**Author's Response**

Reviewers' comment: Blue
Author's Response: Black and white; with added text in italic font

**Review 2**

L20, gives the impression that 3 brands will be tested, but they are all Picarro.

We see why the sentence can be read like that, although we did write in parenthesis (G5131-*i*, Picarro Inc.). We have changed the wording to be more clear. The new sentence reads:

*This study presents a comprehensive and detailed correction and calibration protocol to post-process $N_2O$ isotopic data, exemplified by data obtained from three commercial G5131-i cavity ring-down spectroscopy (CRDS) analysers manufactured by Picarro Inc.*

L24, if only Picarro's were tested, is it fair to say "any" instrument must be routinely characterized? Would another instrument brand be more reliable?

It is true that we do not know this for other instruments, although we suspect this is the case. We have added "G5131-*i*" as exemplified below:

"*Consequently, any G5131-i instrument must be routinely characterised to maintain high-quality data."* (See tracked changes L24-25)

L25-26, The earlier Stephen Harris paper seems to suggest that some instruments have additive effects and that these corrections are possible. Are these non-additive effects specific to Picarro?

Based on our observations in this study, the behaviour described here includes non-additive components that cannot be fully corrected using purely additive approaches. Our earlier work in Harris et al. did not test additive effects but speculates on additivity of corrections. At present, we lack sufficient comparative data across different instrument types to determine whether these non-additive effects are specific to Picarro CRDS instruments or represent a more general characteristic of this detection scheme of wavelength region applied for $N_2O$ isotope measurements. In section 4.1.1 we speculate, that effects might be related to an erroneous $CO_2$ interference post-correction. However, this would require retrieval and analysis of analyser raw data in agreement with recommendations by CCQM GAWG-IRWG. However, this is beyond the scope of the present study, and our data cannot add much substance to draw broader conclusions at present.

L28, the MATLAB code is certainly a nice addition. Of course not necessary, but it would be wonderful if there was an R version available so that everyone can access it.

Both reviewers have flagged this, and in principle, we agree that an R or Python implementation would improve accessibility and inclusion, but rewriting and validating the code in another language is beyond the scope of the present study. The present code was implemented in MATLAB because this is the environment in which the authors were most confident and efficient. The MATLAB source code is openly available under an open license and can be viewed and adapted without MATLAB using any standard text editor. This is a worthwhile direction for future work (see our reply to RC1).

L31, consider reversing the words to state "case study".

We have revised the wording in L31 as suggested by the reviewer.

L39, Consider using "N2O" consistently.

We thank the reviewer for pointing this out. We have made this change in the manuscript.

L465-466, when it is stated that the data fits well with a linear model, how is that assessed?

We used a first-order polynomial (linear) regression model in MATLAB to determine the slope and intercept of the data. The adjusted $R^2$ value was used to estimate the goodness of fit. The adjusted $R^2$ provides a measure of how well the data conform to the linear model while accounting for the number of fitted parameters. The details of the linear fit model and the formulae used for the computation of adjusted R2 is presented in L250.

In our analysis, an $R^2$ value greater than 0.60 is used as a pragmatic threshold to indicate an adequate representation of the data by a linear model. For our dataset (n = 21, p = 1), this corresponds to an adjusted $R^2$ value of approximately 0.58. We therefore use this as a threshold to assess the goodness of the fit. We have now added this information in L265:

*"In this study, values of $R^2 > 0.60$ (adjusted $R^2$ ~0.58 for our dataset) are used as a pragmatic threshold to indicate an adequate representation of the data by a linear model."*

Fig 1. Please include more tick labels in the X and Y axis of figure 1. Unclear what is meant by "sliced" data and it would be helpful to include that in the figure caption.

We have added additional tick labels to both the X and Y axes in Fig. 1 to improve the visual presentation of data. For the Allan variance analysis shown in Figure 1, pressurised air was used as the sample. The detailed description of raw data, sliced data and drift-corrected data is presented in L115-130. We have now added the term 'sliced data' in the Fig. 1 caption for clarity.

Fig 5. There might be value in adding the R2 of the relationships to better understand the linear relationships (see above comment).

Thank you for the very helpful suggestion. The adjusted $R^2$ values corresponding to the data presented in Figure 5 are listed in Table 4. Since the figure already contains multiple data sets, fitted line plots and their 95% confidence bounds, we chose to present the slope, intercept and the adjusted $R^2$ values in a separate table to keep the figures clean and understandable. We have now revised the figure caption to explicitly state that the slope, intercept, and adjusted $R^2$ values for each linear fit are provided in Table 4.

Fig 10, the caption refers to solid gray lines, but I can't find gray lines in the figure other than the dashed ones.

Thanks for spotting this. Given the high information density of Figure 9 and 10, we have decided to remove both the grey and black solid lines from the figure and retain only the symbols. This should reduce visual clutter and improve readability. The captions have been updated accordingly.