text{mole/mole}$ for most of the sites except for SITE04/06/07 where old-type instruments with low accuracy are used. The calibration period is also a little longer for those three sites to ensure the calibration accuracy. The valid calibration are data in last few minutes, which is about 5 min for SITE04/06/07 and 3 min for other sites.

For the trend of CH₄, the transition seems to be more abrupt, and the transition times are shorter, about 0.5 min for all the sites.

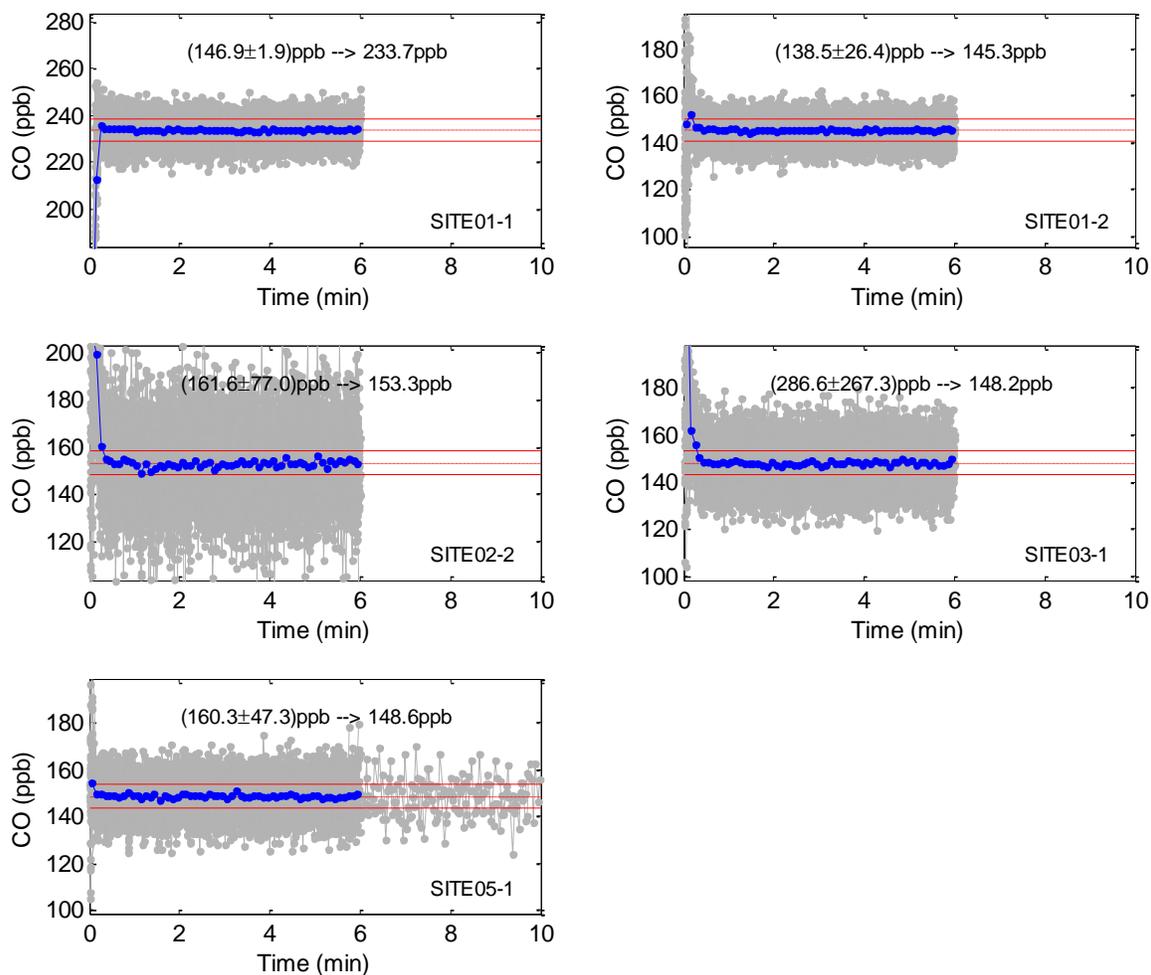




(a) Trend of CO₂



(b) Trend of CH₄



(c) Trend of CO

Figure 12: Trend of calibration data for species at different sites

The blue lines are the median value for the converted CO₂ values which show the trend after “solenoid_valve” changes. The valid time for the calibration has been analyzed by those blue lines. The text in each figure frame gives the difference between concentration before calibration and the reference value. The largest difference is 40 μmole/mole at SITE13, and the smallest difference is 5 μmole/mole for SITE11. The values of those difference seem to have no impact on the trend of time series. To ensure the measurement values have stabilized, the valid calibration times are set to be the last 5 min for SITE04/06/07 and the last 3 min for other sites, as shown in Table 8.

Table 8: Valid time for calibration of different sites

Site No.	Species	Model	valid time (min)
SITE01/02/08	CO ₂ /CO/CH ₄ / H ₂ O	G2401	3.0
SITE03/05/09	CO ₂ /CO/ H ₂ O	G2302	3.0
SITE10/11/13	CO ₂ /CH ₄ /H ₂ O	G2301	3.0
SITE 04/06/07	CO ₂ / H ₂ O	G1301	5.0

3.6.2 Rejection of outliers in the valid time period

For the calibration data of SITE10, there are some low value points, numerous enough to influence the calibration, as shown in Figure 13(a). Due to those abnormal points, the standard deviation of the calibration data for some days can grow over 0.1 $\mu\text{mole/mole}$ which is only 0.02 $\mu\text{mole/mole}$ for the normal days. Those low value points should be flagged and rejected from the valid calibration points as shown in Figure 13(b). The difference between the averaged calibration value with and without those low value points is 0.06 $\mu\text{mole/mole}$ for the case of Figure 13(b).

The criterion for the abnormal points is set at $\pm 0.07 \mu\text{mole/mole}$ from the averaged value of possible normal points. The normal points are those in the last 3 min and within range of 1 sigma (standard deviation). The abnormal points are shown in Figure 13(b) with red.

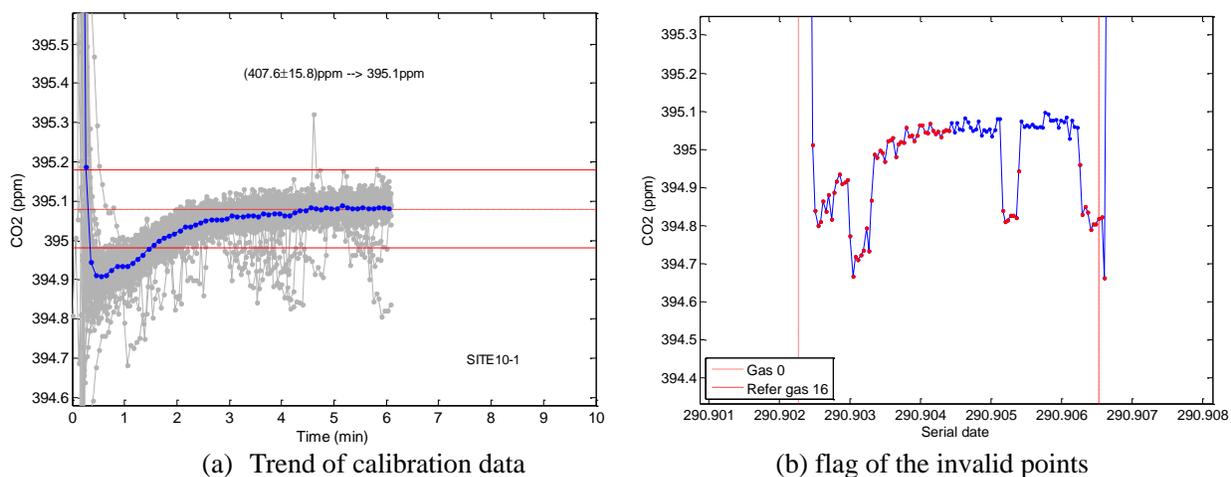
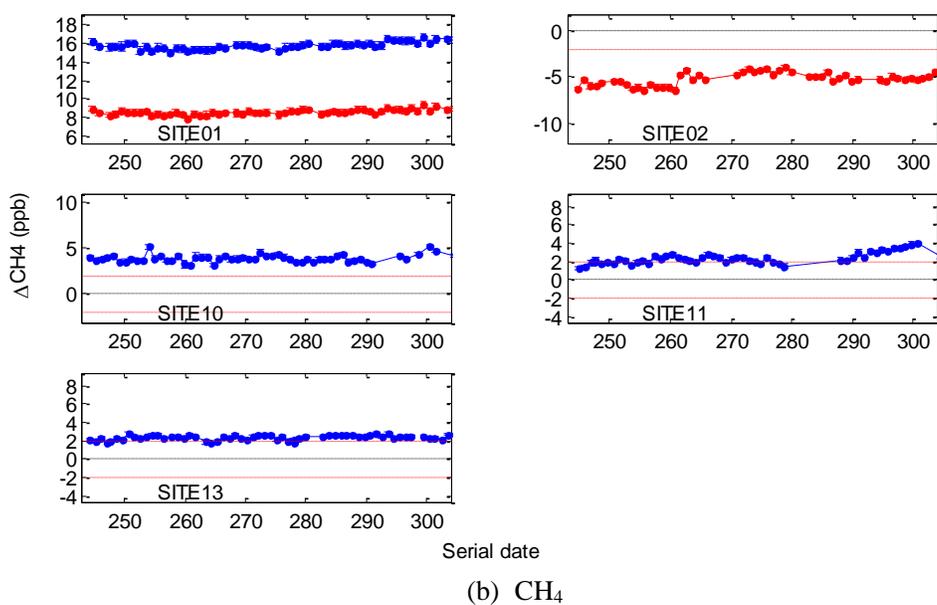
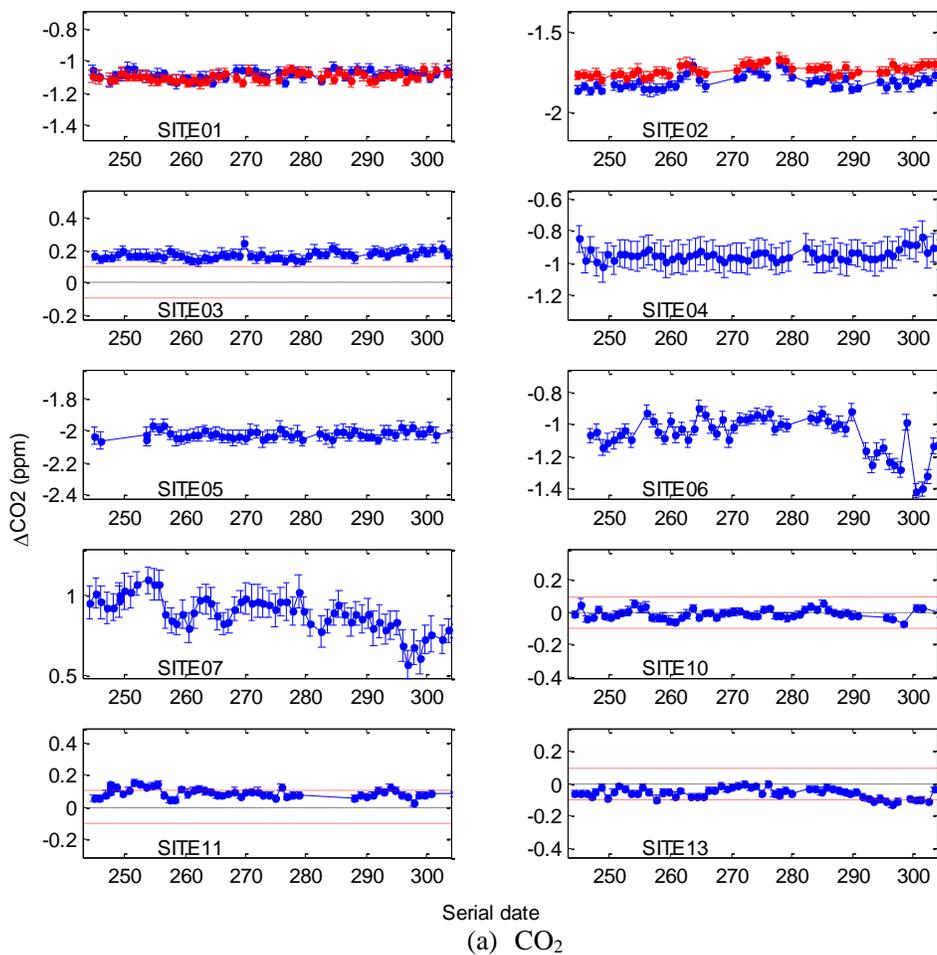
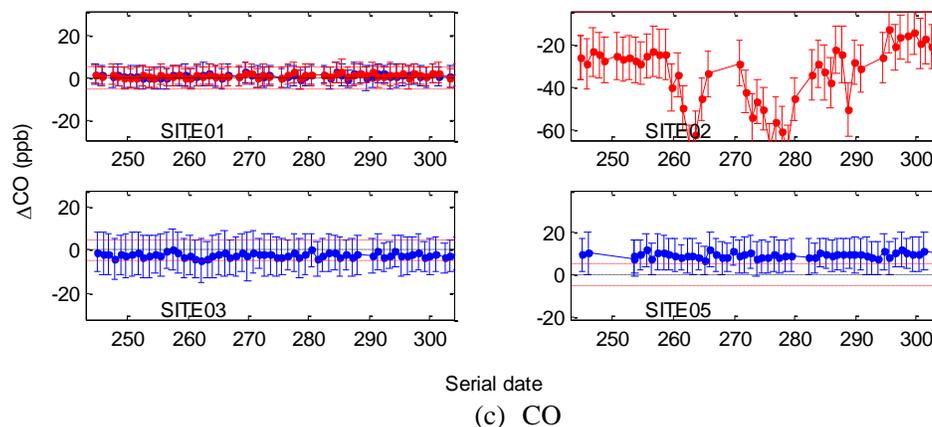


Figure 13: Rejection of the abnormal points for calibration of SITE10

3.6.3 Criteria of the standard deviation

Figure 14 shows the calibration results for the three species at all the sites. The standard deviation of single calibration and the inter-calibration change will be discussed. Criteria of accepting the calibration data needs to consider the high limit of standard deviation and inter-calibration change of them.





(c) CO
Figure 14: Calibration results

From the statistical results shown in Table 9, the criterion of standard deviation for each calibration can be obtained as shown in Table 10. CH₄ calibration is better and CO calibration is a little worse than the WMO standard which is 0.1 μmole/mole for CO₂, 2 ppb for CH₄, 5 ppb for CO. The mean values of correction are very large for some sites although they are very stable. Those calibrations are valid, but the instruments needs to be inspected and the reason of for such large difference should be found.

Table 9: Statistical result of calibration

site	num	CO ₂ (μmole/mole)				CH ₄ (ppb)				CO (ppb)			
		single standard		correction		single standard		correction		single standard		correction	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
SITE01-1	77	0.03	0.00	-1.08	0.03	0.35	0.04	15.81	0.39	4.8	0.4	1.1	0.8
SITE01-2	76	0.03	0.00	-1.10	0.02	0.31	0.05	8.56	0.28	4.7	0.4	1.3	0.8
SITE02-1	77	0.03	0.00	-1.81	0.04	0.21	0.03			11.0	1.0		
SITE02-2	74	0.03	0.00	-1.74	0.03	0.21	0.03	-5.21	0.68	11.0	0.8	-34.5	14.7
SITE03	66	0.03	0.00	0.17	0.02					8.8	1.1	-2.3	1.1
SITE04	136	0.09	0.01	-0.95	0.03								
SITE05	64	0.04	0.00	-2.02	0.02					7.7	0.7	9.2	1.2
SITE06	209	0.04	0.00	-1.06	0.12								
SITE07	186	0.09	0.01	0.88	0.11								
SITE10	53	0.02	0.00	-0.01	0.03	0.21	0.05	3.91	0.41				
SITE11	63	0.02	0.00	0.09	0.03	0.19	0.03	2.37	0.60				
SITE13	53	0.02	0.00	-0.05	0.03	0.17	0.02	2.39	0.26				

Table 10: Criteria of the calibration

Site	CO ₂		CH ₄		CO	
	single standard	correction	single standard	correction	single standard	correction
SITE4/6/7	0.12				15	
Others	0.07		0.5		15	

3.6.4 Waiting time for separation

Waiting time before getting valid data is not only necessary for reference gas but also necessary for sampling gas. Data of sampling gas in the first several seconds should also be ignored when we extract valid data from original files. The waiting time includes the gas flow time and instrument response time. Figure 15 shows the response of CO₂ concentration after the gas source is changed from sampling gas to reference gas. The concentration gradually goes into the range of 0.1 μmole/mole after about 1.5 min, and totally stabilized after about 3.0 min.

Data of last several minutes can be used to catch the averaged value for reference gas whose concentration is constant, whereas more data more representational to get the averaged value for a certain period for sampling gas whose concentration fluctuates markedly. The waiting time for sampling gas needs to consider both good representativeness using more data and abandonment of data with biases. It is 1.5 min for SITE11, which is the time that the concentration just goes into the range of ±0.1 μmole/mole as shown in Figure 15. Table 11 gives the waiting time for all the sites. A large amount of data are involved especially for those sites with multiple sampling heights.

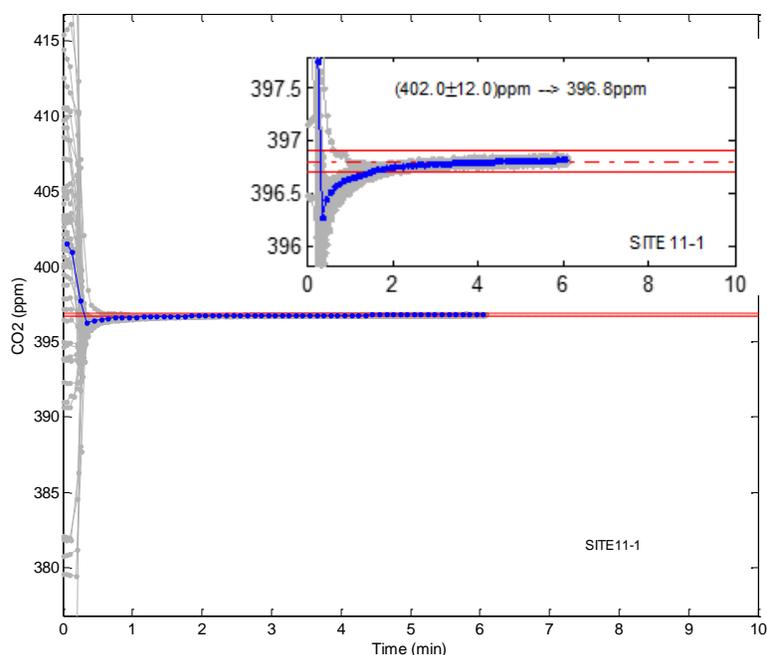


Figure 15: Response after changing to reference gas

Table 11. Waiting time for sampling height transfer

Site No.	Species	Model	Waiting time (min)
SITE01/02/08	CO ₂ /CO/CH ₄ / H ₂ O	G2401	1.5
SITE03/05/09	CO ₂ /CO/ H ₂ O	G2302	1.5
SITE10/11/13	CO ₂ /CH ₄ /H ₂ O	G2301	1.5
SITE 04/06/07	CO ₂ / H ₂ O	G1301	2.0

4 Comparisons with other process

Pennsylvania State University (PSU) is in charge of the tower measurement for the INFLUX project. They have developed a quality control process for the raw data sampled from the towers in Indianapolis. Based on the same raw data of two months in 2013, QC has been done by both of PSU process and NIST process. The hourly averaged data obtained have been compared and analyzed.

4.1 Differences of the QC process

The basic description of PSU process is listed below,

- Script to extract zip files from email.
- Script to reorganize columns to a common format and check for wild values.
- Ignore data for some amount of time (depending on flow rates, etc.) between different heights and sample / reference gas.
- Ignore nonsensical values for flags.
- Compare the reference gas measured values to known value.
- Apply adjustment based on above results to entire day of data.
- Apply adjustment for time for gas to go from top of tower to instrument (at our flow rates, this is about 5-10 min).
- For profile sites, we sample top level for 30-40 min, other levels for 10 min of each hour.
- Calculate hourly average for each level.
- Check flow rates, H₂O, reference gas values (stability).

The processes of PSU and NIST are mainly similar, and there are only some differences in the operation sequences and parameters which result in some differences of the QC results. The main differences of QC process are listed as below.

Table 12: Comparison of QC process between NIST and PSU

Category	Item	NIST	PSU
Data screening	Dry concentration	Abandon the lines with “CO ₂ _dry”, “CH ₄ _dry” or “CO” out of range.	Similar check
	Pressure and temperature in the cavity	Abandon the lines with “CavityPressure” out of (140±4) Pa and “CavityTemp” out of (45±0.02) °F.	
	Outlet proportional valve	Except the range check of [2E4, 5.8E4], the running standard deviation check is also applied with a range of [0.1,1E3]	Just range check of [2E4, 5.8E4]
	SolenoidValves	Unreasonable transition values are amended to be their previous reasonable values.	Delete the points with the incorrect transition values.
	Time gap	Delete the data in the first several seconds after the time gap larger than 1 min	

Separation	Species	Separation depending on the species column of raw data	Update every N lines (N is the species number)
	Transition time	2 min for the sites with old instrument, and 1.5 min for other sites	ignore 4 min after each transition for some sites
Calibration	Valid time	Last 5 min for the sites with old instrument, and last 3 min for other sites	for reference gas, last 2-3 min
	Data range of calibration correction	The current calibration to the next calibration	Apply the calibration for the same day
	standard deviation	Upper limit is 0.07 $\mu\text{mole/mole}$ for CO_2 , and 0.5 ppb for CH_4	
Averaged output	Time lag	Added before averaging	Added after averaging
	Output items	Mean value, standard deviation of concentration, and count of valid points.	Mean value, standard deviation, and uncertainty of calibration

For the sites with only one height and with multiple heights, there are some different problems faced. So the comparison will be done separately.

4.2 Comparison of sites with only one sample height

The comparison of the QC'ed CO_2 hourly data for several days of SITE13 is shown in Figure 16. The trends of CO_2 concentration are consistent with only a few concentration difference at some certain points.

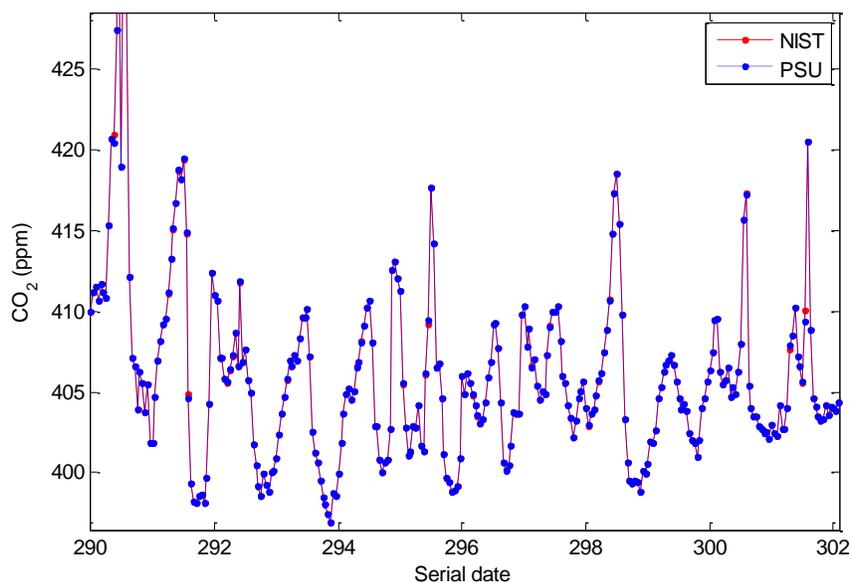


Figure 16: Comparison of the QC'ed hourly data (SITE13)

The concentration differences $C_{NIST} - C_{PSU}$ are plotted directly in Figure 17 to make them clearer. Most of the data points show very good consistency, there are only a few differences in the first

day or scattered along the time series. The differences on the first day are due to the calibration correction. PSU process applies the calibration correction on the data in the same day, whereas NIST process apply the calibration correction after reference gas are measured which makes no calibration correction for the first-day data before reference gas measured. The scattered difference points are usually in the wake of reference gas where the first several seconds of data are ignored. Some of the concentration differences are obvious when the concentrations of ignored data differ greatly from other data in this hour period.

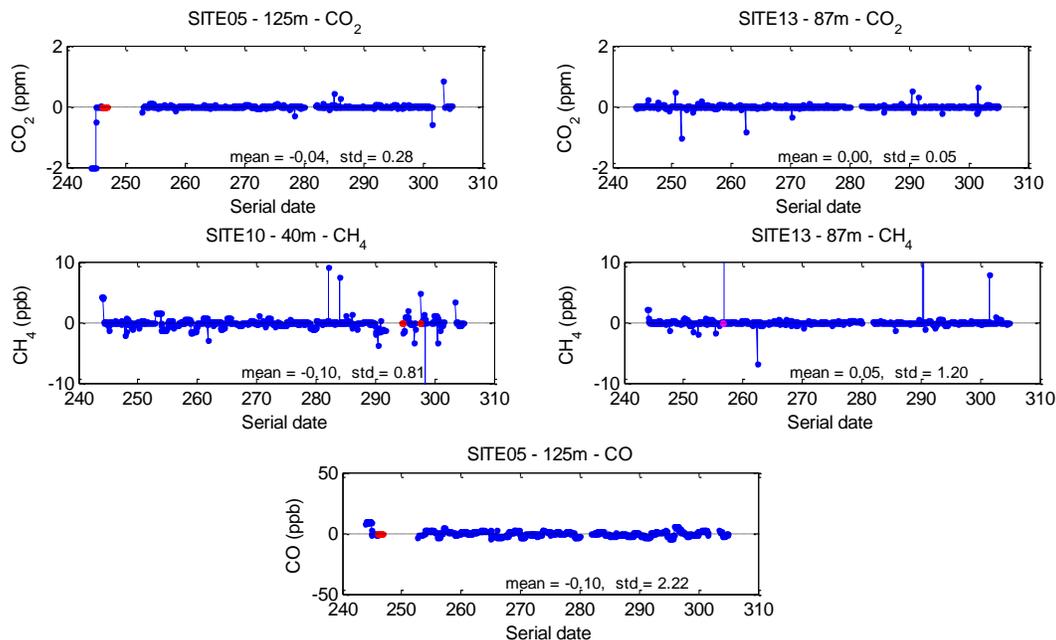


Figure 17: Comparison of the single-height sites

Table 13 gives the mean values and corresponding standard deviation for these difference. Although there are some CO₂ difference of 1~2 μmole/mole at some certain points, means of differences for all the points are very small and close to zero, and the standard deviations are also not very large. Only SITE11 has larger difference for both of CO₂ and CH₄, because there is no original “solenoid_valves” and rebuilding of them results in some difference in calibration correction.

Table 13: Difference of mean value and standard deviation (single-height sites)

Site	CO ₂		CH ₄		CO	
	MEAN	STD	MEAN	STD	MEAN	STD
SITE04	-0.03	0.12				
SITE05	-0.04	0.28			-0.10	2.22
SITE06	-0.04	0.21				
SITE07	0.00	0.13				
SITE10	-0.01	0.08	-0.1	0.81		
SITE11	-0.13	0.04	1.68	0.75		

SITE13	0.00	0.05	0.05	1.20		
--------	------	------	------	------	--	--

4.3 Comparison of sites with multiple sample heights

For the sites with multiple heights, it is more complicated. As shown in Figure 18 and Table 14, the means of difference for sites are still small, but the standard deviations are larger than these sites with only single sampling height. The scattered differences are due to different number of ignored points after “Soleniod_valve” changes. More sampling heights produce more changes of “Soleniod_valve”, and then result in more differences of ignored point number. But the mean value of these differences can stay very small, because influence of the ignored points is neither only positive nor only negative.

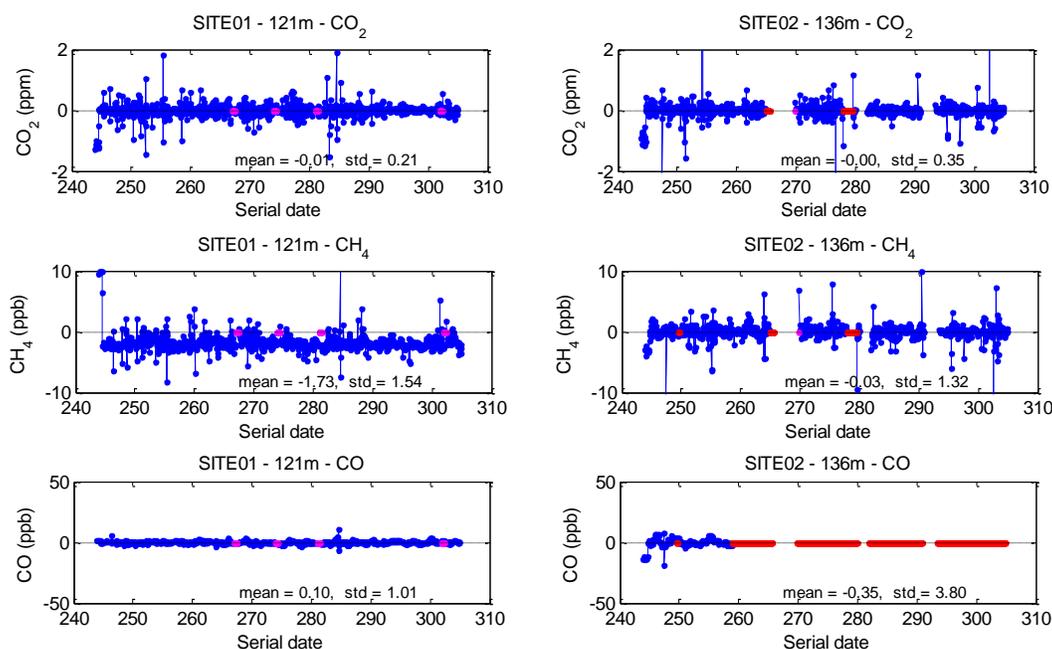


Figure 18: Comparison of multiple-height sites

Table 14: Difference of mean value and standard deviation (multi-height sites)

SiteNo-Height	CO ₂		CH ₄		CO	
	MEAN	STD	MEAN	STD	MEAN	STD
SITE01-121m	-0.01	0.21		1.54	0.10	1.01
SITE01-40m	-0.01	0.39		8.07	0.02	4.10
SITE01-10m	-0.03	0.79		11.05	0.99	37.56
SITE02-136m	0.00	0.35	-0.03	1.32	-0.35	3.80
SITE02-40m	0.01	1.00	0.14	7.96	-0.48	10.42
SITE02-10m	0.01	0.92	0.20	3.42	-0.32	7.40
SITE03-54m	0.00	1.30			0.15	9.22
SITE03-40m	0.00	0.87			0.19	7.41

SITE03-20m	0.02	0.81			0.41	8.02
SITE03-10m	-0.03	1.47			-0.44	13.41

5 Conclusion

Uncertainty of observation is very important for the inversion of emission fluxes. A small bias in the observation data can result in a large bias in the computed fluxes, because anthropogenic sources result in a small enhancement to atmospheric concentrations relative to the background.

The accuracy of GHG observation data can be influenced by the QC. Different QC processes are used by different organizations, which can result in non-ignorable difference. Quality control processes should be standardized to reconcile the possible differences.

A comprehensive QC process is developed, and QC'ed data for two months are compared with that from PSU. The influences of different methodologies in the QC process are analyzed, and the important parameters are optimized to reconcile the differences.

Acknowledgements

Thanks to Kenneth Davis, Natasha Miles, Scott Richardson of Pennsylvania State University for their QC'ed data and useful discussion on the QC issues.

Reference

- [1] Y Tsutsumi, K Mori, T Hirahara, et al. Technical Report of Global Analysis Method for Major Greenhouse Gases by the World Data Center for Greenhouse Gases. World Meteorological Organization, 2009.
- [2] Thomas Lauvaux, Andrew E. Schuh, Marc Bocquet, Lin Wu, Scott Richardson, et al. Network design for mesoscale inversions of CO₂ sources and sinks. *Tellus B* 2012, 64, 17980.
- [2, 3] M. O. L. Cambaliza, P. B. Shepson, D. R. Caulton, et. al.. Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions. *Atmos. Chem. Phys.*, 2014, 14, 9029–9050.
- [4] I Zahumenský. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. WMO, 2004.
- [5] 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011). GAW Report No. 206, 2011
- [6] A E Andrews, J D Kofler, M E Trudeau, et al. CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network. *Atmos. Meas. Tech.*, 2014, 7: 647-687.
- [7] U.S. Environmental Protection Agency. Quality Assurance Handbook for Air Pollution Measurement Systems (Volume II Ambient Air quality Monitoring Program). EPA-454/B-13-003, 2013.

- <http://www.epa.gov/ttnamti1/files/ambient/pm25/qa/QA-Handbook-Vol-II.pdf>
- [8] Kenneth Davis, Maria Cambaliza, Kevin Gurney, et al. Measuring CO₂ and CH₄ Emissions from Indianapolis: A Test Bed for Urban Greenhouse Gas Emissions Monitoring. 14th National Conference and Global Forum on Science, Policy and the Environment, 2014.
- [10] Chris Rella. Accurate Greenhouse Gas Measurements in Humid Gas Streams Using the Picarro G1301 Carbon Dioxide / Methane / Water Vapor Gas Analyzer. Picarro white Paper, 2010.
- [9] L R Welp, W Paplawsky, R F Keeling, et al. A low-maintenance drying system for ambient air greenhouse gas monitoring. NOAA ESRL global monitoring annual conference, 2011.
- [11] P B Shepson, M Cambaliza, K J Davis, et al. The INFLUX Project: Indianapolis as a Case Study for the Accurate and High Resolution Determination of CO₂ and CH₄ Emission fluxes from an Urban Center [poster]. 2010.
- [12] U.S. Environmental Protection Agency. Quality Assurance Handbook for Air Pollution Measurement Systems Volume II Ambient Air Quality Monitoring Program. 2013.
- [13] M.L. Fischer. Carbon Dioxide Flux Measurement Systems (CO₂FLX) Handbook. ARM TR-048, 2004.
- [14] N L Miles. Tower-based CRDS greenhouse gas measurements in an urban environment, Calibration, field tests, and results from the INFLUX project. 2013
- [15] S J Richardson, N L Miles, K J Davis, et al. Field testing of cavity ring-down spectroscopic analyzers measuring carbon dioxide and water vapor. J. Atmos. Oceanic Tech, 2012, 29: 397-406.
- [16] Hyytiälä ICOS Atmospheric measurements. http://www.mv.helsinki.fi/home/oppeltol/Hyytiala_picarro_g1301.html
- [17] Manual for the GAW precipitation chemistry programme (guidelines, Data Quality Objectives and Standard Operating Procedures). GAW Report No. 160, 2004.

Appendix: comparison plot for all the sites

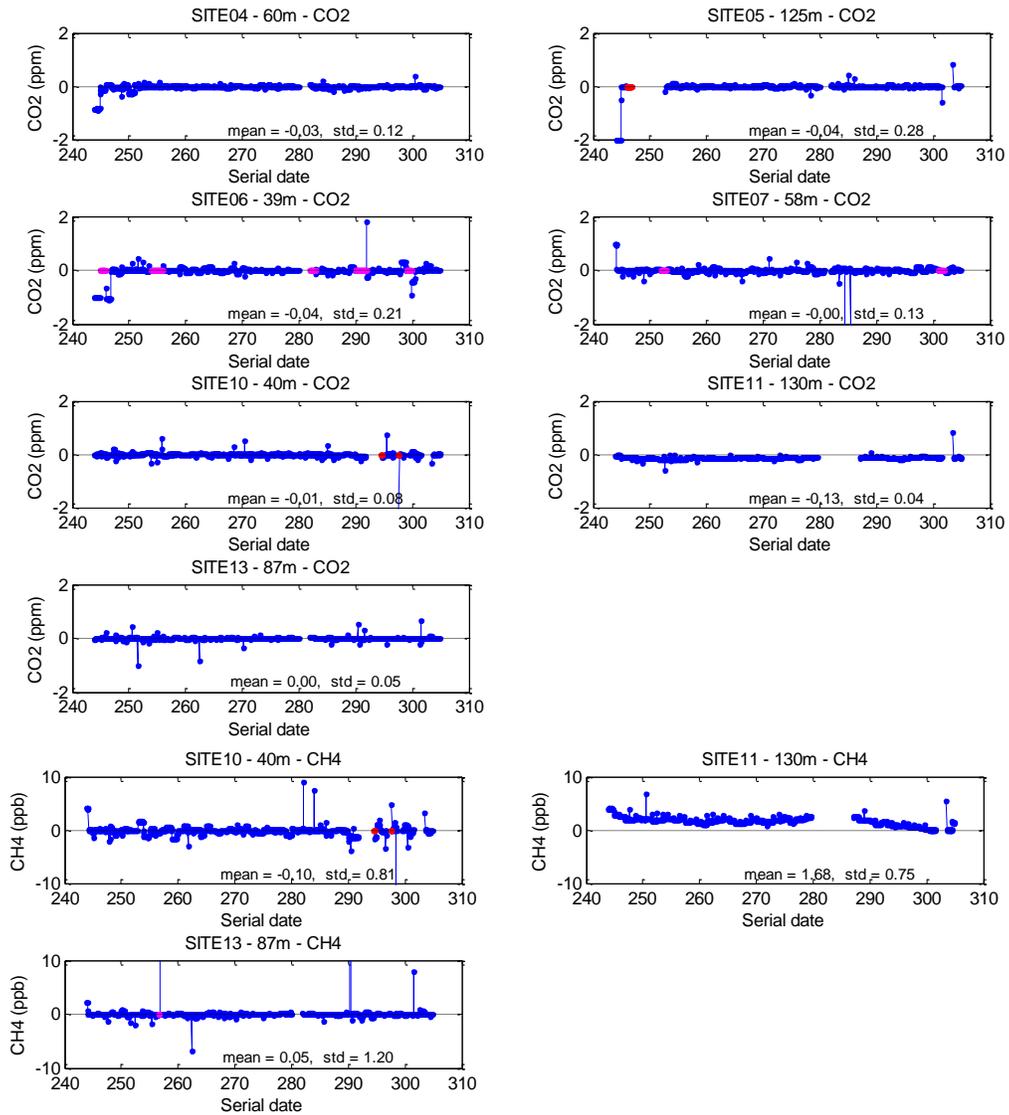
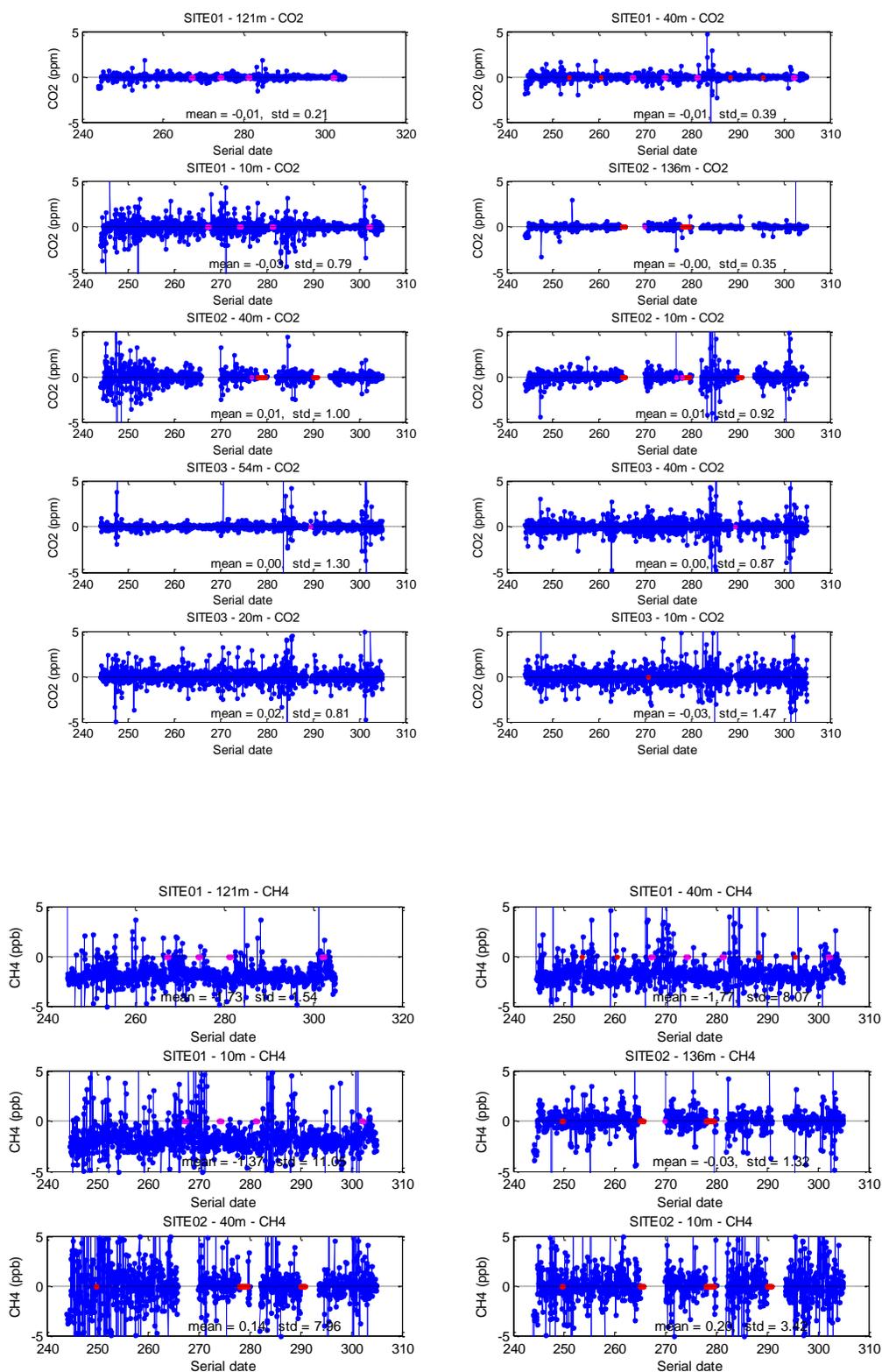


Figure 19: comparison of single height sites



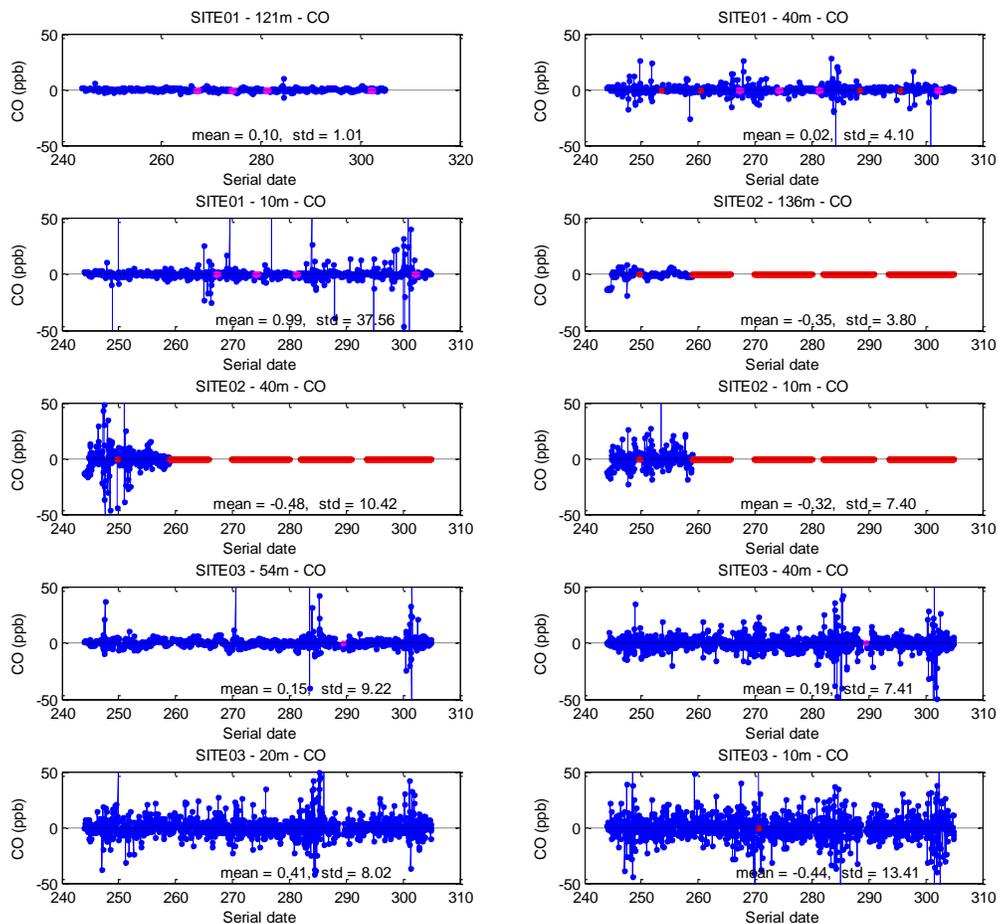


Figure 20: Comparison of multiple height sites