We thank referee #3 for the useful comments and the encouragement for deeper exploration of the data. We have considered each point carefully and we hope that our responses add clarification to the manuscript.

The referee comments are marked in *blue cursive*, the author response in black and text added to the manuscript is **bolded**

The manuscript Laitinen et al. presents experimental evidence that allows to assess the compatibility of two independent measurement programmes at the Pallas measurement site. The provision of longterm quality control (QC) data giving evidence of the data quality of observations is crucial to evaluate the consistency of the global observational data set. The information the authors have compiled for their study is exemplary. The two independent set-ups have each independent methodological choices concerning aspects like drying, calibration approaches or the frequency of QC measurements. The offsets that are visible in the CO2 and CH4 comparisons document a very satisfactory performance. In addititions, both programmes maintain each their proper external auditing entitities. The core data set of this manuscript is the result of a coordinated auditing campaign to have four parallel measurements. This data set provides a great opportunity to learn about causes for biases or methodological weaknesses. So first of all I highly welcome the publication of such a study. On the other hand I feel that currently the manuscript does not evaluate the data set sufficiently. It has drawn just few conclusions without strong ambitions to use the findings to learn about which methodological approaches are more successful than others. The manuscript would be of much more interest if it could give guidance what limitations might be cause for measurement offsets. Therefore, I would strongly encourage the authors to further explore the data available to them as suggested below.

General comments

- *The four parallel measurements done during the audit campaign have very clear differences in their calibration approaches (no. of calibration points between 2 and 9; frequency between daily and every 3 months). They all also have there own internal quality control set-up (mostly using target standard gas analysis to monitor the success of the efforts to achieve maximum accuracy). Those internal QC records should contain information on what internal reproducibility the two observational programmes (GAW and ICOS) but also the two audit programmes (by the ICOS mobile laboratory and GAW WCC) deliver. From my point of view, those target data records should also be included into the manuscript and a thorough data uncertainty assessment for the individual measurement programmes be made (for the continuous station measurements as well as for the auditing programmes). An estimate of the expected measurement compatibility between the various systems based on such an assessment is needed to judge if the observed comparison mean offsets and their spread presented in Table 2 point to additional systematic analytical limitations.*

  Thank you for the comment. This is a very valid point and we have put the effort in analyzing the target cylinders records more carefully. We have addressed the individual comments regarding this in the specific comments below. In addition, we have added a section 3.4: Quality assurance

3.4 Quality assurance

For quality assurance (QA) of the different measurement instruments, we analyzed their respective calibrated target cylinder measurements. Using the calibrated values allows for a comprehensive evaluation of the instrument as well as the calibration cylinders and calibration method, which also varies by instrument. As a measure of long-term repeatability (LTR), we calculate the deviation of the target cylinder measurement from the assigned value of the cylinder, and calculate the standard deviation of these values. This approach is used to account for different target cylinders used over time. For cases where only one cylinder is used, this value is equal to simply taking the standard deviation of the target cylinder measurements. For short-term repeatability (STR), we use the standard deviation of the individual cylinder measurement sequences. In addition, we calculate the mean bias of the measured values to the assigned target mole fractions for each instrument. For stabilization, only the last minutes of the injection are used for the analysis. For ICOS and GAW instruments, each measure is calculated for the whole time period of concurrent measurements as well as for the audit period. For the ICOS instrument, the LT is used for the long-term comparison QA and the ST for the audit period. For GAW instrument, one target cylinder is used for QA and a short-term working standard is used for drift correction. For the audits, one target cylinder is analyzed for each travelling instrument. In order to to get a table measurement, only the last minutes of each cylinder measurement is used for calculating the means. The different cylinders are presented in tables 3 (CO2) and 5 (CH4). The assigned values, measurement times as well as the number of minute data points used for averaging are presented, as well as STR, LTR and mean biases.

- *The focus of the manuscript could be sharpened. There is general site descriptions that have already been published (in a reference that is cited) and are not essential to the understanding of the main topic of the study.*

  While we agree that the focus of the manuscript is quite wide, we feel like having e.g. the general site descriptions still in the manuscript is helpful for people to understand the location and the site specific characteristics. Especially the cited paper describes the whole Pallas area, while we want to focus especially on the atmospheric station.

- *Throughout the document it remains a challenge to understand at which points different names for the site are used as synonyms and where those specific names cover a different meaning. Please try to harmonize and avoid terms that are used as synonyms.*

  We will harmonized the terms used to refer to the station throughout the manuscript.

Specific comments

- *l. 25: It is not clear why exactly those two references have been selected for the use of observational data for atmospheric inverse models and the top-down emission estimates.*
  We have added different citations that hopefully reflect better the need for measurements:
  McGrath et al. (2023)*, https://doi.org/10.5194/essd-15-4295-2023*)
  Petrescu et al. (2023), https://doi.org/10.5194/essd-15-1197-2023)
  Lauerwald et al.(2024), https://doi.org/10.1029/2024GB008141)
  Saunois et al. (2024), https://doi.org/10.5194/essd-2024-115) ;
  ***Friedlingstein et al. (2025),* https://doi.org/10.5194/essd-14-4811-2022)**

- *l. 30: The term supersite does not have a clear definition. Either define it or omit it.*
  Included definition of supersite on line 50 and changed ”supersite” on line 30 to ”site”

- *l. 41: "This data is also available as GAW data" is a bit confusing. Consider rephrasing as:"This data is also submitted to the WMO World Data Centre for Greenhouse Gases..."*
  Changed as suggested

- *l. 54: The following sentence does not fully meet the meaning of WMO Expert Group's recommendations: "To ensure that the station's measurements are compatible with the WMO/GAW goals, they must be compared against other instruments to ensure the differences are within acceptable limits." I suggest to rephrase like "Assessing the compliance to the WMO network compatibility goals requires comparison of station measurements with other laboratory's measurements". Please add Andrews et al (www.atmos-meas-tech.net/7/647/2014/) as reference.*
  Changed as suggested

- *l. 111: "Pallas was labelled as an ICOS Class 1 atmosphere station": The meaning of "labelled as an ICOS class 1 station" ought to be explained and doi.org/10.5194/amt-14-89-2021 should be provided as reference.*
  Added clarification:
  Within the ICOS ATC, the stations are classified into Class 1 and Class 2 stations (Ywer-Kwok et al, 2021).

- *Figure 4a: The caption should indicate that LTT only is displayed. The trend in an increasing DCO2 for D348367 is substantial with respect to the mean bias numbers in Table 2. A discussion of this should be offered some part in the manuscript. As the LTT are analysed only every 15 days after the calibration the STT record would be of greater interest to compare it with the observed measurement system differences.*
  Removed Figure 4 and moved the information to Figures 7 and 13, where also the ST of the ICOS instrument is now presented.

- *l. 134-136: Are the water vapour correction coefficients assumed to be constant over time (i.e. established just once) or are those re-established according to a certain schedule?*
  The correction coefficients are monitored during the audit as well as approximately once per year by the station PI, and updated if deemed necessary by the ATC. Added explanation to the text:
  These coefficients are evaluated during the ICOS audit, as well as approximately once per year by the station PI, and updated if deemed necessary by the ATC.

- *l. 145: The GAW Picarro is calibrated every 2.5-3 months which assumes response stability throughout that period. When was the calibration done during the auditing campaign?*
  The GAW picarro was calibrated at the beginning and end of the audit period. Added in section 3.2:
  The GAW instrument was calibrated at the beginning and at the end of the joint audit period.

- *l. 154: "STT and LTT cylinders.." A similar figure as Figure 4 displaying the GAW programme's analyzer QC record should be presented.*
  Added figures 7 and 13 to show these for GAW as well, removed figure 4.

- *l. 200 ff: Same comment for the Mobile Laboratory measurements: the target measurement results should be presented similarly to Fig. 4.*
  Added figures 9 and 15 to show these measurements

- *l. 213: What are the results of the assessment the validity of the internal water vapor correction coefficient?*
  Here we noticed a mistake; The audit is validating the coefficients determined by the ATC before the instrument is being shipped to the station. Also added assesment of the coefficients.
  [...] the validity of the water vapor correction coefficients **determined by ATC. For Pallas ICOS instrument the coefficient were deemed valid and no change was required.**

- *l. 219: The WCC calibration approach is not fully clear: the description states that there is a two point calibration performed (zero plus one atmospheric reference standard). The function of the working standard to account for "drift" (analyzer response drift or CO_2 drift?) is unclear. Is the working standard effectively being used as a second calibration standard?*
  Changed the explanation the be more detailed:
  The zero reading of the WCC-Empa travelling instrument has been calibrated with CO_2 and CH_4 free air (or nitrogen 6.0) prior to field use by adjusting the offsets in the user calibration file of the instrument. During the field use of the TI, only one working standard (WS) is used to calibrate the instrument for CO_2 and CH_4. In a first step, a Loess function is fitted to the WS (measured every 1445 min) to correct for drift. The resulting drift correction is then applied to all TI data in a second step. The drift corrected WS is then used to apply a calibration factor to the data using the assigned

value of the WS based on calibration against the CCL standards before and after field use. Two target standards are measured to verify the drift correction.

- *l. 221: Same comment as above for the WCC audit measurements: the target measurement results should be presented similarly to Fig. 4.*
  Added figures 9 and 15 to show these measurements

- *Figure 5 caption: "the MBL data" and "the NOAA MBL data" refer to the same data set. Rather use only one term – I think "the NOAA mean MBL data" describes it best.*
  Changed term "MBL data" to NOAA mean MBL data"

- *l. 275-276: I cannot judge if a change from 94.1 % to 95.3 % is significant. The authors should consider adding a figure displaying the inter-instrument bias vs. the measured water content to reveal if there is any influence of the different drying and water vapour correction functions. Likewise a figure displaying the bias dependence on the mole fraction would be appreciated that might give an indication of a potential calibration offset contributing to the offset.*
  We have added discussion on the dependency of the difference on the mole fraction and water vapor concentration, as well as two figures A4 and A6 to illustrate.

  CO_2:
  There is a negative correlation between the DCO_2 and the water vapor concentration of about -0.05 ppm/v% (intercept 0.05 v%, $p < 0.05$, R^2: 0.12) when both of the instruments measure wet air (Fig. A3, (a)). The correlation changes to slightly positive when the ICOS sample air is dried (slope 0.04 ppm/v%, intercept = -0.04, $p < 0.05$, R^2= 0.08). Additionally, only a very weak correlation with mole fraction was observed on the differences (Fig. A3, (b)).

  CH_4
  As for CO_2, there is a negative correlation between DCH_4 and the water vapor concentration of about -0.80 ppb/v% (intercept 1.35 v%, $p < 0.05$, R^2 = 0.40) when both of the instruments measure wet air (Fig. A5, (a)). After the ICOS analyzer sample is dried, the correlation is slightly positive of about 0.22 ppb/v% (intercept -0.02, $p < 0.05$, R^2 = 0.06). When measuring wet samples, there is a slight correlation with mole fraction of about 0.01 ppb/ppb (intercept -9.37, R^2 = 0.06), and when the ICOS analyzer sample is dried no significant mole fraction dependency is observed (Fig. A5, (b)).

- *l. 279: For readability please do not abbreviate "growth rates" ("GRs").*
  Changed instances of GRs to growth rates and GR to growth rate

- *Figure 6 / Figure 12: Please state explicitly which of the two data is subtracted from which (ICOS-GAW or GAW-ICOS) either in the figure caption or in the y-axis label.*
  Added explanation (GAW-ICOS) to captions

- *Figure 8: In the presentation of three records in one graph the visibility of the blue data points is too limited. Three rows of graphs should be considered (e.g. grouping GAW-ICOS and WCC-MobLab / ICOS-WCC and ICOS-MobLab / GAW-WCC and GAW-Moblab.*
  We have added a third row to the figure for visibility.

- *I do not see the additional value of Fig. 9 compared to Table 2. I would see additional information, though, if the offset between the measurement result pairs would be depicted vs. the mole fraction. The same applies to Fig. 14 and Table 3.*
  Moved Fig 9 and 14 to appendix and changed them to show the mole fraction biases.

- *l. 356: The authors reflect on the source of the biases and suggest calibration being one cause. It is unclear why this is written in a bit speculative manner as the audit of the Mobile Laboratory includes a cross-comparison of the reference gases by the station instrument and the Mobile Laboratory (see l. 195). The results of this is calibration gas assignment bias is not presented - why? Were the reference standards also exchanged between GAW, WCC and Mobile Laboratory? This is data that is necessary information and should be provided in an additional table (or in a supplement).*
  Only two cross-comparisons were made: Measuring the WCC TCs with the GAW instrument and the ICOS Mobile Lab TC with the ICOS instrument. Added tables A1 and A2 to present the calibration cross-comparisons of GAW and ICOS calibration gases. Provided discussion in the manuscript:

  CO_2
  To account for possible differences in the calibration standards, the ICOS Mobile Laboratory cylinders were measured with the ICOS analyzer, and the WCC travelling cylinders were measured with the GAW analyzer. The results of the measurements are presented in table A1 and table A2. The calibration cylinder measurements show that for CO_2 the ICOS analyzers measurements differ on average from -0.09 ppm to 0.01 ppm to the assigned cylinder values, and the GAW analyzer differs from -0.05 ppm to 0.03 ppm.

  CH_4
  To quantify this effect, during the audit the ICOS analyzer measured the ICOS Mobile Laboratory calibration standards, and the GAW analyzer measured the WCC travelling standards. The results of the calibrations are presented in tables A1 and A2. For CH_4, the ICOS analyzer measured 0.12 to 0.54 ppb lower values than the assigned values of the cylinder, and the GAW analyzer measured 0.16 to 0.72 ppb higher values.

- *l. 383-384: "In the CO2 data the effect [of the Nafion dryer] is less clear, with the difference before adding the Nafion being 0.02 ppm and -0.02 ppm afterward". Fig. 6a clearly shows that a seasonal cycle of ca 0.1 ppm in the offset was much reduced after using the Nafion on the ICOS system. This has been described for CH4 (l. 332) but should also be mentioned for CO2. It is clear evidence that a problem has been remedied even if the absolute number of the offset has not changed.*

  We have added this discussion to the CO_2 section:

  Before the Nafion was installed, there exists a seasonal variation in the differences between the instrument. With the addition of the Nafion dryer to the ICOS analyzer inlet at the end of 2020**, the seasonal variation is reduced** [...]

- *l. 389-391: "the largest spread (CI range) in the differences was between ICOS and GAW for CO2 and between GAW and the ICOS Mobile Lab for CH4. This is expected, as the GAW system is the only one measuring from its own inlet". This conclusion is not fully convincing as long as the insignificance of other factors is not proven. Factors that might result in GAW measurements being slightly different could be for instance: 1) GAW measurements are also the only ones that are not made with dry air. So the application of the water vapour correction makes GAW different to all others. 2) Respective assignment errors (or instable CO2) of reference standards might result in mole fraction dependent offsets. 3) The reproducibility of the GAW system might be inferior due to less frequent calibrations. This could be easily checked by an evaluation of the target records.*

  We have added discussion on the topic.

  This is **partly** expected, as the GAW system is the only one measuring from its own inlet and all the other systems are connected to the same inlet. **In addition, the GAW instrument is still measuring the sample wet, while the other instruments are drying the sample to different degrees (ICOS instrument and WCC TI with a Nafion and the ICOS Mobile Laboratory with ICOS dryer). No significant differences in the LTR was observed for CO_2 or for CH_4 across the instruments during the audit.**

- *l. 393-394: "Compared to the differences between the ICOS and GAW analyzers over the whole period, the CO2 difference during the audit is slightly higher." It seems more reasonable to compare the audit mean difference to the mean difference since December 2020 when the Nafion has been added to the ICOS system. Else this aspects needs to be mentioned in the discussion here.*

  Changed to:

  During the audit the difference between the ICOS and GAW instruments is comparable to the difference over the whole period, when the ICOS instrument is sampling dried air.

- *l. 401-402: "The audit period also indicates that the ICOS system is performing slightly better". Phrase like: "The better measurement agreement with the two auditing units suggest a better performance of the ICOS system". It would be very useful to add if this corresponded to a superior reproducibility estimate for the ICOS measurements based on the record of its short term target compared to the respective target records of the GAW system.*

  Changed to:
  The better measurement agreement with the two auditing units suggest a better performance of the ICOS system. However, the LTR of the GAW instrument is similar to the LTR of the ICOS instrument, when measuring the ICOS LT cylinders. The LTR of the ICOS ST cylinders is worse, however this could be also caused by cylinder drift.

- *l. 403: "but the effect of theNafion dryer on the differences between the two systems is still unclear". As pointed out above (comment to l. 383) this statement is not justified. The Nafion dryer removed a stronger systematic seasonality in the offset. There might be a smaller seasonality remaining after December 2020 with the phase having shifted but it is difficult to capture this just visually from Fig. 6a. This could point to a remaining smaller humidity related offset now caused by the wet air analysis of the GAW system and also a non-perfect water vapour correction. This is something the authors should check.*

  We have added a figure A3 to show the seasonal variation in the differences before and after the Nafion was taken in use. It shows significantly reduced seasonal variation after the Nafion is installed, especially in CH_4 but also CO_2. There is more relatively more variation in the CO_2 measurements, especially during summer which to some extend mask the effect of the water vapor.

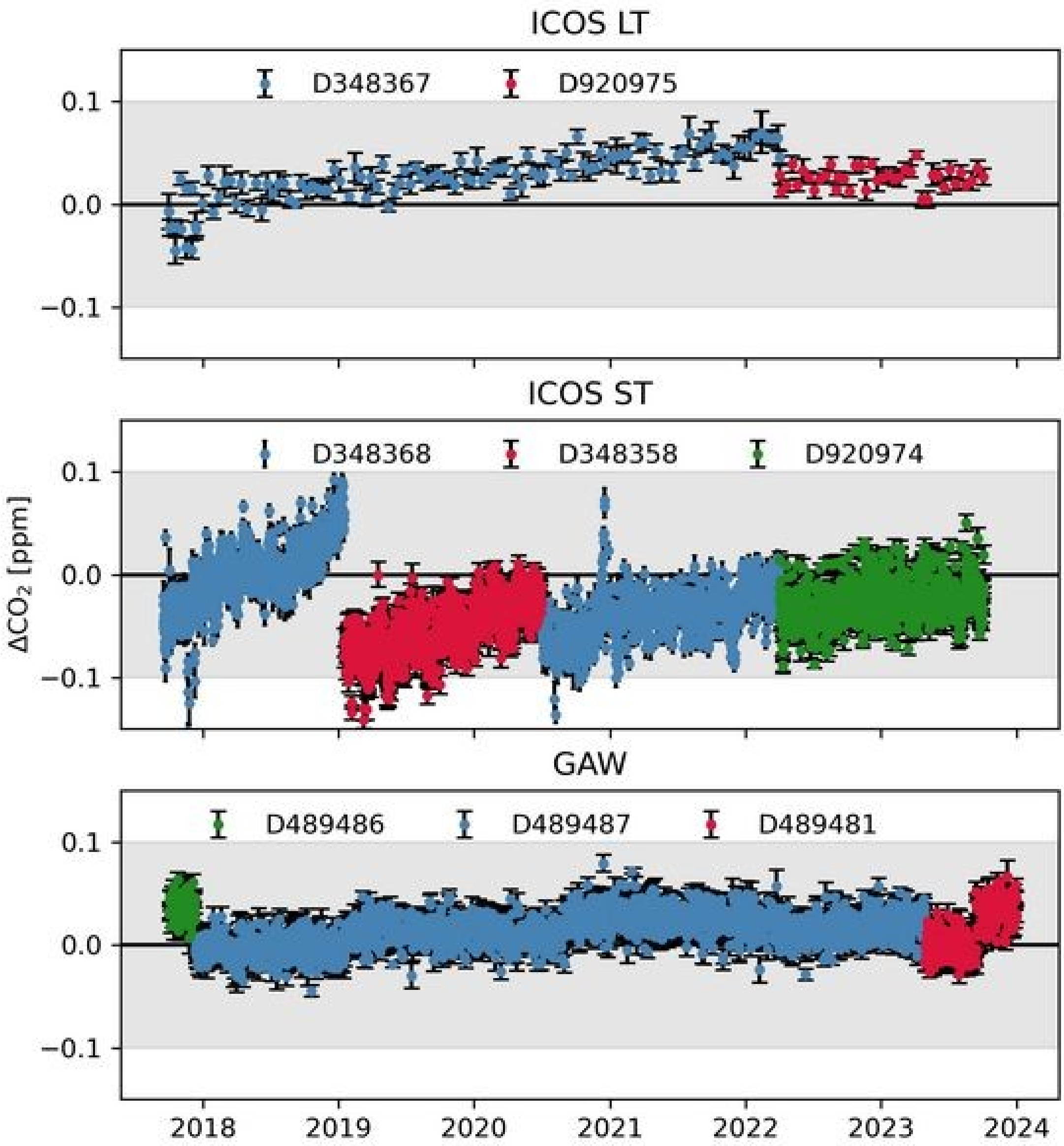
  Added to CO_2

  There could be a remaining seasonal variation after the ICOS sample is dried, however the stronger variation during summer in CO_2 masks this effect.

  Changed in Summary:

  […] addition of Nafion dryer on the intake line of the ICOS instrument. Especially for the CH_4 measurements the improvement is clear, the difference before drying the sample is 0.76 ppb on average and 0.21 ppb after. In the CO_2 measurements the effect is less **pronounced**, with the difference before […]

New figures added:



s

*Figure 7. CO2 target measurements of ICOS LT, ICOS ST and GAW target cylinders over the whole comparison period. Different cylinders
used are marked with distinct colors. The data are given as means of each sequence with the associated standard deviation.*

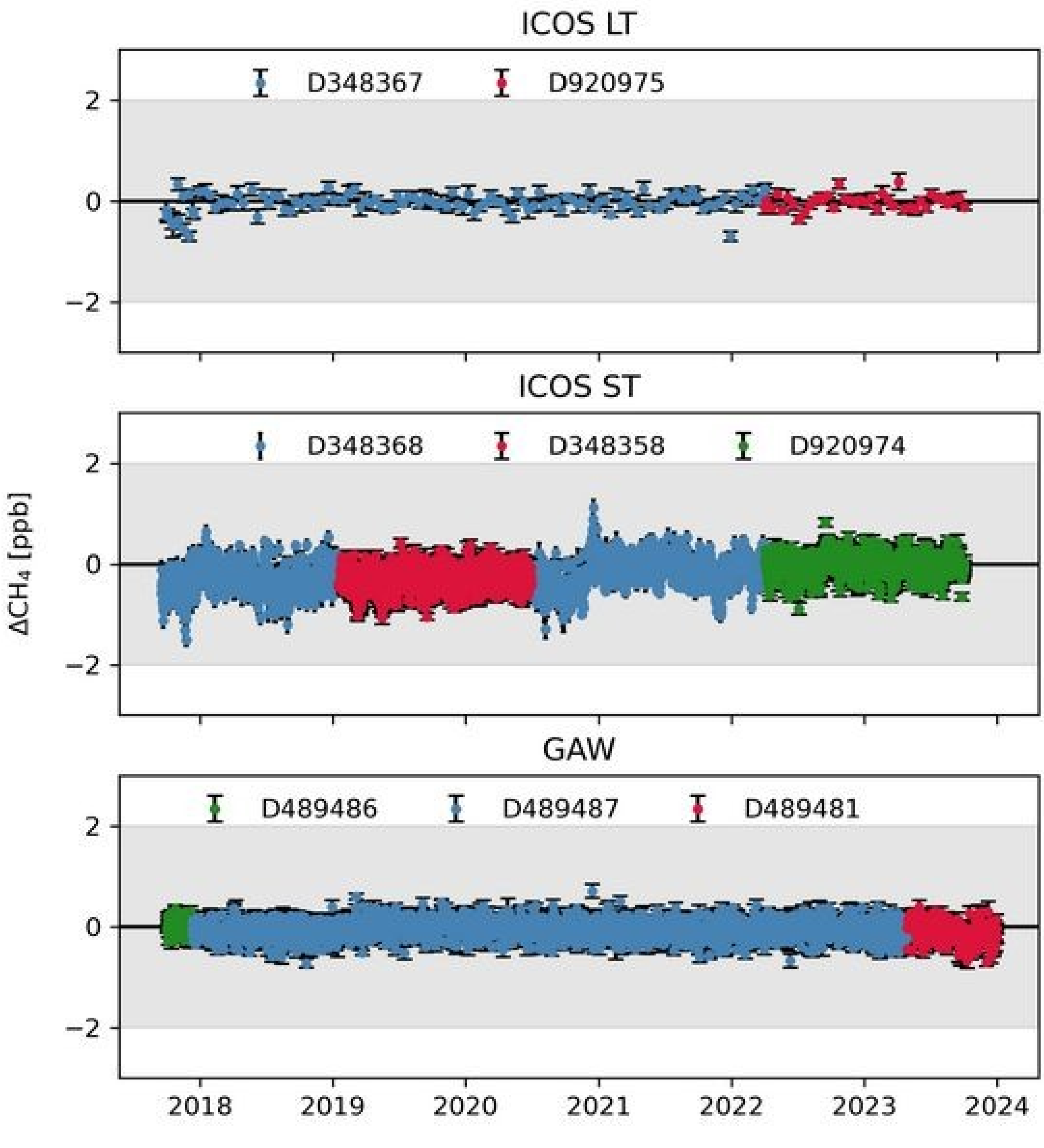


Figure 13. CH4 target measurements of ICOS LT, ICOS ST and GAW target cylinders over the whole comparison period. Different cylinders used are marked with distinct colors. The data are given as means of each sequence with the associated standard deviation

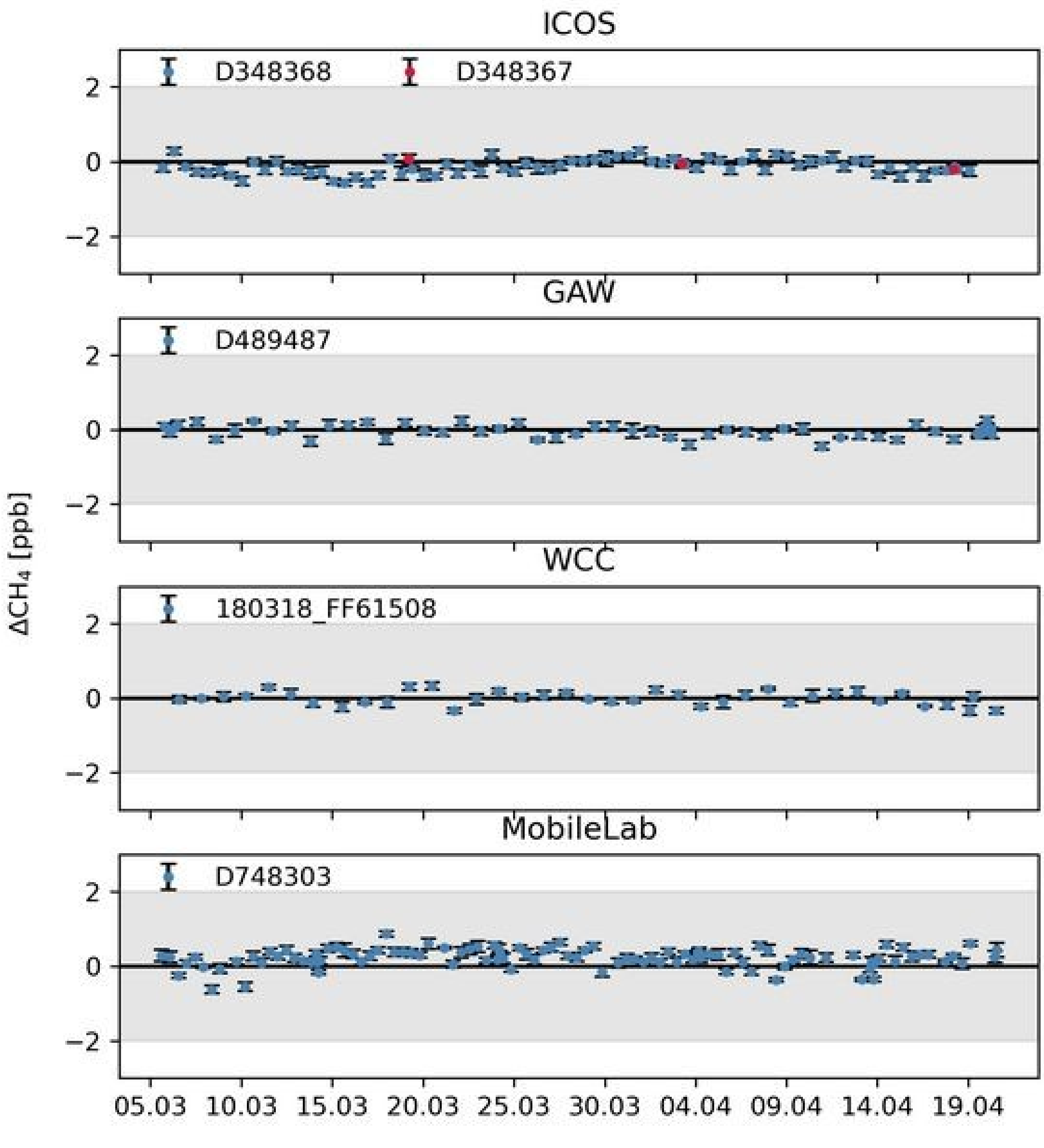


Figure 15. CH4 target measurements for ICOS, GAW, WCC and Mobile Laboratory instruments during the audit period. For ICOS both ST and LT are presented. The data are given as means of each sequence with the associated standard deviation.

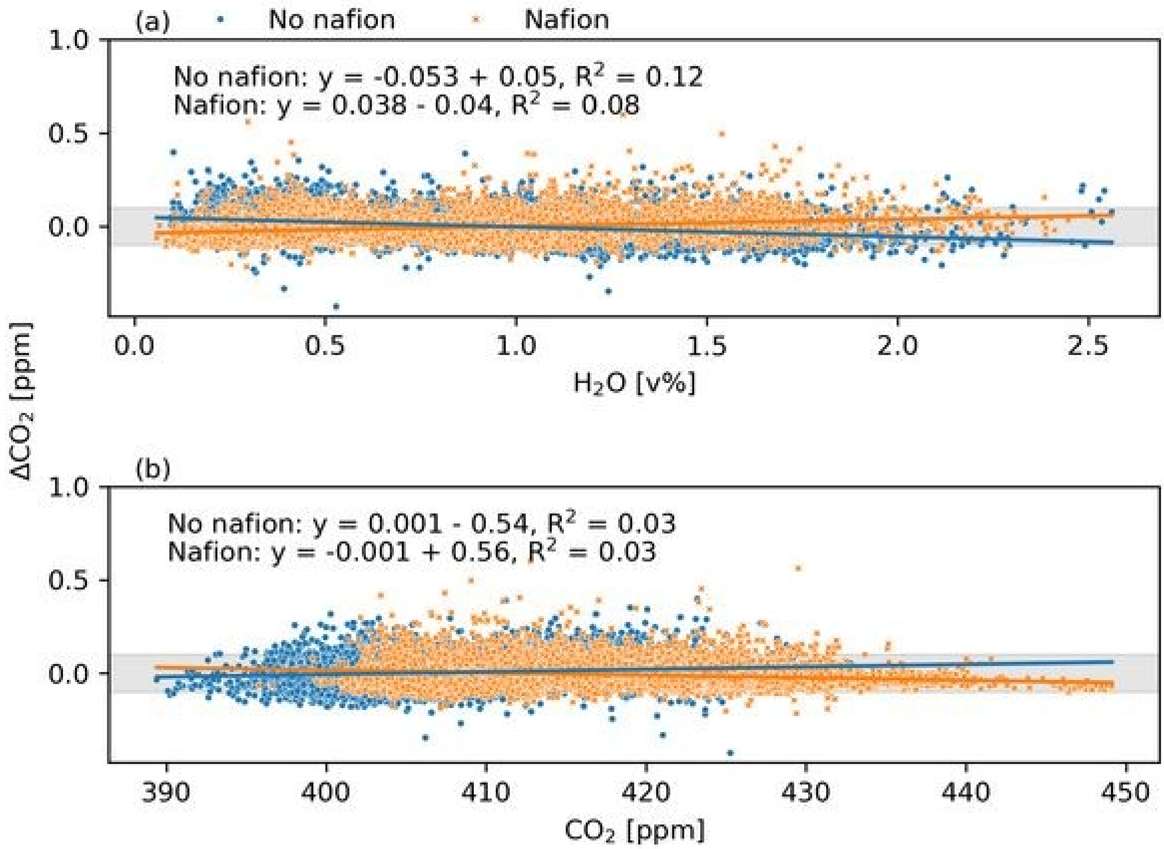


Figure A2. Dependency of the CO2 difference (GAW-ICOS) on water vapor concentration (a) and mole fraction (b). The data is split into two groups: before the installation of the Nafion and after.

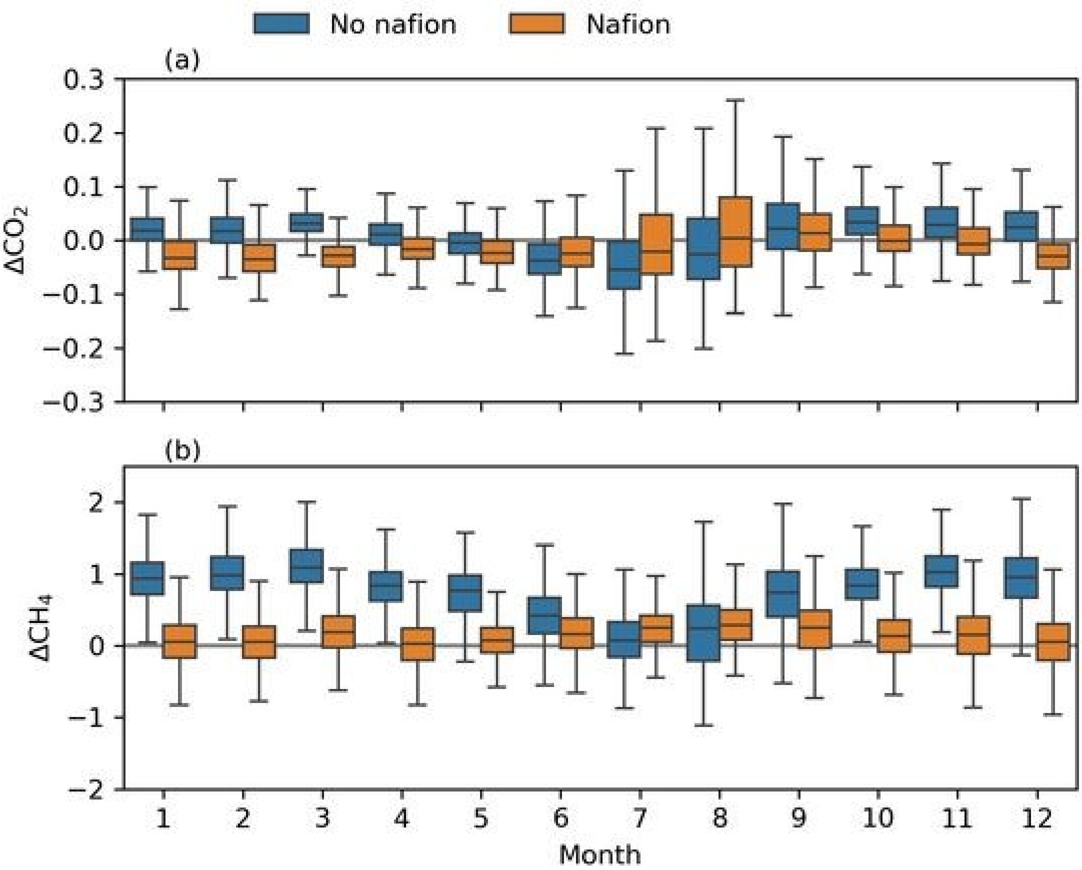


Figure A3. Seasonal variation of the differences (GAW-ICOS) in CO2 (a) and CH4 (b) before and after installation of the Nafion.

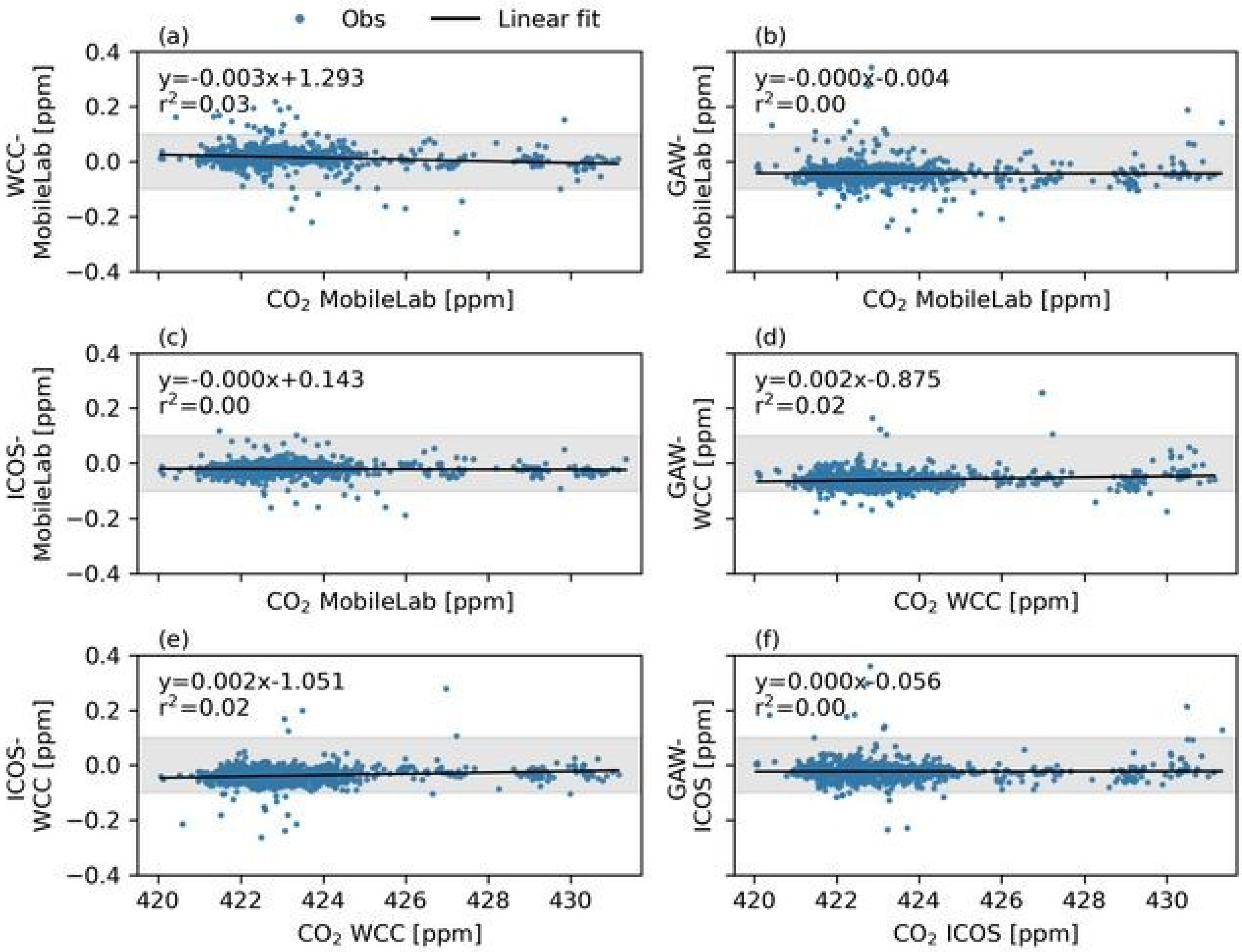


Figure A4. Mole fraction dependency of the difference between each comparison pair for CO2. Linear regression fitted to the data.

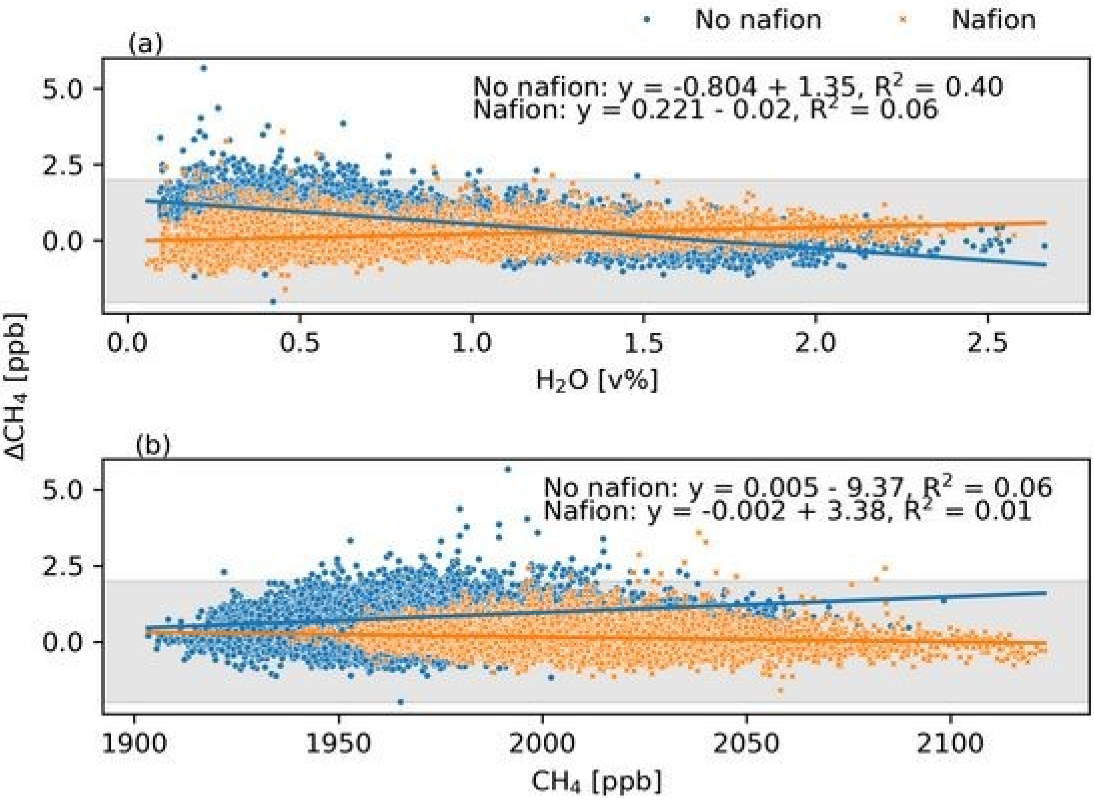


Figure A5. Dependency of the CH4 difference (GAW-ICOS) on water vapor concentration (a) and mole fraction (b). The data is split into two groups: before the installation of the Nafion and after.

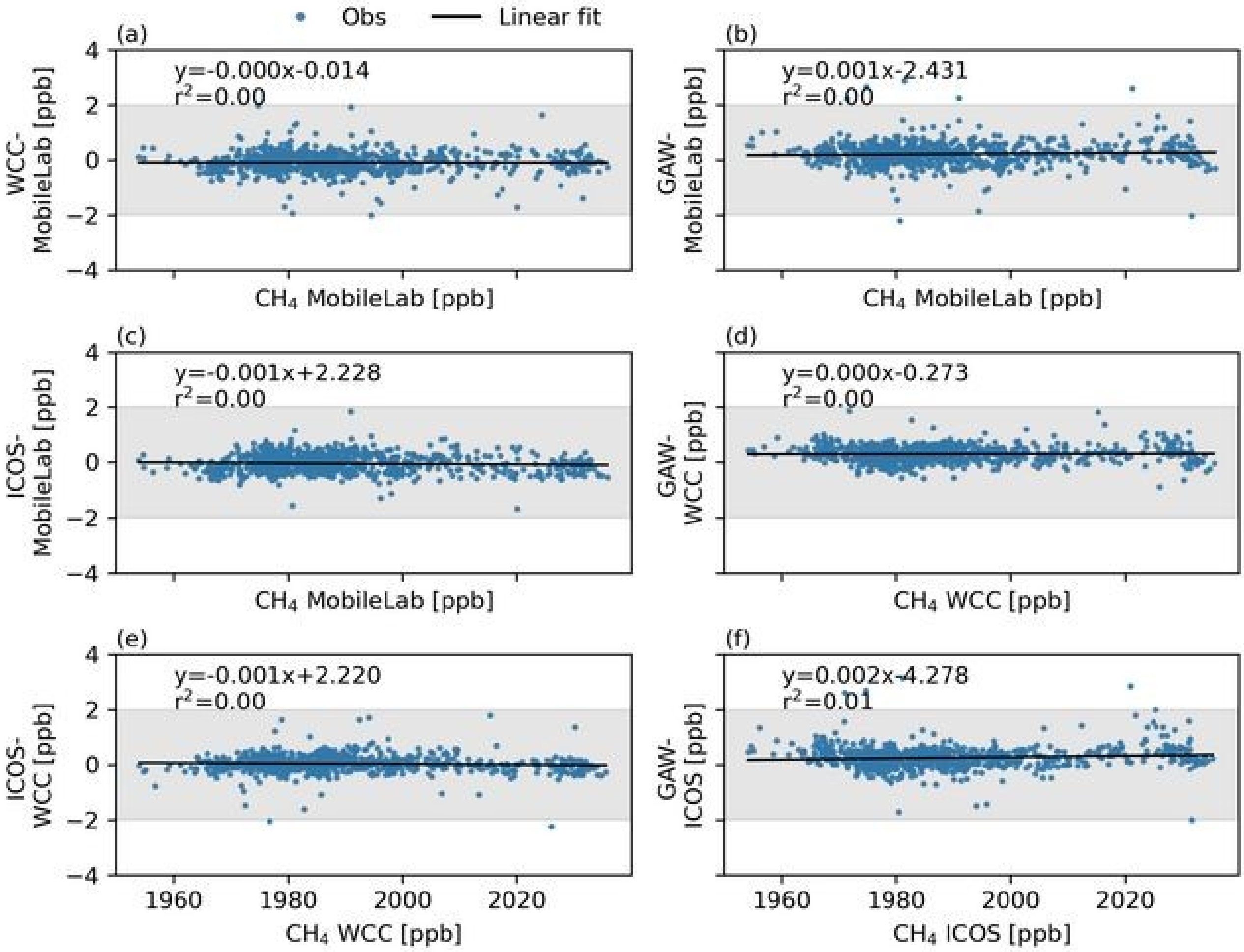


Figure A6. Mole fraction dependency of the difference between each comparison pair for CH4. Linear regression fitted to the data.

Added tables:

	Cylinder	Purpose	LTR (ppm)	STR (ppm)	Bias (ppm)	Conc (ppm)	Nb	Measure time (min)
Full period								
ICOS	D348367	LT	0.02	0.01	0.03	450.80	10	20
	D920975	LT	0.01	0.01	0.03	461.97	10	20
	D348358	ST	0.02	0.01	-0.05	409.76	10	20
	D348368	ST	0.03	0.01	-0.03	399.71	10	20
	D920974	ST	0.02	0.01	-0.03	415.02	10	20
GAW	D489481		0.02	0.01	0.01	418.30	9	18
	D489486		0.01	0.01	0.04	396.61	9	18
	D489487		0.02	0.01	0.01	411.07	9	18
Audit period								
ICOS	D348367	LT	0.02	0.01	0.05	450.80	10	20
ICOS	D348368	ST	0.02	0.01	-0.04	414.64	10	20
GAW	D489487		0.01	0.01	0.03	411.07	9	18
WCC	180318_FF61508		0.01	0.01	-0.05	417.57	4	9
MobileLab	D748303		0.01	0.01	0.02	411.94	8	20

Table 3. Results of the target measurements of CO_2 for each instrument for the full period of comparisons between ICOS and GAW instruments and for the audit period. Nb refers to the number of data points used for averaging and measure time is the total time each cylinder is measured during one injection.

	Cylinder	Purpose	LTR (ppb)	STR (ppb)	Bias (ppb)	Conc (ppb)	Nb	Measure time (min)
Full period								
ICOS	D348367	LT	0.18	0.11	-0.03	2097.11	10	20
	D920975	LT	0.14	0.10	-0.01	2196.16	10	20
	D348358	ST	0.24	0.11	-0.31	1949.49	10	20
	D348368	ST	0.30	0.12	-0.24	1948.65	10	20
	D920974	ST	0.22	0.10	-0.04	1936.69	10	20
GAW	D489481		0.21	0.11	-0.12	1961.42	9	18
	D489486		0.15	0.10	-0.01	1686.86	9	18
	D489487		0.18	0.10	-0.06	1938.79	9	18
Audit period								
ICOS	D348367	LT	0.14	0.11	-0.06	2097.11	10	20
ICOS	D348368	ST	0.20	0.11	-0.12	1974.51	10	20
GAW	D489487		0.17	0.10	-0.04	1938.79	9	18
WCC	180318_FF61508		0.18	0.09	<0.01	1963.81	4	9
MobileLab	D748303		0.25	0.10	0.23	1937.38	8	20

Table 5. Results of the target measurements of CH_4 for each instrument for the full period of comparisons between ICOS and GAW instruments and for the audit period. Nb refers to the number of data points used for averaging and measure time is the total time each cylinder is measured during one injection.

	CAL 1	CAL 2	CAL 3
CO_2, assigned [ppm]	379.24	414.46	449.39
CO_2, measured [ppm]	379.21	414.37	449.40
ΔCO_2 [ppm]	-0.03	-0.09	0.01
CH_4, assigned [ppb]	1985.48	1799.53	2210.77
CH_4, measured [ppb]	1985.36	1798.99	2210.63
ΔCH_4 [ppb]	-0.12	-0.54	-0.14

Table A1. Cross-calibration of the ICOS Mobile Laboratory calibration standards with the Pallas ICOS analyzer: The assigned values of the cylinders, average measured values with the GAW analyzer, and the difference of measured value to the assigned value.

	CAL 1	CAL 2	CAL 3	CAL 4	CAL 5	CAL 6	CAL 7
CO_2, assigned [ppm]	378.12	387.39	406.99	411.21	417.53	412.70	427.81
CO_2, measured [ppm]	387.15	387.39	407.01	411.18	417.48	412.67	427.80
$\Delta CO2$ [ppm]	0.03	0.00	0.02	-0.03	-0.05	-0.03	-0.01
CH_4, assigned [ppb]	1883.44	1890.78	1933.20	1953.82	1963.81	1998.97	2191.22
CH_4, measured [ppb]	1884.16	1891.23	1933.69	1954.13	1964.16	1999.13	2191.50
ΔCH_4 [ppb]	0.72	0.45	0.49	0.31	0.35	0.16	0.28

Table A2. Cross-calibration of the GAW travelling standards with the Pallas GAW analyzer: The assigned values of the cylinders, average measured values with the GAW analyzer, and the difference of measured value to the assigned value.