Response to reviewer 1

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 δ^2H-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 δ^2H-CH_4 and δ^{13}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 δ^2H-H_2O values. Specifically we have done the following: (i) In addition to annual precipitation δ^2H values, we now also analyze growing season precipitation δ^2H, which is defined as the amount-weighted mean δ^2H of months with mean temperature greater than 0º C. This provides an opportunity to assess whether seasonal variation in precipitation in the mid to high-latitudes is important in controlling the environmental δ^2H-H_2O value; (ii) separately analyzing inland water and wetland environments, since these are different hydrological environments and the controls on δ^2H-H_2O are potentially different.

This analysis led to the following key results (see Table R1 below for summary of results): A) growing season modeled precipitation δ^2H is a better predictor of inland water δ^2H-H_2O than annual precipitation δ^2H, in that the regression curve is indistinguishable from the 1:1 line. B) Annual modeled precipitation δ^2H is a better predictor of wetland δ^2H-H_2O, 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±9‰. We interpret this as an indicator of likely widespread evaporative effects on δ^2H-H_2O in wetland environments.

We use these results to then develop a 'best estimate' for comparing δ^2H-H_2O with d2H-CH4. (i) For sites with measured δ^2H-H_2O values we use the measured value. (ii) For inland water sites without measured δ^2H-H_2O we use modeled growing season precipitation, since as discussed above the regression of this against measured δ^2H-H_2O is indistinguishable from the 1:1 line. (iii) For wetland sites without measured δ^2H-H_2O we estimate the δ^2H-H_2O using the regression relationship with annual precipitation δ^2H-H_2O 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 δ^2H-H_2O (29 out of 51 sites) and estimated δ^2H-H_2O based on precipitation isotopic measurements (22 out of 51 sites).

Table R1: Comparison of regression relationships between modeled δ^2H_p and measured δ^2H-H_2O

	Slope	Intercept	R^2	RMSE	*p*	*n*
Inland waters						
Growing season δ^2H_p	1.05±0.09	-0.3±8	0.82	22.3	4.81E-13	33
Annual δ^2H_p	0.85±0.06	-2.1±7	0.84	20.5	3.17E-14	33
Wetlands						
Growing season δ^2H_p	1.24±0.09	14.8±10	0.87	16.5	4.46E-13	28
Annual δ^2H_p	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 δ^2H-H_2O on the regression between δ^2H-H_2O and δ^2H-CH_4. To do this carefully we performed the regression analysis using four different estimates of δ^2H-H_2O: (i) the 'best-estimate' of δ^2H-H_2O as described above in Planned Major Revision 2; (ii) measured δ^2H-H_2O, only analyzing sites with this measurement; (iii) modeled annual precipitation δ^2H; and (iv) modeled growing season precipitation δ^2H. 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 δ^2H-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 δ^2H-H_2O (29 out of 51 sites). A summary of the results of this analysis is 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 δ^2H-H_2O does not have a significant effect on the estimate of the relationship between δ^2H-H_2O vs. δ^2H-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 δ^2H-H_2O and δ^2H-CH_4 using different estimates of δ^2H-H_2O

	Slope	Intercept	R^2	RMSE	*p*	*n*
Best Estimate δ^2H-H_2O						
All	0.44±0.05	-298±5	0.42	27.4	1.33E-16	129
Wetlands	0.51±0.06	-300±5	0.59	23.7	7.85E-12	55
Inland Waters	0.42±0.07	-295±7	0.34	29.1	6.72E-08	74
Measured δ^2H-H_2O						
All	0.5±0.08	-292±8	0.43	28.6	1.22E-08	61
Wetlands	0.53±0.11	-298±13	0.44	26.2	6.58E-05	28
Inland Waters	0.42±0.1	-291±10	0.37	28.9	0.000156	33
Modeled Annual δ^2H_p						
All	0.42±0.04	-293±5	0.44	26.9	1.17E-17	129
Wetlands	0.57±0.06	-287±6	0.65	21.9	1.01E-13	55
Inland Waters	0.37±0.06	-293±7	0.36	28.4	1.25E-08	74
Modeled Growing Season δ^2H_p						
All	0.51±0.05	-292±5	0.41	27.6	2.55E-16	129
Wetlands	0.71±0.07	-285±6	0.63	22.4	4.05E-13	55
Inland Waters	0.44±0.07	-294±8	0.33	29.3	1.03E-07	74
Waldron et al. (1999)						
All data	0.74±0.1	-284±6	0.5	26.3	6.56E-09	51
Measured δ^2H-H_2O only	0.79±0.2	-279±10	0.44	29.6	9.21E-05	29

4) We then used the 'best-estimate' δ^2H-H_2O values and the regression based on those values, shown in Table R2, to calculate a revised δ^2H-$CH_{4,w0}$ 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 δ^2H-H_2O, using the values and regression curve for those sites.

Notably, for the comparison between δ^2H-$CH_{4,w0}$ 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 δ^2H-H_2O. 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 δ^2H-CH_4 generally, as suggested by reviewer 1. Instead we will discuss four processes or variables that have the potential to influence δ^2H-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 δ^{13}C-CH_4 or α_C cannot fully resolve the effects of these processes on δ^2H-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 δ^{13}C-CH_4 dataset. We want to make clear that to our knowledge this is the largest database of freshwater methane δ^{13}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 δ^{13}C-CH_4 database is not comprehensive (unlike the δ^2H-CH_4 database), in that it does not include many measurements that are not paired with δ^2H-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 δ^2H-CH_4, it is not within its scope to provide a comprehensive database of freshwater δ^{13}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 C_4 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 CH_4 isotopic composition, differentiated by latitude, produced in this study to estimate global source δ^2H-CH_4 and δ^{13}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 δ^2H-CH_4, we are not attempting to estimate atmospheric values, but instead the integrated source δ^2H-CH_4 and δ^{13}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 δ^2H-CH_4 and δ^{13}C-CH_4 clearer in this figure.

Specific Responses to Reviewer 1: Reviewer comments are in plain text. **Responses are in bold text.**

We thank Dr. Waldron for her careful and thorough review. We appreciate that this review presented a challenging situation, and we value her honesty and openness. We are confident that we can address her concerns in the revised manuscript.

Substantial conclusions are reached, but the interrogative approach has weaknesses that propagate through substantial analytical reasoning and so the integrity of the conclusions is questionable. I detail this further below, but until the analytical approaches are reconsidered the conclusions are not securely reached

We understand this critique, and in response we will strengthen the statistical analyses and interrogation of the data, in particular with regards to the 1) comparison of measured and modeled δ^2H-H_2O values, 2) the inference of a regression slope between δ^2H-H_2O and δ^2H-CH_4, and 3) the application of carbon isotope fractionation factors to evaluate the potential effects of methane oxidation, methanogenesis pathway, or other biogeochemical effects on δ^2H-CH_4. See the planned major revisions 2, 3, and 4 above.

With respect to understanding isotopic compositions: the methods are not all valid, particularly the reconstruction of missing δD-H_2O for a field measurement of δD-CH_4. The interrogation of this relationship (Fig. 2) lacks statistical rigour, and its propagation - a relationship that has bias and significant variability - is unconsidered in all analysis thereafter (as represented by Figs. 3-9 and possibly 10) and so this reasoning is flawed and the interpretations may be wrong.

We will thoroughly re-analyze the relationship between measured and modeled δ^2H-H_2O in the revised manuscript. See Planned Major revision 2 above. We will then carry this revised analysis forward to the remainder of the manuscript. See Planned Major Revisions 3 and 4.

The authors are not consistent in identifying when processes they are interpreting are based on hypothesised relationships and the impression is given such processes are certain (detailed below).

We will make it clearer in the revised manuscript where we are discussing hypothesized relationships, as discussed in more detail below. In particular regarding hypotheses regarding the effects of methanogenic pathway on δ^2H-CH_4 we will be more circumspect in the revised manuscript, and discuss alternate hypotheses in greater detail. See Planned Major Revision 4.

I found it difficult to follow the calculations behind αC – an important part of the manuscript – when I was trying to compare other data sets with their approach.

We are uncertain what aspect of this calculation was unclear, but we will endeavor to make the description of this calculation clearer.

Largely bit not always, for example there is a large section in 4.31. that is repeating suggestions made in section 1.1. of Waldron et al 1999, but this work is unreferenced and so as written implies the review m/s is the first to have suggested this; the abstract does not make clear refining an existing phenomena observed and described similarly previously.

We regret the omission of references and acknowledgment to Waldron et al., 1999, in the ideas presented in section 4.3.1. We will thoroughly revise this section to provide proper credit for these ideas, and integrate the discussion with that previously published by Waldron et al., (1999).

Broadly but not sure how "geographic variability in freshwater methane hydrogen isotope ratios has implications for microbial biogeochemistry" - the microbes are active with no knowledge δD…so this can be refined.

We will follow this suggestion and modify the title to: *Geographic variability in freshwater methane hydrogen isotope ratios and its implications global isotopic source signatures*.

It is a paper with a lot of detail and so to follow it all the reader has to concentrate deeply for the results section. As such, and maybe in addition, the discussion from section 4 onwards seems in places repetitive.

We agree the results are very detailed. Given the critiques of the reviewers we will need to add additional statistical analyses and data interrogation to the results,

leading to an overall increase in detail. However, we will present this in as streamlined and clear manner as possible. We will also carefully review and revise the discussion sections to avoid repetition.

Broadly yes – I suggest a group whose work may be missing in the intro.

We will add the suggested citations in the introduction

Yes, very helpful, but sheet 2 could make it clearer if the data offered is used in α_C or these are summarised data from other sources.

Sheet 2 will be omitted from the revised manuscript, as multiple reviewers have questioned the value of the predicted fields for pathway and oxidation dependent isotopic variation. See Planned Major Revision 4.

The substantive conclusions in this manuscript rely on a data set where δD-H_2O does not exist for more than half the data: 53% of the sites do not have field measured δD-H_2O (L88). In these cases, δD-H_2O is inferred from a reputable global precipitation database and a correlation observed for sites where measured values exist. The authors consider this relationship sufficiently robust to proceed to use the reconstructed δD-H_2O where measured values do not exist. I disagree this is the case.

We acknowledge this is an important critique. In the revised manuscript we will take steps to strengthen the analysis of the relationship between modeled and measured δ^2H-H_2O, and to carefully evaluate if using modeled values of d2H-H2O leads to a bias in the inferred relationship between δ^2H-H_2O and δ^2H-CH_4. See Planned Major Revisions 2 and 3. Our analysis indicates that using modeled δ^2H-H_2O does not lead to a significant difference in the regression relationship with δ^2H-CH_4. See Table R2 above.

The statistical integrity shown elsewhere in the manuscript is lacking in this section on reconstructing δD-H_2O, with the authors describing their predictive relationship as showing "generally good agreement" and proceeding to use it. The bias and variability in a predictive δD-H_2O_p and thus how far it may be from the true δD-H_2O appear unconsidered in any further analysis (no errors propagated through for estimated δD-H_2O?).

As described above, in the revised manuscript we will more carefully evaluate this relationship. See Planned Major Revision 2.

Further, I note that the data in table S3 supplementary information for which there are measured δD-CH_4-δDH_2O fit closely to the in-vitro line from which Waldron et al 1999 project a global relationship - but the data with estimated δDH_2O in table S3 do not. This is important for two reasons:

The reviewer has pointed to an interesting observation. The difference observed by the reviewer is at odds with Figure 3A in the original manuscript, which clearly showed that the regression lines for δ^2H-H_2O vs δ^2H-CH_4 fully overlap whether modeled (black regression

line) or measured (blue regression line) δ^2H-H_2O is applied, and was not in agreement with the in vivo line of Waldron et al., (1999).

On further analysis of the data, we identified that the discrepancy observed by the reviewer was probably caused by two factors:

1) the weighted regression method that we used in the original manuscript was strongly influenced by a few sites at high latitudes that both (a) have a large number of measurements (and therefore a low standard error and a higher weight); and b) relatively high δ^2H-CH_4 values. We infer that this strong weighting at these sites leads to a strong bias in the regression. Based on statistical research involving environmental samples (Fletcher and Dixon, 2012) we infer that unweighted regression is preferable for this dataset (See Planned Major Revision 3).

2) The unweighted regression performed by the reviewer was likely strongly influenced by an outlier site from the Amazon with very high δ^2H-CH_4. When this point is removed, as suggested by reviewer 1, the regression slope becomes flatter.

The two factors above effectively cancel each other out. In re-analyzing the data after accounting for these two changes (see Table R2 above) we find that (a) there is not a large or significant difference in the regression slope if measured δ^2H-H_2O, 'best-estimate' δ^2H-H_2O, or modeled precipitation δ^2H-H_2O is used and (b) all of these regression slopes are flatter than that of Waldron et al., (1999).

1. It confirms the predictive relationship in Waldron et al 1999 for δD-CH_4 from δD-H_2O still has integrity, more so by adding in another methane-producing environment (innocula), a significant time gap, and another geographic locality.

As mentioned above, our revised regrewssion analysis continues to result in flatter slopes than the predictive relationship proposed by Waldron et al., (1999), regardless of the method of estimating δ^2H-H_2O. However, as we also discuss above, given the wide confidence intervals of these relationships we do not find a significant difference with the prediction of Waldron et al., 1999 using multiple group comparison ANCOVA. However, given that every analysis of the larger dataset presented here results in a flatter slope, we think it is highly likely that the global relationship has a somewhat flatter slope than that inferred by Waldron et al., (1999).

2. If statement 1 is considered sound, then the poor fit of paired δD-CH_4- δD-H_2O with predicted δD-H_2O supports the assertion above that the relationship the authors are using here to reconstruct δD-H_2O is questionable.

As noted above, we do not see evidence that there is a significant difference in δ^2H -CH_4 vs. δ^2H -H_2O based on the method of estimating/inferring δ^2H-H_2O. Therefore we disagree with the assertion that the methods used to reconstruct δ^2H-H_2O are questionable. However, we have revised our approach to predicting δ^2H-H_2O, as explained in more detail in Planned Major Revisions 2 and 3.

We think it is important to note here that the in-vivo relationship of Waldron et al., (1999) is not based purely on sites with measured δ^2H-H_2O. Instead, that study used a combination of sites with δ^2H-H_2O measurements (57%) measurements and estimates based on precipitation isotope measurements (43%).

Quoting from that study: "*Where paired δD(CH4)–δD(H2O) measurements were not published δD(H2O) was sourced from measured precipitation values for the area, for example, the weighted mean of the precipitation samples collected in south Florida over a 3-yr period (Swart et al., 1989) was used as an appropriate value for δD(H2O) for St. Marks Swamp, Florida (Happell et al., 1994).*

Other unknown δD(H2O) signatures (e.g., the Alaskan Lakes; Martens et al., 1992) were estimated from the weighted mean value of sites close to the area sampled, that participated in the global network, Isotopes in Precipitation (IAEA, 1992) or from the meteoric water line (Craig, 1961). We are aware that δD(groundwater) can differ by up to 30‰ frommeasured δD(precipitation) (e.g., Hornibrook et al., 1997; E. R. C. Hornibrook, pers. comm.), but such fractionation is difficult to quantify and the logical approach we have adopted provides the best estimate for dD(H2O) where measured values are unavailable."

We agree with Waldron et al., (1999) that differences between precipitation and groundwater (or lake water) can be large, and these differences can be difficult to quantify. However, we do not think these potential differences negate the value of estimating δ^2H-H_2O using available estimates of precipitation δ^2H, along with accounting for the effects of precipitation seasonality and evaporation, as described in Planned Major Revision 2. Indeed, we think our revised approach of combining measured and estimated water δ^2H into a 'best-estimate' is a logical extension of the approach used by Waldron et al., (1999). However, we agree that a more careful evaluation of this approach is warranted, and we are planning to add this to the revised manuscript as described above (Planned Major Revision 2).

With the greatest of respect, using the predicted data produces an outcome that is like a 'house of cards' – all subsequent analysis using this data is built on a shaky foundation. I therefore think that incorporating paired $\delta D-CH_4$ -$\delta D-H_2O_p$ in further analysis is flawed and offer two examples why:

1. It creates a new global line for $\delta D-CH_4$ -$\delta D-H_2O$ that may be wrong.
2. It could lead to artefact in interpretation, which indeed may be 'visible' in the dependent analysis. For example, the data in Fig. 3b visually also appears to separate between paired $\delta D-CH_4$ -$\delta D-H_2O$ data that are predicted (inland waters) and measured (wetlands), and if this is the case interpreting a biome difference here, and later in the paper, is also questionable.

We do not agree with the house of cards analogy, but we do agree that it is imporotant to provide more confidence in our underlying analyses. As mentioned above (Planned Major Revision 3; Table R2), we do not observe a significant difference in δ^2H-CH_4 vs. δ^2H-H_2O whether modeled or measured $\delta^2H-H_2O$is used. Therefore we disagree that there is a 'shaky foundation' to our subsequent analysis.

However, as discussed above we have revised our data analysis to use the 'best-estimate' δ^2H-H_2O value, including measured values where available.

We note that in Figure 3B the reviewer likely misinterpreted the data presented. All of the data shown in this figure are based on modeled δ^2H-H_2O. Therefore the observed difference between inland waters and wetlands cannot be because of differences in the source of δ^2H-H_2O data. See Table R2 above, which shows that there are consistent differences between wetlands and inland waters in the slope of the regression line regardless of the method used to estimate δ^2H-H_2O, but also that these differences are small and statistically insignificant. Therefore in our revised manuscript we will state that we cannot confidently infer a difference in the relationship between these environments.

With respect to the redefining of a new global $\delta D-CH_4-\delta D-H_2O$ and consideration of how this has changed from the relationship offered in Waldron et al 1999: unless the authors can produce a more robust estimation of $\delta D-H_2O_p$, the data that uses $\delta D-CH_4$ paired with predicted $\delta D-H_2O$ needs to be removed - for as noted earlier, there is insufficient confidence this is an accurate representation of the field situation and may create a false outcome. I suspect this will change the global relationship and increase the slope as paired data with $\delta D-H_2O_p$ visually appears to dominate the enriched samples.

We believe the approach taken by our Planned Major Revision 3 effectively addresses this critique. We now take an approach similar to that of Waldron et al., (1999), namely combining measured and modeled δ^2H-H_2O values to produce a best-estimate value for each site.

As noted above, we do not observe a significant difference in the δ^2H-CH_4 vs. δ^2H-H_2O relationship whether modeled or measured δ^2H-H_2O is used (See Table R2)

Please plot both the in-vitro and in-vivo relationship, and for the former its prediction intervals - which are missing from 3b and so give the sense of a poorer fit of Waldron et al 1999 to the bgd expanded field data set here.

We are planning to include a more robust comparison with the data from Waldron et al (1999) in Figure 3 in the revised manuscript, including confidence intervals. We argue that confidence intervals are the more appropriate metric, since this gives the uncertainty of the regression relationship, as opposed to the predicted range of observations. We are more interested in comparing the underlying regression relationships, as opposed to the predicted range of observations.

Also, the new figures will be highly complex with multiple regression relationships and therefore we think that including the in-vitro relationship, which does not significantly overlap with the data, would only further complicate the figures. However, we will include a comparison with the in-vitro relationship in the revised version of Figure 6, as was done in the original manuscript.

Compare whether the in-vivo line is statistically different to the relationship generated from the data set presented in the bgd manuscript. This will allow confidence in any further discussion on how the relationship has been redefined (than just comparing slopes etc). If the two relationships are indistinguishable statistically, nuanced statements about differences in slope etc are meaningless – all that has happened is that the expanded data set has redefined better the field relationship for δD-CH_4 -δD-H_2O (as indicated likely in Waldron et al, 1999) - noting that this field relationship does not wholly reflect the relationship at production (see next point).

We will use analysis of covariance (ANCOVA) to statistically compare differences in regression relationships. As discussed in Planned Major Revision 3, multiple group comparison with ANCOVA does not indicate a significant difference in slope between our dataset and that of Waldron et al. (1999). However, regardless of the method of estimating δ^2H-H_2O used, the analysis of the larger dataset produces a flatter slope. Therefore we infer that the 'true' global slope is likely to be flatter than that inferred by Waldron et al., 1999, but also that more data and further analysis is needed to confirm this and reduce the uncertainty of the slope.

Assess whether the expanded field data set is predominantly $_{13}$C-enriched compared to the in-vivo relationship described in Waldron et al 1999, and therefore consistent with an interpretation that differences in field δD-CH_4 may be an artefact of fractionating processes post-production than pathway per se This is advocated as I am still unaware of experimental evidence methanogenic pathway in shallow freshwaters changes δD-CH_4, but there is evidence of processes, oxidation and mixing, causing enrichment, and so this approach is consistent with scientific principle of parsimony and interpreting data using the simplest approach.

We have assessed this and in fact the opposite is the case. The data from sites included in Waldron et al., (1999) is somewhat higher in $\delta^{13}C$-CH_4 (-60.8±0.9‰ SEM) relative to the total dataset (-62.6±0.6‰) or the sites that were not included in Waldron et al., (1999) (-63.4±0.8‰). We do not think it is likely that there is systematic difference in these sets of sites in terms of post-production processes, which we take to mean oxidation, diffusive fractionation, and mixing of different CH_4 reservoirs. We agree that such processes can lead to variation in δ^2H-CH_4 (see planned major revision 4 above), but see no evidence that this explains the difference between our dataset and that of Waldron et al., (1999).

To explore why the paired δD-CH_4-δD-H_2O measurements are not fully described by the best fit line, the authors explore whether a difference in (dominant) methanogenic pathway is evident in the data. With no evidence from paired δD-CH_4-$\delta_{13}CH_4$ the authors draw on a_c as a proxy for methanogenic pathway to assess this. Step-wise regression is used to explore this. I think this is interesting and something to revisit when the paired data relying in predicted δD-H_2O has been removed, but currently it is the next floor in the 'house-of-cards', reliant on data that we do not know to be accurate, and therefore the significant relationships that the authors infer changes in methanogenic pathway from, we do not know to be true.

We have re-assessed the step-wise regression relationship between α_C and d2H-CH4,wO using revised values for the latter as described above (Planned Major Revision 4). Indeed to ensure this relationship is robust we tested it using two different approaches: 1) δ^2H-$CH_{4,w0}$ using the 'best estimate' for δ^2H-H_2O, as described above; 2) δ^2H-$CH_{4,w0}$ using only sites with measured δ^2H-H_2O. Both of these approaches indicated a step-wise linear relationship. However, the two relationships generated were not consistent in the breakpoint, and the linear relationships were not all statistically significant. Given this

result we agree that it is prudent to focus less on methanogenic pathway as an explanation of residual variability in δ^2H-CH$_4$, and instead discuss the complex interrelationship of multiple variables and processes that can influence δ^2H-CH$_4$, namely i) methanogenic pathway; (ii) methane oxidation; (iii) isotopic fractionation due to diffusion; and (iv) differential thermodynamic favorability of methanogenesis, or differential enzymatic reversibility.

The authors in their revision should be careful in the value of thinking about a$_c$ for the following reasons: some of the literature generating a$_c$ relies on assumption of differences in methanogenic pathway interpreted from differences in δD-CH$_4$, but there is competng evidence δD-CH$_4$ cannot be interpreted in this way (so a$_c$ using a$_c$ to infer methanogenic pathway in δD-CH$_4$ when δD-CH$_4$ has been used to infer methanogenic pathway becomes a circular, self-supporting and flawed approach).

We do not agree with the reviewer's contention of circular reasoning here. While both α_c and δ^2H-CH$_4$ have in the past been used to infer methanogenic pathway, the use of α_c is primarily based on theoretical predictions of fractionation factors for these pathways, and to our knowledge has not been 'validated' via analysis of δ^2H-CH$_4$. There is evidence from culturing studies (e.g. Valentine et al., 2004;Penning et al., 2006a), and from studies that isolate specific pathways in the environment (e.g. Penning et al., 2006a,b; Galand et al., 2010), that α_c varies in relation to differences in methanogenic pathway.

However, we do note that other variables related to methanogenesis have the potential to influence α_C, including enzymatic reversibility and the thermodynamic favorability of methanogenesis, as well as diffusive fractionation and methane oxidation. In addition, sources and sinks of CO$_2$ in natural environments that are independent of methanogenesis will also influence measured α_C. We discussed this possibility in the original manuscript (Lines 510 to 518), but we will give more emphasis to this in the revised manuscript. See Planned Major Revision 4 above.

To help here I would advise the authors to consider Waldron et al 1998 (Geomicrobiology, 15, 157-169), which contributes to the in-vitro line in Waldron et al 1999, but the authors do not cite so I am unsure if they are aware of the detail in this.
Here dominance of methanogenic pathway was changed in mixed culture (as would be found in the field) incubations, and δD-CH$_4$ monitored with time – so not just one measurement as may be misinterpreted from Waldron et al 1999. Except for one measurement broadly within analytical uncertainty, δD-CH$_4$ remained constant. However, δ$_{13}$CH$_4$ did change and consistently with fractionation ranges for the methanogenic pathways thought to be dominant (as assessed from independent measurements of substrate turnover). I advise the authors to consult Waldron 1998 for two reasons:
1. The authors approach in the bgd paper to draw on δD-CH$_4$ to represent differences in methanogenic pathway would be stronger if they can provide an explanation for the constancy in δD-CH$_4$ while δ$_{13}$CH$_4$ changes.

This is an important study and we appreciate the reviewer highlighting it. We will certainly cite and discuss it in the revised manuscript. The finding of constant δ^2H-CH$_4$ is intriguing. Recent pure culture studies have clearly shown that acetoclastic methanogenesis differs in hydrogen isotope fractionation from hydrogenotrophic methanogenesis under the same conditions (i.e. Gruen et al., 2018), implying that acetate-methyl hydrogen does not fully

exchange with water during methanogenesis. Therefore we infer that the effect observed in Waldron et al., (1998) likely results from hydrogen isotope exchange with water during fermentation of acetate from butyrate or other substrates. The constancy of the δ^2H-CH_4 would therefore imply that the isotopic fractionation of H-exchange between water and the acetate methyl group effectively compensates for the difference in hydrogen isotope fractionation between acetoclastic methanogenesis and hydrogenotrophic methanogenesis. Clearly, this is an interesting result that merits further study to resolve with the results of pure culture experiments. However, we do not feel that this study on its own negates the potential for differential net hydrogen isotope fractionation between acetoclastic methanogenesis and hydrogenotrophic methanogenesis.

As discussed above in Planned Major Revision 4, in the revised manuscript we will focus less on the role methanogenic pathway in controlling δ^2H-CH_4, and emphasize the complexities induced by multiple mechanisms influencing hydrogen isotope fractionation during and after methanogenesis.

2. Waldron et al can also be used to calculate a_c (both from CO_2 and from estimated substrate composition). a_c CO_2-CH_4 generates values of 1.057 for the period when CO_2 reduction is considered dominant (i) and 1.055 when acetoclastic methanogenesis is considered dominant (ii). These are very similar and it would be valuable to understand how the authors interpret this when they infer much wider ranges in a_c. For clarity $\delta_{13}CO_2$ and $\delta_{13}CH_4$ respectively for (i) were -8.3 ‰ and -62‰ , and for (ii) were 1.55‰ and -47.5‰

This is also an interesting result. We think there could be some complicating factors that influence δ^{13}C-CO_2 in this study in particular. We note that the headspace concentration of CO_2 decreased through the experiment, which would not be the expected stoichiometric result of a net shift from hydrogenotrophic to acetoclastic methanogenesis. This suggests that there were additional sinks of CO_2 in the experiment that became more prevalent as the experiment proceeded, and this may have led to the observed enrichment in δ^{13}C-CO_2. In particular we are curious about the possible role of increased homoacetogenesis, although this is difficult to evaluate based on the results of the study.

Overall, we do not think this finding necessarily negates the use of α_C as an indicator of differences in methanogenic pathway, which is supported by other studies (e.g. Penning et al., 2006a,b; Galand et al., 2010). But it does point to the potential for other variables to complicate the relationship between α_C and the relative proportion of different pathways. In our revised manuscript we will highlight these complications, including citing this paper. See planned major revision 4.

Abstract: Is clear and summarises the paper but projects a future methane emissions scenario (L25-26) before the modelling and assessment of how well this approach can reconstruct current estimates (L27-30) and this seems in the wrong order to me, given the former has a reliance on the latter. Further, the abstract does not acknowledge this research is augmenting the research that historically first documented the global relationship between δD-CH_4 and δD-H_2O easily addressed for example by changing L12 to 'We have refined the existing global relationship between δD-CH_4 - δD-H_2O by the compilation of a more extensive global dataset…."

We agree with these suggestions, and we will revise the abstract accordingly. We note that based on the suggestions of the other reviewers there will be other changes to the abstract, including a modification of the description of the upscaling component of the manuscript.

L28: The authors postulate the mismatch is dependent only on the work of others (emission inventories, etc) and not possibly an error in their approach. Scientifically this is not correct – both 'sides' could have errors.

We will change this as part of revising the upscaling results and analysis.

L19: results do not imply; one interprets data to generate a 'result'.

We will change the language here to make clear this is an interpretation of the data.

L22: high (more ^{13}C-enriched) in rivers and bogs - this is the dataset that has more $\delta D-H_2O$ projected, so is this an artefact of the modelling than a real biome-specific difference?

We are not sure what the reviewer means by 'this is the dataset that has more d2H-H2O projected.' 81% of bog sites have δ^2H-H_2O measurements, while 37% of river sites have this measurement. For the dataset as a whole the percentage is 48%.

As discussed above (Planned Major Revision 3) we carefully assess the use of modeled precipitation δ^2H-H_2O, and find it does not have a major impact on the regression relationship between δ^2H-H_2O and δ^2H-CH_4. Therefore we do not believe this result is an artifact. Regardless, we will revise our analysis of differences by ecosystem using the 'best-estimate' δ^2H-H_2O and resulting $\delta^2H-CH_{4,w0}$.

L27: integrated (by mass balance) not combined (which is used when sources are added) – which I know the authors have done (L204) but the descriptor is incorrect here.

We will change the wording here.

L36: I think the following references is missing: Variability in Atmospheric Methane From Fossil Fuel and Microbial Sources Over the Last Three Decades. / Thompson et al: Geophysical Research Letters, Vol. 45, No. 20, 28.10.2018, p. 11499-11508 (and I invite the authors to wonder if also some of the work from the Royal Holloway group should augment L47-51)

We thank the reviewer for this suggestion, and we will add the suggested reference and also see if other work from the Royal Holloway group would be good to add in the suggested part of the introduction.

L59 & L83 Citations are given in chronological order of 1999b and 1999a which seems not typical convention to me (uncertain of the referencing convention for BG but for example the two references for Walter K are not in chronological order in the reference list so the in-paper citations would not be b then a due to this convention in the reference list?)

We were relying on EndNote for citation management, and there may have been some errors with the citation format in the software. We will check this carefully.

L68: Logic only follows that impact on $\delta^{13}CH_4$ can affect geographic provenancing if reader knows it can also affect $\delta D-CH_4$, so does this need to be made explicit?

We do not understand the reviewer's comment here. The hypothesized geographic variation in $\delta^{13}C-CH_4$ is independent of variation in $\delta^{2}H-CH_4$, as they are controlled by different mechanisms. We will make this clearer in the revised text.

L70: this implies that different ecosystems have different methanogenic pathways. More accurate text would be "differentiated geographically based on ecosystem differences in the relative strengths of different methanogenic pathways and $\delta_{13}C$ of source organic matter" (as per the introduction of the Ganesam paper). Noting relative strengths is important, as a common mistake propagated in the literature and again here (L???) is to assume methanogenesis proceeds by one methanogenic pathway only – this would be rare, with field-based methane production contemporaneous from CO_2 and acetate, and varying temporally in strength as input of fresh OM changes seasonally (or not).

The reviewer raises an important point here, and we will make it clear that the difference is in the relative strength of the pathways operating in difference ecosystems, and not different pathways per se.

L84-85: sounds a bit defensive? How about "We have advanced existing compilations of freshwater $\delta D-CH_4$ by 1,2,3 …? I would remove significantly (statistical connotations) and just say larger as the number speak for themselves.

We agree with this suggestion and will make the proposed change.

L91: The aims are clear (good) but 'then' and 'potential' not needed – the latter as embedded in implications that there is a potential for impact

We will make this change

L106 & L117, 9L206 and possibly elsewhere): small w for where, as this follows from an unfinished sentence in both cases with the equation used in between

We will make this change

L136: the five ecosystem categories are not clear from this sentence: 'lakes' and 'rivers' and then there are five wetlands listed. Further, it is debatable that floodplains are aligned with rivers as CH_4 production would only occur when sediments are deoxygenated from standing water. So I would say more with ponds as the recession of water can be slow and could be like a pond drying in some situations. Noteworthy here is that gas loss from rivers is velocity dependent (see Long et al (2015) Hydraulics are a first order control on CO_2 efflux from fluvial systems Journal of Geophysical Research – Biogeosciences, 120, (doi:10.1002/2015JG002955), and similar references. This will also be the case with methane – possibly more so as insoluble, and may cause an isotope fractionation independent of degassing, and may also be a reason the Amazon rivers in Fig. 5 plot differently.

We note that there was an error in this text and it is actually six categories. We will add a numbered list to make this clearer. We note that essentially all of the river data come from floodplain lakes or deltas, with one exception, and most of the data are from the Amazon. We think it is valid to continue to differentiate these environments from other lakes and ponds, since they are different from typical lakes and ponds in a number of ways (overturning and redox regimes, nutrient inputs, dynamics of gas loss and hydraulics). We will make the nature of the river/floodplain/delta sites clearer in the revised manuscript.

The reviewer mentions an important point about degassing dynamics in rivers, and we will include this in our discussion of differences between environments in Section 4.4.

L139: Similarly, I question the scientific integrity in lumping lakes with rivers here – gas loss from river systems is controlled by hydrological processes primarily and there could be fractionations during emission from lotic systems that are different to lentic systems where diffusion and wind of lake thermal orographic processes control turnover. This starts to become important where these mean sources are used to simulate a resultant atmospheric composition e.g. L227. Thus, the authors should think about how to provide added confidence of the robustness of their catergorisation.

See above. We are primarily analyzing floodplain lakes and deltas in the river category. We will make this clearer in the revised manuscript.

However, the basis of this categorization is essentially to align with flux inventories (Saunois et al., 2020), which specifically define an 'inland water' category that includes rivers and lakes, as well as reservoirs. We keep to this categorization in order to be able to compare with the flux estimates. We do clearly discuss possible differences between the river and lake sites in the Discussion.

L145-148: Such categorisation is good, and the open access data set is very welcome. This categorisation relies on the integrity of the interpretation, but this integrity is important as the data analysis relies on this. With 131 sites it is impossible for the reviewer to know each site and so as a 6 check I can only look at my own data: L61 in the excel files. These methane samples were collected in-situ from porewater diffusing into samplers embedded in the peat (the GBC abstract notes in-situ and the methods clarifies at depth sampling) so I would classify as more aligned with dissolved porewater than diffusive flux (which is normally associated with the potential for oxidation and change in δ _v_a_l_u_e_s_ _). Further I comment in the GBC paper there is a dynamic zone and interpret that is the section from which gas can be emitted. Mean $\delta D\text{-}CH_4$ here is -332 ± 17‰, more depleted the -294 ± 39‰ used in the table and subsequent data analysis. Thus, some feedback from the authors in the revised manuscript that their interpretations are not sensitive to the variation their interpretation of environment and which data to use would be valuable.

The reviewer raises a valid point about the complexities of each site. This complements the comments of reviewer 3 about the validity of including data from deep peat samples. As discussed in our response to reviewer 3 we think this is a valid concern and we have decided to limit our data from peatlands to the uppermost 50 cm. See Planned Major Revision 1. This coincides approximately with the dynamic zone mentioned here.

We note that peatlands (bogs and fens) were the only environments to be sampled on depth gradients, and that similar issues are unlikely to affect the interpretation of data from other ecosystem categories. We will change the entry in the Table for this Waldron et al. 1999b to Dissolved-Pore Water. Since we group diffusive flux and dissolved pore water, this distinction does not make a difference to our analysis.

L152 – typically small – as this manuscript relies on several source of data estimation (here, δ_2H_2O, it would be good to provide estimates as to what the maximum is this would manifest in δD (recognising that it changes with resolution and scale of figure and so this is challenging, but saying small is insufficient).

We will provide a quantitative estimate of the likely error produced by this analysis, as a percent error, and then translate that to δ^2H values.

L177: the authors need to unpick for the reader the statement more as they have with L179 onwards. I am thus left to interpret the reasoning. I assume it is based on considerations that methanogenic pathway influences δD-CH_4? If so please see earlier substantive comments on this and decide whether to proceed in the revised manuscript.

Based on the reviewers comments here and below we plan to thoroughly revise section 2.3.2 to discuss in more detail how δ^2H-CH_4 is predicted to vary with carbon isotope composition, and the different processes that would influence that co-variation, considering an expanded set of processes. See Planned Major Revision 4 above.

L200: Clarify where the flux estimate comes from at this point – I presume from Saunois et al as in L209, but this should be clarified when first introduced. I am not expert enough to judge if the methodology for the bottom up flux section is sound, but it seems reasonable to me.

We will clarify the source of the flux estimates earlier in this section. It is indeed Saunois et al., 2020.

L 267: given the statistical approaches such as Monte Carlo bootstrapping used with the flux estimate section previously I would have expected more rigorous comparison should be undertaken here to show if there is a statistical offset between measured and predicted δD-H_2O than relying on descriptors of "generally good agreement" and using RMSE. The RMSE is a red herring if the lines generating 19 and 23 ‰ do not overlap - ?

As discussed above, we will provide a more detailed analysis of the comparison of measured and modeled δ^2H-H_2O. See Planned Major Revision 1.

Fig 2: Should the predicted (postulated and therefore dependent) not be regressed onto the measured (the true field value, so measured and independent and as a control of δD-CH_4 the one to get as close to the true value as possible)?

This depends on the goal of the regression. In this case we are attempting to develop a regression relationship that predicts measured δ^2H-H_2O as a function of modeled δ^2H-H_2O, in order to use this as a predictive tool for sites without δ^2H-H_2O measurements, and to assess the goodness of fit. Therefore it makes more sense to have measured δ^2H-H_2O on the y-axis in this case, and we will keep this orientation for the revised Figure 2.

Fig 3B: this needs revisited once the δD-CH_4 -δD-H_2O predicted data has been removed as described above. There may still be an inland water specific difference here, but again that this may not be controlled by anything more complex than lentic and lotic freshwater systems having generalised differences in gas transport mechanism (ebullition or diffusion). These would be influenced by atmospheric and sediment interface boundary layer dynamics, transit time, depth of oxidative zone, lake stratification, and surface roughness, with the latter in turn influenced by wind speed, depth of water, and river flow velocity, slope. In other words, considerable methane isotope fractionation (enrichment) is possible, or not.

Based on our revised analysis, we find that we cannot detect a significant difference in the regression relationship between inland waters and wetlands (See Planned Major Revision 3

above). The difference inferred in the original manuscript is likely partly a result of the hydrological differences in these environments, and resulting differences in the regression of modeled vs measured δ^2H-H_2O (see Planned Major Revision 2). Therefore we will revise the manuscript here to reflect this revised understanding.

Fig 4. It is good to see this plotted but not surprising given δD-H_2O varies with latitude and δD-CH_4 varies with δD-H_2O. The same difficulties in estimating field δD-H_2O from modelled δD-H_2O are evident when considering δD-CH_4 as a function of predicted δD-H_2O. The authors need to note here that there may be an imbalance of where methane is sampled from globally and so if more measurements existed from the higher latitudes then there may be as much scatter as with the lower latitudes.

We are glad the reviewer agrees with us on the utility of plotting the data in this way. We did note the uneven geographic distribution of data at several points in the manuscript, but can further emphasize the likelihood of similar scatter at all latitudes with more sampling. We are assuming the reviewer meant to say there is greater scatter at high latitudes, which is what we observe.

Section 3.4 jumps to something completely different with L313 “shifts to being controlled by changes in methanogenic pathway to being controlled by ….”. There has not been clear discussion from the authors to date they are considering changes in methanogenic pathway of δD-CH_4 so this seems out of context. And yet L317 goes on to consider this in more detail. The key message in the Waldron et al 1999 paper is that considering methanogenic pathway a control on δD-CH_4 is misplaced and that “that 50% of the variation in natural δD-CH_4 samples can be explained by δD-H_2O, with isotopic fractionation post-production, or mixing with gas already fractionated likely responsible for most of the noise in the natural system”. The analysis prior to section 3.4 may be more likely to support this interpretation than refute it, particularly when the data in Fig. 3.2. is appropriately compared (as described earlier), and so now considering data as a function of methanogenic pathway seems to be ignoring this. Indeed the authors observe they find no relationship between $\delta_{13}C$-CH_4 and δ_2H- $CH_{4,W0}$ which would be expected if δ_2H-CH_4 was influenced by methanogenic pathway as $\delta_{13}C$-CH_4 is (Fig. 5a). Thus, the authors should not make clearer statements such as L312 of “shifts from being controlled by variation in methanogenesis pathway” are inferred controls.

This is clearly a key point of concern for the reviewer, and we understand the reservations about inferring that variability is a function of methanogenic pathway. However, we think it is unlikely that all of the remaining variability not explained by δ^2H-H_2O is controlled by “isotopic fractionation post-production, or mixing with gas already fractionated”. First, it is important to be clear about what these post-production processes are. To our knowledge there are two key post-production processes that can affect methane isotopic composition: methane oxidation (either aerobic or anaerobic) or isotopic fractionation caused by diffusion. We are unaware of other important processes. Both of these processes would be likely to lead to higher δ^2H-CH_4, and lead to positive co-variation with $\delta^{13}C$-CH_4. Oxidation will also lead to negative co-variation with α_c, because CH_4 is invariably oxidized to CO_2, leading to a smaller isotopic difference between these gases. However, diffusion would lead to positive co-variation with α_c, because diffusion is expected to have a smaller isotopic effect on CO_2, as a result of smaller mass difference between CO_2 isotopologues (Chanton, 2005). Mixing effects will depend on the mixing end-members. Unless there is a large proportion of non-microbial methane present, which we argue is unlikely in most circumstances, mixing will not alter the overall isotopic signature of microbial methane in

the ecosystem. It is possible to have mixing with 'gas already fractionated', but in this case the underlying fractionation is the key process controlling the isotopic composition of the resulting gas, and again to our knowledge this would have to be the result of oxidation or diffusion.

It is unclear on what basis Waldron et al (1999) ascribed the remaining ~50% of variability in δ^2H-CH_4 to these post-production processes, and we would argue that this assertion is untested.

We do agree that our focus on methanogenic pathway did not include other plausible mechanisms for co-variation between δ^2H-CH_4 and α_C. We plan to expand our discussion to take other processes, namely (i) diffusion and (ii) differences in enzymatic reversibility, into account.

Figs. 5b=c. The uncertainty around what a_c should be for different methanogenic pathways has been described earlier in this review. But additionally, although breakpoint analysis was used, there is a high dependence in this on data set that has enriched δ_2H-CH_4 to generate opposing trends. The eye is drawn by the projected pathways, but if these was not included as we cannot be sure it is oxidation[1] and all the remaining data was considered in a weighted regression would there be trends?

[1] If the high δ_2H-CH_4 is from the Amazonian rivers, there are shales in this basin that fuel C cycling (Vihermaa et al) and this could be thermogenic: δ_2H-CH_4 is also consistent with this. Vihermaa L.E., Waldron S. , Garnett M.H., and Newton J. (2014) Old carbon contributes to aquatic emissions of carbon dioxide in the Amazon. Biogeosciences, 11, 3635-3645. (doi: 10.5194/bgd-11-1773-2014).

We acknowledge concerns about the 'predicted trends', both by reviewer 1 and 2, and therefore we will remove them from the revised manuscript. We will instead focus on the patterns of co-variation, and potential explanations for them. As discussed above (Planned Major Revision 4) we will focus less on methanogenic pathway, and increase our focus on other mechanisms.

We agree that the one outlying point with very high δ^2H-CH_4 (and δ^{13}C-CH_4) is questionable, and may be thermogenic methane. It is indeed from the Amazon. We will therefore remove this from our dataset and repeat the analyses (Planned Major Revision 1)

As noted above, weighted regression is leading to biases in this analysis, and is not generally preferable to unweighted regression (Fletcher and Dixon, 2012), and therefore we are proceeding with unweighted regression in the revised manuscript.

It is remarkable Fig 7 is so consistent – this is very interesting. Is it what we would expect?

We assume the reviewer is referring to Figure 7B. This is not necessarily what we would expect based on other studies. We have already provided some discussion of this in section 4.4, but plan to revise this in response to questions from reviewer 2, especially focusing on possible biases in the δ^{13}C-CH_4 dataset.

L370 discussion is over-interpretations given the differences between sites are not statistically significant. It would be ok to say the prevalence of more depleted CH_4 is greater in the ecosystems sampled but for example this could represent accessibility of field sites, or differential investment into research measurements in these areas, than group compositional differences per se. Ecosystem types are not evenly distributed by latitude (L370) – nor is resource for investment in field research with tropical regions of the Earth lacking measurement due to access or financial constraints – we need to start recognising what we have not measured is as important as what we measure.

We agree that this analysis is preliminary given the small sample sizes for each ecosystem. We tried to emphasize this in the original manuscript, and noted that the possible differences represented hypotheses that merited further testing. But we will further emphasize this. We will also note clearly here that more investigation of tropical ecosystems is especially important.

Fig. 10 is tiny and needs to be bigger

We are going to revise Figure 10 to simplify it based on comments from reviewers 2 and 3 on the upscaling exercise. We will reduce it to a single panel (equivalent to Figure 10C), which will make it more legible.

L426 "roughly as strong a predictor". Too big a leap: explain how – from ice core gases "roughly is a colloquialism"

We do not fully understand this comment, but we agree that this language is imprecise, and we will make a more quantitative statement. We are not sure what the reviewer is saying about ice-core gases.

L487 – as noted earlier, the paired measured values plot on Waldron et al 1999 In-vitro line, consolidating further the significant of this line. Please acknowledge this.

We are assuming the reviewer meant the in-vivo line here. As discussed above, we will provide a more thorough comparison of the Waldron in-vivo line with the results of this study, and will represent this in the revised discussion. Our analysis (Table R2) shows that the paired measured values do not plot on the Waldron et al (1999) in vivo line, and have a flatter slope. This difference is not significant based on ANCOVA, but we infer that the larger dataset implies a flatter global slope.

L508 – in the revised manuscript please detail the % variation explained by δD-H_2O and then additionally by a_c should this prove to still be important

We will perform a re-assessment of this analysis, and include it if α_c is still a significant predictor of δ^2H-CH_4. If so, we will clearly detail the % variation that is explained by these two variables.

L510 – this is the crux of what is new to explore in isotope biogeochemistry of methane and also the role of methanol substrates.

We agree with the reviewer that CH_4 isotopic variability related to enzymatic reversibility is an important topic, and based on other comments we will expand discussion of this here, as well as in the methods and results. At this point there is little we can say about methanol substrates, but we will mention it briefly as another variable that merits consideration.

L519 – same comments as before about is there really a relationship, but why more points classified as oxidised with this pairing than with a_c?

This is an interesting question, and we don't know the answer. We would speculate that it is because sources of CO_2 can be very variable, and this may be adding noise to Figure 5C that is not present in figure 5B. As discussed above we will be substantially revising this section, but will point out the differences in these two plots as an indicator of complex interactions of different biogeochemical processes.

L551- Much of 4.31. is repeating statements first described in Waldron et al 1999 section 1.1., paragraph starting "In addition…" but this is not referenced and as written implies the authors are the primary source of this thinking. This is not the case and should be referenced appropriately to indicate this was first noted 20+ years ago.

We regret that we did not acknowledge the earlier statement of these ideas. We will thoroughly revise this section to provide credit to Waldron et al., (1999) for the ideas that are presented there.

L564 – please note pure cultures are not representative of the field processes of methane production and thus the batch cultures and other experimental data collated in Waldron et al 1998, 1999 are. This is not clear from the statement.

We agree that pure cultures are not representative of methanogenesis in the environment. We are not sure that batch cultures or incubations are truly representative either, in that they do not necessarily fully represent the processes occurring in natural environments, but agree they are clearly a closer approximation than pure cultures. We did try to make this distinction clear in the original manuscript, but will further clarify in the revised manuscript. However, we do feel that inferences from pure cultures are important for understanding the more complex processes that occur in batch cultures or natural environments. For example, differential hydrogen isotope fractionation between water and methane by methanogenic pathway has been clearly observed in pure culture experiments (i.e. Gruen et al., 2018). This implies that acetate methyl hydrogen is not fully equilibrated with water during the methanogenesis reaction itself.

L569, please reverse the order of the references or remove Whiticar 1999. The Waldron 1999 paper is the one that is particularly focussed on the global relationship between $\delta D\text{-}CH_4\text{-}\delta D\text{-}H_2O$, and constructs the first global relationship, which this paper finds with new data is similar. This gives appropriate credit to the conceptual understanding. The Whiticar paper coplots $\delta D\text{-}CH_4\text{-}\delta D\text{-}H_2O$ but does not assert that " $\delta_2H\text{-}H_2O$ is a primary determinant of $\delta_2H\text{-}CH_4$ on a global scale", rather the focus is on the interpretation of how $\delta_2H\text{-}CH_4$ reflects methanogenic pathway or marine vs. freshwater.

There seems to be an error in the page numbering, and we are not sure which citation the reviewer is referring to. However, we will make clear in the revised manuscript that Waldron et al., (1999) first proposed and found evidence for the global relationship between $\delta D\text{-}CH_4\text{-}\delta D\text{-}H_2O$.

To conclude: this has been an uncomfortable review for me to undertake as my position of not anonymising the review puts me up for public scrutiny, and a misinterpreted that I am trying to defend my work and am unwilling to accept an addition to this. This does not represent my professional scientific principles, I would urge the authors to accept this is not the case - indeed in the 1999 GCA paper I welcome refinement of my work. However, the authors have still not presented here compelling evidence that δD-CH_4 can represent well different methanogenic pathways and so the reliance of this in the manuscript I find troubling. I consider the a_c approach may be valuable in helping constrain the signal in δD-CH_4 that is not defined by δD-H_2O, but the current manuscript is not constraining uncertainty sufficiently and the approach is therefore flawed. I would urge the authors to find a way to better constrain projected δD-H_2O and revisit this, or work with only measured data and revisit this. Their refined analysis should undertake rigorous statistical comparison with the existing field δD-CH_4 -δD-H_2O relationship from Waldron et al 1999 to say whether it is different (although the new larger dataset will likely be a more representative relationship that the community can go forward with), and adopt a parsimonious interpretation of variation within the data set, as that is least likely to induce an erroneous interpretation. The biome specific considerations and upscaling should also be revisited if the removal of biased and inaccurate data pairings changes the source bulk compositions, and further thought should be given to the basis for source differentiation based on scenarios of methane production and loss in this upscaling.

Once again, we regret that this has been an uncomfortable review process. We appreciate the frank and detailed signed review, and the collaborative nature of the comments. We do not agree with all arguments made by the reviewer, but do agree with many of the suggested improvements to the manuscript, and will make these changes in a thoroughly revised manuscript. See the Planned Major Revisions above for a summary of these changes.

We believe that these changes will address the reviewer's concerns and will greatly strengthen the conclusions of the manuscript.

References Cited:

Chanton, Jeffrey P. "The effect of gas transport on the isotope signature of methane in wetlands." *Organic Geochemistry* 36.5 (2005): 753-768.

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.

Galand, P. E., Kim Yrjälä, and Ralfi Conrad. "Stable carbon isotope fractionation during methanogenesis in three boreal peatland ecosystems." *Biogeosciences* 7.11 (2010): 3893-3900.

Gruen, Danielle S., David T. Wang, Martin Könneke, Begüm D. Topçuoğlu, Lucy C. Stewart, Tobias Goldhammer, James F. Holden, Kai-Uwe Hinrichs, and Shuhei Ono. "Experimental investigation on the controls of clumped isotopologue and hydrogen isotope ratios in microbial methane." *Geochimica et Cosmochimica Acta* 237 (2018): 339-356.

Penning, H., Claus, P., Casper, P., & Conrad, R. (2006). Carbon isotope fractionation during acetoclastic methanogenesis by Methanosaeta concilii in culture and a lake sediment. *Applied and Environmental Microbiology*, *72*(8), 5648-5652.

Penning, H., S. C. Tyler, and R. Conrad. "Determination of isotope fractionation factors and quantification of carbon flow by stable carbon isotope signatures in a methanogenic rice root model system." *Geobiology* 4.2 (2006): 109-121.

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.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., and Patra, P. K.: The global methane budget 2000–2017, Earth System Science Data, 12, 1561-1623, 2020.

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.

Valentine, D. L., Chidthaisong, A., Rice, A., Reeburgh, W. S., & Tyler, S. C. (2004). Carbon and hydrogen isotope fractionation by moderately thermophilic methanogens. *Geochimica et Cosmochimica Acta*, *68*(7), 1571-1590.

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