Reply to the referee comments of Anonymous Referee #1

The submitted manuscript presents 3 years of continuous molecular hydrogen (H2) measurements at a tall tower in the Netherlands. The H2 mixing ratios are alternately monitored at four heights from 20m to 200m above ground. This is a very interesting and rather unique dataset that allows to gain insight into various source and sink processes determining the local H2 concentrations at the sampling site. Hence, the paper is within the scope of 'Atmospheric Chemistry and Physics' and is of sufficient originality to merit publication in this journal. However, the manuscript is a bit wordy, too detailed in some parts but also misses some (mainly technical) information. I think the papers could become more attractive when being shortened. Some suggestions for shortening are listed below (see the specific comments).

According to the authors, the flux estimates are pretty uncertain, partly due to lack of reliable Radon fluxes. The measurements in different heights above ground should allow determining fluxes without Radon flux information, only based on H2 gradients and meteorological information. I would be curious to see if the results would match the H2 fluxes determined with the Radon technique.

Finally, some conclusions to the representativeness of the data/results should be included.

Point-by-point reply to the specific remarks

Abstract:

1. the abstract is way too long and too detailed. It has about 450 words (e.g. some journals restrict the abstract to 250 words).

A. We agree that a shorter abstract would be better and adjusted the abstract according to the reviewer's suggestions.

Suggestions to shorten the abstract are:

2. Remove page 5590, lines 8 – 11 'From the three-year ...determine multi-annual trends'.

A. The text "From the three-year ...determine multi-annual trends" has been removed.

3. Remove page 5590, line 17 'which differentiates ... measured'

A. The text has been removed.

4. Remove page 5590, lines 18 – 23 'that is, ... of the soil uptake'.

A. The text has been removed.

5. Remove page 5590, lines 25 – 28 'Local soil and weather ... in our study domain'.

A. The text has been removed.

6. Remove page 5591, line 4 'In contrast'

A. The text has been removed.

7. Remove page 5591, lines 5 – 6 'are very similar to results of previous studies'.

A. The text has been removed.

Other suggested abstract changes:

8. page 5590, line 3: add 'October' after 'since'

A. The word "October" has been added.

9. page 5590, line 7: add 'long-term' after 'quasi-continuous'

A. "long-term" had been added.

10. page 5590, line 11: 'Seasonal H2 cycles are observed [at all heights] or [with decreasing? amplitude with sampling height]'.

A. Both statements are true. We included "at all heights" in the text.

11. page 5590, line 11: add 'unique' in front of 'behaviour'.

A. The text has been changed.

12. Page 5591, line 2: what do you want to say with 'scattered'? H2/CO ratios are more variable than estimated?

A. We replaced "scattered" with "variable".

13. Page 5591, line 3 -4: replace 'a different driving regime due to frequent traffic jams' with 'frequently occurring congested traffic'.

A. We prefer to leave this statement as it was originally written. The congested traffic is the indirect cause - i.e. the emissions do not depend on the traffic density, but on the way the engines function. Replacing as the reviewer suggests could lead to the misinterpretation that the traffic density is the direct control on the H_2:CO ratios.

Main text:

14. Page 5596, lines 12 – 15: details about the GHG measurements are not needed.

A. The details here only specify the instruments used; we think this is necessary since we later mention these instruments to explain how the whole system is organized.

15. Page 5597, lines 8 – 14: what's the loop size? Is the loop temperature-stabilized?

A. The sample loop size is 5 ml and it is temperature stabilized. We added this information in the text.

How does the temperature vary in the lab? Temperature and pressure changes are linearly influencing your amount of H2 molecules in the loop and thus your result, right? Did you correct for that?

A. The lab temperature varies by up to 2°C on daily time scale, and there is a seasonal variation with a summer – winter difference of about 10°C.

The temperature and pressure variations do indeed influence the amount of molecules in the sample loop and this translates into variation of the sensitivity (i.e. area or height of the peak per unit of mole fraction). This effect is corrected by the ratio-ing of the raw data to the interpolated results for the Working tank (which is measured every half-hour), because the variations in the sensitivity are similar. After this correction we do not detect any

significant dependence on ambient temperature or pressure, as verified using the Target tank results.

You performed the calibration daily. Always at the same time of the day?

A. The calibration is performed at different hours, usually during night, at 25-hour intervals.

How often did you measure the working and the target tank?

A. The Working and the Target tank are measured every half hour (the measurement sequence takes half hour, and includes measurements of air from the four heights of the tower, followed by measurements of the Working and Target tanks). We changed the text in Sect. 2.2 to state more clearly the period of target and Working tank measurement.

Later (page 5599) you mention that WT measurements are used for correcting for short term sensitivity changes. What do you mean with short term?

A. The daily calibration corrects for sensitivity variations on daily and longer time scales. We use the WT which is measured every half hour to correct for sensitivity variations that occur between two calibrations.

16. Page 5597, lines 15 – 22: aluminium cylinders are often subject to H2 drifts. Did you check for drifts?

A. We did not notice drifts in H_2 in the cylinders that are consumed the fastest (the target and working tanks, which have a lifetime of up to 2 months). For only one cylinder we did notice instabilities in CO at low cylinder pressure, but H2 was not affected.

Calibration gases: so far we did not notice drifts when comparing the calibration gases with each other. Moreover, we performed intercomparisons with other laboratories using travelling cylinders at different times (2008 - 2009) and the results did not significantly change in time, which shows that our calibration cylinders did not drift within our detection limit. (see also Sect. 3.2 Precision and accuracy).

17. Page 5598, lines 10 – 11: 'the only difference … T1 and T2' could be removed.

A. The text "the only difference … T1 and T2 (Fig. 2)" has been removed.

18. Page 5598, lines 18 – 21: Radon monitor: add some information such as: manufacturer; how is the calibration done, where is the Rn monitor on Fig. 2, does it have separate inlet lines?

A. The ^{222}Rn monitor is a completely separated system and is not included in Fig. 2.

We extended the text in Sect. 2.2 as follows:

" ^{222}Rn is measured in air sampled via dedicated inlet lines at two heights of the tower, 20 and 200 m, by two independent dual flow loop, two-filter radon instruments, designed and constructed at the Australian Nuclear Science and Technology Organisation (ANSTO) (Whittlestone & Zahorowski, 1998). The inlet flow rate is 80 L/min and each detector has a delay volume of 1500L. From each height there is one data point every half hour, with the exception of few hours per month when the calibration is performed. The lower limit of detection, defined as the radon concentration corresponding to a relative counting error of 30% for a one hour count, is equal to or less than 30 mBq/(m^3). Calibrations of the radon detectors are performed monthly by injecting known amounts of radon into the inlet air stream from a calibrated radon source (Model 2000A, Pylon Electronics Inc., Canada).

Calibration accuracy of the source is +- 4% at a 1-sigma confidence level. For concentrations higher than 700 mBq/(m^3) the overall uncertainly is predominantly attributed to the accuracy of the source."

19. Page 5600: sections 'Precision' and 'Accuracy' could be merged into one.

A. We think it would not be clear which part refers to precision and which refers to accuracy.

20. Page 5601, Cabauw MPI-BGC difference in CO. Did you do the CO comparison also with your GC-FID measuring CO? If so, how did the offset look like?

A. We did not yet include the FID results in this comparison. At Cabauw, the FID CO data were at times of a poorer quality than the RGA data. For this comparison, we will first have to do a careful calibration and a detailed quality check of the FID data, which has not been completed yet.

21. Page 5602, equation 1: how does it look like when you include the linear term? I don't really see the point to exclude the linear term. When the time series is short and the trend is small, the linear term will be simply insignificant.

A. When the time series is short and the intra- and inter-annual variability is much larger than the possible trend, the risk is to obtain a false linear trend.

We did however test the fitting procedure with and without a linear term. If included, the computed linear trend in indeed insignificant (i.e. the error is much larger than the trend) but it is not zero, and it does affect the results for the other terms of the equation. This effect is different for the different sampling heights. Thus including a linear term in the equation does not bring additional information, but it does increase the overall errors and it may induce false differences between the seasonal cycles of the different sampling heights.

22. Page 5602, line 19 – 21: calculation of the baseline: '... averages of the 5th percentile of the weekly afternoon data...' I don't get it. You take all afternoon data within one week and calculate the 20% percentile? But what do you average? Please clarify. How is afternoon defined? What is your averaging interval? 5th percentile is the 20% percentile? Do you apply running averages? Same for CO.

A. There was no averaging; there was an error in the text. We selected all the afternoon data for each week, after that the 5th percentile was calculated for each week from the selected afternoon data. We corrected the text as follows: "these were computed from the 120 m level data by fitting the same seasonal function to the 5th percentile of the weekly afternoon data for H$_2$, and to the 3rd percentile of the monthly afternoon data for CO".

23. Page 5606: reword 'High altitude pollution signal'. 200m isn't a high altitude.

A. We replaced with "pollution events detected at the highest sampling altitude".

24. Page 5606, line 9: do you have any information on H2 emissions from airplanes? Please add a reference.

A. Actually we are not aware of a study that investigated airplane emissions, and it would be a nice future study. We suspected emissions from airplanes or from other high sources, because the signal came from above and it was transmitted towards the lower sampling heights.

25. Page 5607, line 10: 'We initially suspected a contamination ...' Why? Because stronger gradients were expected? If so, which kind of?

A. In some cases we observed stable H_2 gradients with the highest mole fractions at the highest sampling height (e.g. Fig. 5e). At the same time, the other species had no noticeable gradients, or the gradients were opposite, with the highest mole fractions at the lowest sampling height. In this situation, when the H_2 data were unusual, and knowing that air in the lab has high H_2 concentration due to a device producing H_2, the natural concern was the possibility of an undetected leak in the system.

26. Page 5610, lines 17 – 20: move this paragraph up (e.g. to the very end of Section 2.1?)

A. This paragraph explains shortly the background of the radon tracer method which is only used in this section. We think that moving it to another section would result in a more fragmented and difficult reading.

27. Page 5611, lines 11 – 20, equations 2 and 3: to my mind, this part is not necessary as the radon tracer method isn't new and is already described in various publications (that are already cited).

A. We think this part is necessary for understanding the explanations that follow on the calculation and errors, and the results of this section. Also, unlike some of the cited papers, which make an average correction for the ^{222}Rn decay, we had to compute this correction for each case, as we take time intervals of different length (Eq. 3).

28. Section 4.5 Soil sink: you mention that your flux estimate is rather uncertain because there are no radon flux estimates/data available. As you measure in several heights above ground, did you try to apply the gradient method to determine the H2 fluxes? (see e.g. Fick's first law, gradient transport theory or K-theory) It would be interesting to see if the results of the different approach are similar.

A. We agree that it would be interesting to use a different approach for estimating the H_2 soil uptake fluxes, and to compare with the results of the radon tracer method. Applying the flux gradient method at Cabauw tall tower would however be a work in itself and it is beyond the purpose of this paper. One complication of this approach is that we only observe significant H_2 gradients due to soil uptake in very stable conditions, and in this situation one or both of the two lowest sampling heights at Cabauw (20 and 60 m) are above the atmospheric surface layer (constant flux layer). A possibility for a future study would be to perform H_2 measurements in air sampled via additional inlet lines closer to the ground, below the 20 m level. For comparison with the radon tracer method results we recently started a one-year campaign of flux measurements with a soil chamber.

29. Conclusions: the conclusions are so far just summarizing what was already said before. Can you generalize/upscale the results or look at it from a broader perspective? It should be possible since you highlight the large footprint at Cabauw on page 5596, lines 3 – 6. Are the results supposed to be characteristic for Dutch conditions?

A. We included in the last section a paragraph on the representativeness of our results.

30. Acknowledgements: which CO analyzer from MPI Mainz you refer to? The CO data shown in the paper are measured with the RGA?

A. We refer to the RGA instrument described in this paper – the text had been corrected.

Minor comments

31. Page 5616, line 18: 'Gif-sur-Yvette' is misspelled.

A. The text has been corrected.

32. Page 5621, line 16: change 'the introduction on larger scale' to 'a large scale introduction'

A. The text has been changed.

References

33. Bond et al. (second one) ... was published in 2011

A. The year has been corrected.

34. Fabian et al., use subscripts for all species

A. The text has been corrected.

35. King, remove symbol after 'Mycobateria'

A. The text has been corrected.

36. Liebl et al., check author names

A. We checked and corrected the names.

37. Neubert et al., affiliations became part of author names, right?

A. The text has been corrected.

38. Rockmann et al., check author names (who is T.S.?)

A. The author names are as given in ACP (T. S. Rhee). The initials are written after the names.

39. Szegvary at el., check superscripts (^{222}Rn, for both references)

A. Both references have been corrected.

40. Tromp et al., complete page numbers

A. The page numbers have been corrected.

Figures

41. Fig. 4: why do you only show two years of data?

A. The purpose of Fig. 4 is only to show the features of the seasonal cycles. We show two full years for visual clarity - with only one year the maxima would be more difficult to see. It is not necessary in this case to show the rest of the time series.

42. Fig. 5, caption: replace 'high altitude pollution events' with 'top level pollution events'?

A. We replaced with "pollution events detected at the highest sampling altitude".

43. Fig. 6: why don't you show all 5 events according to Fig. 5. It would make it easier for the reader.

A. Fig. 5b shows a rather complicated pollution event, with highly variable vertical gradients – there is no particular moment when the vertical gradient would be representative for this type of event.

44. Figs. 8 – 10: what do the error bars show?

A. The error bars represent the absolute error of the results.

In Fig. 8, the error is the composite of the ^{222}Rn flux error, the $H_2/^{222}$Rn slope errors and the error resulted from the correction for the concurrent emissions (see also Sect. 4.5).

In Fig. 9 and 10, the error bars show the absolute errors of the computed H_2/CO ratios, i.e. the error of the slope of the linear fit function.