Interactive comment on "A new estimation of the recent tropospheric molecular hydrogen budget using atmospheric observations and variational inversion" by C. Yver et al.

Anonymous Referee #2

General Comments:

This paper addresses an interesting question on the global budget of H_2, especially for its soil uptake which has the greatest magnitude and uncertainty. This reviewer is not familiar with the variational inversion system, but inverting so many processes (soil uptake, emissions from fossil fuel combustion, biomass burning, and N_2-fixation, and photochemical production of HCHO) in each model grid cell seems challenging, considering the sparseness of the observations over regions outside Europe. The impact of the spatial coverage and resolution of the observations on the variational inversion results at model grid cell resolution needs to be discussed. Only the posterior flux maps for Europe are shown because of the relatively sufficient constraints for this region. How about the global posterior flux maps? How did the spatial distributions of the fluxes change between posterior and prior globally?

The paper is well structured, the observations, method, and input data are well introduced, and the inversion results and main conclusions are clearly presented. The paper would benefit if the authors would explain why the inverted fluxes behave as they do (e.g., the causes of the seasonality of the soil uptake, see specific comments). The paper also needs some reworking to improve graphical presentation, and clarity in text (see specific comments) before publication in ACP.

Specific Comments:

p. 28966, l. 5-11: Although a list of references are mentioned beforehand, it would be better to know where these budget magnitudes for each process come from.

p. 28967, l. 15-16: Sometime you use numbers (e.g., 10 sites), and sometime you use words (e.g., nine additional sites).

p. 28967, l. 22: You mention here that the soil sink ranges from 40 to 90 Tg yr^{-1}, but earlier on p. 28966, l. 15 you also give a 65 to 105 Tg yr^{-1} range. Please clarify.

p. 28968, l. 7: Are there 25 flasks sampling sites or 24, according to the introduction on p. 28967, l. 11-16 (5+10+9)?

p. 28968, l. 11: Why not starting from January 2005 (see l. 8 on the same page)?

p. 28968, l. 25: Table 2 should be Table 1.

p. 28968, l. 26: a -> an?

P. 28969: For both RAMCES and EUROHYDROS networks, did you treat sites sampling clean or polluted air masses differently? How?

p. 28971, l. 14-20: How about the seasonal cycle in the Southern Tropics?

p. 28971, l. 21-22: peak-to-trough (not through)?

p. 28971, l. 28: Please explain (or cite from literature) why the soil uptake is strongest in autumn in the HNH.

p. 28971: What is the time resolution of the observations used in the variational inversion?

p. 28972, l. 3: Does each data point represent each site? Is it 3-year average? How to calculate the error bars?

p. 28972, l. 9: The gradient should have a unit of ppb/degree, change it to difference or change the unit.

p. 28973, l. 13-16: It would be helpful to state explicitly that this work will also use the indirect estimation method as the references. One realizes this until p. 28974, l. 10.

p. 28973, l. 17: What does INCA stand for? Should probably add reference here.

p. 28973, l. 22: Here the deposition velocities for H_2 are given by LMDz-INCA, are they the same or different from the soil deposition velocity measurements on p. 28969, l. 27?

p. 28973, l. 24-25: You mention SACS can be used to estimate the sources and sinks of CH_4, CO, HCHO and H2. How (with the use of PYVAR?)? Any reference?

p. 28974, l. 9: What measurements are used to constrain the HCHO flux? Is HCHO production in each cell in the 3-D space inverted? If you only use the surface measurements of H_2, are they sufficient to constrain so many variables in the state vector including soil uptake, emissions related to biomass burning, fossil fuel combustion, and N_2 fixation, and HCHO production in each grid cell? This needs to be discussed.

p. 28974, l. 9: Why eight-day frequency of the state vector? What is the time frequency of the observations used in the inversions?

p. 28974, l. 25: The diagonal elements of R should be the squares of the standard deviations (variances) of the measurements.

p. 28974, l. 27: One just realizes you also use MCF measurements to constrain OH until this sentence. Probably should state that more explicitly in text on p. 28973, l. 12-16. More information about the MCF measurements should be provided. How are the uncertainty values of H_2 and MCF measurements derived?

p. 28975, l. 7: Is soil sink or dry deposition velocity of H_2 used from Hauglustaine and Ehhalt (2002)?

p. 28975, l. 21-22: Is soil uptake or v_{dep} in the state vector? Since the soil uptake rate also depends on the mixing ratio of H_2, is the H_2 field also optimized in the framework and how?

p. 28975, l. 21-27: Resolving the soil uptake, fossil fuel emissions, biomass burning emissions within each model grid seems difficult, given the sparseness of the observations. As you mentioned, different processes cannot be completely independently inverted. Are the inversion results robust regarding different settings of scenarios and prior conditions?

p. 28975, l. 25-27: Here you mention using the bottom-up CO emission inventory from Lamarque et al. (2008), but earlier on the same page, l. 3-4, you mentioned CO emission distribution from other references. If they represent different scenarios, please state it explicitly.

p. 28976, l. 14: how is the value 1.28 derived?

p. 28976, l. 16-17: Again, is the soil sink or deposition velocity inverted?

p. 28977, l. 7-10: Please comment on the reasons for this large scale seasonality.

p. 28977, last paragraph: Please comment on the hotspots in S0.

p. 28980, l. 1-2: It seems that the difference in photochemical production between S0 and other scenarios is greater than 5 Tg / yr.

p. 28980, l. 2: others -> other

p. 28980, l. 8: For the statement "… for S3 and S4, the maximum is shifted in August", one cannot see the S4 line in Fig. 6, and can see the S5 line instead. Do S4 and S5 overlap? It would be helpful to mention S5 explicitly, as later on you conclude that S5 is the most pertinent scenario.

p. 28980, l. 8: Please comment or cite references on why soil uptake is a maximum in August/September in the HNH.

p. 28980, l. 18: To make it clear, the 30% increase should be for the a priori estimate of the soil sink in S1 and S2. Again, do you modify the deposition velocity or the soil sink? Is the deposition velocity or the soil sink inverted as in the state vector?

p. 28980, l. 22-23: Please comment on other causes.

p. 28981, l. 6: expect for S0 (and S1)?

p. 28983, l. 26-27: Please comment on the reason for the smaller spring soil uptake than the autumn uptake.

p. 28984, last paragraph: It would be helpful to find a way to add those percentage contribution values to Table 5.

p. 28985, l. 6: Should be "soil uptake of ~ 59 Tg yr^{-1} and emissions of ~ 36 Tg yr^{-1}", to be consistent with p. 28982, l. 12-13 and Table 4.

Fig. 2: There are so many sites. Would be helpful to show the figures in a way such that it is easier to see in which semi-hemispheres the sites are located.

Fig. 6: Why are there two light green dotted lines in the Emission panel? Should the one with the larger magnitude be light green solid line representing the total prior emissions?