Author's response to the comments of Dr. Anja Werner

April 17, 2012

We thank referee Anja Werner for her insightful comments to this paper.
Dr. Werner pointed out that in her view, the scientific relevance and significance of our findings are insufficiently addressed in the introduction. We will try to remedy this by adding some sentences about the most interesting findings of this paper (i.e. the relation between CH_3 and $\delta D(H_2)$ in the LMS, and the δD decrease in the Indian monsoon) to the introduction.

1 Responses to specific comments

p.590 l.1-2 : "...in the upper troposphere lower stratosphere (UTLS) region around the tropopause (TP)..." Do you want to emphasize anything special with the "around the tropopause" add-on ? If not, delete it!
Will be deleted.

p.591, l.8 "...it will lead to large-scale leakage of H2 into the atmosphere..." Do you really expect large-scale leakages? Isn't this too strong ? Certainly, considering the high volatility and an intensified use of H2, increased H2 emissions into the atmosphere are to be expected - an worth to be discussed. However, due to safety and probably financial reasons there will be a strong interest in minimizing leakages.
We agree that the magnitudes of future H_2 leakages are highly uncertain and depend, among other things, on the measures taken to minimize leakage. Different authors have made very different estimates of the leakage rates that are to be expected. Tromp et al. (2003) expected leakage rates between 10 and 20%, whereas Warwick et al. (2004) assumed leakage rates between 1 and 12% for their calculations. Schultz et al. (2003) claimed that leakage rates between 0.1 and 20% are possible, and used a value of 3% as a reasonable upper limit for their simulations. With this 3% loss, the tropospheric burden increased by 30% in their model. We feel that this is a large disturbance of the present-day H_2 cycle that justifies the use of the adjective "large-scale". However, this sentence in the paper does perhaps not reflect the uncertainties in the estimates well enough. We will therefore rephrase it to: "It is expected that when this technology comes into wide use, it will lead to leakage of H_2 into the atmosphere, which may lead to considerably increased H_2 mixing ratios. This may affect ...".

Table 1: Please explain the subscript and superscript numbers. Are these marking the errorbars ? Why are there no errors on the production terms ? All in all, the averge tropospheric mixing ratio of H2 you cite from the literature is 530 ppb - however, does this mixing ratio apply also to the upper troposphere, where the soil sink is neglectable? Your data imply much higher average H2 mixing ratios!

This table shows the budget estimate from Pieterse et al. (2011) as an example of recent global budget estimates. Pieterse et al. (2011) based the estimates of the magnitudes and isotopic signatures of the different surface sources on previously published studies. The super- and subscripts indicate the upper and lower bounds of these estimates. The values for photochemical production, photochemical removal and uptake by soils were calculated from their model (TM5) output. In this first H_2 study with TM5, no extra runs were made to provide an error estimate for these numbers. We will provide a more informative caption for the final revised paper.

The average mixing ratio in the tropospheric CARIBIC samples is 550.5 ± 0.9 ppb (after the iterative outlier removal procedure, error indicates one standard error). The quoted value of 530 ppb is based on the findings by Novelli et al. (1999). While quoting this number, we forgot to translate this number to the new mixing ratio scale introduced by Jordan and Steinberg (2011). On the new scale, this number would be closer to 547 ppb, which largely resolves the discrepancy.

It is possible that mixing ratios in the upper troposphere are slightly larger than at ground level. Rhee et al. (2006) found an average mixing ratio of around 560 ppb for their tropospheric CARIBIC samples (calculated to the new H_2 scale). A reduced influence of the soil sink at this altitude seems a plausible explanation. Model results indicate that a vertical gradient may exist, especially in the Northern Hemisphere extratropics (Hauglustaine and Ehhalt, 2002; Price et al., 2007; Pieterse et al., 2011). In the final paper, we will quote the Novelli value on the new H_2 scale, and add a note that this is a value based on ground stations that may not be fully representative for the upper troposphere.

Section 1.2. What effects or consequences has this dD enrichment of the troposphere by STE? Just to study STE, there is a whole set of well studied more stable, long-lived trace gases which are easier to measure. Please, point out why it is worth to study the isotope effects, as well.

The reason for incorporating the isotopic composition into models of the H_2 cycle is that this may provide an extra constraint on the different flux terms in the H_2 budget. For this constraint to have real added value, it is necessary that the isotope effects of the different processes are well known. STE is one of the processes that affect the tropospheric isotope budget, and according to modeling studies by Price et al. (2007) and Pieterse et al. (2011), its effect is quite large. Furthermore, Pieterse et al. (2011) showed that the modeled magnitude of its effect depends quite sensitively on its parameterization. Hence, a good understanding of the δD effects of STE is necessary to model the isotope budget of H_2 well, which in turn may help to constrain the global H_2 budget.

Besides the importance of STE for the tropospheric H_2 cycle, the stratospheric cycle of

H_2 is closely linked to that of methane and water and therefore important. Since the mixing ratios of H_2 vary little in the stratosphere, they provide little information about the cycling of H_2 that is going on. The isotope effects in H_2, on the other hand, are quite strong and as a result δD contains more information.
We will modify this paragraph to emphasize these considerations more.

Though Sections 1.2 and 1.3 give a nice introduction into dD issues, the information seems to be disordered. Just a suggestion: Re-order Sections 1.2 and 1.3; one section regarding basic dD issues and information in the troposphere as well as in the stratosphere (e.g. p.592 l.2 - 11 + parts from 1.3) followed by an overview on yet performed studies and the Caribic project (e.g. p.592 l.14 -p.593, l.26). Maybe, you could even add the background values of dD (tropospheric and stratospheric) in Table 1 ?
We have considered this, but still prefer to keep the tropospheric and stratospheric issues in different sections. Adding background values to Table 1 is certainly a good idea.

p.595 Description of the sampling unit: Why are there two units - are parallel samples taken or are the units used on after the other ? Why does the sampling time (duration) change? Is it to adapt to the different pressures at different altitudes?
During the first part of a flight, from Frankfurt to the destination airport, the canisters in both units are filled with sample air. During the return flight, the canisters in one of the units are vented and filled again with sample air. The aim of this procedure is to ensure that all canisters are filled when the aircraft arrives in Frankfurt again, even if the flight schedule is changed or the equipment malfunctions. If no such problems occur, one of the units will contain 14 samples from the first flight leg, and the other will contain 14 samples from the return flight (Schuck et al., 2009).
The sample collection time does indeed depend on ambient pressure and therefore on flight altitude. This is not an active adaptation, but a result of the requirement that all samples should have a certain minimum final pressure; it takes longer to reach this pressure at lower outside pressures.
We will critically re-evaluate this section and adapt it where the paper itself and the papers it refers to (Brenninkmeijer et al., 2007; Schuck et al., 2009) are insufficiently detailed.

p.595,l.13-14: "Due to the regularly spaced sampling, the distribution of samples is likely representative of the different air masses encountered by the aircraft." I would say this is strongly dependent on the actual distribution (horizontal and vertical) of air masses during the flight. If there is strong filamentation or fine scale structures are observed, the sampling is not necessarily representative even though or maybe especially since you are sampling in a regular interval. However, the number of flights and samples, regions captured by the flight tracks (...) should yield relatively robust statistics.
The "regularly spaced" phrase was written with a comparison to other possible (event-driven) sampling schemes in mind. We agree that in this context, it is not the best choice of words. The first part of the sentence will be replaced with "With this method

of regular, automated, non-event-driven sampling, ...".
The air samples collected by CARIBIC are, in general, very representative for background air. For example, Schuck et al. (2009) showed that this was the case for the CO_2 data from CARIBIC samples.

p.595, l. 22 averaging over 15 trajectories. Are fifteen trajectories calculated for each sampling location? Or the fifteen closest trajectories ? Please, explain.
Fifteen 5-day trajectories are calculated for each sample. All 15 trajectories start at the same time, but at slightly different pressures and/or horizontal positions. There are three different starting pressures (the sampling pressure, and the sampling pressure offset by + or - 3%) and five different horizontal positions (the sampling position, and the sampling position offset by + or - 0.3° in latitude or longitude).
One mean trajectory is calculated by averaging over the 15 trajectories at each timestep (one hour). The trajectory parameters for the sample are then calculated from that mean trajectory.

p.597, l.20 "This estimate can be made for any sample with an m(O3) higher than at the local thermal TP". If I understand the procedure in Sprung and Zahn (2010) correctly, the O3 mixing ratio has to be higher than the average O3 mixing ratio at local thermal TP of the corresponding adjacent balloon sounding sites !? Which sounding sites have been used to derive the vertical distance to the TP, especially, for the flights over Asia? Is the procedure to derive the vertical distance above the tropopause from O3 still applicable for flights in the subtropics (lat < 35) ? Could there be any problems with the analysis if strong horizontal gradients (streamers etc.)occur e.g. in spring to early summer !?
It is correct that the O_3 mixing ratio has to be higher than both the average O_3 mixing ratios at the thermal TP of the two adjacent sounding sites to make the estimate of the height above the TP (and in this case, to label the sample as stratospheric). The two adjacent stations were always selected by latitude, not longitude.
The 12 sounding stations used for the estimates are listed in Table 1. Two stations, Sapporo and Tateno, are asian. It can be seen in Figure 2 of (Sprung and Zahn, 2010) that there are no obvious longitudinal differences in the monthly mean O_3 mixing ratios discernible over the longitude range from 114° W to 140° E. Sprung and Zahn (2010) even attribute a large part of the scatter between stations to the limited number of available soundings. This makes this procedure a robust method for quantifying a "chemical" vertical distance around the tropopause.
Since there are no stations south of 35° N in this set, the procedure may indeed be less suited to make estimates for samples taken here. We note that this potentially affects only five samples in the presented dataset. We agree that the accuracy of the procedure could be improved by including more ozone sounding locations and soundings.
Importantly, this O_3-derived altitude above the thermal TP is not a measure of the instantaneous distance above the TP, but an indication of the degree of mixing between tropospheric and stratospheric air. As such, it is independent of and not influenced by streamers, tropopause folds, etcetera (in such cases it is, however, especially important

not to confuse it with instantaneous altitude). In the final version, we will emphasize the difference with instantaneous altitude more.

Table 1: Ozone sounding stations used for calculating the O_3-based distance above the TP

Station	Latitude	Longitude
Alert	82° N	62° W
Ny Alesund	79° N	12° E
Resolute	75° N	95° W
Sodankyla	67° N	27° E
Churchill	59° N	147° W
Edmonton	53° N	114° W
Goose Bay	53° N	60° W
Hohenpeissenberg	48° N	11° E
Sapporo	43° N	141° E
Boulder	40° N	105° W
Wallops Island	38° N	76° W
Tateno	36° N	140° E

p.598,l.4-5 "This was repeated until the averages did not change anymore..." Just to understand the procedure correctly: What do you exactly mean? You did not find any more data points outside the +-3s range, thus, the average was constant? Or, averages n and n+1 did not show a statistically significant difference!?
It was repeated until no more data points were found outside the $\pm 3\sigma$ range and the average was constant. We will rephrase this.

p.598, l.23: Why is it representative?
Perhaps "representative" is not the correct word here. We will replace the first part of this sentence with "As an illustration of the features that are typically observed in the CARIBIC dataset, ...".

p.599,l.5ff "The difference between the m(H2) averages is less than 10 ppb, which is small compared to the 17 ppb standard deviation on the averages." What are the actual average tropospheric and stratospheric mixing ratios of H2 on this route ? Elevated H2 mixing ratios in tropospheric samples - many in the southern part of the flight - could those be influenced by southern hemisphere or tropical air masses (Fig.2) ? Did you calculate the tropospheric average using the complete latitude range ? Some of the stratospheric samples (at 40° N) seem to exhibit elevated H2 mixing ratios, though (Fig.2) - possible reasons!?
Calculated over the whole dataset without the iteratively selected outliers, the average $m(H_2)$ value and standard deviation are 550.5 ± 16.6 ppb (311 samples, standard error

0.9) for the troposphere and 559.8 ± 16.1 ppb (120 samples, standard error 1.5) for the stratosphere. If I select from these data only the data from the Caracas route, these values are 550.1 ± 15.0 ppb (133 samples, standard error 1.3) for the troposphere and 558.2 ± 14.3 ppb (36 samples, standard error 2.4) for the stratosphere. These averages are all calculated over the complete latitude range. The mixing ratio difference between the tropospheric and stratospheric samples is statistically significant when tested with a T-test or Kruskal-Wallis test. Pieterse et al. (2012) observed events at Mace Head where H_2 mixing ratios were elevated, which they attributed to stratospheric subsidence. So, it seems that H_2 mixing ratios are slightly elevated in the lower stratosphere with respect to the troposphere, but the effect is small compared to the isotope effect. A vertical gradient in the mixing ratios above the tropopause was not observed.
The samples with elevated mixing ratios in the southern part of the flight could be influenced by tropical airmasses, as their sampling locations are in or almost in the tropics. Southern Hemisphere air masses are a possibility, but at least for the two samples in the summer 2009 flights with the largest mixing ratios, the calculated trajectories do not indicate a very southern air mass origin.

Section 3.1. Nice discussion of possible pollution effects! However, you write that a possible contamination by water vapour during the take off and landing phase is unlikely - why ? The fact, that intermediate canisters do also show pollution signatures could be due to other effects. As you very nicely point out, there are several possible causes. It is not necessarily only one single cause that is responsible for every contaminated flask. Fig. 3: What if one ore more of the data points with dD < 0 and 1/H2 < 1 are ommited ? What source signature do you yield in that case?
We consider a pollution with H_2 formed from water vapor on a metal surface unlikely, because the actual metal surface area that the sample air is exposed to is quite small. Another consideration is that in the past, in-cylinder formation of H_2 happened in specific individual cylinders, and not in others of the same type (see (Gerst and Quay, 2000)). Our pollution effect, however, did not take place in specific cans of the sampling units. We agree that a combination of causes is possible.
Omission of the three datapoints with an inverse mixing ratio below one in the Keeling plot increases the source signature from -420 ± 31 ‰ to -364 ± 60 ‰, and omission of the datapoints with a negative δD increases the signature to -316 ± 59 ‰. Hence, omitting the most "extreme" samples leads to a slightly higher source signature, but it is still on the low end of currently used estimates for fossil fuel combustion.

Will there be a further examination of dD signatures from aircraft exhaust?
We are currently not planning such an investigation, although we think it could be interesting.

p.601 l.25 "The O3-derived altitude can be considered as a measure of the degree of mixing between low-O3 air from the troposphere and high-O3 air from the stratosphere. When it is used to plot m(H2) and dD against mean trajectory latitude and height (Fig.

4a,b)..." Sounds like you actually plot O3 vs the mean trajectory height. Maybe reformulate. Further, are you referring to the average of 15 trajectories as mentioned in Section 2.1 or as described in Fig 4 a 5-day trajectory? Do you average the complete trajectory - sorry, guess, I did not understand your procedure! (See also under Technical Corrections referring to Fig. 4)
We will remove the potential ambiguity by replacing the second sentence with: When $m(H_2)$ and δD are plotted against mean trajectory latitude and this O_3-derived altitude above the TP ..."
The mean trajectory latitude mentioned here is the average over 15 5-day trajectories as mentioned in section 2.1 (see above). We will try to make these points more clear.

p.602, l.5 - 19 The correlations are compact, since - as you mention - lifetimes of the considered species are long compared to the transport times, but what does this have to do with the age of air ? Especially, since you link non-conserved trace gases N2O and CH4 (neither in the troposphere nor in the stratosphere) as an "indicator of the average age of air". What do you want to tell ?
We will use " average degree of stratospheric processing" instead of "average stratospheric age" in the revised version.

What would be an advantage using dD instead of other long-lived tracers?
We do not pretend that δD may be a better tracer of stratospheric transport than other more widely used, tracers. But it is an additional one that contains information about the oxidation reactions of H-containing species (H_2, CH_4, H_2O,HCHO, ...) that took place in the air mass; more information than, for example, H_2 mixing ratio.
The importance of studying δD effects of STE has been outlined above.

p.602, l.14-15: lower stratosphere or lowermost stratosphere ?
The statement is most likely true for the whole lower stratosphere. From the tight correlations found in the CARIBIC data, we can conclude that the lowermost stratosphere is no exception. We prefer to keep the term "lower stratosphere" in this sentence, as we also show results from literature collected above the lowermost stratosphere

p.603, l. 6 25ppb is still significantly larger than the troposheric increase of 15 ppb observed by Duglogencky et al.(2009). Tropospheric CH4 has a latitudinal gradient. How do the CH4 mixing ratios compare when you look at a similar latitude range (like 43 N, where the balloon samples from Röckmann et al. (2003) were taken)?
A fit to the data between 39 and 49° N yields almost the same difference ($\approx$ 1.6 ppb smaller) between the Röckmann et al. and the CARIBIC data, so latitudinal variation does not seem to be the cause of the discrepancy.
A possible but speculative explanation for the discrepancy is that Röckmann et al. (2003) sampled "older" stratospheric air than CARIBIC, as they sampled higher in the stratosphere. Their air masses may have been introduced into the stratosphere before 2000,

when tropospheric methane levels were even lower, whereas CARIBIC probably samples more recently imported stratospheric air masses.

p.605, l. 10 560ppb; is this the average mixing ratio of the non-monsoon data ?
The 560 ppb is an estimate by eye of the baseline of the $m(H_2)$ data from the flights to Chennai. We will replace it with the average mixing ratio of the non-monsoon data in the final version.

2 Responses to some of the technical corrections

p.596, l.16 ff use "∼"
We switched to the use of "≈" instead of "∼" on recommendation of a previous paper's editor. In his words: "The "approximately equal" sign is "≈", i.e. two wavy lines, not the tilde sign. Use of the tilde sign (∼) to represent approximations is due to limitations of early typewriters and should be abandoned."
According to Wikipedia, "≈" means "approximately equal", which is the meaning we intend. "∼" can mean "approximately", "similar to" or "of the same order of magnitude as" and is therefore a more ambiguous symbol. For these reasons, we prefer ≈.

p.618 Fig 4, Caption: delete several ":" The unit of "height above TP" is missing. How is the colour of squares derived - by averaging the samples within each square? Please, describe!
The color of the squares is indeed derived by averaging the samples within each square. We will add this to the caption.

References

Brenninkmeijer, C. A. M., Crutzen, P., Boumard, F., Dauer, T., Dix, B., Ebinghaus, R., Filippi, D., Fischer, H., Franke, H., Frieß, U., Heintzenberg, J., Helleis, F., Hermann, M., Kock, H. H., Koeppel, C., Lelieveld, J., Leuenberger, M., Martinsson, B. G., Miemczyk, S., Moret, H. P., Nguyen, H. N., Nyfeler, P., Oram, D., O'Sullivan, D., Penkett, S., Platt, U., Pupek, M., Ramonet, M., Randa, B., Reichelt, M., Rhee, T. S., Rohwer, J., Rosenfeld, K., Scharffe, D., Schlager, H., Schumann, U., Slemr, F., Sprung, D., Stock, P., Thaler, R., Valentino, F., van Velthoven, P., Waibel, A., Wandel, A., Waschitschek, K., Wiedensohler, A., Xueref-Remy, I., Zahn, A., Zech, U., and Ziereis, H.: Civil Aircraft for the regular investigation of the atmosphere based on an instrument container: The new CARIBIC system, Atmos. Chem. Phys., 7, 4953–4976, doi:10.5194/acp-7-4953-2007, 2007.

Gerst, S. and Quay, P.: The deuterium content of atmospheric molecular hydrogen: Method and initial measurements, Journal of geophysical research, 105, D21 26 433–26 445, doi:10.1029/2000JD900387, 2000.

Hauglustaine, D. A. and Ehhalt, D. H.: A three-imensional model of molecular hydrogen in the troposphere, J. Geophys. Res., 107, D17 4330, doi:10.1029/2001JD001156, 2002.

Jordan, A. and Steinberg, B.: Calibration of atmospheric hydrogen measurements, Atmos. Meas. Tech., 4(3), 509–521, doi:10.5194/amt-4-509-2011, 2011.

Novelli, P. C., Lang, P. M., Masarie, K. A., Hurst, D. F., Myers, R., and W.Elkins, J.: Molecular Hydrogen in the troposphere: Global distribution and budget, J. Geophys. Res., 104, 30 427–30 444, doi:10.1029/1999JD900788, 1999.

Pieterse, G., Krol, M. C., Batenburg, A. M., Steele, L. P., Kummel, P. B., Langenfelds, R. L., and Röckmann, T.: Global modelling of H_2 mixing ratios and isotopic compositions with the TM5 model, Atmos. Chem. Phys., 11, 7001–7026, doi:10.5194/acp-11-7001-2011, 2011.

Pieterse, G., Krol, M. C., Batenburg, A. M., Brenninkmeijer, C. A. M., Popa, M. E., O'Doherty, S., Grant, A., Steele, L. P., Krummel, P. B., Langenfelds, R. L., Wang, H. J., Vermeulen, A. T., Schmidt, M., Jordan, A., Engel, A., Fisher, R. E., Lowry, D., Nisbet, E. G., Reimann, S., Vollmer, M. K., Steinbacher, M., Levin, I., Hammer, S., Sturges, W. T., and Röckmann, T.: Modelling the global and regional scale variability in atmospheric H_2 with the two-way nested TM5 model, in preparation, 2012.

Price, H., Jaeglé, L., Rice, A., Quay, P., and Novelli, P. C.: Global budget of molecular hydrogen and its deuterium content: Constraints from ground station, cruise and aircraft observations, J. Geophys. Res., 112, D22 108, doi:10.1029/2006JD008152, 2007.

Rhee, T. S., Brenninkmeijer, C. A. M., and Röckmann, T.: The overwhelming role of soils in the global atmospheric hydrogen cycle, Atmos. Chem. Phys., 6, 1611–1625, doi:10.5194/acp-6-1611-2006, 2006.

Röckmann, T., Rhee, T. S., and Engel, A.: Heavy hydrogen in the stratosphere, Atmos. Chem. Phys., 3, 2015–2023, doi:10.5194/acpd-3-3745-2003, 2003.

Schuck, T. J., Brenninkmeijer, C. A. M., Slemr, F., Xueref-Remy, I., and Zahn, A.: Greenhouse gas analysis of air samples collected onboard the CARIBIC passenger aircraft, Atmos. Meas. Tech., 2, 449–464, doi:10.5194/amt-2-449-2009, 2009.

Schultz, M. G., Diehl, T., Brasseur, G. P., and Zittel, W.: Air pollution and climate-forcing impacts of a global hydrogen economy, Science, 302(5645), 624–627, doi: 10.1126/science.1085169, 2003.

Sprung, D. and Zahn, A.: Acetone in the upper troposphere/lowermost stratosphere measured by the CARIBIC passenger aircraft: Distribution, seasonal cycle, and variability, J. Geophys. Res., 115, D16 301, doi:10.1029/2009JD012099, 2010.

Tromp, T. K., Shia, R.-L., Allen, M., Eiler, J. M., and Yung, Y. L.: Potential environmental impact of a hydrogen economy on the stratosphere, Science, 300, 1740–1742, doi:10.1126/science.1085169, 2003.

Warwick, N. J., Bekki, S., Nisbet, E. G., and Pyle, J. A.: Impact of a hydrogen economy on the stratosphere and troposphere studied in a 2-D model, Geophys. Res. Lett, 31(5), 2–5, doi:10.1029/2003GL019224, 2004.