Response to Editor

We appreciate the insightful comments from the Editor and the reviewers. We have taken serious consideration of these comments and we provide a revised version of our manuscript. We identified that most of the comments were focused on clarification issues for the application of the tuHLs approach.

We have edited figures and made substantial edits in the text to improve clarity. The figures that were in the Appendix have been moved to Supplementary Material as we believe that section is more appropriate for that information. We hope you share our enthusiasm for this study and consider it for publication in Biogeosciences.

Thank you for your clear responses to the reviewer comments. I now invite you to submit your revised manuscript. In addition to the reviewer comments, please also address the following question:
Your analyses indicate the advantage of the tuLHs approach over the FTS approach. In your study, you can conclude this because you have high frequency data available to compare both approaches. As indicated by you and by the reviewers, the optimal sampling design will likely differ between years, sites, ecosystems, ...

Response: We appreciate the support from the Editor.

What is not clear to me is how the tuHLs approach can then be used in experiments where high-frequency data are not available.

Response: In theory, the tuHLs approach can be applied to any length of a time series, and the results will reflect an optimization scheme based on the probability distribution and temporal information of the time series used in the tuHLs. The more available information, the more accurate an optimization will be to reflect the "reality" of the physical world.

The tuHLs approach can be applied to experiments where no high-frequency data is available. Consequently, the tuHLS will provide information about which subsamples are the most relevant to reproduce the probability distribution and temporal dependency of the information. What is essential is to question if a few measurements from an experiment represent the reality of the physical world because if limited information is available, then the actual probability distribution and temporal dependence of the phenomena could be an unknown-unknown. In other words, with few measurements, we may not be aware, and we will not be able to know which is the actual probability distribution and temporal dependence of the studied phenomena. To address this challenge, we tested the tuLHs approach with high-temporal frequency information representing the probability distribution of multiple soil GHG fluxes at the daily time-step across a calendar year.

Action: We have added and edited this discussion in lines 402-413 and 449-459.

That are the situations where manual sampling is needed, but I am not yet convinced that the tuHLs approach really provides a more reliable sampling design then. Isn't there a great risk of missing out on important sampling moments because of false assumptions?

Response: We fully agree with this comment. The tuHLs approach cannot tell you which is the ideal sampling design if there is no available information (see lines 402-413). The tuHLS only identifies subsamples to represent the probability distribution and temporal dependency of the data used to inform the tuHLS. What we think is important is that when only a few manual measurements are possible, we advocate performing these measurements in an optimized way informed by a tuHLS approach. The optimization can be performed using high-temporal information of the actual target variable, forecast scenarios, or a proxy (417-420). We postulate that a few measurements properly distributed across time provide better agreement with information derived from automated measurements than measurements performed using an FTS approach.

Action: We have edited the discussion in lines 391-401, 402-413, 422-443. Furthermore, we have added a comment that co-location of manual and automated measurements is important as suggested by the reviewers (lines 421-423)

What are the criteria that need to be met to obtain a reliable sampling design based on tuHLs?

Response: We clarify that the tuHLs is a statistical method for generating subsamples of parameter values (i.e., soil GHG gas fluxes in this case study) to reproduce the probability distribution and the temporal dependence of each original time series of GHG fluxes (see lines 91-92). The tuHLs only provides information about which subsamples are the best (i.e., provides an optimization) to represent the probability distribution and the temporal dependency of the data used to parameterize the tuHLs. If the input data is biased, then the tuHLs results will be biased. If the input data represents the “true” probability distribution and the temporal dependency of the phenomenon of interest, then the tuHLs results will aim to approximate that information. To address this challenge, we tested the tuLHs approach with high-temporal frequency information representing the probability distribution of multiple soil GHG fluxes at the daily time-step across a calendar year.

Action: We have clarified the purpose of the tuHLS in lines 91-93.
We have edited the discussion in lines 391-401, 402-413, 422-443.

Response to reviewer #1

Comment: The authors present an interesting and novel evaluation of the bias inherent in sampling strategies at relevant timescales to capture estimates of GHG fluxes. As the authors point out, the ability to measure all three GHGs (CO2, CH4 and N2O) simultaneously is now more common and has advanced understanding of the complex drivers of these important gases. Discrete, manual flux chamber sampling in which all three GHG fluxes from soils as tends to be the most common method, with good spatial but limited temporal representation. For convenience and cost effectiveness, discrete sampling strategies simultaneously measure all three GHGs, however, this strategy relies on the underlying assumption that each GHG responds similarly to biological and physical drivers at these same fixed temporal steps. Systems that automated the GHG flux sampling process are becoming more common but are still limited in application due to the costs associated with them, limiting spatial representation but providing high temporal sampling frequency. Automated, continuous measurement of all three GHG fluxes as high temporal frequency is better able to capture their temporal response to drivers that may not be co-occurring and offer a better understanding of the underlying drivers of each GHG flux as well as estimates of annual GHG budgets.

In this work, the authors aim to address how discrete manual flux sampling strategies in which all three GHGs are measured simultaneously at fixed temporal stratification (FTS) may violate the underlying assumption of co-occurring responses at temporal timesteps and bias the interpretation and understanding of each GHG. The authors utilize a dataset in which all three GHGs were sampled at hourly timesteps via an automated sampling system, for one year (Sept 2014-Sept 2015) in a temperate forest. By extracting subsets from this dataset at discrete timesteps, the authors create a series of examples of FTS at common sampling strategies (12, 24, 48 sample dates per year). The authors then utilize a novel technique, temporal univariate Latin hypercube sampling (tuLHs) to subsample the same annual dataset at the same temporal frequency (12, 24 and 48 annually). tuLHS optimizes the temporal selection of these subsets to reflect the same statistical properties and temporal patterns specific to each individual GHG reflective of the yearly GHG dataset. The authors argue that optimizing the sampling strategy for each GHG (tuLHS) is needed to avoid bias that may be inherent in FTS, particularly when the annual sample size is small (for example monthly, 12)

Response: We appreciate the detailed summary of this study by the reviewer.

Action: We have revised the text to improve clarity and addressed the comments from two reviewers.

Comment: The authors carefully show that measuring GHGs at common FTS biases estimates at annual timesteps, for this specific dataset, and that the tuLHS method produces a more representative reflection of yearly patterns of GHG fluxes providing a proof of concept for this novel method.

Response: We appreciate the supportive comment to recognize the novelty of our work.

Action: We have revised the text to improve clarity and addressed the comments from two reviewers.

Comment: This work is useful and informative and will provide a method (tuLHS) to aid researchers when developing a discrete manual sampling strategy for each GHGs. My concern is how easily this method is implemented broadly, either across years at the same site or how representative a tuLHS derived sampling strategy may be across similar ecosystems.

Response: We appreciate the supportive comment to recognize the novelty of our work.

We clarify that the main goal is to introduce the application of the tuLHS and show that the underlying assumption that each GHG responds similarly to biological and physical drivers may not be universal and should be tested. We provide a case study to introduce tuLHS as a proof-of-concept to show how the method works and to show that the general assumption that sampling at the same time all GHGs may not be appropriate to represent the probability distribution nor the temporal dependency of each GHG.

Action: We have edited the manuscript introduction and discussion to make it clear that this is a proof-of-concept to provide insights for monitoring purposes (e.g., lines 30-32; 83-93; 346-348; 381-383; 469-475). We have added a paragraph in the discussion section addressing the applicability of this study and the limitations of this proof-of-concept manuscript (lines 402-413).

Comment: The authors acknowledge that the tuLHS method needs to be site specific, but a minimum of 1 year of automated continuous GHG fluxes (one without large data gaps) is needed to determine the optimal sampling strategy for each GHG using tuLHs. This also assumes that one year is representative of annual and interannual variation in each GHG flux patterns. Although this may be sufficient for CO2, CH4 and N2O are more variable at sub-daily to annual timesteps. A strategy developed in one year, may not be appropriate for the following year, especially if there are shifts in climate. It would seem that multiple years of site specific automated GHG measurements would be needed to determine if there are any wide variations in the optimal sampling strategy under different climate conditions.

Response: We appreciate the insightful discussion provided by the reviewer. Here we provide responses on the main points.

1- Our goal is not to prescribe a universal sampling time for each one of the GHGs, but to introduce the tuLHS approach and show that sampling all GHGs at the same time using discrete measurements may result in bias estimates. This is because a fixed temporal sampling is not able to capture the probability distribution nor the temporal dependency of each GHG when compared with automated measurements.

2- In theory the tuLHS can be used with any length of a time series as the method aims to optimize a sampling that represents the probability distribution and the temporal dependency. We present our case study with a 1-year time series as an example but there is no specific requirement for a time series length. That said, the longer the time series the better and if multiple years are available, then (arguably) the optimized sampling design could be more representative of the natural variability of the ecosystem.

3- The tuLHS could be applied to sub-daily time series but this was not the goal of the case study as we focused on daily time steps to simplify the example and present the case study. That said, this is possible, and the method could shed light on how to optimize measurements in a sub-daily time scale.

4- We agree that a strategy developed in one year may not be appropriate for the following year. We clarify that the goal of this study is to introduce the application of the tuLHS and show that the underlying assumption that each GHG responds similarly to biological and physical drivers may not be universal and should be tested.

Action: We have edited the manuscript introduction and discussion to make it clear that this is a proof-of-concept to provide insights for monitoring purposes (e.g., lines 30-32; 83-93; 346-348; 381-383; 469-475). We have added a paragraph in the discussion section addressing the applicability of this study and the limitations of this proof-of-concept manuscript (lines 402-413). We also clarified that the year used for this study has a typical mean annual temperature and annual precipitation (lines 113-115).

Comment: Further, the tuLHS method may produce an optimal sampling strategy for each GHG, which logistically may be unreasonable to pursue given time, labor and cost constraints. To me this work highlights the need to either have an automated sampling system or co-locate automated and manual sampling strategies to truly capture temporal and spatial GHG fluxes from sites.

Response: We fully agree with this comment. First, co-location of automated and manual sampling strategies would be the best approach to capture the temporal and spatial variability of GHG fluxes from sites. Second, this study shows that the optimal sampling strategy is not to sample all GHG at the same time with a few discrete measurements. This is an important result because a few discrete measurements cannot reproduce the probability distribution nor the temporal dependency of the time series of GHG fluxes.

Action: We have added the reviewer's suggestion of co-locating automated and manual sampling strategies (lines 421-423)

Comment: As a "proof of concept" in this site-specific case, the authors clearly show that FTS does produce bias in magnitudes and temporal patterns compared to tuLHS, which optimizes the sampling strategy, when compared to a one-year automated GHG flux dataset. More analysis, at multiple sites and conditions, is needed to ascertain the broad applicability of tuLHS.

Response: We thank the reviewer for the support of this study. We agree that this method should be tested in different ecosystems.

Action: We have added a paragraph in the discussion section addressing the applicability of this study and the limitations of this proof-of-concept manuscript (lines 402-413).

Comment: I recommend minor revisions.

Response: We thank the reviewer for the support of this study.

Specific comments and questions:

1. Was Sept 2014-Sept 2015 a typical climate (temperature and precipitation) year at the site? Can the authors provide insight on how deriving a sampling strategy from one year, particularly if it is not a normal climatic year, and utilizing that strategy in subsequent years may impact results?

Response/Action: We clarified that the year used for this study has a typical mean annual temperature and annual precipitation (lines 113-115). We have added a paragraph in the discussion section addressing the applicability of this study and the limitations of this proof-of-concept manuscript (lines 402-413).

2. Do the authors think there was any influences in results due to missing automated GHG flux data, which appears predominately in the winter- early spring? It seems curious for N2O to have tuLHS select predominately in the fall/winter period as representative of annual N2O flux temporal and statistical characteristics.

Response: Unfortunately, missing data due to quality assurance/quality control, electrical power or mechanical failure are common in automated measurements. The tuLHS will optimize the sampling approach based on the statistical properties of the time series, and our assumption is that the data presented is representative of the statistical properties of the time series analyzed.

Action: We have added the explanation that …"For the case of soil N_2O fluxes, the variogram shows a constant temporal variability, that is there is no temporal dependence. Therefore, the optimized measurements are concentrated within the fall season due to their distribution probability (Fig. 1a)." (lines 388-391). We also discussed that the specific results could vary depending on the length of the time series because of how tuLHS works (lines 403-411).

3. Lines309-311: the results show that tuLHs provided closer estimates of cumulative sums and uncertainty ranges than FTS. Were these estimates significantly better?

Response: Figure 5 show the result of how the different sampling designs influence the cumulative sum and uncertainty ranges. We did not perform a formal test for significant statistical differences, but in all cases the annual sum and uncertainty ranges derived using the tuLHS are closer to those from automated measurements (Figure 5). We highlight that sampling N2O using FTS results in the largest bias (150%) in cumulative sums.

Action: We have revised the text throughout the manuscript to improve clarity.

4. Overall, since the means for FTS and tuLHs were not statistically different, if a researcher's goal is only to estimate an annual GHG flux, is FTS, particularly at bi-weekly time steps, a sufficient strategy?

Response: The means from FTS and tuLHS were not statistically different but that does not imply that cumulative sums nor uncertainty is similar (see Figure 5). Our results show that the cumulative sums and uncertainty derived from FTS are biased for all GHGs (Figure 5). The tuLHS approach consistently provided closer estimates for cumulative sums and uncertainty ranges than FTS for all GHG fluxes.

Action: We have revised the manuscript to improve clarity about this point. We highlight that it is possible to have a similar mean with the wrong reasons (i.e., not reproducing the probability distribution nor the temporal dependency); which is the case for the FTS approach (lines 246-250).

"Although this appears promising, more than a simple comparison of the means is needed to evaluate the information derived from different sampling approaches. In other words, it is possible to have a similar mean value without reproducing the probability distribution nor the temporal dependence of the original time series (i.e., correct answer but for the wrong reasons)."

Technical corrections/comments:

1. In the graphs the authors use Time (days) from 1-365. I assume that is DOY and 1 is Jan 1. The data collected by the automated chambers is Sept 2014-Sept 2015 and I just want to clarify that day 1 is not Sept 2014 and the year follows that timeline.

Response: We appreciate this comment as it was a mistake from our part. Although we have data since September 14, for this study, we only analyzed data from January to December 2015. We did not include data from September 2014 to January 2015 because of large gaps in the dataset. At that time, we were testing the instrumentation and the electrical power supply, and we had several days without electrical power to our system.

Action: We revised the methods section to clarify that the measurements were between January and December 2015 (lines 112-113). We have also edited the legends of Figures 1, 2, S3 and S4 by adding: "Time (x-axis) represents days from January 1 to December 31 of 2015."

2. Figure 2: The blue line is very difficult to see. Perhaps make the open black circles smaller, thicken the horizontal lines for better clarity.

Response/Action: We thicken the horizontal lines in Figure 2, Figure S3, and S4.

3. Figure A1, A2, A7 and A8: These figures are too small to read when printed.

Response/Action: We change the orientation of these figures and made them larger.

4. What program did the authors use to apply the tuLHs to their automated dataset and can they provide that code alongside their already referenced dataset?

Response/Action: We used the R program and the code is available online. (see lines 205-206; 479-480).

Response to reviewer #2

Comment: Title of the manuscript: "The paradox of assessing greenhouse gases from soils for nature based solutions" addresses an important topic and will help to improve our understanding of the greenhouse gas fluxes from the soils. Manual chamber techniques are currently widely used for measuring the three GHG fluxes from soils, since they allow parallel deployment of multiple treatments and lands. However, it requires a lot of care and post-field lab analyses thus limiting temporal representations due to its labour-intensive nature. Since soil N2O and CH4 exhibit sporadic peaks due to their time resolution, a significant problem may arise here; however, CO2 may not be a big concern since it tends to be highly autocorrelated. The availability of automatic chamber sampling thus improves this time resolution concern but they are quite pricey.

Response: We appreciate the detailed summary of this study by the reviewer.

Action: We have revised the text to improve clarity and addressed the comments from two reviewers.

Comment: In this manuscript, as compared to a fixed sampling, the author presents a novel approach for monitoring soil GHG fluxes using temporal univariate Latin Hypercube sampling. The authors used an annual dataset (Sept 2014-Sept 2015) for the three GHGs monitored at 45-minute intervals in a temperate forest. By using temporal univariate Latin Hypercube sampling, each subset of GHGs in the annual dataset is selected based on its statistical properties and temporal patterns. This method reduces bias introduced by fixed sampling, especially for small samples size. In the end, the authors conclude that while these results are crucial for assessing GHG fluxes from soils and reducing uncertainties concerning soils' role in nature-based solutions in the future, the approach needs to be tested across different ecosystems, which may result in different site-specific recommendations.

Response: We appreciate the detailed summary of this study by the reviewer.

Action: We have revised the text to improve clarity and addressed the comments from two reviewers.

Comment: I thus believe that the topic is very interesting and of great relevance to Biogeosciences. The manuscript is well written and has a good structure in terms of design and evaluation results. There is a great deal of work done by the authors in discussing the results, and they have well referenced them. Apart from a few minor changes to the manuscript, I believe that the work is very relevant and very important.

Response: We appreciate the detailed summary of this study by the reviewer.

Action: We have revised the text to improve clarity and addressed the comments from two reviewers.

Comment: For example, the authors should briefly explain the annual weather pattern for the study area. It would be interesting to see how this vary annually to relate with the trend pattern of the gases.

Response: Our goal was not to relate weather variability to GHG fluxes because the tuHLS approach can be performed independently of weather data.

Action: We clarified that the year used for this study has a typical mean annual temperature and precipitation (lines 112-113). We also included in the discussion how temperature variability may be related to the temporal variability of GHG (lines 356-370).

Comment: Since means from univariate Latin Hypercube sampling and fixed sampling did not differ statistically, is it possible to estimate annual GHG fluxes by adjusting weekly fixed sampling?

Response: The means from FTS and tuLHS were not statistically different but that does not mean that cumulative sums nor uncertainty are similar (see Figure 5). Our results show that the cumulative sums and uncertainty derived from FTS are biased for all GHGs (Figure 5). The tuLHS approach consistently provided closer estimates for cumulative sums and uncertainty ranges than FTS for all GHG fluxes.

Action: We have revised the manuscript to improve clarity about this point. We highlight that it is possible to have a similar mean with the wrong reasons (i.e., not reproducing the probability distribution nor the temporal dependency); which is the case for the FTS approach (lines 246-250).

"Although this appears promising, more than a simple comparison of the means is needed to evaluate the information derived from different sampling approaches. In other words, it is possible to have a similar mean value without reproducing the probability distribution nor the temporal dependence of the original time series (i.e., correct answer but for the wrong reasons)."

Specifically

LN 106: What is the reason for using 45 minutes rather than hourly intervals?

Response: This is a mistake in the methods section, and we appreciate the reviewer for identifying this typo.

Action: The time step is 1 hour and was edited in line 112.

LN 117: Could a flux calculation that only considers the highest R^2 eliminate low fluxes?

Response: Not necessarily. This is a common approach to deciding if a flux should be calculated using a linear or exponential fit. Low fluxes can also have high R^2 values. We have followed standardized QA/QC protocols described in previous studies (lines 120-129).

LN 232: Does this site's N_2O lack a temporal dependency for any biological reason?

Response: The site is an upland forest where no additional fertilization is applied. In all our measurements we have found that N_2O emissions are low and do not have clear seasonal patterns nor diel variability (Petrakis et al 2018, Barba et al 2019). There are not many automated measurements of N_2O in upland forests to compare our estimates, but we are aware that in agricultural systems there may be a stronger temporal pattern of N_2O.

Petrakis, S., J. Barba, B. Bond-Lamberty, and R. Vargas. 2018. Using greenhouse gas fluxes to define soil functional types. Plant and soil 423:285–294.

Barba, J., R. Poyatos, and R. Vargas. 2019. Automated measurements of greenhouse gases fluxes from tree stems and soils: magnitudes, patterns and drivers. Scientific reports 9:4005.

LN 243: Include the CO_2 unit after 5.9, also LN 257 include unit of CH_4 after -0.93,

Response/Action: Done

LN 545: Figure A1 does not indicate the graph for soil CO2 (FA CO2), but repeats soil N2O (FA N2O) fluxes.

Response: We are confused about this comment.

Action: We edited the figure following the comments from Reviewer #1.

LN 569: The horizontal blue line is not clear. Could you consider using brighter green instead?

Response/Action: We thicken the horizontal lines in Figure 2, Figure S3, and S4.