

Response to reviewer 2

All of the reviewers provided excellent suggestions and feedback on the paper, and we think that by addressing their concerns the paper will be greatly improved. Many of their comments were complementary. Therefore we will first summarize the major revisions we plan to make to the paper before responding to each reviewer in detail:

Planned Major Revisions

1) We have revised the $\delta^2\text{H-CH}_4$ dataset in response to comments from Reviewer 1 and Reviewer 3. (i) For peatland sites with depth stratified sampling we have decided to only include samples from the upper 50 cm, as suggested by reviewer 3, since this is the depth range that is most likely to emit CH_4 to the atmosphere. This affects a total of 8 sites. (ii) Reviewer 1 noted that an outlier sample from the Amazon River with very high $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$ could be derived from thermogenic methane. We agree that this outlier is suspect, and therefore have decided not to include it. (iii) We also noted that one site (Mirror Lake, Florida, USA) was analyzed in two separate studies, and therefore was included twice in the dataset. We have combined the data from the two studies into one site entry.

2) As suggested by all three reviewers, we have performed much more rigorous analysis of the relationship between measured and modeled $\delta^2\text{H-H}_2\text{O}$ values. Specifically we have done the following: (i) In addition to annual precipitation $\delta^2\text{H}$ values, we now also analyze growing season precipitation $\delta^2\text{H}$, which is defined as the amount-weighted mean $\delta^2\text{H}$ of months with mean temperature greater than 0°C . This provides an opportunity to assess whether seasonal variation in precipitation in the mid to high-latitudes is important in controlling the environmental $\delta^2\text{H-H}_2\text{O}$ value; (ii) separately analyzing inland water and wetland environments, since these are different hydrological environments and the controls on $\delta^2\text{H-H}_2\text{O}$ are potentially different.

This analysis led to the following key results (see Table R1 below for summary of results): A) growing season modeled precipitation $\delta^2\text{H}$ is a better predictor of inland water $\delta^2\text{H-H}_2\text{O}$ than annual precipitation $\delta^2\text{H}$, in that the regression curve is indistinguishable from the 1:1 line. B) annual modeled precipitation $\delta^2\text{H}$ is a better predictor of wetland $\delta^2\text{H-H}_2\text{O}$, in that the slope of the regression is indistinguishable from 1, and the R^2 value is higher. However, the regression line is offset from the 1:1 line by $18.6\pm 9\%$. We interpret this as an indicator of likely widespread evaporative effects on $\delta^2\text{H-H}_2\text{O}$ in wetland environments.

We use these results to then develop a 'best estimate' for comparing $\delta^2\text{H-H}_2\text{O}$ with $\delta^2\text{H-CH}_4$. (i) For sites with measured $\delta^2\text{H-H}_2\text{O}$ values we use the measured value. (ii) For inland water sites without measured $\delta^2\text{H-H}_2\text{O}$ we use modeled growing season precipitation, since as discussed above the regression of this against measured $\delta^2\text{H-H}_2\text{O}$ is indistinguishable from the 1:1 line. (iii) For wetland sites without measured $\delta^2\text{H-H}_2\text{O}$ we estimate the $\delta^2\text{H-H}_2\text{O}$ using the regression relationship with annual precipitation $\delta^2\text{H-H}_2\text{O}$ shown in Table R2. We feel this approach combining measured and modeled data is

most consistent with that of Waldron et al., 1999, who we note also analyzed a combination of sites with measured $\delta^2\text{H-H}_2\text{O}$ (29 out of 51 sites) and estimated $\delta^2\text{H-H}_2\text{O}$ based on precipitation isotopic measurements (22 out of 51 sites).

Table R1: Comparison of regression relationships between modeled $\delta^2\text{H}_p$ and measured $\delta^2\text{H-H}_2\text{O}$

	Slope	Intercept	R ²	RMSE	<i>p</i>	<i>n</i>
Inland waters						
<i>Growing season $\delta^2\text{H}_p$</i>	1.05±0.09	-0.3±8	0.82	22.3	4.81E-13	33
<i>Annual $\delta^2\text{H}_p$</i>	0.85±0.06	-2.1±7	0.84	20.5	3.17E-14	33
Wetlands						
<i>Growing season $\delta^2\text{H}_p$</i>	1.24±0.09	14.8±10	0.87	16.5	4.46E-13	28
<i>Annual $\delta^2\text{H}_p$</i>	1.057±0.08	18.6±9	0.88	15.7	1.20E-13	28

3). As suggested by all three reviewers, it is important to consider the effects of modeled $\delta^2\text{H-H}_2\text{O}$ on the regression between $\delta^2\text{H-H}_2\text{O}$ and $\delta^2\text{H-CH}_4$. To do this carefully we performed the regression analysis using four different estimates of $\delta^2\text{H-H}_2\text{O}$: (i) the ‘best-estimate’ of $\delta^2\text{H-H}_2\text{O}$ as described above in Planned Major Revision 2; (ii) measured $\delta^2\text{H-H}_2\text{O}$, only analyzing sites with this measurement; (iii) modeled annual precipitation $\delta^2\text{H}$; and (iv) modeled growing season precipitation $\delta^2\text{H}$. We think it is valuable to continue to include the regression relationships for modeled precipitation because these relationships could be used in future studies using Earth Systems Models to predict the distribution of $\delta^2\text{H-CH}_4$. For each of these cases we analyzed all sites, inland waters, and wetlands. We also compare these relationships with those of Waldron et al., (1999), both for the total dataset in that study, and for the dataset that only includes sites with measurements of $\delta^2\text{H-H}_2\text{O}$ (29 out of 51 sites). A summary of the results of this analysis are shown in Table R2 below.

A key point is that we have decided to use unweighted, as opposed to weighted, regression. Comments by Reviewer 1 made us realize that weighting by standard error was causing a few sites to strongly bias the regression results. Statistical research has found that for environmental data with poorly constrained error variance unweighted regression is frequently less biased than weighted regression (Fletcher and Dixon, 2012). Using a statistical test proposed by that study we find that unweighted regression is a good choice for our dataset. Note that in Table R2 we apply unweighted regression to the dataset of Waldron et al., (1999), in part because the specific weighting methodology was not specified in that study. This produces a small difference in the regression relationship shown in Table R2 with that reported by Waldron et al., (1999), but the two regression relationships are within error.

We then used analysis of covariance (ANCOVA) to examine differences between the regression relationships shown in the table. Based on a multiple comparison test, none of

the regression relationships shown in Table R2 are significantly different from one another. Therefore we conclude that (i) using modeled $\delta^2\text{H-H}_2\text{O}$ does not have a significant effect on the estimate of the relationship between $\delta^2\text{H-H}_2\text{O}$ vs $\delta^2\text{H-CH}_4$; (ii) Differences in the slope of this relationship between inland waters and wetland sites are not conclusive; and (iii) that since all of the regression relationships using the larger dataset produce a flatter slope than that of Waldron et al., (1999), the true global slope is likely to be flatter than inferred in that study, but confirmation of this flatter global slope will require more data and further analysis.

Table R2: Comparison of regression relationships between $\delta^2\text{H-H}_2\text{O}$ and $\delta^2\text{H-CH}_4$ using different estimates of $\delta^2\text{H-H}_2\text{O}$

	Slope	Intercept	R ²	RMSE	<i>p</i>	<i>n</i>
Best Estimate						
$\delta^2\text{H-H}_2\text{O}$						
<i>All</i>	0.44±0.05	-298±5	0.42	27.4	1.33E-16	129
<i>Wetlands</i>	0.51±0.06	-300±5	0.59	23.7	7.85E-12	55
<i>Inland Waters</i>	0.42±0.07	-295±7	0.34	29.1	6.72E-08	74
Measured						
$\delta^2\text{H-H}_2\text{O}$						
<i>All</i>	0.5±0.08	-292±8	0.43	28.6	1.22E-08	61
<i>Wetlands</i>	0.53±0.11	-298±13	0.44	26.2	6.58E-05	28
<i>Inland Waters</i>	0.42±0.1	-291±10	0.37	28.9	0.000156	33
Modeled						
Annual						
$\delta^2\text{H}_p$						
<i>All</i>	0.42±0.04	-293±5	0.44	26.9	1.17E-17	129
<i>Wetlands</i>	0.57±0.06	-287±6	0.65	21.9	1.01E-13	55
<i>Inland Waters</i>	0.37±0.06	-293±7	0.36	28.4	1.25E-08	74
Modeled						
Growing						
Season $\delta^2\text{H}_p$						
<i>All</i>	0.51±0.05	-292±5	0.41	27.6	2.55E-16	129
<i>Wetlands</i>	0.71±0.07	-285±6	0.63	22.4	4.05E-13	55
<i>Inland Waters</i>	0.44±0.07	-294±8	0.33	29.3	1.03E-07	74
Waldron et al.						
(1999)						
<i>All data</i>	0.74±0.1	-284±6	0.5	26.3	6.56E-09	51
<i>Measured $\delta^2\text{H-H}_2\text{O}$</i>						
<i>only</i>	0.79±0.2	-279±10	0.44	29.6	9.21E-05	29

4) We then used the ‘best-estimate’ $\delta^2\text{H-H}_2\text{O}$ values and the regression based on those values, shown in Table R2, to calculate a revised $\delta^2\text{H-CH}_{4,w0}$ value for each site. These

analyses were then applied in the subsequent analyses in the paper shown in Figures 5,8 and 9. We also calculated an alternate value for sites with measured $\delta^2\text{H-H}_2\text{O}$, using the values and regression curve for those sites.

Notably, for the comparison between $\delta^2\text{H-CH}_{4,w0}$ and α_C we have found that there continues to be evidence for a segmented linear relationship. However, the breakpoint of this relationship is not consistent when analyzing all sites or only sites with measured $\delta^2\text{H-H}_2\text{O}$. Furthermore, the regression relationships for the two components of the segmented linear relationship were weaker than in our original analysis, and were not consistently statistically significant. Therefore in our revised analysis we will place less emphasis on this result, and less emphasis on the relationship between methanogenic pathway and $\delta^2\text{H-CH}_4$ generally, as suggested by reviewer 1. Instead we will discuss four processes or variables that have the potential to influence $\delta^2\text{H-CH}_4$ in freshwater environments: (i) differences in methanogenic pathway, including possible use of methanol as a substrate; (ii) methane oxidation; (iii) isotopic fractionation due to diffusion; and (iv) differential thermodynamic favorability of methanogenesis, or differential enzymatic reversibility. Ultimately, our conclusion is that $\delta^{13}\text{C-CH}_4$ or α_C cannot fully resolve the effects of these processes on $\delta^2\text{H-CH}_4$ on a global basis, and other approaches will be necessary to determine their relative importance, or the possible importance of other processes.

We will continue to present these results in Figure 5, given that we feel it is important to show co-variation, or lack thereof, between these isotopic measurements. Given the findings mentioned above, we will substantially revise Figure 6. Instead of distinguishing samples by inferred methanogenic pathway in this figure, we will distinguish samples by environment (wetland vs inland water), and also show available data for cow rumen and landfills. We may reverse the order of Figures 5 and 6.

5) Reviewer 2 made numerous comments about the representativeness of our $\delta^{13}\text{C-CH}_4$ dataset. We want to make clear that to our knowledge this is the largest database of freshwater methane $\delta^{13}\text{C-CH}_4$ currently compiled. For comparison, the second largest dataset, that of Sherwood et al., (2017), includes 48 freshwater sites (including rice paddies), of which 16 are also included in our database. However our $\delta^{13}\text{C-CH}_4$ database is not comprehensive (unlike the $\delta^2\text{H-CH}_4$ database), in that it does not include many measurements that are not paired with $\delta^2\text{H-CH}_4$ measurements and that have not yet been compiled into a database. It is also probably not representative, because some important environments, namely C_4 plant dominated ecosystems, are not well represented.

Since the primary focus of this paper is $\delta^2\text{H-CH}_4$, it is not within its scope to provide a comprehensive database of freshwater $\delta^{13}\text{C-CH}_4$, although that would be a worthwhile goal for future research. In order to make our analysis as complete as possible, in our revised manuscript we will include the 32 freshwater sites from Sherwood et al., (2017) that were not included in our original analysis in our calculations for the upscaling exercise, as well as Figures 7, 8, and 9. We will also carefully discuss the likely biases in

this dataset, especially in terms of C₄ plant environments, and their implications for our interpretations.

6) Both reviewers 2 and 3 expressed some concerns with the upscaling analysis. We acknowledge that the upscaling analysis is relatively simplistic, and that some of the interpretations were speculative. However, we still think it is valuable to use the estimates of freshwater CH₄ isotopic composition, differentiated by latitude, produced in this study to estimate global source $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$, and to compare that with other estimates. We wish to make clear that given uncertainties and complexity in estimating sink fractionations, particularly for $\delta^2\text{H-CH}_4$, we are not attempting to estimate atmospheric values, but instead the integrated source $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$ prior to sink fractionations. We think there is value in comparing this with (i) previous bottom-up estimates of these values; and (ii) with the top-down estimates reported by Rice et al., (2016). We concur with Reviewer 2 that the discussion of alternate emissions scenarios is too speculative and simplistic, and therefore we will remove this discussion. Instead, we will focus on likely sources of error in the isotopic source signatures, and the best ways to address these errors in future studies.

We disagree with Reviewer 2 that the error estimates for isotopic source signatures are generally too optimistic, which we will discuss in more detail in our response to that reviewer.

Given comments from all three reviewers we will revise Figure 10 to only include panel C, and make the comparison with other estimates of global source $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$ clearer in this figure.

Specific Responses to Reviewer 2: Reviewer comments are in plain text. **Responses are in bold text.**

The paper investigates the relation between the hydrogen isotopic composition of methane emitted from freshwaters on the global scale and the isotopic composition of water and/or modeled precipitation, as well the carbon isotopic composition of methane and carbon dioxide. The authors analyze data from a large number of previous studies and apply statistical methods in order to evaluate correlations between the various signatures. The statistics are applied in a straightforward manner.

We thank the reviewer for their assessment.

I am missing a more detailed/critical scientific analysis of differences between the results of this study and previous studies. This has two aspects: 1) The study uses more sites than previous studies for dD, and it uses modeled fields of dD in precipitation. Which of these differences is primarily responsible for the differences to the previous literature (or is it both)?

This is a good question and similar questions were raised by reviewers 1 and 3.

In response to these questions we will present a much more detailed comparison of the previous literature (Waldron et al., 1999) in comparison with our study. See Planned Major Revisions 2 and 3 for more details on this. The short answer is that regardless of which water isotope values are used, our dataset produces a flatter slope between $\delta^2\text{H-H}_2\text{O}$ and $\delta^2\text{H-CH}_4$ than that of Waldron et al., (1999). However, analysis of covariance (ANCOVA) indicates this difference in slope is not significant. We ascribe this difference to the inclusion of many more sites from high-latitude environments in this study. Our analysis is that the relatively small number of high-latitude sites analyzed by Waldron et al., (1999) were skewed toward relatively low $\delta^2\text{H-CH}_4$ values. We will expand on this explanation in the revised manuscript.

2) The study uses less sites than previous studies for $\delta^{13}\text{C}$. Are the results from these sites still adequate to be used in a global extrapolation?

These are important points for clarification. However, we disagree that this study uses less sites than previous studies for $\delta^{13}\text{C-CH}_4$. See our comments on Planned Major Revision 5. We noted that the dataset was not comprehensive for $\delta^{13}\text{C-CH}_4$ (i.e. it does not include all published data), whereas it is comprehensive (to the best of our knowledge) for $\delta^2\text{H-CH}_4$. However, our $\delta^{13}\text{C-CH}_4$ dataset for freshwater environments is substantially larger than the largest previously published dataset that we are aware of (Sherwood et al., 2017). We include $\delta^{13}\text{C-CH}_4$ data for 129 freshwater sites, whereas the database of Sherwood et al. (2017) included 48. Of these, 16 are included in both databases. In order to make our $\delta^{13}\text{C-CH}_4$ analysis more accurate we will include all sites from Sherwood et al., (2017) in our analysis of $\delta^{13}\text{C-CH}_4$ variability. This expands the number of sites included to 161. There is a clear need for a larger effort to compile freshwater CH_4 $\delta^{13}\text{C-CH}_4$ data into a comprehensive database, but such an effort is beyond the scope of this paper. We will highlight the importance of this for future research in our revised discussion.

The derived global average ^{13}C source signature derived by the authors is almost certainly too light, given what we know about the fractionation in the sinks. Furthermore, I think that the errors assumed for the bottom-up determination of the global average the source signatures are too optimistic, and the discussion on the implications for the atmospheric isotope budget in section 4.6 and too simplistic. See detailed comments below.

We agree that it is too light, which was a key point of our analysis in the original Discussion (Line numbers 617-638). Based on the comments of reviewer 2, as well as reviewer 3, it is clear that the upscaling exercise in the current version of the paper is too limited to provide new insights into atmospheric methane budgets. However, we also feel that a more detailed upscaling exercise is beyond the scope of this paper, which as mentioned by Reviewer 1 is long and ambitious in scope. We think it is still worthwhile to perform the mixing model calculations for global methane source isotope signatures, and to compare these with previous estimates. See Planned Major Revision 6 for more details on this. Instead of the comparison to atmospheric budgets that we originally discussed, our revised discussion will focus on the likely

sources of error or bias in isotopic source signatures, and make recommendations to improve isotopic source signal estimates.

We disagree in general that our uncertainties for the isotopic source signatures are too optimistic. We will provide more details on this below.

L37: I suggest citing Worden et al., 2017, where this point is shown particularly well.

We thank the reviewer for bringing this to our attention, and we will cite this paper and modify the text accordingly.

L64: Maybe you want to include here, or later in the discussion section, that there are also other lines of evidence that the hydrogen isotopic composition of CH₄ (and other trace gases) depends on the isotopic composition of the precipitation, e.g., CH₄ from biomass burning across climatic zones (Umezawa et al.2011), CH₄ produced by UV irradiation of leaves that were grown with isotopically distinct waters (Vigano et al., 2010) or molecular H₂ produced in the combustion of wood from different climatic zones (Röckmann et al., 2010).

We appreciate this suggestion. We will reference these studies in both the introduction and the discussion

L109: Replace the factor 1000 by 1, the delta value is defined the correct way in line 105, and no factor 1000 is necessary.

Thanks for this reminder, we will make the suggested change.

L136: What are the 5 categories? This is not clear, to me it sounds like 4 categories.

The list of categories will be clarified in the manuscript with a numbered list. In fact it is six categories: 1) lakes and ponds; 2) rivers and floodplains; 3) bogs; 4) fens; 5) swamps and marshes; and 6) rice paddies.

L159: Is the annual average dD value of precipitation really the best estimator for a source that very likely has a strong seasonality?

This is an important question, and given this comment as well as those of reviewer 1 clearly needs more attention. See our detailed comments on Planned Major Revision 2 that discuss this at length. In short, we plan to take seasonality into account in our revised manuscript, and we find that it is important for inland water environments in particular.

L253, Figure 1: Many of the sites are hidden behind others so I cannot see the colors. Would this improve if the figure is enlarged? It may be useful to show by color or shape for which of the sites you have measured dD-H₂O and for which not.

This is challenging because many of the sites are very close to one another, and it is difficult to resolve the individual sites, while also showing the global distribution. We included the colors to give a sense of how the values vary globally, but for a more in-depth picture of geographic variability Figure 4 may be more useful.

To respond to the reviewer's comment we will provide inset maps of specific areas with many measurements, specifically Eastern North America and Alaska. We will also show the sites with water $\delta^2\text{H}$ measurements with a different shape, i.e. a triangle.

L244, Table 1: The $\delta^{13}\text{C}$ signatures for wetland have an opposite "latitudinal order" compared to what is usually assumed, i.e. they are higher at high latitudes and lower at low latitudes. The data in Table 1 for wetlands do not agree with the data presented in Figure 7. Please explain the difference. You mention that the dataset evaluated here is different from what other studies have used for $\delta^{13}\text{C}$, so is your dataset now representative? Should this limited set of values be used in the upscaling later? The errors presented for the different source categories are too optimistic, especially for the fossil sources at the bottom of the table, but probably also for the wetland category.

The reviewer raises some key aspects of the table that are not clear.

The opposite order of the $\delta^{13}\text{C}-\text{CH}_4$ data in the wetlands is simply what the data indicate. The uncertainties overlap, and our analysis therefore implies that we cannot confidently infer a latitudinal difference in $\delta^{13}\text{C}-\text{CH}_4$ in wetlands based on currently compiled data. This is also shown in Figure 7. We note here and elsewhere in our response that there is an important absence of data from C_4 plantecosystems in this dataset and other databases. Including more data from such ecosystems would probably lead tropical sites to have a higher $\delta^{13}\text{C}-\text{CH}_4$ value. We plan to discuss this in more detail in the methods, the results, and discussion. As discussed in Planned Major Revision 5, we will include addition $\delta^{13}\text{C}-\text{CH}_4$ data from Sherwood et al., (2017). However, our analysis indicates this will not change the observation of no significant latitudinal differences in wetland $\delta^{13}\text{C}-\text{CH}_4$ values.

The differences between Table 1 and Figure 7 are a result of the Table presenting mean values, whereas Figure 7 presents median values. We presented mean values in Table 1 because it is simpler to express uncertainty for the mean, and because when thinking about atmospheric contributions we think the mean is the best estimate of the isotopic source signal. In boxplots like Figure 7 it is more common to depict the median value. However, to avoid confusion and for the sake of comparison we will also plot the mean and its standard error in Figure 7 (and also do so in Figures 8 and 9).

It is not clear to us what the reviewer means when they say the errors are too optimistic for the fossil fuel categories. The error estimates are 95% confidence intervals for the mean values for these categories based on the fossil fuel database of Sherwood et al., (2017). We consider the 95% confidence interval of the mean to be

a well-established metric for characterizing the uncertainty in the mean value of these sources. We have categorized the fossil fuel sources slightly differently than Sherwood et al., (2017), to align with the inventory categories of Saunio et al., (2020), but our uncertainty estimates are essentially the same as, and actually somewhat larger than, those of the original study (see Table 5 in Sherwood et al., 2017). Note Sherwood et al., (2017) presents standard errors of the mean. 95% CI is derived by multiplying this value by 1.96. In addition, our uncertainties for the $\delta^{13}\text{C-CH}_4$ source signal for fossil fuels is very similar to those used by Worden et al., (2017). Without further details, it is unclear why the reviewer considers these error estimates to be too small or optimistic.

We used the same approach in our estimates of uncertainty in the wetland source signatures, and other source categories, and therefore also disagree that these estimates are too optimistic.

L276, Fig 2 and related text: This is a key figure for the following analysis. In principle it is an interesting approach to use modeled dD values in case measurements are not available, but it is also a source of error. Although there is a generally good agreement, the slope is lower than 1 and this may contribute to the differences and thus may affect some of the further analysis.

We agree this is a key figure and requires more in-depth analysis, which we will provide in the revised manuscript. See our Planned Major Revision 2. We agree with the reviewer that the slope being lower than 1 is concerning. In our revised analysis we find that applying annual precipitation $\delta^2\text{H}$ to wetland environments, and growing season precipitation $\delta^2\text{H}$ to inland water environments, results in slopes that are within error of 1.

L284: Maybe you could state briefly whether you can reproduce the slope of Waldron et al. when you use the same dataset. Just as a baseline.

This is a valuable suggestion. Please see our response in Planned Major Revision 3. We have included a much more careful comparison of our dataset with that of Waldron et al (1999). It is important to note that the analysis of Waldron et al. (1999) also included key assumptions that influence the regression relationship produced with that dataset. Specifically, that study included sites with measured water $\delta^2\text{H}$ (57%) and sites with estimated water $\delta^2\text{H}$ based on regional precipitation measurements (43%). To perform a robust comparison we re-analyze the Waldron et al dataset, which is discussed in our Planned Major Revision 3. Because the exact details of the weighted regression method used by Waldron et al., 1999 are not provided, we did not precisely reproduce their regression relationship [see Table R2]. But using unweighted regression we produced a relationship that is statistically indistinguishable.

L292: Figure 3a: It looks like the lower slope is caused by a lot of points where you have only modeled but no measured dD data near the low dD-H₂O end. And these

are mostly inland waters (Figure 3b). Can you evaluate this in more detail? Can this be caused by a bias in the modeled dDp? Probably not, but it is useful to investigate further to strengthen your argument.

The reviewer correctly noted that the reported regression line for inland waters was not a good visual fit to the data, and this influenced the overall regression line. This was also noted by reviewer 1. We note that in the original Figure 4a the two regression lines were very similar, so this effect was not a result of bias in modeled $\delta^2\text{Hp}$, since a very similar regression was produced when only analyzing sites with measured water $\delta^2\text{H-H}_2\text{O}$.

After analyzing this more closely we realized that this is a result of the weighted regression methods we were using. Specifically, a few high-latitude sites with 1) many measurements (and therefore a low standard error) and 2) high $\delta^2\text{H-CH}_4$ values, were heavily weighted and had a large effect on the regression relationship. We therefore decided that a more accurate regression relationship would be produced using unweighted regression. This is supported by studies on the efficacy of unweighted regression in analyzing environmental data, which in many cases is less biased than weighted regression (Fletcher and Dixon, 2012). See more details in Planned Major Revision 3.

The unweighted regression provides a somewhat steeper slope for the overall dataset, as well as for inland waters. It also indicates there is not a significant difference in the regression whether measured $\delta^2\text{H-H}_2\text{O}$ or modeled $\delta^2\text{H-H}_2\text{O}$, or a combination of the two (i.e. a ‘best-estimate’) is used. See Planned Major Revision 3 and Table R2 above.

L308: Would you find a correlation if you took the slope of Waldron et al. for calculating CH_4, W_0 ?

We have significantly revised this analysis, as discussed in Planned Major Revision 4 above. The result of this is that the relationship between α_C and $\delta^2\text{H-CH}_{4, \text{w}0}$ is observed regardless of how $\delta^2\text{H-H}_2\text{O}$ is estimated. However, the specifics of this relationship are not robust to the method of estimating $\delta^2\text{H-H}_2\text{O}$, and therefore we will emphasize this relationship to a lesser degree in the revised manuscript.

The slope of Waldron et al., (1999) is not a good fit to the overall dataset, and therefore we do not think it makes sense to apply this to calculate $\delta^2\text{H-CH}_{4, \text{w}0}$. However, we have performed the suggested analysis as a test. It still results in a segmented relationship with α_C , but this relationship is less strong than the one presented in the original manuscript. This finding has contributed to our decision to focus less on the relationship between α_C and $\delta^2\text{H-CH}_{4, \text{w}0}$ as a signal of differences in methanogenic pathway and methane oxidation.

L323, Figure 5: Does it make sense that in b) only few points are classified as oxidation

influenced and in c) many more points? Does it make sense that in c) the very lowest dD value is in the group of the oxidation influenced points? I find the “pathway trend” concept a bit confusing, this indicates a smooth transition of dD-CH₄, W0 with alpha_C or d13C_CO₂. Is this a real trend, or rather a consequence of two different groups of data (acetoclastic and hydrogenotrophic sites)? Wouldn't it be useful in this case to show these two groups with two different colors, separated by the potential break points, rather than the trend areas?

The reviewer raises important questions about the predicted trends that we presented in Figure 5. Reviewer 1 also raised important questions about this, and given the overall lack of agreement on the predicted patterns we have decided that we should not present predicted trends, as there is not a strong consensus on these predictions. Instead, we will focus on the co-variance (or lack thereof) between $\delta^2\text{H-CH}_{4,w0}$, $\delta^{13}\text{C-CH}_4$, α_C , and $\delta^{13}\text{C-CO}_2$, and multiple mechanisms that could influence this co-variation in freshwater ecosystems. Our ultimate conclusion is that patterns of co-variation cannot definitively resolve which mechanisms for $\delta^2\text{H-CH}_4$ are most important when comparing between sites.

L350 and Figure 7b, wetlands: These numbers do not agree with the data in Table 1.

As noted above, these are median values, whereas Table 1 presents mean values. To clarify this we will also plot mean values in Figure 7.

L374-379: I get a bit confused by the diverging statements on significance with different tests, please try to reformulate, or add a sentence to synthesize.

We will re-write to clarify the significance tests, focusing on the pair-wise comparison between wetlands and inland waters first (Mann-Whitney test), and then the multiple group comparison (Kruskal-Wallis test).

L395-397: See points above: Are the uncertainties for the different categories adequate? Is there an issue with the difference between values in the text and table 1? Is the rather heavy d13C value for high latitude wetlands appropriate?

See our response to comments on Table 1. It is unclear what difference between the table and text is being referred to- we assume this is the difference between median values (Figure 7) and mean values (Table 1). The heavy value for high latitude wetlands is the mean value of this dataset, and therefore we argue it is appropriate. In our revised manuscript we will include additional data from Sherwood et al., (2017), as discussed above, which includes 5 additional high latitude wetland sites. This makes the mean $\delta^{13}\text{C-CH}_4$ value 0.5‰ lower, but does not change the median value. We will include this value in a revised Monte Carlo analysis, but in essence this additional data does not change our conclusion. Based on our analysis, an assumption of low $\delta^{13}\text{C-CH}_4$ in high latitude wetlands is not supported by the available data, and we think this assumption requires further empirical validation.

L431 ff: The differences to the previously published values from Waldron et al. should be discussed in some more detail. E.g., is there an influence from the modeled dDp values, or a certain sampling region? L439 ff: Same for the discussion of the environment type

See our responses above and Planned Major Revisions 2 and 3. Our conclusion is that the difference is largely controlled by the small number of high-latitude sites in the Waldron et al (1999) dataset, and that those sites were skewed towards relatively low $\delta^2\text{H-CH}_4$ values. We do not observe a significant difference in the regression relationship when modeled or measured $\delta^2\text{H-H}_2\text{O}$ values are used (see Table R2 above, as well as Figure 4a in the original manuscript).

L465, section 4.2.1: See comments above on the representativeness of the dataset analyzed here and possible consequences. You write that the dataset is not comprehensive or d13C, so should it be considered as representative? In this case, what have other studies potentially missed?

See Planned Major Revision 5 above. As mentioned above, it is the largest compiled dataset available, but it is not comprehensive because there is a large amount of $\delta^{13}\text{C-CH}_4$ data that has not yet been compiled into a database. It is also probably not representative, with a notable lack of data from C_4 plant ecosystems. Given that it is the largest dataset available, we proceed with analyzing it. However, in the revised manuscript we will give more attention to the likely sources of error, and key data gaps that should be addressed.

L483 ff: You may want to refer here to the studies I mentioned in the beginning that looked at other (non-microbial) sources.

Thanks for this suggestion, we will mention these studies here.

L519 ff: The authors state that they do not observe a correlation between dD and d13C of CH_4 . Nevertheless, the vast majority of the points in Fig 5a seem to fall in the range of the “pathway trend” (I find the term misleading, see comments above). Does this not mean that the two groups (acetate fermentation and CO_2 reduction) still form distinct distributions?

As mentioned above, there are concerns with the ‘pathway trend’ noted by Reviewer 2, as well as Reviewer 1 and we have decided to omit this from the revised manuscript. Our primary concern is whether $\delta^{13}\text{C-CH}_4$ is a strong predictor of $\delta^2\text{H-CH}_4$, and our analysis indicates that it is not. It is still possible that different pathways form different distributions in terms of $\delta^{13}\text{C-CH}_4$, but these distributions do not correspond to clear differences in $\delta^2\text{H-CH}_4$.

L549: the remark on the intercepts does not add much and is rather trivial when the slope is different.

This discussion will now be heavily modified, as discussed in Planned Major Revision 5. We will not focus on the role of methanogenic pathway as much in the revised manuscript. We will use analysis of covariance (ANCOVA) for any comparison of regression relationships in the revised manuscript.

L555 - 561: I am also not aware of dD measurements in natural acetate, but the method from Greule et al. (2008) has been used in Vigano et al. (2010) to measure dD in methoxyl groups which were compared to produced CH₄ and modeled dD in water.

We appreciate these suggested references. We will include them in our revised discussion.

L574 – 578: Why do you explain the variability for bogs by the pathway difference, and the high values in rivers by oxidation. Can oxidation not also cause large differences for bogs?

This inference was based on the differences in both $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$. Since bogs have higher $\delta^2\text{H-CH}_4$ on average, but lower $\delta^{13}\text{C-CH}_4$, we inferred this was related to a pathway difference. We were also influenced by previous studies (i.e. Ganesan et al., 2018) that had suggested bogs have a higher proportion of hydrogenotrophic methanogenesis. In contrast, rivers are higher in both $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$, which we inferred to be a signal of oxidation. We will make this analysis clearer in the revised manuscript. In addition, as shown in Figure 5a, co-variation in $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ is not necessarily indicative of mechanisms for isotopic variability, so we will moderate our interpretations here in the revised manuscript.

L599: Why should the oxidation signal only be apparent for dD and not for d13C (L603-604)?

Overall dissolved CH₄ from inland waters is also shifted to higher $\delta^{13}\text{C-CH}_4$ values, although this is not a significant difference. We will note in the revised manuscript that greater oxidation would be expected to lead to higher $\delta^{13}\text{C-CH}_4$ values, and the absence of a strong signal in $\delta^{13}\text{C-CH}_4$ may be inconsistent with our hypothesis. We will also discuss other possible mechanisms for the observation of high $\delta^2\text{H-CH}_4$ in dissolved inland water samples, including different water sources and effects of diffusion on isotopic fractionation.

L606: I do not understand how you can conclude that “: : that the relative balance of diffusive vs. ebullition gas fluxes should not have a large effect on the isotopic composition of freshwater CH₄ emissions.”. The chance for oxidative effects is much larger for a slow process like diffusion compared to the fast process of ebullition.

This statement is simply a reflection of the available data, as shown in Figure 9a and b, which do not show a clear difference between these two gas sample types in their

isotopic composition. We note below this (lines 607-610) several caveats that moderate this conclusion, and that the question deserves more study. We will add the likely greater effect of oxidation on diffusive fluxes as an additional area that requires further empirical validation.

L611: The analysis in this section has much less scientific rigor than the previous sections and presents some sensitivity calculations involving highly improbable assumptions, see following points.

We acknowledge that the sensitivity calculations and scenarios are somewhat simplistic and loosely defined. As discussed above, we think the solution to this is to scale back this section to focus on the results of a global source mixing model calculation, to compare that with previous estimates of global source signals, and to discuss key data gaps that are likely leading to biases in this estimate (See planned major revision 6). Therefore the revised manuscripts would not include the sensitivity calculations, which would be left for future work.

L619 ff: See comments above on the depleted d13C source signature. Here you argue that three factors may explain this difference. I am quite convinced that the first one (errors in the sink fractionation factors) cannot explain the large difference. The two published studies for the fractionation in the CH₄ + OH reaction (Cantrell et al, 1990, Saueressig et al, 2001) are 5.4 and 3.9 per mill, respectively. A contribution from Cl may increase this a bit, but not enough to support a global average source signature of -56.4 per mill. So I think that the reason should come from the other two processes mentioned. Given the discrepancy to previous studies I wonder whether it is not mainly the choice of signatures in this study. In line 625 you already show that changing one parameter leads to a change of the global average source signature of 1.3 per mill, which is almost the entire uncertainty range reported.

We acknowledge the point the reviewer is making. As discussed above, we will revise this section to limit our interpretation to comparison with previous estimates and possible biases in isotopic source signals, and not focus on sink fractionations, which are not a focus of this study. We will mention errors in flux inventories, which we think is probably partly responsible for the discrepancy.

L628: Rather arbitrarily changing big sources by a factor of 2 is a huge adjustment of the atmospheric CH₄ budget. This investigation on the effect on the atmospheric isotopic composition is too simplistic.

We understand this critique, and as discussed above we will avoid performing this analysis in the revised paper. This analysis was based on the work of Schwietzke et al., (2016), who make a similar, but more precise adjustment. We will mention the possibility of higher fossil fuel emissions than in inventories, as discussed by Schwietzke et al., (2016), but leave a detailed analysis resolving this with $\delta^2\text{H-CH}_4$ measurements to future studies.

L634 ff: Same comment for the bb source, this should be discussed in a more detailed way. Worden et al. (2017) illustrate the strong influence of the bb source.

As discussed above, we feel it is best to omit the discussion of specific different emissions scenarios from the discussion. We will briefly discuss the results of the Worden et al., (2017) study, and mention biomass burning emissions as an influential variable for isotopic source signatures that merits further study, particularly in terms of $\delta^2\text{H}$.

L660f: The statement “This flatter slope may be the result of the inclusion of a greater proportion of inland water sites in our dataset.” requires more underlying analysis. I think that the “may be” can be replaced by “is likely”, but this should be investigated. See also other points above.

Based on the comments of all three reviewers we will thoroughly revise our comparison of our results with that of Waldron et al (1999). Therefore this part of the conclusions will be changed to reflect this revised comparison, and likely causes of the different slope. Our revised analysis implies that differences between inland waters and wetlands is probably not primarily responsible for this difference (see Table R2), and that a greater amount of data from high-latitude environments is more important.

L662: If possible make more concrete after reevaluation of the impact of modeled data.

We will also revise this statement after a more thorough analysis of the differences in the regression relationship for modeled and measured $\delta^2\text{H-H}_2\text{O}$. Our revised analysis shows that using modeled $\delta^2\text{H-H}_2\text{O}$ provides a good estimate of the relationship between $\delta^2\text{H-H}_2\text{O}$ and $\delta^2\text{H-CH}_4$, and supports the use of isotope-enabled Earth Systems Models to predict $\delta^2\text{H-CH}_4$.

L686: Here the second argument of the three presented before (see comment on L619) has disappeared, but as argued above it may be the most important one and particularly the sink argument does likely not explain (at least exclusively) the difference.

As discussed above, we will substantially revise and scale back the upscaling estimates. Therefore these conclusions will be thoroughly changed. We will focus primarily on the uncertainties in the source signature estimates.

References Cited:

Fletcher, D., & Dixon, P. M. (2012). Modelling data from different sites, times or studies: weighted vs. unweighted regression. *Methods in Ecology and Evolution*, 3(1), 168-176.

Rice, A. L., Butenhoff, C. L., Teama, D. G., Röger, F. H., Khalil, M. A. K., and Rasmussen, R. A.: Atmospheric methane isotopic record favors fossil sources flat in

1980s and 1990s with recent increase, *Proceedings of the National Academy of Sciences*, 113, 10791-10796, 2016.

Schwietzke, S., Sherwood, O. A., Bruhwiler, L. M., Miller, J. B., Etiope, G., Dlugokencky, E. J., Michel, S. E., Arling, V. A., Vaughn, B. H., and White, J. W.: Upward revision of global fossil fuel methane emissions based on isotope database, *Nature*, 538, 88-91, 2016.

Sherwood, O. A., Schwietzke, S., Arling, V. A., and Etiope, G.: Global inventory of gas geochemistry data from fossil fuel, microbial and burning sources, version 2017, *Earth System Science Data*, 9, 2017.

Waldron, S., Lansdown, J., Scott, E., Fallick, A., and Hall, A.: The global influence of the hydrogen isotope composition of water on that of bacteriogenic methane from shallow freshwater environments, *Geochim Cosmochim Acta*, 63, 2237-2245, 1999a.