Response to comments by reviewer #1

In this manuscript, Zhang et al propose a new proxy to reconstruct fluvial organic matter inputs to coastal marine settings. They suggest that brGMGTs are produced in-situ in rivers and estuaries and that the distribution of brGMGTs is principally controlled by salinity. Based on these facts they generate a new Riverine Index (RIX) using the fractional abundances of H1020c and H1034b versus H1020a and H1020b. To validate the RIX in deep time they compare RIX values to the BIT index and terrestrial pollen/spore deposits deposited during the PETM from IODP Expedition 302 Hole 4A. They report a closer relationship between RIX and terrestrial pollen abundance than BIT and terrestrial pollen abundance, indicating that at least in this site RIX outperforms BIT in accurately reconstructing riverine inputs. In all, this is an interesting study that will likely be of interest to BG readers. I have a number of comments that aim to strengthen the manuscript.

We would like to thank the reviewer for all the constructive comments and the positive assessment on the significance of our manuscript. A point-by-point reply to all the reviewer's comment is provided below and is colored blue. The text which would be added into the revised manuscript is shown in orange italics.

General comments:

In some sections (see specific comments) the use of English in the paper is poor and obfuscates the meaning of the text. I suggest that the authors carefully read through the manuscript to catch all typos and grammatical errors. Likewise figure quality varies considerably. In some cross plots, it is impossible to see the data as the marker size is so small (see specific comments). Characters that should be superscripted/subscripted are left as regular text (see specific comments). Lines of best fit are drawn through data but there is no information as to how these lines were constructed (see specific comments). As such, this manuscript would benefit greatly from more attention to detail from the authors.

In a revised version, we would carefully check the manuscript for any typos and grammatical errors to improve its readability and clarity. Regarding the figure quality, we would address all the issues highlighted by the reviewer, including marker size, superscripting/subscripting, and would provide detailed information (e.g. lines of best fit) in the figure caption.

Additionally, as the authors are proposing a new GDGT salinity index, I would like them to calculate and report previously formulated salinity indices from their samples. Specifically, the ACE index (Turich and Freeman 2011) and the IR6+7me (Wang et al 2021) both have been calibrated against water salinity in marine saline ponds and hypersaline lakes respectively. I know the author's brief touched on comparing IR6me from this study to values from Wang et al (2021) in the text but a more thorough examination of prior GDGT-derived water salinity reconstructions would strengthen the manuscript. Readers will be interested to see how these indices compare against RIX in reconstructing salinity from an estuarine environment.

We agree with the suggestion by the reviewer. In a revised version, we would calculate the ACE and IR_{6+7me} indices and would present them in a supplementary figure (presented below), allowing the comparison of the corresponding values with those of the RIX. The ACE and IR_{6+7me} indices do not correlate with salinity in the Seine River basin.
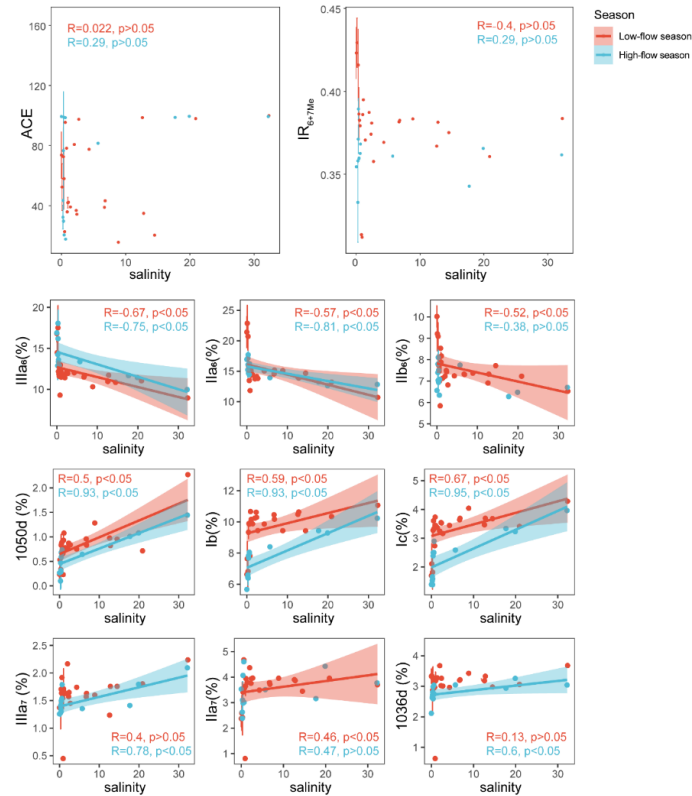
Supplementary figure. Salinity plotted versus ACE, IR_{6+7Me}, relative abundance of 6-methyl and 7-methyl brGDGTs ($IIIa_6$, IIa_6, IIb_6, $IIIa_7$ and IIa_7) as well as compounds 1050d, 1036d, Ib, and Ic through the linear regression. Shaded area represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM.

We would add the following text in the revised manuscript:
"Compared with RIX, other salinity proxies, including the ACE index (Turich and Freeman, 2011) and IR_{6+7Me} (Wang et al., 2021), do not show significant correlations with salinity in the Seine River basin ($p > 0.05$, Wilcoxon test; Fig. S7). Although ACE has been successfully applied in hypersaline systems (Turich and Freeman, 2011), it performs less effectively in certain saline settings due to the complex sources of archaeol and GDGTs (Huguet et al., 2015) and/or distinct ionization efficiencies between these compounds (He et al., 2020; Wang et al., 2021). Similarly, IR_{6+7Me} may be influenced by the preferential production of 6-methyl brGDGTs related with nitrogen nutrient loadings in a specific region of the estuary, as discussed in 4.1.2. Consequently, only RIX successfully tracks salinity variations in this basin, while ACE and IR_{6+7Me} show relative insensitivity."

Additionally, the evidence for in situ production of brGDGTs and brGMGTs in downstream estuary sites is pretty weak. This is demonstrated by Fig 2 where we see that distributions of d13Corg and d15N in soils and downstream estuary samples are very similar in addition to Fig 5 where your PCA on sample brGDGT and brGMGT distributions cannot separate out soils from downstream estuary samples. Yes you see (on average) higher concentrations of brGDGTs and brGMGTs in downstream estuary samples than in soils but the actual distributions of brGDGT and brGMGT abundance in soils are pretty large, indicating that some soils have pretty substantial quantities of these compounds. A great way to add more clarity to this sourcing issue is to train a random forest model using a similar method to Martinez-Sosa et al (2023)

on your brGDGT and brGMGT samples (and isoGDGTs as these should be available to you). If the random forest model can accurately separate out soils from downstream estuary samples then you can be pretty sure that your downstream estuary samples were produced in situ. This won't require much additional work and can be implemented easily using python (https://scikit-learn.org/) or another language of your choice.

We thank the reviewer for this suggestion. In a revised manuscript, we would use a similar approach as the one proposed by Martinez-Sosa et al. (2023) by applying a random forest model to our brGDGT and brGMGT datasets. As the objective of our manuscript is to investigate the sources and controlling factors of brGDGTs and brGMGTs, we would use brGDGTs and brGMGTs separately for training the machine learning model. The application of this model to brGDGT and brGMGT datasets accurately separates downstream (SPM and sediment) estuarine samples from soil ones, indeed supporting *in situ* production of these lipids in downstream Seine Estuary.

In the material and methods, we would add a new machine-learning section describing the model, as follows:

"Our lipid dataset was split into a training set (75%) and a test set (25%). We then used a supervised machine-learning algorithm (random forest) to train models. This algorithm was applied to classify the downstream estuary and soil samples based on brGDGTs or brGMGTs as input, implemented using the scikit-learn library (https://github.com/scikit-learn/) (Pedregosa et al., 2011) in Python (version 3.10.12). Hyperparameter tuning was conducted using a randomized search approach implemented through the RandomizedSearchCV function in scikit-learn.

SHapley Additive exPlanations (SHAP) is a game-theoretical method used to interpret machine learning models (Lundberg et al., 2020). SHAP analysis was applied to identify which compounds were important for the classifications, implemented by the SHAP library in Python. A higher SHAP value indicates a more substantial contribution of the feature (brGDGTs or brGMGTs) to the predicted outcome (downstream estuary or soils)."

Two figures (one for brGDGTs, another for brGMGTs) showing the performance of the model would be added to the supplement:
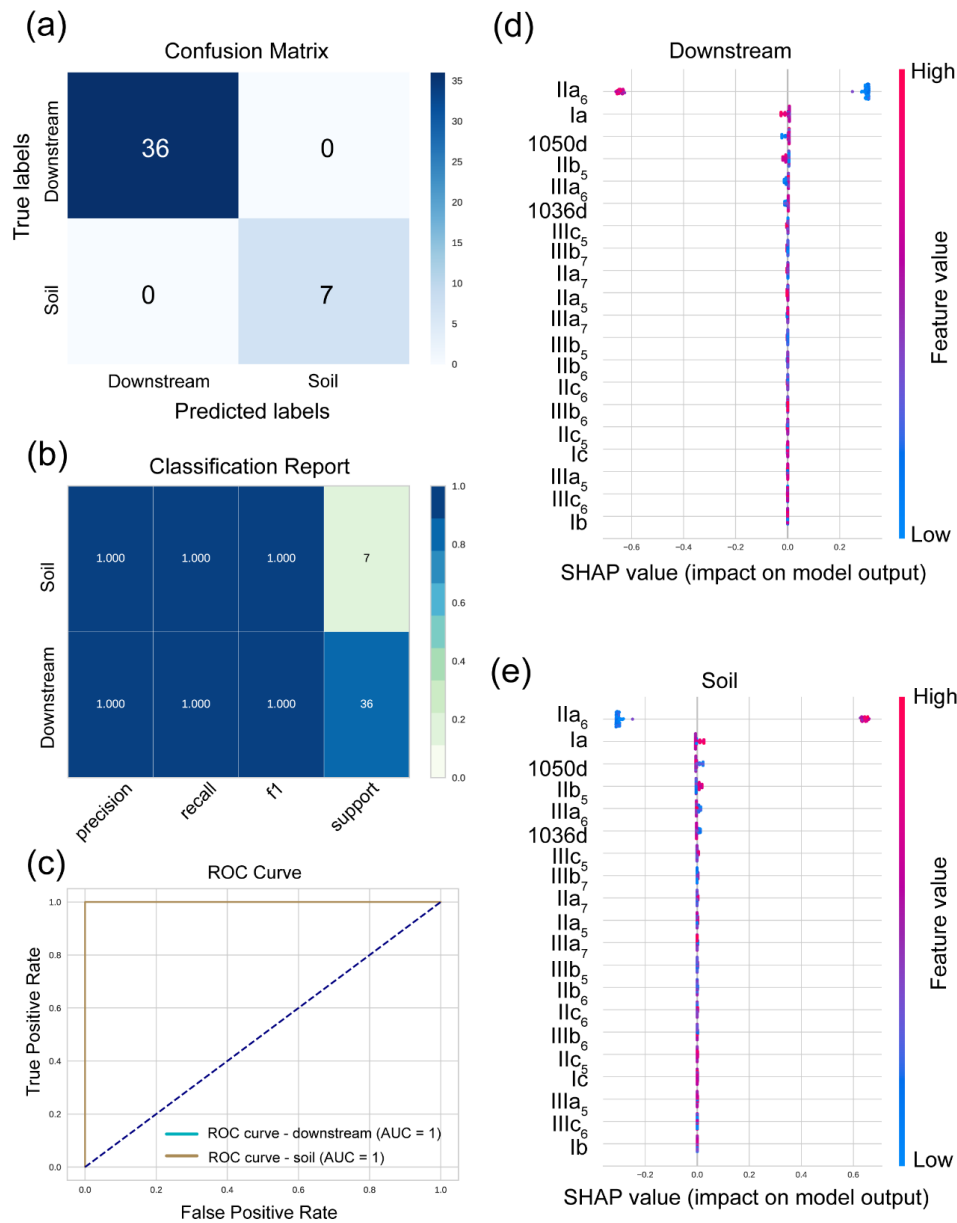
Supplementary Figure (brGDGTs). Evaluation of the random forest model based on brGDGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. Each bullet in the plot represents a single sample in the training set, with the color indicating the feature value (fractional abundance of the brGDGTs) from low (blue) to high (pink). The bullets positioned on the right side of the SHAP summary plot correspond to positive SHAP values, indicating a positive effect on the model output (downstream estuary or soils). The bullets on the left side of the plot indicate negative SHAP values, suggesting a negative effect on the model output. The variables (brGDGTs) with higher impact on the model performance are shown at higher positions. Training sets include downstream SPM and sediment samples (d) and soils (e).
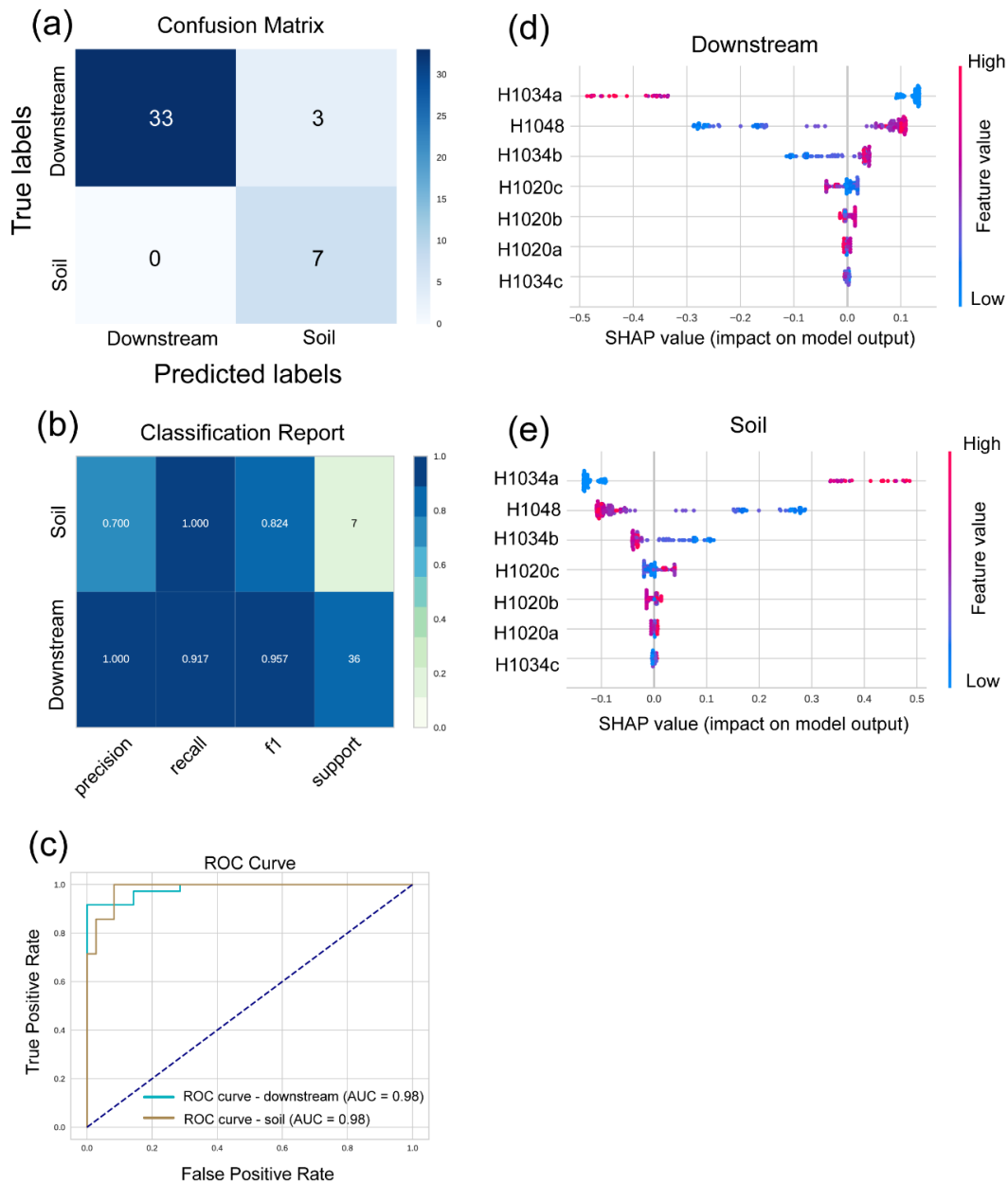
Supplementary Figure (brGMGTs). Evaluation of the random forest model based on brGMGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. Each bullet represents a single sample within the training set, with the color representing the feature value (fractional abundance of the brGMGTs) ranging from low (blue) to high (pink). The bullets positioned on the right side of the SHAP summary plot correspond to positive SHAP values, indicating a positive effect on the model output (downstream estuary or soils). The bullets on the left side of the plot indicate negative SHAP values, suggesting a negative effect on the model output. The variables (brGDGTs) with higher impact on the model performance are shown at higher positions. The training sets include downstream SPM and sediment samples (d) as well as soils (e).

We would add the following text in a revised manuscript to describe and discuss the results related to the application of the model to the brGDGT dataset:

5

"Martínez-Sosa et al. (2023) showed the efficiency of random forest algorithm to distinguish the origin of GDGTs from a global dataset comprised of soil, peat, marine and lake samples. Consequently, we chose to use this algorithm to train and classify brGDGTs from downstream estuary and soil samples in the Seine River basin. Binary classification (downstream estuary vs. soils) was performed using a random forest model based on fractional abundances of brGDGTs. This trained model (Fig. S8) indicated distinguishable brGDGT distributions between downstream estuary (SPM and sediments) and soil samples, supporting in situ production of brGDGTs in the downstream estuary."

We would also add the following text in a revised manuscript to describe and discuss the results related to the application of the model to the brGMGT dataset:

"As with brGDGTs, we applied a random forest algorithm to distinguish brGMGT distributions between downstream estuary and soil samples. This trained model accurately distinguishes soils from downstream estuarine samples (Fig. S11), indicating in situ production of brGMGTs in the downstream estuary. Given the significantly low brGMGT concentrations in soils (p<0.05, Wilcoxon test; Fig. S5b) and the distinct distributions between brGMGT in soils and aquatic settings identified through PCA (Fig. 4) and machine learning (Fig. S11), it can be assumed that the impact of soil-derived brGMGTs on the observed RIX signal in the water column of the Seine basin is low."

Last, isoGDGT data could be discussed in our future work, but this is beyond the scope of the present manuscript.

Specific comments

Line 35: This complicates paleoenvironmental interpretations in SOME aquatic settings not ALL aquatic settings
We agree with this suggestion, this would be corrected.

Line 37: "all along this basin, from land to sea" awkward phrasing
We would rephrase this sentence as follows:

"BrGDGTs and brGMGTs were analyzed in soils, Suspended Particulate Matter (SPM), and sediments (n=237) collected along the land-sea continuum of the Seine basin."

Line 40: "Redundancy analysis further shows that both salinity and nitrogen loadings dominantly control the brGDGT distributions." No, the loadings indicate that SALINITY (not salinity loadings) controls the brGDGT distribution.
This would be corrected.

Line 40-43: "Furthermore, the relative abundance of 6- methyl vs. 5-methyl brGDGTs (IR6Me ratio), Total Nitrogen (TN), δ 15N and chlorophyll a concentration co-vary in a specific zone with low salinity" Is this zone geographical, in your redundancy analysis, or something else?
This zone is geographical. We would propose the following sentence in a revised manuscript:
"Furthermore, the relative abundance of 6-methyl vs. 5-methyl brGDGTs (IR_{6Me} ratio), Total Nitrogen (TN), $\delta^{15}N$ and chlorophyll a concentration co-vary in a specific geographical zone with low salinity,

suggesting that 6-methyl brGDGTs are preferentially produced under low-salinity and high-productivity conditions."

Line 44-45: "Salinity is positively correlated with homologs H1020a and H1020b, 45 and negatively correlated with compounds H1020c and H1034b." Is this in soils, sediments or SPM?
This correlation was found in SPM. This would be specified as follows:
"Salinity is positively correlated with homologues H1020a and H1020b, and negatively correlated with compounds H1020c and H1034b in SPM."

Line 45: "This suggests that bacteria thriving…" thriving is not the correct word (carries implications of a value judgment) replace with "living".
This would be corrected.

Line 45-47: It seems like you aren't mentioning results from soils and sediments, only from SPM? Or maybe all your sediment samples are exclusively from rivers? The reader is unclear on this.
Correlation between salinity and lipid distribution is based on SPM samples. We would modify the abstract as follows:
"Both types of compounds (i.e. brGDGTs and brGMGTs) are shown to be produced in situ, in both freshwater and saltwater, based on their high concentrations and distinct distributions in aquatic settings (SPM and sediments) vs. soils.

Line 51-52: "a paleorecord across the upper Paleocene and lower Eocene," You should name this record and say where it is.
We would add the name (Arctic Coring Expedition) and location (Lomonosov Ridge) of this record.

Lines 51: "showing its potential applicability in both modern samples and in paleorecords." Perhaps you could evaluate its usage in both these cases e.g. - we successfully/unsuccessfully applied RIX in …
We would rephrase this sentence as follows:
"We successfully applied RIX to the Godavari River basin (India) and a paleorecord across the upper Paleocene and lower Eocene from the Arctic Coring Expedition at Lomonosov Ridge, showing its potential applicability in both modern samples and in paleorecords."

Line   55: ", although some of them were attributed to the phylum Acidobacteria" Imprecise language.
Thank you for the comment. We would rephrase this sentence into:
"Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are membrane lipids produced by bacteria, some of them belonging to the phylum Acidobacteria."

Line 57-58: "The distribution of brGDGTs (number of cyclopentane moieties and methyl groups; cf. structures in Fig. S1) was empirically linked with pH and Mean Annual Air Temperature" Again, imprecise language. The phrase "empirically linked" doesn't convey much useful information.
We would replace "empirically" by "has been":
"The distribution of brGDGTs (number of cyclopentane moieties and methyl groups; cf. structures in Fig. S1) has been linked with pH and Mean Annual Air Temperature (MAAT) in soils."

Line 60: Should really cite some earlier lake GDGT papers in addition to Martinez-Sosa et al., 2021.
Thank you for this suggestion. We would cite some earlier work: Tierney et al., 2010 GCA and Russell et al., 2018 OG.

Lines 60-61: "The brGDGT-based proxies (i.e. MBT'5ME and CBT') have been largely applied to reconstruct MAAT and pH from sedimentary archives (Coffinet et al., 2018; Harning et al., 2020; Wang et al., 2020)." Not quite true - Martinez-Sosa et al (2021) and Dearing Crampton-Flood et al (2020) generated Bayesian linear regressions between the Mean temperature of months above freezing and MBT'5me. These BayMBT models have been used widely in the community since their publication.
We agree that the new models were not applied to records available before their publication. We would modify this sentence as follows:
"The brGDGT-based proxies (i.e. MBT'$_{5ME}$ and CBT') have been largely applied to reconstruct paleoclimate from sedimentary archives (Coffinet et al., 2018; Harning et al., 2020; Wang et al., 2020)."

Line 62-63: "In aquatic settings, brGDGTs were initially suggested to be predominantly derived from watershed soils and transported by erosion in the sediments (Hopmans et al., 2004)." Maybe you mean "transported by erosion to the sediments"?
We agree and would rephrase this sentence as follows:
"In aquatic settings, brGDGTs were initially suggested to be transported by erosion to the sediments."

Lines 63-78: The use of English throughout this paragraph is poor and hard to follow. Needs copyediting.
We would rephrase this paragraph as follows:
"In aquatic settings, brGDGTs were initially suggested to be transported by erosion to the sediments (Hopmans et al., 2004). Based on this assumption, the Branched and Isoprenoid Tetraethers (BIT) index was defined as the abundance ratio of the major brGDGTs to crenarchaeol (isoprenoid GDGT mainly produced by marine Nitrososphaerota). The BIT index ranges between 0 and 1, with high BIT values (around 1) reflecting a higher contribution of terrestrial organic matter compared to marine organic matter (Hopmans et al., 2004).
Over the last few years, the BIT index has been broadly used to quantify the relative contribution of terrestrial organic matter in aquatic systems (Xu et al., 2020; Yedema et al., 2023) and to evaluate the reliability of the TEX$_{86}$ palaeothermometer (Cramwinckel et al., 2018). However, several studies have shown that brGDGTs can also be produced in situ in aquatic settings, including rivers (e.g., De Jonge et al., 2015; Freymond et al., 2017; Kim et al., 2015; Zell et al., 2014, 2013), lakes (Tierney and Russell, 2009), and marine environments (Dearing Crampton-Flood et al., 2019; Zeng et al., 2023). This adds complexity to the identification of brGDGT sources in aquatic ecosystems and to the application of brGDGTs as (paleo)environmental proxies, including the BIT index.
The BIT values have all the more to be carefully interpreted, especially considering the potential influence of the selective degradation of branched vs. isoprenoid GDGTs (Smith et al., 2012). Thus, complementary molecular proxies for quantifying the input of terrestrial organic matter to aquatic settings are still needed. These proxies may cross-validate other available terrestrial proxies, such as the $\delta^{13}C$ of organic carbon (Lamb et al., 2006), heterocyst glycolipids (Kang et al., 2023), and long-chain diols (Lattaud et al., 2017)."

Lines 63-78: You should read and cite Martinez-Sosa et al (2023) here for their work on a Random Forest

approach to classify GDGT sources (i.e. Marine, Soil, Lake etc).

Thank you for this suggestion. We would refer to the work by Martínez-Sosa et al. (2023) as follows:

"Recently, a machine-learning approach (BIGMaC model) was proposed to infer the origin of environmental samples (e.g. soil, peat, marine and lake settings) based on their GDGT distribution (Martínez-Sosa et al., 2023). While such an approach shows potential for differentiating distinct sources of GDGTs, its application to aquatic systems has not yet been extensively explored."

Line 80-83: "The improvement of analytical methods allowed the separation and quantification of 5-, 6- and 7-methyl brGDGTs (methyl groups at the fifth, sixth, and seventh positions; Fig. S1), that in previous chromatographic protocols co-eluted (De Jonge et al., 2013, 2014; Ding et al., 2016)." No real link between the previous paragraph and this one. Also, which methods? How did they improve?

This comment would be taken into account as follows:

"The improvement of chromatographic methods allowed the separation and quantification of 5-, 6- and 7-methyl brGDGTs (methyl groups at the fifth, sixth, and seventh positions; Fig. S1) that previously co-eluted (De Jonge et al., 2013, 2014; Ding et al., 2016). This led to the development of new brGDGT-based proxies based on these specific brGDGT isomers (De Jonge et al., 2014)."

Lines 86-87: "In addition to temperature and pH, other environmental factors may influence brGDGT distributions in terrestrial and aquatic settings and hence the application and interpretation of brGDGT-derived proxies" This is a repetition from earlier in the introduction.

This sentence would be removed from the revised manuscript.

Lines 91-99: You should mention that brGMGTs have previously been called H-brGDGTs in the literature.

We would mention this as follows:

"Compared with brGDGTs, the branched glycerol monoalkyl glycerol tetraethers (brGMGTs, also referred as H-brGDGTs) are a much less studied group of lipids."

Lines 91-111: This paragraph was very well written and is an example of the standard the entire manuscript should meet.

As said above, we would carefully check the English quality of our revised manuscript.

Lines 117-123: You go from talking about the hypothesis you aim to test in the paper to talking about the aims of the paper. Surely your aim is to test the hypothesis you have just laid out - why do we need to talk about more aims here?

Thank you very much for this comment. We consider as appropriate to transition from stating the hypothesis to clearly discussing the aims of the paper in this paragraph.

Line 125-126: "by high population density". High population density of what?

Population density refers to the number of human beings who live in a given region. We would remove this as it is redundant with the next half sentence.

Line 127: "macrotidal". Please define this term.

Macrotidal means large tidal range. We would define this term in a revised manuscript.

Figure 1: I really like this figure - it nicely summarizes your water sampling campaign.
Thank you for this comment.

Line 167: "Both decarbonated and non-decarbonated samples (~6 mg for SPM and ~20 mg for soils) were enclosed in a tin capsule" You should mention that you will split your samples and decarbonate one aliquot. Otherwise, the reader is confused as to where your non-decarbonated samples are coming from.
Thank you for the suggestion. We would add the following sentence:
"The samples were split, and one aliquot was decarbonated."

Line 172-174: "The isotopic composition (δ 13C or δ 15N) was expressed as the relative difference between isotopic ratios in samples and in standards (Vienna Pee Dee Belemnite for carbon or atmospheric N2 for nitrogen)" Should be "and atmospheric N2…".
This would be corrected.

Line 176: What were these "additional…analyses"? Do you mean the same analyses aforementioned but on different samples, or different analyses on different samples?
The elemental and isotopic data of the SPM and sediments collected in 2015 and 2016 (*n*=84) were published by Thibault et al. (2019). To avoid confusion, we would rephrase this sentence as follows:
"Additional elemental and isotopic data based on SPM and sediments collected in 2015 and 2016 (n=84) were obtained from Thibault et al. (2019)."

Line 177: "(4-20g, n=51)" Looks to me like you've used the minus sign, not the en dash here. If so use the en dash.
This would be corrected.

Line 180-183: "The total lipid extracts were then separated into fractions of increasing polarity on an activated silica gel column, using (i) 30 mL of heptane, (ii) 30 mL of heptane:DCM (1/4, v/v), and (iii) 30 mL of DCM/MeOH (1/1, v/v) as eluents." That seems like a nonstandard amount of solvent. Are you using very large columns here? If so state how many g of silica gel were used.
We use glass columns with a total volume of ca. 10 mL to separate the lipids. The amount of solvent is three times the column volume. We would modify the text as follows:
"The total lipid extracts were then separated into fractions of increasing polarity on an activated silica gel column (ca. 10 mg), using (i) 30 mL of heptane, (ii) 30 mL of heptane:DCM (1/4, v/v), and (iii) 30 mL of DCM/MeOH (1/1, v/v) as eluents."

Line 233: Vegan should be vegan. No capital V.
This would be corrected.

Lines 240-243: I don't think you effectively explain how your hierarchical partitioning method actually works here. As some readers won't be familiar with this method, more details are needed.
We agree with the reviewer and would add more information about the hierarchical partitioning method as follows:
"Briefly, this approach suggests that shared variance can be decomposed into equal components based on the number of involved predictors (environmental factors), allowing for the estimation of the relative

Figure 2. I really don't like how the axis of these plots has been extended to include chart labels. The top left panel scale is completely distorted by the addition of these labels. You should also define the features of your "boxes" in your box plot. These comments apply to all boxplots in the manuscript.

We agree with the reviewer and would modify the plots (notably by changing the scales) and captions of the figures accordingly.

Line 268: "The different brGDGTs were detected in all studied samples" Which brGDGTs?

We would specify all the brGDGTs which were detected in the following sentence:

"The different brGDGTs ($IIIa_5$, $IIIb_5$, $IIIc_5$, IIa_5, IIb_5, IIc_5, $IIIa_6$, $IIIb_6$, $IIIc_6$, IIa_6, IIb_6, IIc_6, $IIIa_7$, $IIIb_7$, IIa_7, Ia, Ib, Ic, 1050d, and 1036d) were detected in all studied samples."

Line 275: "The relative abundances of the brGDGTs were determined all along the Seine River basin" I feel like this sentence should be at the start of this section not in the second paragraph.

We would prefer to keep the current arrangement. We still prefer placing the description of chromatogram at the beginning of the section and then fractional abundances, which emphasizes the logical flow of information.

Line 290: "which explained 40.9% of the variance in two dimensions" Which two dimensions are these?

Thank you for the comment. These are the first two dimensions. We would correct a typo here and modify this paragraph as follows:

"A Principal Component Analysis (PCA) was performed to statistically compare the fractional abundances of brGDGTs from different location (river, upstream and downstream estuary, based on SPM and sediments collected in the river channel), which explained 54.1% of the variance in the first two dimensions (Fig. 4a). The first axis (PC1) explained 40.9% of the variance, with negative loadings for most of the 6-methyl brGDGTs and positive loadings for the remaining brGDGTs (Fig. 4a)."

Line 291: "Samples from the downstream estuary clustered well" Colloquial language, you should describe the data using words that don't convey a value judgment.

We would remove the word "well" from this sentence.

Line 314-315: "It allowed to explain 39.79% of the variability through two dimensions." Doesn't make sense - please proofread your manuscript.

We would remove this sentence.

I feel like you have just randomly placed the figures in the text. You should line up the first in-text citation of a figure with the location of the figure in the manuscript. Currently, the text and the figures are out of sequence which makes reading this document a challenge.

We have carefully checked our figures and in-text citations. They were appropriately positioned. We

present the PCA (RDA) plots for brGDGTs and brGMGTs together to avoid adding too many figures to the manuscript. In the text, we describe the results related to brGDGTs first and then those related to brGMGTs. The in-text citations indeed correspond to the order of the figures.

Figure 5: Visually this figure is quite busy. I don't think having the brGDGT names in blue (the same colour used for the downstream bubble) helps. I would use black for these names and also the arrows.
We agree with the reviewer and would change the color of these names and arrows into black.

Line 336: "The brGMGTs identified in previous studies" Which brGMGTs and which studies? This lack of precise usage of language is present throughout the text.
This would be corrected in the following sentence:
"The brGMGTs (H1020a, H1020b, H1020c, H1034a, H1034b, H1034c, and H1048) identified by Baxter et al. (2019) were detected in the samples collected across the Seine River basin."

Line 343-345: "In SPM and river channel sediments, the total brGMGT concentration was observed to be slightly higher in the riverine part (0.26 ± 0.24 µg/g Corg) than in downstream (0.20 ± 0.13 µg/g Corg) and upstream estuary samples (0.17 ± 0.18 µg/g Corg; Fig. S4b)." Slightly higher but not significantly higher. If it's not significant you should say so.
The difference in brGMGT concentration along the estuary is not significant. This would be acknowledged as follows:
"In SPM and river channel sediments, the total brGMGT concentration was observed to be slightly (but not significantly) higher in the riverine part (0.26 ± 0.24 µg/g C_{org}) than in downstream estuary (0.20 ± 0.13 µg/g C_{org}) and upstream estuary samples (0.17 ± 0.18 µg/g C_{org}; Fig. S5b). The total brGMGT concentrations were the lowest in soils (surficial soils and mudflat sediments) all over the basin (0.07 ± 0.09 µg/g C_{org}; Fig. S5b)."

Line 346: "The PCA analysis based on the brGMGT relative abundances (Fig. 5b) explained 70 % of the variance". I'm unsure what the authors are trying to say here but I think they mean that the first two PCs sum to 70%. The second half of the sentence "which allows to observe that samples from the different parts of the basin clustered well apart from each other." doesn't make sense and I'm unsure what the authors are trying to say.
Yes, the first two PCs sum to 70%. To clarify this point, the sentence would be rephrased as follows:
"The PCA analysis based on the brGMGT relative abundances (Fig. 4b) explained 70 % of the variance in the first two dimensions, which separate samples from different parts of the basin."

Line 357: "allows to explain" This phrase doesn't make sense in this context - please remove all uses of it from the manuscript.
This would be corrected.

Lines 406-408: "The similarity in distributions between soils and downstream samples may be due to the overrepresentation of downstream soil samples, as 82% of the soils were collected downstream (Fig. 1a and Table 1)." I don't understand your point here. Are you saying that the similarity between downstream estuary brGDGT distributions and soil brGDGTs is because the downstream estuary predominantly receives brGDGTs from downstream soils?

We thank the reviewer for this comment. We would rephrase the sentence as follows to clarify this point:
"The similarity in brGDGT distributions between soils and downstream samples may be due to the influx of brGDGTs from the downstream soils into the downstream estuary, as 82% of the soils were collected downstream (Fig. 1a and Table 1)."

Lines 409-412: "Nevertheless, the soil-derived brGDGT contribution to the downstream samples is expected to be much lower than the autochthonous one, as the average brGDGT concentration in soils was ca. 3 times lower than the one in downstream (i.e. SPM and river channel sediment) samples (Fig. S4a)." Right, but it's curious that the distributions are so similar between brGDGTs in soils and downstream estuaries. To bring more clarity to this point it would be interesting to see you attempt a machine learning approach (see Martinez-Sosa et al 2023, PP) to investigate whether (or not) a random forest model can distinguish soil samples from downstream estuary samples.
As previously said, we applied a machine learning approach, similar to that of Martinez-Sosa et al. (2023), to our dataset. Additional figures would be added to the supplementary material, as well as text to the discussion (cf. reply to main comments above).

Lines 426-429: It would be great to see you calculate and report IR6+7me following Wang et al (2021) to determine if these indices correlate to salinity in an estuarine location.
We have calculated IR_{6+7me} as suggested by the reviewer. We would modify the figures and main text accordingly (cf. reply to main comments above) and would notably add the following sentence:
"The salinity proxy (IR_{6+7me}) proposed by Wang et al. (2021) does not show significant correlation with salinity in this study (p>0.05, Wilcoxon test; Fig. S9). This suggests that IR_{6+7me} is relatively insensitive in the Seine Estuary, potentially due to the preferential production of certain brGDGTs in specific estuarine regions."

433-436: "The distinct behavior of 6-methyl brGDGTs between lakes and the Seine river-sea continuum might be due to the lower salinity range in the Seine River basin (0-32 psu) vs. the lakes (0-376 psu) 435 investigated by Wang et al. (2021). This suggests that the limited range of salinity variation in the Seine River basin might be insufficient to trigger significant 6-methyl brGDGT production, as observed in hypersaline lakes." This is actually incorrect. Wang et al 2021 report that IR6me is sensitive to salinity in the range of 5-1000 (mg/L) but relatively insensitive beyond this range.
We agree with the reviewer and would modify the text accordingly by removing the reference to the publication by Wang et al. (2021) here:
"Indeed, the significant negative correlations between the salinity and the relative abundance of 6-methyl brGDGTs is observed in the Seine basin (Fig. S9), which suggests that the bacteria producing 6-methyl brGDGTs are preferentially present in the low salinity area of the estuary."

458-460: "As the nutrient concentration is higher in the upstream part of the Seine estuary (Wei et al., 2022), the substantial 6-methyl brGDGT production observed in the aforementioned zone (260 460 < KP < 340, Fig. 8)" Right but why would the nutrient runoff be higher for this specific section of the basin? Do we see more agricultural activity here or something? It would be good to try and flesh out this point.
This specific region of the estuary is indeed characterized by intense agricultural activity, which could at least partly explain the high nutrient concentration in this zone, especially during the low-flow season. The text of the manuscript would be revised as follows:

"As the nutrient concentration is higher in the upstream part of the Seine estuary (Wei et al., 2022), and this zone is characterized by high proportions of agricultural land use (Flipo et al., 2021), the substantial production of 6-methyl brGDGT observed in the aforementioned zone (260 < KP < 340, Fig. 8) during low flows could be attributed to elevated nutrient levels, particularly nitrogen, resulting from intense agricultural activities."

Figure 8 and throughout: Make sure to superscript 15 in d15N and subscript 6 in IR6me.
This would be corrected.

509-510: "The current knowledge on the parameters controlling the brGMGT distributions in the terrestrial and marine realm is still limited." Why is it limited? Be specific.
Thank you for the comment. This group of lipids (brGMGTs) has only recently gained attention. Consequently, there are many aspects (e.g. controlling factors) still unknown for brGMGTs compared to brGDGTs. To be more specific, we would rephrased our sentence as follows:
"The current knowledge on the parameters controlling the brGMGT distributions in the terrestrial and marine realm is still limited, as there is little literature available (Kirkels et al., 2022a)."

Fig 9: Almost impossible to see the data points on some of the figure panels (e.g. panel e). Make the points bigger. Also, keep a consistent label text size to make the figure look neater. Also, you should say in the caption how you constructed the straight lines drawn through the data in some panels (e for instance). I'm assuming this is a linear regression but you have to inform the reader of your methods.
Thank you for this suggestion. To enhance visibility, we would increase the size of the data points in figure panels, especially in panel e. Additionally, we would standardize the label text size across all figure panels. Furthermore, we would provide more information in the figure captions.

557-558: " However, the average concentrations of brGMGTs are an order of magnitude lower in the soils than in the river channel sediments and SPM samples of the Seine basin (Fig. S4b)." Maybe it is, but visually it doesn't look like that, so include the numbers in this sentence. You can also argue that the brGMGT abundance within soils varies by an order of magnitude. Do you know what is driving such a large variance in the soil brGMGT abundance?
We agree with the reviewer that the brGMGT concentration in soil samples shows large variance. This highlights the need for further investigation into the environmental controls on brGMGT concentration and distribution in soils. However, as shown by the boxplot, the upper and lower quartiles as well as the median value of the soil brGMGT data are low compared to the river, upstream, and downstream samples. In any case, downstream (SPM and sediment) samples and soils display distinct distribution and concentrations, also captured by the application of the machine learning model to the brGMGT dataset (cf. reply to the main comments above).
We would consider the comment of the reviewer in a revised manuscript through the following sentence:
"A large variance in the soil brGMGT concentration was observed (Fig. S5b), suggesting that further investigation is needed to better understand the environmental controls on the brGMGT production in soils."

589: Missing the word "index" after BIT
This would be corrected.

You need a map showing the location of IODP 302 Hole 4A

We would add the following map showing the location of the core in the supplement:
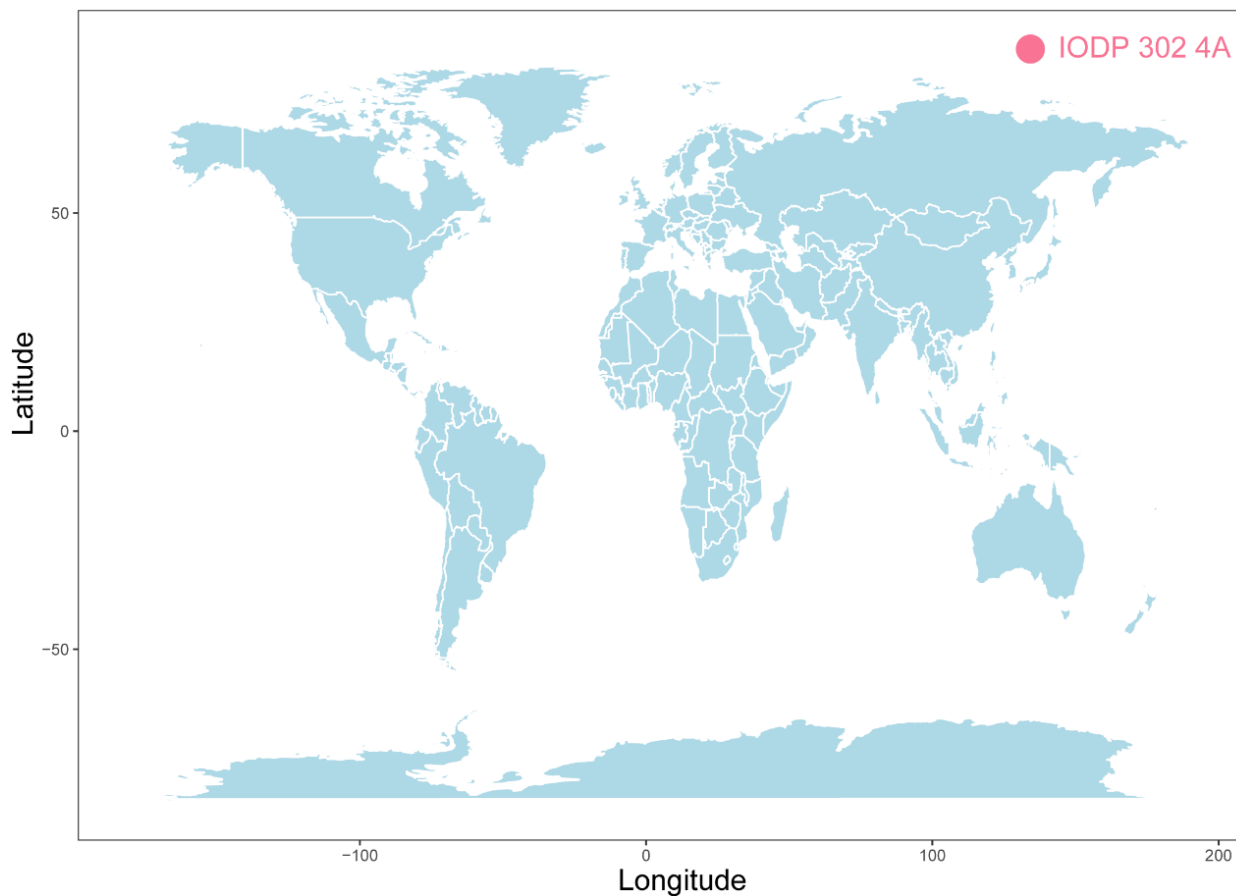
Lines 605-607: "This core is considered proximal to the coast and has considerable changes in terrestrial inputs (i.e. continental spores and pollen) over time (Sluijs et al., 2009, 2006), making it a suitable paleorecord for testing runoff proxies." Again would be great to have some specifics. The readers will be interested in how close this core site was to the coast around the PETM. You should also say why there was a considerable change in terrestrial inputs (I'm assuming large changes in sea level are responsible).

We thank the reviewer for this comment. The changes of sea level are indeed responsible for the changing terrestrial inputs. We would rephrase this sentence as follows in a revised manuscript:

"This core is considered to record significant changes in terrestrial inputs (i.e. continental spores and pollen) due to sea level changes over time (Sluijs et al., 2009, 2006), making it a suitable paleorecord for testing runoff proxies."

Lines 616-617: "Such decreased runoff during the PETM body was previously attributed to a local sea level rise" Ah here is the explanation - this should have been in the previous paragraph. Also, be specific, are you saying there was decreased runoff during the PETM, OR did your sediment core record decreased runoff due to a change in sea level? These are two different things.

In addition to this core (Sluijs et al., 2008), a rise in sea level during the PETM has been recorded in many other sites worldwide (Speijer and Morsi, 2002; Harding et al., 2011; Sluijs et al., 2014). We would rephrase this sentence to clarify this point in a revised manuscript:

"Such decreased runoff during the PETM body was previously attributed to a rise in sea level (Sluijs et

al., 2006), which has been recorded in many other sites worldwide (Speijer and Morsi, 2002; Harding et al., 2011; Sluijs et al., 2014)."

References:

Martínez-Sosa, P., Tierney, J. E., Pérez-Angel, L. C., Stefanescu, I. C., Guo, J., Kirkels, F., ... & Reyes, A. V. (2023). Development and application of the Branched and Isoprenoid GDGT Machine learning Classification algorithm (BIGMaC) for paleoenvironmental reconstruction. Paleoceanography and Paleoclimatology, 38(7), e2023PA004611.

Wang, H., Liu, W., He, Y., Zhou, A., Zhao, H., Liu, H., Cao, Y., Hu, J., Meng, B., Jiang, J., Kolpakova, M., Krivonogov, S., and Liu, Z.: Salinity-controlled isomerization of lacustrine brGDGTs impacts the associated MBT5ME' terrestrial temperature index, Geochimica et Cosmochimica Acta, 305, 33–48, https://doi.org/10.1016/j.gca.2021.05.004, 2021.