Gomez-Pelaez, A. J., Ramos, R., Gomez-Trueba, V., Novelli, P. C., and Campo-Hernandez, R.: A statistical approach to quantify uncertainty in carbon monoxide measurements at the Izaña global GAW station: 2008–2011, Atmos. Meas. Tech. Discuss., 5, 6949-6989, doi:10.5194/amtd-5-6949-2012, 2012

Author response to the referee comments

We thank the two anonymous referees for their comments on the manuscript. In the responses given below, the comments are in blue italics and our replies in black. Referee comments are labelled as R1 or R2 following the journal's identification. The changes made to the manuscript in response to specific referee comments are listed below. The revised manuscript is attached as pdf, in which the insertion of text is noted by a blue arrowhead and removed text appears crossed out. The new text and label of the corresponding referee comment can be viewed by moving the mouse pointer over the arrowhead or the crossed out text.

Reply to Anonymous Referee #1

1) General comments:

R1-1. The manuscript presents a detailed statistical study of various contributions to the uncertainty associated with the continuous CO measurements at the station Izaña. The approach goes beyond the usually estimated uncertainties of measurement results. As such, it is very informative, and the paper presents new insights. The calculations would be applicable to other stations, notably within the GAW network. Therefore the paper might be of interest to a wider readership, notably readers working in the field of high-quality trace gas monitoring.

The paper is well structured and clearly organized. The outcome is well summarized. The abstract covers all relevant findings. The style of tables and figures is appropriate. Nonetheless, there are a number of details, mostly technical in nature, which should be improved. These are specified below. In summary, the paper should be accepted for publication after minor revisions have been made.

2) Specific comments

R1-2. page 6953, line 6: The diameter and length of the columns should be given.

Agreed. In the revised manuscript, we have added the sentence: "For both columns, the outer diameter is 3.2 mm and the length is 76.8 cm".

R1-3. p 6953, l 22: Delete "primary". According to the recommended GAW terminology a laboratory standard is the standard of the highest rank at a GAW station.

We agree and have deleted "primary" in the revised manuscript.

R1-4. p 6955, lines 7 and 22: While the CO mole fractions are known to drift in high-pressure cylinders, the usually observed drift is rather slow. Here the authors observe negative drifts ranging from 0.58 to 1.63 nmol/mol per month (!) and a positive drift of 2.75 nmol/mol per month (!). Could the authors comment on the reasons for this rapid drift?

The rapid drift rates are most likely due to the rapid consumption of the working gases. Their internal pressure decreases from 125 to 25 bar during the few months they are in use. Change in CO contained in cylinders is also related to cylinder materials. Those

used to hold the Izaña working standards may have an aluminium alloy prone to greater CO drift.

In the revised manuscript, we have added a sentence in page 6954 line 4:
“The 29-litre cylinders containing the laboratory standards were obtained from Scott-Marrin, Inc, whereas the 20-litre cylinders containing the working standards were obtained from Air Liquide Spain. They may differ in the type of aluminium alloy used and their internal conditioning.”

We have also rewritten the sentence of page 6955 lines 23-25 as follows:
“These rapid changes likely result from the interaction of CO with the internal surface of the cylinders and the rapid decrease in their internal pressure (from 125 to 25 bar) during the few months they are in use.”

R1-5. p 6959, paragraph from line 12 to line 19: This is a very important paragraph with its reference to Fig. 5 and Table 3. Here major results are summarized, which demonstrate the success of improved analytical quality.

R1-6. p 6959, l 10: "peak baseline": Is this really a well-known expression? You probably mean the imaginary baseline connecting peak start and peak end.

Yes, the referee is correct. We mean the imaginary straight line which connects the peak beginning and the peak end. Dyson (Dyson, 1998, “Chromatographic integration methods”, Second edition, The Royal Society of Chemistry) calls it both the “peak base” and the “baseline placement under the peak”. In the revised manuscript, we use “peak base”.

R1-7. p 6962, l 4: Could the authors give a short explanation why u<par> behaves as random for the annual means in contrast to the other averaging periods?

The response function parameters are determined every two weeks. For each parameter there is a time series of parameter residuals (one value every two weeks) with respect to the considered linear drift or constancy in time of the parameter. In this series, which could be considered as the time series of parameter error, we do not observe significant autocorrelation for time-lags larger than two months. Therefore, we expect random cancellation of error when computing annual means. However, for periods one month or smaller, the parameter error is poorly monitored (only 2 values per month). We stay on the safe side considering it as autocorrelated in time. Therefore, we consider the parameter error behaves as a systematic error for periods one month or smaller.

This is a rather long explanation, and we think it is better not to implement it in the revised manuscript.

R1-8. p 6962, l 22: CCGG stands for Carbon Cycle Greenhouse Gases Group (see website).

In the revised manuscript, we have added the missing word “Gases”.

R1-9. p 6969, l 15-17: Have the changes in sampling time been accounted for in NOAA’s analyses of the time series?

NOAA tends to focus on CO in the remote marine boundary layer. Novelli et al. (1992, 1998, 2003) do not include the Izaña station due to its high altitude.

R1-10. p 6976, Table 3: This is indeed an important Table, which nicely shows the different contributions to the uncertainty and the progress achieved for the analytical quality.

3) Technical corrections

R1-11. Hyphenation (entire manuscript): There are a number of cases where hyphens are lacking according to the rules of grammar, which contrasts with the correct use in many cases. Examples for missing hyphen: p 6952, l 27. Should read: 300-ml glass flask p6953, l 10: Should read "High-purity synthetic air" p 6954, l 3: Should read: "highpressure tanks" p 6955, l 5: Should read: "The time-dependent response . . ." and several similar cases.

In the revised manuscript, we have added hyphenation as needed.

R1-12. page 6950, line 3: The abbreviation GAW should be spelled out when used for the first time.

Agreed, we have added this in the revised manuscript.

R1-13. p 6950, l 6 and same word on other pages: Remove "'s" from "Izana's"

The revised manuscript includes this correction.

R1-14. p 6950, l 14/15: Remove plural "s" from "parameters"

In the revised manuscript, the structure of the sentence has been changed. Now, the word "parameters" does not act as an attributive noun, and therefore, it remains in plural form (also in p 6956, l 25; p 6969, l 26).

R1-15. p 6951, l 5: Replace "role on the cycles" with "role in the cycles" (or "for")

Done.

R1-16. p 6951, l 5: Consider writing "hydroperoxy radical (HO2)"

As suggested, we added "hydroperoxyl radical". The "l" is necessary according to Seinfeld & Pandis ("Atmospheric chemistry and physics", Wiley-Interscience, 1998).

R1-17. p 6951, l 22, p 6958:l 5: Typo, remove space from "can not" (cannot)

Done.

R1-18. p 6952, l 24: Consider replacing "a.g.l." with "above ground level".

Done.

R1-19. p 6952, l 26-27: The order should be rearranged to yield: ". . . general inlet to the analytical system using a KNF . . ."

Agreed, in the revised manuscript we have rearranged the sentence as suggested.

R1-20. p 6953, l 10: Should better read "We used a stainless . . ."

Agreed.

R1-21. p 6953, l 12: Should better read "positions of the valve, . . ."

This sentence has been removed in the revised manuscript due to the comments of referee 2.

R1-22. p 6953, l 14: Style, should be rephrased to, for example: ". . .chromatogram, where the H2 peak appears first, followed by the CO peak"

Agreed, in the revised manuscript we have rephrased the sentence as suggested.

R1-23. p 6954, l 5: Rephrase to ". . . the procedure for conditioning described by Lang . . ."

Done.

R1-24. p 6954, l 14: Rephrase to ". . . are obtained by fitting . . ."

Done.

R1-25. p 6955, l 28: One should better say: "We begin by considering . . ."

Agreed, we have made the suggested change.

R1-26. p 6958, l 19 and similar cases: Should read: ". . . these parameters with respect to . . ."

Agreed, we have made the suggested change.

R1-27. p. 6960, l 23: Style. rather say " . . . the exact value of . . .does not matter."

Done.

R1-28. p. 6968, l 11: Style; rather say: ". . . increased by 4.0 . . ."

Done.

R1-29. p. 6968, l 16: Style, rather say: " . . . is in the middle of August." or " . . . is in mid-August."

Done.

R1-30. p 6969,l 15: Style, better say: " . . . at Izana in late 1991, . . ."

Agreed. The text was changed as suggested.

R1-31. p 6971, l 2: Style, rather say: " . . . are highly biased with respect to . . ."

We have not applied the proposed change because it would alter the meaning of the sentence. The actual sentence “… are biased high with respect to …” means that there is a positive bias, whereas the proposed sentence “… are highly biased with respect to…” means that there is a large bias.

R1-32. p 6971, l 21: You might consider just saying: " . . . Zellweger for advice during . . .".

Agreed, in the revised manuscript we have re-written the sentence as suggested by the reviewer.

R1-33. p 6975, Table 2, header: It should read "Residuals with respect to . . ."

In the revised manuscript, Table 2 has been removed (see referee comment R2-20).

R1-34. p. 6982, Fig. 2, caption. Style, rather say. "The fit is plotted . . ." or " The fitting curve. . ."

Done, the revised manuscript says “The fitting curve …”.

Reply to Anonymous Referee #2

R2-1. The manuscript ‘A statistical approach to quantify uncertainty in carbon monoxide measurements at the Izaña global GAW station: 2008–2011’ by A. J. Gomez-Pelaez et al. presents a thorough estimation of the uncertainty of carbon monoxide measurements carried out at Izaña. The methodology is sound and covers the most of the relevant contributors to the overall uncertainty. Similar approaches might be used at other stations and for other measurement parameters. Izaña continuous CO data is compared to collocated flask measurements, and the additional uncertainty of the discrete flask sample due to smaller data coverage is also estimated. As such, the manuscript is a valuable contribution for AMT.

R2-2. However, the readability and presentation quality needs further attention before publication in AMT. I recommend publication after addressing the following points: The manuscript is rather difficult to read over large parts. I recommend that the language of the manuscript is checked by a native English speaker, or that the manuscript receives copy-editing for English. It should also be checked if clearer (and shorter) formulations can be used. The presentation quality needs to be improved before publication in AMT.

Following the referee advice, Paul Novelli (a co-author who is a native English speaker) has checked the language and presentation of the manuscript. Details of the changes made to the manuscript are provided in the attached pdf. The statistical treatment of uncertainty is a demanding topic. We have tried to make the formulations as concise and clear as possible. A shorter presentation may not contain enough information to allow other scientists to apply our methods.

Specific comments:

R2-3. P6950 L18-19: I understand what you mean here, but this nevertheless needs to be re-phrased.

This referee comment is probably related to comment R2-17. We have proceeded in a similar way than for that comment. In the revised manuscript, we have removed “exactly” and added “temporal”. The sentence now reads “A fifth type of uncertainty we call representation uncertainty is considered when some of the data necessary to compute the temporal mean are absent”. We have applied the same changes in page 6970 line 13.

R2-4. P6950 L24: ‘The uncertainty in the differences is determined and whether these are significant.’ This is difficult to understand.

In the revised manuscript, this sentence has been rewritten as follows: “The uncertainty in the differences is computed and used to determine whether the differences are significant.”

R2-5. P6951 L5: ‘hydroperoxy’ should be ‘hydroperoxy radical’.

Agreed. We include the term “hydroperoxyl radical”. The “l” is necessary according to Seinfeld & Pandis (“Atmospheric chemistry and physics”, Wiley-Interscience, 1998).

R2-6. P6951 L14: ‘the quality objectives’ replace with ‘the data quality objectives’.

Done.

R2-7. P6952 L11: ‘might’ instead of ‘my’.

In the revised manuscript, we have used “may”.

R2-8. P6953 L10: ‘steel’ instead of ‘steal’.

Done.

R2-9. P6953, description of the RGA system: This needs to be revised. E.g., sentences such as ‘The system is controlled by a computer’ contain no relevant information. Consider to delete this, or alternatively give more (relevant) details.

In the revised manuscript, several sentences of this section have been removed as suggested by the referee.

R2-10. P6953 L20: It should be mentioned here that NOAA/GMD is the CCL for CO.

Agreed, this is included in the revised manuscript.

R2-11. Section 3: Do you have any indication of the stability of the standard gases over time? Have they been re-calibrated at NOAA?

The laboratory standards have not yet been recalibrated by NOAA. We believe there is sufficient evidence that our laboratory standards have not drifted, as discussed in the following paragraph, which has been added to the revised manuscript (page 6953 line 25).

“Stability of the Izaña laboratory standards was evaluated in two ways, being both compatible with no drift in the mole fraction of these gases. First, in 2009, the WMO World Calibration Centre (WCC) for CO, which is hosted by EMPA, carried out an audit at the Izaña station (Zellweger et al., 2009) which included a blind analysis of five WCC travelling CO-in-air mixtures with mole fraction ranging from 88 to 201 nmol/mol. In each analysis, repeated injections of travelling cylinder gas alternate with working gas injections. The WCC assignments initially used were on an earlier version of the WMO scale, WMO-2000 (Zellweger et al., 2009}. When the WCC travelling standards were revised to the WMO-2004 scale used at Izaña, the differences in the mole fractions assigned by Izaña and the WCC ranged from -1.69 to 2.63 nmol/mol (Zellweger, 2010, private communication). If we consider only the three travelling cylinders within the ambient range at Izaña (~ 60-150 nmol/mol), the differences range from -1.69 to 0.45 nmol/mol. The later values are compatible with the uncertainty in the Izaña RGA measurements (Section 4). Second, the stability of the laboratory standards was also evaluated by comparing Izaña in-situ measurement results with results from air

samples collected weekly in flasks at Izaña and analysed by NOAA-ESRL-GMD. The annual mean differences between CO results by the two laboratories are not significant for the years 2009, 2010, and 2011, and show no significant trend over this period (Sect. 5), indicating there was no significant change in the Izaña laboratory standards relative to their NOAA assignments."

R2-12a. Section 4: Overall, the uncertainty estimation carried out in this section is sound. However, the uncertainty of ust does not account for potential drift in the NOAA standards. I think that this should be included in a comprehensive uncertainty estimation, although it might be difficult to quantify. Maybe relative stability of the standards over time could be used for an estimate.

The referee is correct that potential drift in the laboratory standards should be considered in determination of u_st. As stated above, within the uncertainty of the measurements we do not observe drift in the NOAA standards and cannot quantify an uncertainty in the assumption of stability. We therefore proceed on the assumption that the uncertainty associated with the laboratory standards has remained constant with time. We have added several sentences in the revised manuscript to clarify this point. The first part of the second paragraph of Sect. 4 has been rewritten as follows:

"The term u_st (Eq. 4) was obtained through a least-squares fit of the standard uncertainties for the WMO standard gases of the Izaña station (Table 1). Therefore, u_st represents the mole fraction dependent uncertainty due to the WMO standard gases assuming they have been stable over time. As stated in Sect. 3, within the uncertainty of the measurements we do not observe significant drift in our laboratory standards. In a more general case of Eq. (4), u_st would account for: 1) laboratory standard gas uncertainties increasing linearly in time due to undetermined potential drifts in the laboratory standards; 2) the uncertainty in laboratory standard drift rates in case significant drifts had been determined."

R2-12b. I didn't understand the difference between the uncertainty of the fit (ufit) and the uncertainty of the response function parameters (upar). Aren't these two uncertainties highly correlated?

They describe different uncertainty components in the response function. To clarify the difference between them, in the revised manuscript we have introduced the following text in Sect. 4 (at the end of page 6958): "Note that the term u_par has been obtained propagating only the parameter repeatabilities. u_par does not include other components of the parameter uncertainties. For example, it does not include the parameter uncertainties that could be estimated for each calibration following Sect. 8.1.2 of Martin (1971). Also, it does not include the whole uncertainty of r_wg (the mole fraction of the working gas). Therefore, what u_par takes into account is the temporal consistency of the response function. The term Sqrt(u_st^2+u_fit^2) provides the uncertainty in the response function for each calibration event (every two weeks). However, for the rest of time instants, the response function is used without performing any calibration, and therefore the term Sqrt(u_st^2+u_fit^2+u_par^2) provides the uncertainty in the response function."

In the published AMTD manuscript, u_par is called "the propagated uncertainty related to the response function parameter uncertainties". In the revised manuscript, u_par is called "the propagated uncertainty related to the temporal consistency of the response function parameters".

R2-13. P6958, L8/9: The values of these standards are different compared to the standards listed in Table 1. Was this a set of different standards? Since the RGA is

non-linear, sufficient coverage of the relevant mole fraction range is very important to obtain adequate calibration functions. Your approach to use the uncertainty of the fit with 5 standards as the minimum uncertainty makes only sense if the calibration functions were similar. Was this the case?

Yes, this is a different set of laboratory standards. We clarify this in the revised manuscript. Yes, the response functions determined in the calibrations were similar. We checked this by comparing the response function parameter results determined in two set of calibrations performed in 2009 using the two different set of laboratory standards but the same working standard.

R2-14. P6958, L9: re-phrase (avoid abnormally)

Done.

R2-15. P6958, L14/16: Sentence 'Taking differentials in . . .' is a fragment / makes no sense.

Agreed. In the revised manuscript, we have rewritten it as: "Taking differentials in Eq. (1), we obtain the equation …", where "we obtain the equation" is the new text included.

R2-16. P6958, L19: 'parameters respect' should be 'parameters with respect'.

Done.

R2-17. P6959 L25: '. . .number of data necessary to compute exactly the mean": This need to be re-phrased. The mean can always be calculated 'exactly' (i.e. the mean of 1 and 2 is exactly 1.5). You probably mean that you increase the uncertainty of the mean if you have a smaller number of data points. With 'exactly the mean' you probably think of the true value. Also 'actual mean' should rather be true value, which is in fact always unknown.

We agree with the referee's comment that the expressions "… number of data necessary to compute exactly the mean" and "actual mean" are confusing. However, what the expressions "exactly the mean" and "actual mean" intended to represent is not equivalent to "true value" at all. In the revised manuscript, we have rephrased all these expressions to clarify the discussion. The first part of the first paragraph in Sect. 4.1 is now written as follows (the revised text is identified by red italics):

"There is a fifth type of uncertainty we call representation uncertainty, u_rs. This is present when computing *a temporal* mean from a number of available data (n) that is smaller than the *theoretical maximum number of independent data (N) within the time interval in which the temporal mean is defined. The temporal mean computed from the n available data may* be different from the *mean determined from the N data (unknown)*. The representation uncertainty quantifies this difference *statistically*. In time series analysis a hierarchy of data assemblages are possible (e.g. hourly mean, daily mean, monthly mean, annual mean), each being computed from the means of the previous level. An additional representation uncertainty is associated with each assemblage. For example, an additional representation uncertainty will appear when computing a daily mean from only 22 available hourly means (N = 24, and n = 22). The value N is known *precisely* for each level except for the first. *For example, in the Izaña data, the lowest ensemble is the hourly mean, for which n=3 and N is unknown but certainly greater than n.*"

R2-18a. Section 4.1: The distribution of missing values has an influence on the uncertainty and the resulting bias. E.g. it makes a difference if the first three hours of a day are missing or if the three missing hours are evenly distributed over the day in case of a significant diurnal cycle. This is not reflected in the 'representation uncertainty'.

The referee is correct. The representation uncertainty has a statistical nature and assumes the missing values are randomly distributed with respect to the mean. If this is not the case, the actual representation error could be larger than what the computed representation uncertainty predicts. In the revised manuscript, we have added the following sentences at the end of page 6960 to clarify this point:
"It assumes that the missing values are randomly distributed with respect to the mean. If this is not the case, the actual representation error could be larger than what the computed representation uncertainty predicts. For example, if three consecutive hours of a day are missing and they are located at an extreme of a significant diurnal cycle, the representation uncertainty will underestimate the actual representation error."

R2-18b. This section should also be condensed; it is too long and difficult to follow.

This section presents the concepts of the representation uncertainty and the propagation of uncertainty in temporal means. Both are rarely discussed and therefore require sufficient explanation. In the revised text, we edited and rephrased certain text to make the discussion easier to follow.

R2-18c. The last paragraph (starting P6962 L21) should be moved to section 5.

This paragraph deals with the computation of the representation uncertainty for the NOAA flasks, a concept presented in Sect. 4.1. Section 5 deals with the comparison between NOAA flasks and in-situ measurements. It would make little sense to compute the representation uncertainty of the NOAA flasks in Sect. 5. Since NOAA flasks are treated first in Sect. 4.1., it is in this section where they should be introduced. However, we have moved the description of the quality control procedure applied to the flasks to Sect. 5. Additionally, we realise that the continuity in the text of section 4.1 is broken when this paragraph appears. To prevent this, in the revised manuscript we have changed the way in which this paragraph is introduced: "As an example of the large representation uncertainty introduced when using very sparse data, we consider the ambient air samples collected weekly at Izaña Observatory since 1991 for analysis at NOAA-ESRL-GMD Carbon Cycle Greenhouse Gases Group (CCGG) as part of …"
Also, in section 5 we indicate that the NOAA flask samples collected at Izaña are introduced in the last paragraph of Sect. 4.1.

R2-19. Section 5 also needs to be shortened.

This section performs a comparison between NOAA flasks and in-situ measurements, but also deals with two additional topics: the comparison uncertainty and the weighted means. We believe the paper should cover these additional topics. A shorter presentation would not contain enough information to allow other scientists to apply our methods.

R2-20. Table 1: Is this table relevant for the manuscript? Consider removing, or combine with Table 2.

Yes, we think Table 1 is relevant for the manuscript. However, in the revised manuscript, Table 2 has been removed, since the residuals are now plotted in Fig.2.

R2-21. Figure 1 is not really necessary. Consider removing.

Figure 1 shows a typical chromatogram of the Izaña RGA. We think that some readers interested in chromatography would be interested in this small figure to see the shape of the peaks and the separation between them.

R2-22. Figure 2: Consider showing the residuals to the fit, e.g. on the second y-axis

In response to this suggestion, a revised Fig. 2 shows the residuals plotted using a second y-axis.

R2-23. Figure 3: It seems many of the working tanks are slowly drifting downwards with time. This is very unlikely, since drift in CO tanks is usually in the opposite direction. Could this be due to an upward drift in the NOAA tanks over time?

As it is stated in page 6955 lines 20-22, six of the sixteen working gases show significant drift. The drift rate is negative for five of the six drifting working gases. The NOAA standards would have to be drifting very rapidly to explain the changes observed in the working standards (see reply to referee comment R1-4). Rapid positive drift in the laboratory standards would lead to a significant positive drift in the differences of the NOAA/Izaña comparison (Section 5, Figure 6) and this is not observed. Indeed, there is sufficient evidence that our laboratory standards have not drifted, as it is argued in our reply to referee comment R2-11. To point out the occurrence of negative drifts and argue their origin, in the revised manuscript we have added the following text in page 6955 line 25:

"According to the experience of other Laboratories, CO-in-air mixtures stored in aluminium tanks usually present positive drift rates, whereas drift in our working gases is primarily negative. This may be due to the type of tanks used to store the working gases or to issues related with the Izaña station filling system. We consider the drifts are accurately determined and accounted for in the data processing."

Note that in the revised manuscript, we have added a sentence in page 6954 line 4 (see our reply to referee comment R1-4):

"The 29-litre cylinders containing the laboratory standards were obtained from Scott-Marrin, Inc, whereas the 20-litre cylinders containing the working standards were obtained from Air Liquide Spain. They may differ in the type of aluminium alloy used and their internal conditioning."

R2-24. Figure 5 shows the individual contributions of the different uncertainty components to the total uncertainty u tot. The total uncertainty must always be larger than any of the individual contribution. However, this is not the case for a large period in 2008, where u pr is larger than u tot. Please comment / correct.

The referee is correct in saying that u_tot must always be larger than any of its four components (u_st, u_fit, u_rep, and u_par), as Eq. (3) states. Note that Fig. 5 fulfils this property. However, Eq. (6) allows u_pr and u_pbeta to be larger than u_par (for a negative large enough covariance term c) and even larger than u_tot. This is what happens in Fig. 5 during part of the first half of 2008. There is nothing wrong with this fact.

To prevent any misunderstanding, in the revised manuscript we have added the following text at page 6959 line 19: "Note that u_pr was larger than u_tot during part of 2008. There is nothing wrong with this fact. According to Eq. (3), u_tot is larger than any of its four components (u_st, u_fit, u_rep, and u_par). However, u_pr and u_pbeta

can be larger than u_par and even u_tot for a negative large enough covariance term c (see Eq. (6)).”

R2-25. Figure 7 is not needed; the annual cycle can be seen in Figure 4.

Agreed. In the revised manuscript Fig. 7 is removed.