Response to comments by reviewer #2

The authors analyzed brGDGTs and brGMGTs in soils, suspended particulate matter, and river sediments in the Seine River basin to evaluate the environmental controls on and sources of these lipids. The basin ranges from freshwater to estuarine, allowing the authors to evaluate the effects of salinity on the GDGT compositions. The major motivation seems to be development of a new GMGT index, called "RIX", to detect terrestrial inputs of GMGTs to marine environments. The authors test this index through application of Cenozoic sections of an IODP site.

There is now a relatively large literature on the environmental controls on GDGTs, though there is less on GMGTs, and combining these across a riverine salinity gradient is a strength of the paper. Overall I think the paper does provide some novel contributions and findings that merit publication. That said, there are a number of technical problems that will require major revision before the paper can be published.
We thank the reviewer for his/her detailed comments and for recognizing the novelty and strength of our work. 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.

First, the Seine basin is complicated by the presence of a dam that separates sections of the river influenced by tides (salinity) from sections upstream. The dam also presumably traps upstream sediment and likely presents a barrier for transport of GDGTs (other than SPM). The authors also have relatively few soil sampling sites – there are only 5 sites and the soil samples are dominated by downstream estuarine soils. I don't think these challenges are adequately discussed in the paper. The dam may be a good thing for the study, since it establishes clear environmental boundaries, but it could be tricky to apply a GDGT index from this environment to other sites/time periods.

The dam of Poses (cf. location on the revised map below) is the frontier between the Seine river and estuary. It represents the upstream limit of the fluvial estuary and of the tidal propagation. It was built in 1887 to regulate the water level and to allow navigation of the ships up to Paris, whatever the season. Indeed, the average water flow of the Seine River measured at Poses is ~ 470 $m^3$ $s^{-1}$, with marked intra-annual differences between winter and summer flows (~ 250 $m^3$ s−1 in the summer and over 700 $m^3$ $s^{-1}$ in the winter). Whatever the period of the year, at least part of the water from the Seine river upstream Poses flows to the estuary. Therefore, the dam should not prevent (part of) the riverine GDGTs associated to SPM to arrive to the estuary. Nevertheless, it cannot be excluded that part of the riverine sediments are trapped by the dam. This would be mentioned in a revised manuscript.
Regarding the estuary itself (downstream Poses), it comprises two major sections: the upstream, freshwater section (from site 5 to 12) and the lower, downstream section influenced by salinity (from site 12 to the coastal zone). All our estuarine samples were (logically) collected downstream of the dam of Poses. Therefore, the observed changes in brGDGT/brGMGT distribution and abundance all along the estuary, with distinct signal in the upstream and downstream estuarine zones, are intrinsic to the biogeochemical functioning of the Seine estuary and cannot be attributed specifically to this dam. This would be specified in a revised manuscript.

Regarding the soils, we agree with the reviewer and acknowledge the limitations of our sampling strategy, with a low number of sampling sites, mainly located downstream. We cannot exclude that the overlay in

brGDGT/brGMGT distribution between the soils and the downstream estuary SPM and sediment samples is partly due to the sampling approach. This would be specified in a revised manuscript with the following sentence: *"The similarity in 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)."*
Nevertheless, the comparison of the brGDGT/brGMGT concentrations and distributions between soils and downstream estuary samples allows distinguishing the two types of samples, as captured by the application of a machine learning approach to our brGMGT/brGDGT datasets.

Last, we kindly disagree with reviewer 2 when saying that "it could be tricky to apply a GDGT index from this environment to other sites/time periods." The RIX index was developed based on samples from the Seine estuary. Nevertheless, it was successfully tested in both modern (Godavari River basin) and past settings (marine sedimentary core IODP 302-4A), showing its potential general applicability.

Second, there are a lot of data / statistical difficulties with this paper, the details of which are discussed below. At times the authors compare concentrations of GDGTs to evaluate in situ production, which is generally not a good way to do this due to the effects of sediment transport from soils to river to estuaries – concentrations may be higher in SPM than soils, for instance, as SPM contains less coarse-grained particles. Although the writing is a bit unclear, the authors appear to compare results of two PCAs, one on soils and one on aquatic samples, to differentiate these two sample types, which is not possible given how PCA works.

In order to better evaluate the *in situ* production of brGDGTs/brGMGTs in the estuary, a machine learning approach (similar to the one proposed by Martínez-Sosa et al. (2023) was applied to the brGDGT/brGMGT datasets, as suggested by Reviewer #1. We have now several lines of evidence supporting the *in situ* production of brGDGT/brGMGTs in the Seine estuary:
1) higher brGDGT/brGMGT concentrations in aquatic environments compared with soils.
2) distinct distributions between soils and aquatic settings (riverine and upstream estuarine samples) identified by PCA.
3) the application of the machine learning approach, which allows distinguishing downstream estuary and soil samples based on brGDGT/brGMGT distributions. This is addressing the overlap observed between downstream estuary and soil samples in the PCA biplot based on brGDGT/brGMGT distributions.
As detailed in the reply to reviewer 1, additional discussion on the *in situ* production of brGDGTs/brGMGTs (based on the above mentioned points) would be added in a revised manuscript.

In addition, we would like to clarify that the PCAs of soils and aquatic samples were not done separately. The biplots do not correspond to a simple overlay. Only active individuals (river, upstream, and downstream estuarine samples) were used for principal component analysis. The coordinates of passive individuals (also known as supplementary individuals) (i.e. soil samples) were just **predicted**/projected using the **existing** PCA information obtained with active ones. This is actually a widely used approach which can be implemented by the R package FactoMineR (https://cran.r-project.org/web/packages/FactoMineR). It has also been used in a recent GDGT paper (Kirkels et al., 2022 *Biogeosciences*), which aims to compare GDGT distributions in soils and aquatic settings. We prefer this approach as it effectively delivers the key information: brGDGT/brGMGT distributions in riverine

and upstream estuarine samples are distinguishable from those in soils. However, in the PCAs based on brGDGT/brGMGT distribution, soils partly overlay with downstream estuary samples. This similarity may be at least partly attributed to our sampling strategy, given that most of the soils were collected around the downstream estuary, as mentioned in the manuscript. Nevertheless, we can efficiently distinguish brGDGT/brGMGT distributions in downstream estuarine samples from those in soils by using a machine learning approach as said above.

In the revised manuscript, we would modify the figure caption of the PCAs to better illustrate our methodology:
*"The coordinates of soils (passive individuals) are added as an overlay and are predicted based on the information provided by the existing PCA performed on SPM and sediments (active individuals)."*

Third, Section 4.4 compares the application of the RIX to IODP site 302 to results from other measurements, such as the BIT and % terrestrial palynomorphs. The comparison is largely qualitative, and it's hard to tell from Figure 11 how well these compare in a statistical sense. Could the authors provide correlation coefficients to show that the RIX captures terrestrial inputs?
We thank the reviewer for this comment. We would provide the correlation coefficients between RIX and % terrestrial palynomorphs as well as between BIT and % terrestrial palynomorphs in a revised manuscript. The corresponding figure is provided below:

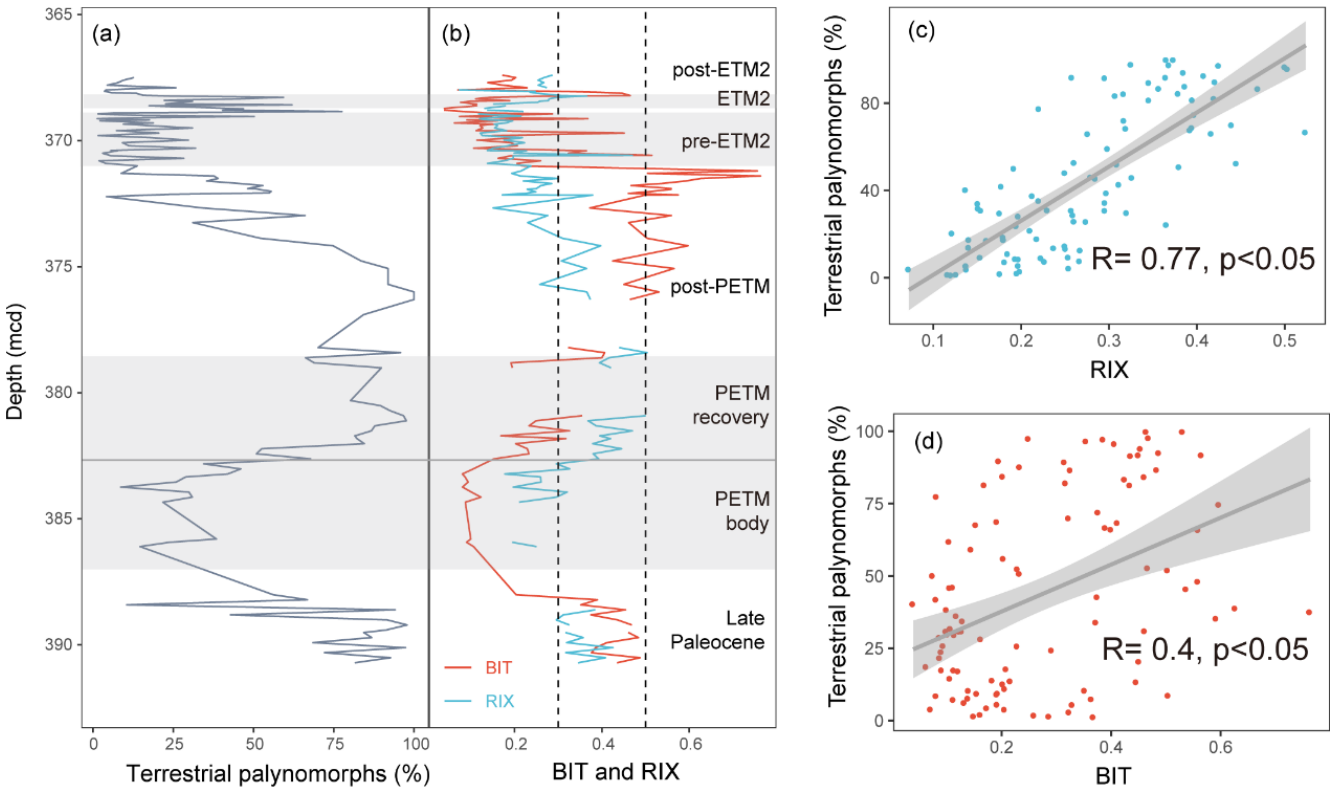


*Figure: Comparison between (a) terrestrial palynomorphs (%) and (b) BIT and RIX across the upper Paleocene and lower Eocene between 391 and 367 meters composite depth below sea floor (mcd) of IODP Expedition 302 Hole 4A. Terrestrial palynomorphs data are from Sluijs et al. (2006) and Sluijs et al. (2009). RIX and BIT were calculated using data from Sluijs et al. (2020). Grey shading represents Eocene Thermal Maximum 2 (ETM2), pre-ETM2 interval, and Paleocene-Eocene Thermal Maximum*

*(PETM). Dotted line represents cutoff values of RIX (below 0.3 for marine contribution and above 0.5 for riverine contribution). Linear regression of the RIX (c) and BIT (d) against the terrestrial palynomorphs. Shaded area represents 95% confidence intervals.*

We would also add the following sentence in a revised manuscript:
*"RIX performs better in this core compared with BIT, which is further supported by a higher correlation coefficient observed between RIX and terrestrial palynomorphs (R=0.77; Fig. 10c) compared with BIT and terrestrial palynomorphs (R=0.4; Fig. 10d)."*

Detailed comments:

Section 2.2. It is a bit hard to tell from this description and the table exactly what samples were collected and analyzed. I take it from the description that 1) subsurface SPM was collected from every green dot (correct?). 2) deeper water SPM was filtered from 5 sites (perhaps these could be indicated in the table), 3) Sediment samples from 8 cores were collected. I cannot tell from the description what depth in the core these samples were taken from (10 cm?), nor how 8 cores yielded n = 68.

Perhaps the dots could be color coded to indicate what types of samples exist (surface SPM, subsurface SPM, these + sediment). It might also be helpful to designate the environment type (river, upstream estuary, downstream estuary) on the map. It would be particularly helpful to indicate the city of Poses/location of the dam on this map.

We agree with the reviewer here. We would change the color of the dots in the map to indicate the different sample types. Locations where only soils were collected would be indicated in black; those where only SPM were collected would be indicated in green; those where both SPM and sediments were collected would be indicated in red. In addition, the location of the dam, as well as information about the environmental type (river, upstream estuary, downstream estuary), would be added to the map. The revised map and caption are proposed below:

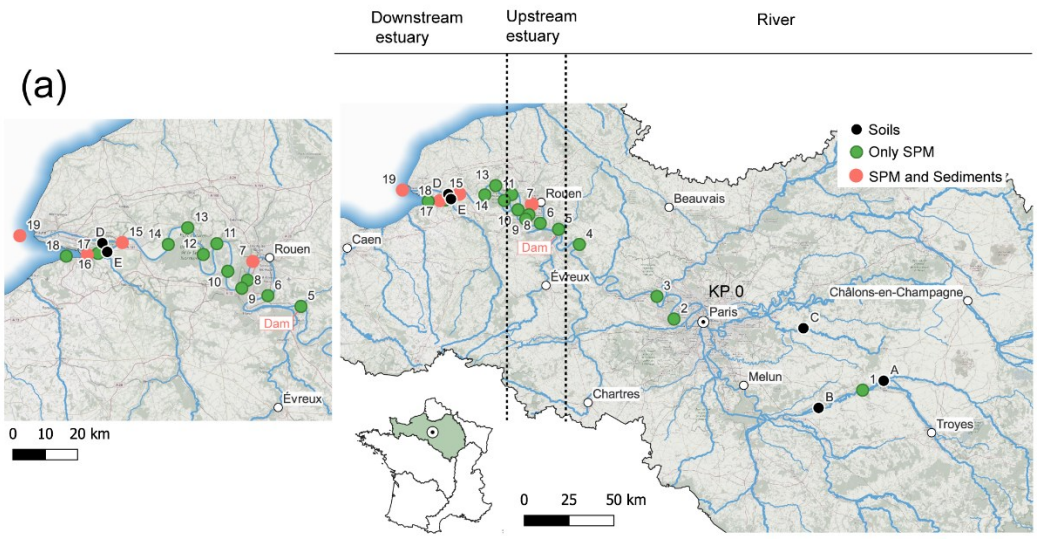


*Figure: Geographical locations of sampling sites in the Seine River Basin (KP: kilometric point, the distance in kilometers from the city of Paris (KP 0)). The sampling sites from upstream estuary and*

*downstream estuary are shown in the zoom-in figure. Sub-surface SPM was collected for all sites from site 1 to site 18, while both sub-surface and bottom SPM were collected at sites 4, 6, 10, 13, and 15.*

To maintain the readability of the map and avoid too many colors, additional details would be provided in the caption as well as in Table 1. Table 1 would allow distinguishing 5 categories of sites depending on the type of samples collected: 1) only soils; 2) only subsurface SPM; 3) subsurface SPM and sediments; 4) subsurface and bottom SPM as well as sediments; 5) subsurface and bottom SPM. We would differentiate subsurface and bottom SPM samples in this table.

Regarding the sediment samples, they were collected from 7 cores (and not 8). This typo would be corrected in the revised manuscript. We would add further details on the sampling strategy as follows:
*" Sediments (n=68) from 7 cores (10-cm depth) were collected in the river channel at the same sites as these SPM samples in 2015 and 2016 using a UWITEC corer as described by Thibault et al. (2019) (Table 1). These sediments were further sliced (1-cm thickness) and freeze-dried. For each core, ten samples were analyzed for brGDGTs and brGMGTs, except for the one collected at site 17 in April 2016, where no lipids were detected between 4-5 and 5-6 m depth."*

What differentiates "upstream estuary" and "downstream estuary"? Is this salinity? Or judgement?
The river and upstream estuary are differentiated by the dam located at Poses. The tide influences the estuary up to Poses, where the dam prevents further tidal propagation. The upstream and downstream estuary are differentiated based on spatiotemporal variations of salinity. The upstream estuary corresponds to the freshwater tidal sector, whereas the downstream estuary is affected by a salinity gradient (e.g. Romero et al., 2015, Environmental Science and Policy; Druine et al., 2018, Marine Geology). This would be clearly specified in a revised manuscript.

Line 237: "correlations" here should be "relationships". These are not correlations in the statistical sense.
This would be corrected.

Line 271? Should this be "decreased in the downstream estuary" samples (not just "downstream")? Having defined upstream estuary and downstream estuary it is good to stay with these terms.
The term "estuarine" would be added in this sentence as well as in other sentences throughout our manuscript.

Line 290. "Negative loadings" is confusing. On which axis? Both? I suggest describing the results by axis – first axis 1, then 2.
To clarify this point, this sentence would be split 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)."*

Figure 3 is not particularly helpful to the reader. If the authors wish to retain it, I suggest moving it to supplemental text.

As suggested by the reviewer, we would move this figure to the supplement (Fig. S2).

Results:

The results of the "bulk parameters" describes the elemental and bulk stable isotopic composition of the solid samples. Nowhere does the paper describe results of other environmental parameters – temperature, etc. It would be helpful to have at least a table indicating the mean and range of these. I expect, for instance, that there is a large range of salinities associated with these samples and a very narrow range of temperatures (they are all close to each other).

We thank the reviewer for this comment. Our revised manuscript would include a new supplementary table (proposed below) to describe the available environmental parameters (temperature, salinity, and water discharge):

**Table S1. Description of available environmental parameters**

| | River | Upstream estuary | Downstream estuary |
|---|---|---|---|
| Number of samples | 6 | 42 | 59 |
| Min temperature (°C) | 20 | 8.49 | 6.4 |
| Max temperature (°C) | 23.41 | 24.4 | 23.38 |
| Mean temperature (°C) | 21.51 | 20.07 | 18.23 |
| Min salinity | 0 | 0 | 0.1 |
| Max salinity | 0.3 | 0.32 | 32.3 |
| Mean salinity | 0.2 | 0.23 | 3.77 |
| Min discharge ($m^3/s$) | 99 | 99 | 99 |
| Max discharge ($m^3/s$) | 138 | 978 | 978 |
| Mean discharge ($m^3/s$) | 116.67 | 180.5 | 222.97 |

The treatment of the soils samples in the analysis and results is difficult to understand. It appears that a large number of soils (up to 34) was taken from some sampling sites, whereas at others 1 sample was taken. These data were then analyzed via PCA separately from the aquatic samples, and the PCA was overlayed onto the PCA of the aquatic samples. The authors conclude that the PCAs show that the GDGT distribution of soils overlap with the SPM and channel sediments. It the PCAs were done separately, one cannot simply overlay the biplots and conclude that they overlap – the PCAs may capture different variance structures such that the PCA axes are not the same. If the authors wish to compare the soils and aquatic samples, do a PCA on all the data together. It's always possible to do a second PCA excluding the soils to evaluate the variance structure of the aquatic samples alone.

This comment was addressed above.

Line 290: "explained 40.9% of the variance in two dimensions". What is meant by this? Based on the plot, axis 1 captures 40.9% of the variance and axis 2 13.2%.

We thank the reviewer 2 for this comment, which was also made by reviewer 1. We would rephrase 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 346: Similar problem. I think the authors mean that axes 1 and 2 capture 71%. The PCA will capture more than this on axes 3 - ???
We agree with the reviewer. The first two dimensions explain 70% of the brGMGT variations. The corresponding 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 separates samples from different parts of the basin."*

Similar problems exist in the description of the RDA, Section 3.3
We would specify that 30.2% of the variance was captured from the first two axes in the revised manuscript.

4.1.1. Why do the authors focus on the 6-methyl brGDGTs here?
We start this section by discussing 6-methyl brGDGTs, as this group of compounds is typically produced in rivers. Nevertheless, this section is also mentioning and discussing the variations of the relative abundances of other types of brGDGTs, especially 7-methyl brGDGTs, across the salinity gradient.

Line 390: The authors suggest that the higher abundances of 6-methyl brGDGTs in upstream vs. downstream samples may reflect degradation:
"It may reflect the fact that riverine 6-methyl brGDGTs are more easily degraded than soil-derived homologues and only partially transferred downstream."
Why would 6-methyl brGDGTs produced in a river degrade faster than those produced elsewhere? The authors argue that this could reflect attachment to particles – but how do these particles differ in upstream vs. downstream river environments.
It seems likely that production of the 6-methyl compounds is suppressed in downstream environments and the dam traps the upstream sediments (and lipids). Can the authors show that this is not the case?

The decrease in the abundance of 6-methyl brGDGTs from the upstream estuary to the downstream estuary cannot be explained by the dam located at Poses, as the latter is separating the riverine part of the Seine and the upstream part of the estuary. There is no dam between the upstream and downstream parts of the estuary (cf. revised version of the map above). Therefore, we favor other hypotheses discussed in the manuscript to explain the changes in 6-methyl brGDGT abundances along the estuary, including 1) preferential degradation of labile (riverine) 6-methyl brGDGTs, as notaby proposed by De Jonge et al. (2015) and 2) dilution by brGDGTs from other sources during downstream transport.

Regarding the first hypothesis, the higher degradation of 6-methyl brGDGTs upstream could indeed be due to the different attachment to particles upstream *vs*. downstream. The median diameter of the SPM was monitored between February 2015 and June 2016 in the upstream (sites 7 and 10) and downstream

(sites 15 and 17) parts of the Seine Estuary (Druine, 2018: https://theses.hal.science/tel-01896520). Upstream, the size of the particles showed only a slight dispersion (80-110 µm) whatever the hydrological conditions. The homogeneity of the size of the particles in the upstream estuary likely reflects their predominant continental origin (i.e. Seine river before the dam of Poses). In contrast, a large variability in the size of the SPM particles was observed in the downstream estuary (15-20 µm to 80-90 µm), related to the complex flocculation and defragmentation processes of the particles occurring in this part of the estuary (Druine, 2018). Therefore, the variability in the size of the SPM particles from upstream to downstream could have an influence on the brGDGT distribution in the Seine estuary. This point would be discussed in a revised manuscript using the aforementioned data.

Line 405. Here the authors suggest that the brGDGT distributions in estuarine soils is similar to that of the downstream samples, based on the PCA (see comment above about the PCA). In the next section (4.1.2), this is not discussed and instead production of the brGDGTs in saline environments is the primary factor accounting for compositional differences in upstream vs. downstream samples. Please coordinate these ideas.

Since PCA alone does not allow distinguishing brGDGT distributions between soils and downstream estuary samples, we further applied a machine learning approach as suggested by Reviewer #1. This method supports the *in situ* production of brGDGTs by effectively distinguishing the brGDGT distributions between downstream estuary and soil samples. As brGDGTs are produced *in situ*, we can explore the compositional differences of these compounds from upstream to downstream and investigate the controlling factors. This would be discussed in a revised manuscript, as also detailed in the reply to reviewer 1.

Line 487, 559: One cannot conclude from concentrations alone that the GMGTs are produced in aquatic environments. Soils contain abundant coarse clastic material that may be lost in the fine SPM and river sediment. The distributions (relative abundances) of GMGTs are key to identifying in situ production.

We fully agree with reviewer 2. The relative abundances of brGMGTs are essential to identify *in situ* production in the estuary, especially through machine learning approach. This comment was addressed above.