Answer to Referee #2

We would like to thank Referee #2 for his/her positive and constructive comments and suggestions. We have studied comments carefully and made corrections, which we hope meet with approval. Comments and responses are listed as follows. In order to facilitate the reference to the questions and proposed changes, we use the following color coding:

Color coding:

Referee comment

Our answer

Proposed change in manuscript

The paper by Zhou et al. reports on fuel sulfur content (FSC) compliance monitoring of sailing ships with unmanned aerial vehicles (UAV). Measurements were carried out in the Yangtze River Delta close to Shanghai in China, which is selected as an domestic emission control area (DECA) in China having a FSC limit of 0.5% (m/m). Since measurements of the FSC and therefore compliance monitoring from sailing ships are sparse, the topic is of interest not only for the scientific community. The manuscript is in general clearly written and I recommend it for publication in AMT. However, to better demonstrate the quality of the instrumentation and the methods used in this study more details should be given in the paper.

Thank you for the comments, we are very encouraged.

Instrumentation:

- There are no details on the custom sensors for SO2 and CO2 given. At least a link to a data sheet (in English) or better a table with specifications is needed.

OK, I have added a detail table (Table 1).

- I am wondering about the short response time of the modules (T90< 1s) which is much better than every electrochemical sensor I know. Looking to figure 5 this response time is very unlikely. SO2 and CO2 measurements of the same plume are out of phase (at least 10 to 15 s) having also completely different gradients.

We don't make sensors, and sensors were purchased from Shenzhen Singoan Electronic Technology Co., Ltd., China.

I looked up the relevant materials (Mellqvist et al., 2017), the t_{90} of CO_2 sensor based on NDIR (LI-COR 7200) is 0.1s. I also consulted other relevant literature (Alföldy et al., 2013, Beecken et al., 2014, Balzani Lööv et al., 2015), the t_{90} of CO_2 sensor is about < 1-5 s. The CO_2 sensor used by us is also base on NDIR principle. It seems reasonable that the CO_2 response time is less than 1 s. However, it can be clearly seen from Figure 5 that the response time of SO_2 is significantly faster than that of CO_2. Does this mean that the t_{90} of SO_2 sensor is also <1s?

To ensure the accuracy of parameter information, I contacted one technical support of Shenzhen Singoan Electronic Technology Co., Ltd., China. I double confirm the relevant parameter information of the sensor. I got the detailed information sheet of these two sensors. The response time (defined as T_{90}) is <30 s. "The t_{90} represents the time taken to reach 90% of the stable response following a step change in the sample concentration". I guess "a step change" does not mean from 0 to full range. If it is from 0 to full range, the response time (T_{90}) is <30 s. The information is list as follow.

Figure picture of the sensor

The type of SO_2 sensor is SGA-700A-SO_2. Link: http://www.singoan.com/article/detail/id/6233.htm

The type of CO_2 sensor is SGA-700A-CO_2. Link: http://www.singoan.com/SGA_400_700_CO2.htm

I fill the main information into Table 1 as bellow. In the original manuscript, the range of SO_2 sensor and accuracy of CO_2 sensor are wrong. I have modified and cross-checked to make sure the information was consistent as the product information.

Table 1: Parameters of the UAS

	Parameter	Value
UAV	Symmetrical motor wheelbase	1133 mm
	Size	1668 mm × 1518 mm × 727 mm
	Weight	9.5 kg
	Recommended maximum take-off weight	15.5 kg
	Hovering accuracy(P-GPS)	Vertical: ±0.5 m, Horizontal: ±1.5 m
	Maximum rotational angular velocity	Pitch axis: 300°/s, Heading axis: 150°/s
	Maximum pitch Angle	25°
	Maximum rising speed	5 m/s
	Maximum rate of descent	3 m/s
	Maximum sustained wind speed	8 m/s
	Maximum horizontal flight speed	65 km/h (no wind environment)
	Hover time	Non-loaded: 32 min; load 6 kg: 16 min
SO_2 sensor	Type	SGA-700A-SO2
	Principle	Electrochemistry
	Measuring range	0–10 ppm
	Diameter and height	33.5 mm; 31 mm
	Weight	30 g
	Accuracy	≤ ±3 % (0.3 ppm)
	Linear error	≤ ±2 % (0.2 ppm)
	Repeatability	≤ ±2 % (0.2 ppm)
	Power consumption	≤ 50 mA
	Response time (T_{90})	≤ 30 s
CO_2 sensor	Type	SGA-700A-CO2
	Principle	Non-Dispersive InfraRed
	Measuring range	0–10000 ppm
	Diameter and height	33.5 mm; 31 mm
	Weight	30 g
	Accuracy	≤ ±3 % (300 ppm)
	Linear error	≤ ±2 % (200 ppm)
	Repeatability	≤ ±2 % (200 ppm)
	Power Consumption	≤ 100 mA
	Response time (T_{90})	≤ 30 s

Alföldy, B., Lööv, J. B., Lagler, F., Mellqvist, J., Berg, N., Beecken, J., Weststrate, H., Duyzer, J., Bencs, L., Horemans, B., Cavalli, F., Putaud, J.-P., Janssens-Maenhout, G., Csordás, A. P., Van Grieken, R., Borowiak, A., and Hjorth, J.: Measurements of air pollution emission factors for marine transportation in SECA, Atmos. Meas. Tech., 6, 1777–1791, https://doi.org/10.5194/amt-6-1777-2013, 2013.

Balzani Lööv, J. M., Alfoldy, B., Gast, L. F. L., Hjorth, J., Lagler, F., Mellqvist, J., Beecken, J., Berg, N., Duyzer, J., Westrate, H., Swart, D. P. J., Berkhout, A. J. C., Jalkanen, J.-P., Prata, A. J., van der Hoff, G. R., and Borowiak, A.: Field test of available methods to measure remotely SOx and NOx emissions from ships, Atmos. Meas. Tech., 7, 2597–2613, https://doi.org/10.5194/amt-7-2597-2014, 2014.

Mellqvist, J., Conde, V., Beecken, J., and Ekholm, J.: Certification of an aircraft and airborne surveillance of fuel sulfur content in ships at the SECA border, CompMon (https://compmon.eu/, last access: 6 November 2018), 2017b.

Beecken, J., Mellqvist, J., Salo, K., Ekholm, J., and Jalknen, J.-P.: Airborne emission measurements of SO2, NOx and particles from individual ships using a sniffer technique, Atmos. Meas. Tech., 7, 1957–1968, http://dx.doi.org/10.5194/amt-7-1957-2014, 2014.

- Please give more information on the method water vapor was filtered out. What about contamination with particles/soot ...?

The pod is a convenient, lightweight device. It does not have the same complex gas filtration as shore-based equipment. A hose filter valve was used to filter out the water vapor, particles and soot. Its figure is as follow.

Figure Different type of the hose filter valve, the length is about 4-20mm

- Have the authors investigated cross-sensitivities to e.g. NO and NO2?

For the sake of lightweight and convenience, the second-generation pod is only equipped with SO_2 and CO_2 sensors. In the research of Mellqvist et al. (2017), they proposed a treatment for cross-sensitivities SO_2 and NO_2. But their sensor is based on the principle of fluorescence (Thermo 43i-TLE), and our sensor is based on electrochemistry. I have looked up the relevant materials, the cross-sensitivities of SO_2 and NO_2 do exist when using the electrochemical sensor. I guess more experiments and researches are needed to eliminate the effects of cross-induction.

Mellqvist, J., Conde, V., Beecken, J., and Ekholm, J.: Certification of an aircraft and airborne surveillance of fuel sulfur content in ships at the SECA border, CompMon (https://compmon.eu/, last access: 6 November 2018), 2017.

- Measurements were carried out close to the funnel. Therefore, temperatures of the air sucked into the system are highly variable. How this is accounted for?

The UAS has a gas pump, gas circuit, filter which can cool the gas to a certain extent.
I know that the equipment used in relevant research work has a constant temperature and pressure detecting environment. The weight of the equipment is usually tens of kilograms. But there are limits to the weight and size of the equipment that UAV can carry.
The conversion relation of units is as follow:

$$\mathbf{X}\ (mg/m^3) = \frac{\mathbf{M}}{22.4} * \mathbf{Y}\ (ppm) * \left[\frac{273}{273+\mathbf{T}}\right] * (\mathbf{Ba}/101325)$$

M is molecular weight of the gas, T is the temperature, and Ba is air pressure.
I guess it has to redesign the pod and laboratory experiments to reduce the impact of this factor.

- Give information on the calibration methods used in this study to ensure long-term data quality. Since sensors used in this study are completely different to those used in Zhou et al. 2019 a simple plot showing the outcome of the sensors when using standard gas mixtures (e.g. 5 and 0 ppm SO2) would be nice.

The calibration process is the same, I have added description.

The details of calibration are as follows: We bought the sensors and then designed and built the pod. Then we send the pod to a third-party inspection agency for certification. Content of verification is mainly about the accuracy. We only know the results and we don't have detailed data. I know that you may interest in the sensors. But the information available to me is limited. I have attached the validation report at the end (the original is in Chinese).

- The UAV used in this study is the same as in Zhou et al., 2019. What does it mean for the off-shore measurements? What is the operation time under typical weather conditions having e.g. a wind speed of 5 m/s? What is the maximum reasonable distance to a sailing ship? Please add a table with specifications of the whole UAV system.

The UAV is a product of DJI and is a relatively common UAV model. We have successfully applied it offshore measurements. But there are still have limitations, it is more affected by the weather compare with other measurement platform. For safety reasons, we can't use it when it rains or when the wind is high. Nevertheless, this is basically the best and most suitable civilian UAV that we can find on the market.
About 15-20 min.
This question is hard to say. We did not measure the distance by instruments. At sea, the human sense of distance is very weak. My personal feeling is about 500 to 3000 meters.
I have added the Table 1.

Methods and uncertainties:
- The authors used peak values of SO2 and CO2 after applying a running mean of 10s to the measurements to calculate the FSC. Looking to Figure 5 b it is not clear to me, how this could give reasonable results. As already mentioned above the gradients (and therefore real response times of the sensors) look completely different even for averaged values. At least one example proving and illustrating this method is needed.

The data in Figure 5 is of good-quality and the difference is not significant. This is even more pronounced if the data is of poor-quality. I've added good and poor typical figures and related discussion. To illustrate how to evaluate data quality, I have supplemented the manuscript with typical good- and poor- data, and the relevant description as follow.

The continuous measurement data for two typical plumes (2019-4-15B and 2019-3-29A) are exhibited in Fig. 5. The data for plume 2019-4-15B (Fig. 5a) were considered to be of a “good” quality, whereas those for plume 2019-3-29A (Fig. 5c) were considered to be of a “poor” quality. Data were determined to be of a good-quality when obvious, easily distinguished peak values were observed, whereas less obvious peaks that still corresponded to a result were considered as poor-quality data. The selection of peak values leads to uncertainty because when the area ratio is selected for the calculation, the starting and ending time points of the area are still associated with substantial uncertainty. Figure 5b and 5d depict the average concentrations of the SO_2 and CO_2 measurements (in Fig. 5a and 5c, respectively) for 10 s periods. The peak value of each average concentration was selected for the calculation. This process is equivalent to selecting the area ratio of SO_2 to CO_2 within 10 s for the calculation, as shown in Eq. (2).

$$FSC[\%] = 0.232\frac{\int(SO_{2,peak}-SO_{2,bkg})dt[ppb]/10}{\int(CO_{2,peak}-CO_{2,bkg})dt[ppm]/10}[\%] + R \approx 0.232\frac{AVG(SO_{2,peak})-AVG(SO_{2,bkg})}{AVG(CO_{2,peak})-AVG(CO_{2,bkg})}[\%], \quad (2)$$

where $AVG\,(\cdot)$ is the calculated function for the average measurement value within 10 s; hence, the data in this study are the average values of measurements in 10 s. When the UAV took off from the patrol boat and flew high into the air, the SO_2 and CO_2 concentrations were relatively low. The background values were obtained at this stage as the minimum SO_2 and CO_2 concentrations. As the UAV flew into the plume, the measured concentrations of SO_2 and CO_2 increased. The obvious, stable maximum values in the observations of the average measurement values should be selected as the peak values. It can be seen that using the average values of measurements within 10 s makes it easier to select the peak values, especially with respect to poor-quality data. However, as there can still be several options for peak values, the

data treatment methods reported by Zhou et al. (2019) were incorporated in this study to select the most appropriate peak values. In Fig. 5b, the time point of selected peak values is at 10:19:11. The measurement values from 10:19:57 to 10:20:15 were not used because the CO_2 concentration covered the full range. In Fig. 5d, the time point of the selected peak values is at 10:38:27. The measurement values from 10:39:57 to 10:41:41 were not used because we ruled out data exhibiting either dramatic changes or errors in continuous observations. The details for selecting the peak values are listed in Table 2.

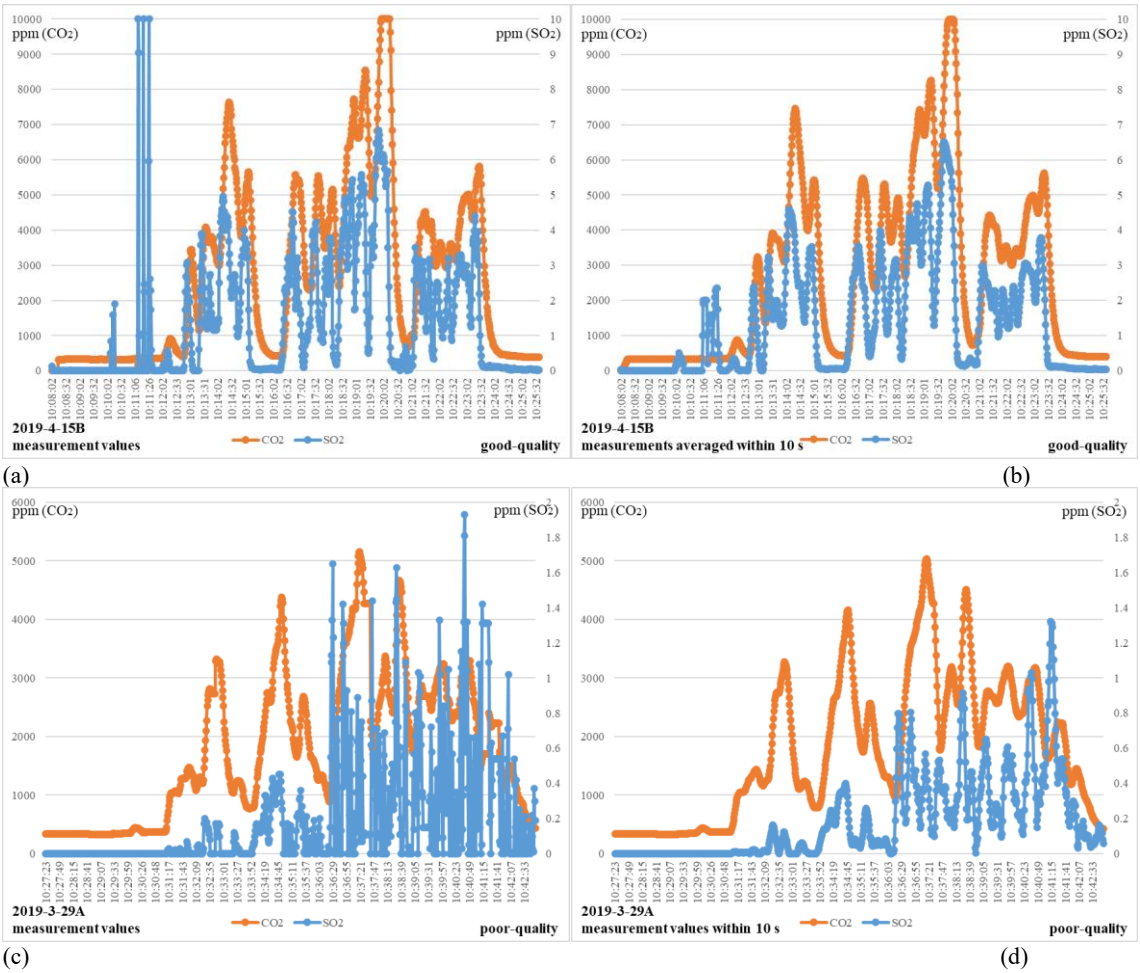

Figure 5. Typical measurement data for SO_2 and CO_2 concentrations, and their corresponding average values within 10 s. (a) and (b) good-quality data from plume ID 2019-4-15B. (c) and (d) poor-quality data from plume ID 2019-3-29A. There are some errors in the measurements from 10:11:06 to 10:12:02 in (a), which may have been caused by sensor uncertainty. These data were ruled out and did not affect the calculation results.

2.5 Uncertainties

In previous research (Zhou et al., 2019), the main uncertainties of UAV measurements were summarized as sensor uncertainty, measurement uncertainty, calculation uncertainty, and exhaust uncertainty. The instrument calibration method, UAV flight procedures, and data treatment methods were designed to reduce these uncertainties. However, some uncertainties remain, as discussed below.

The average gas concentration within 10 s was chosen for the FSC calculations; however, this does not mean that 9 s or 11 s could not have been selected. To demonstrate this, a comparison calculation was carried out using both 9 s and 11 s, which showed that these led to very little differences in the results. However, it is necessary to ensure that the gradient of the gas measurements is stable within the sampling time (the interval length of the integral). Moreover, the interval length cannot be too short (e.g., 2 s) or too long (e.g., 20 s). If the time is too short, it is difficult to determine whether the measurements are stable and undisturbed over time. Similarly, if the time is too long, it is also difficult to ensure that all of the measurements in the integral interval are stable and undisturbed. In addition, during the flight of the UAV in this study, the time available for measuring the plume was ~5 minutes. As both the ship and the UAV were moving at this time, it was virtually impossible to ensure that the UAV was flying consistently within the plume and obtaining

stable measurements. Accordingly, 10 s is also a relatively appropriate value for the measurement process.

Nevertheless, there is also some uncertainty associated with choosing the peak values. After ruling out the peak values across the full range as well as those corresponding to dramatic changes, the global maximum values were selected as the peak values to calculate the FSC. The maximum values probably correspond to the measurements taken in the center of the ship's plume. At that location, the measurement values were relatively stable, and the probability of interference from other factors was lower. Furthermore, the higher the peak value is, the greater the proportion of exhaust gas is; hence, the impact from the incomplete mixing of the exhaust gas with clean air is relatively small.

In summary, the obvious and stable maximum values are selected as peak values to calculate the FSC. There are, of course, situations where multiple similar peaks can occur simultaneously. In this case, their calculated FSCs may be very similar, and the results obtained by the calculation of the highest peak should have high credibility, for instance, the measurements of plume 2019-4-15B.

Table 2: All peak values and their corresponding FSC results. The background values of plume 2019-4-15B were 0 ppm and 310 ppm for SO_2 and CO_2, respectively. The background values of plume 2019-3-29A were 0 ppm and 329 ppm for SO_2 and CO_2, respectively. The remarks indicate the reason for choosing or not choosing the peak. It can be seen that the peak value of plume 2019-4-15B was more obvious and that the results obtained by multiple alternative peaks were similar. The peak of plume 2019-3-29A was less obvious and there were fewer alternative peaks. This was also the basis for distinguishing data as being of a "good"/"poor" quality. The FSC result of selected peak values are marked as "√".

Plume ID	Time point	Peak value of SO_2 and CO_2 (ppm)	Estimated value of FSC (% (m/m))	True value of FSC (% (m/m))	Remark
2019-4-15B	10:12:52	2.406, 2020	0.326	0.168	Reject; less obvious peak values
	10.13.23	3.235, 2372	0.364		
	10.14.07	4.594, 4665	0.245		Non-maximum peaks of alternative peak values
	10.14.57	3.529, 4872	0.179		
	10.16.39	3.549, 4444	0.199		
	10:17:27	3.989, 3911	0.257		
	10:18:01	3.159, 4607	0.171		
	10:18:47	4.757, 6895	0.168		
	10:19:11	5.287, 7634	0.167 (√)		Maximum peak of the alternative peak value
	10:19:46	6.515, 8100	0.194		Reject; measurements exceeded the range
2019-3-29A	10:34:41	0.399, 3880	0.026	0.035	Reject, less obvious peak values
	10:35:19	0.258, 2011	0.036		Non-maximum peaks of the alternative peak values
	10:37:15	0.567, 4994	0.028		Reject; less obvious peak values
	10:38:27	0.913, 4022	0.057 (√)		Maximum peak of the alternative peak value
	10:40:37	1.031, 2996	0.090		Reject; error in the measurement data
	10:41:13	1.321, 1700	0.224		

I have rewritten these parts of 2.4 Calculation and 2.5 Uncertainties to illustrate the problem and related problem.

- Simply taking into account the given accuracies for the sensors (5 and 3 % on measuring ranges respectively) an error of at least 0.03 % (m/m) can be calculated. Within this calculation no other errors e.g. due to the measurement procedure are included. Therefore, the reported total uncertainty of 0.03 for low FSC levels sounds quite optimistic to me.

According to the literature available, the main method to measure the ship plume are land-based and airborne-based method. UAV measurements are indeed more accurate than these approaches. I guess this is mainly because UAV measurements are taken at close range. However, please note that this accuracy is the measurement result of the berthing ships. One is unable to obtain samples of fuel from sailing ships normally. We attempted to measure more than 40 ship plumes in open water; however, only 27 of them resulted in good- or poor-quality data, i.e., usable data. The success rate is not very high.

- Please give details or an illustration what is meant with good- and poor-quality data.

As mentioned above, I have added good- and poor-quality.

Minor corrections:
- 2.1. Instrumentation: I guess, dimensions of the pod are given in cm.

Yes, you are right, it is cm. Thank you for pointing out this error.

- Figure 5: More details needed, which values are used for the calculation of the FSC (see above)

As mentioned above, I have added.

- Tables 1 and 2: Please always give FSC Values in % (m/m) using only the number of meaningful digits. “True value” implies that the analysis of fuel samples have no error which is of course not the case. Please refer to e.g. fuel sample analysis and give the typical error for this method (should be roughly 0.01 % (m/m))

Yes, I have modified the unit.
Maritime authorities send fuel samples to third-party testing institutions, and the test results can be used as a basis for law enforcement. I don't have the information about the error of their detection methods. But I guess the accuracy of direct fuel detection is definitely far higher than the FSC result estimated from gas measurement. It is appropriate to take it as “True value”.

- Table 2: Add values for the fuel sample analysis when available.

OK, I have added.

The following is the third-party inspection report of the pod.

上海市计量测试技术研究院
华东国家计量测试中心
中国上海测试中心

检 测 报 告

Test Report

委托者 Customer	上海安馨信息科技有限公司
委托者地址 Address of customer	上海临港新城海基六路218弄13号楼3楼
样品名称 Name of sample	船舶尾气检测吊舱
制造厂 Manufacturer	上海安馨信息科技有限公司
型号/规格 Model/Specification	AX-YD
样品编号 No. of sample	1145003

（机构检测专用章）

批准人/职务 Approved by / Functions	郝玉红 质量主管
核验员 Checked by	核验员标记
检测员 Tested by	检定员标记

检测日期 Date for test 2019 年 Year 03 月 Month 11 日 Day

地址：上海市张衡路1500号(总部) Address No.1500 Zhangheng Road，Shanghai(headquarters)
电话：021-38839800 Tel.
传真：021-50798390 Fax
邮编：201203 PostCode
客户咨询电话：800-820-5172 Inquire line
投诉电话：021-50798262 Tel. for complaint

未经本院/中心批准，部分采用本证书内容无效。
Partly using this report will not be admitted unless allowed by SIMT.

国家法定计量检定机构计量授权证书号(中心/院):(国)法计(2017)01039号/(2017)01019号
The number of the Certificate of Metrological Authorization to The Legal Metrological Verification Institution is No.（2017）01039/ No.（2017）01019

本次检测所依据的技术规范（代号、名称）：
Reference documents for the test (code 、name)

参照JJG 635-2011《一氧化碳、二氧化碳红外气体分析器检定规程》
参照JJG 551-2003《二氧化硫气体检测仪检定规程》

本次检测所使用的主要测量仪器：
Main measuring instruments used in this test

名称 Name	型号规格 Model	编号 Number	测量范围 Measurement range	不确定度或准确度等级或最大允许误差 Uncertainty/Accuracy Class/Maximum Permissible Error	证书编号/有效期限 Certificate No./Due date
氦中二氧化碳气体标准物质	/	770357/201803	4.97×10^{-2}mol/mol	U_{rel}=2.0%（k=2）	770357/2019-03-29
氮中二氧化硫标准气	/	L00502040	100×10^{-6}mol/mol	U_{rel}=2.0%（k=2）	L00502040/2019-03-29
精密气体稀释仪	MGB1000	06132	稀释比：1~1000	±3%	2019I30-30-1727259001/2020-02-14
/	/	/	/	/	/

检测地点及环境条件：
Location and environmental condition for the test

地点： 张衡路1500号理化东楼125室
Location

温度： 21℃ 湿度： 63%RH 其它： /
Ambient temperature Relative humidity Others

备注： /
Note:

本报告提供的结果仅对本次被测的样品有效。
The data are valid only for the sample(s).

检测结果/说明：
Results of test and additional explanation

委托日期	2019.03.07	样品状态描述	正常	受样方式	客户送样

检测项目	标准气体浓度（μmol/mol）	仪器显示值（μmol/mol）
CO_2	10.0	9.84
	900	924
SO_2	1.0	0.96
	8.0	8.21

检测结果内容结束