We thank the reviewer for their very detailed comments, which prompted us to reflect deeply on several of the points raised. We hope that our responses are sufficiently precise and well-supported. The original comments are in bold font, and the responses in normal font.

General Comments:

This manuscript presents an approach to reduction of large quantities of spatially characterized isotopic data obtained by raster scans of samples using LA-ICPMS. Image reduction programs already exist. The main novel feature of this one seems to be the ability to average data from virtual spots in order to achieve an optimal balance between precision and spatial resolution. As such, it is of potential use to a large number of geochemists and geochronologists. I found it to be very interesting, although I have no experience in using a femtosecond laser and have never tried elemental imaging.

We thank you for this overall positive comment.

Specific Comments:

The manuscript could use a more extensive and clearer discussion of limitations and how these might be improved.

The main limitation of our approach concerns samples with low U and/or Pb concentrations (low number of counts), for which biased ages may be obtained depending on the regression method used (TW or Wetherill). These points are discussed in more detail below. We will place greater emphasis on this issue and propose a workflow for applying the approach, following a suggestion from Barbara Kunz and Igor Figueiredo (reviewers 2).

The authors use a single collector mass spectrometer so masses are measured at different times. The lateral movement between mass scans is about 1.4 mic while the pulse separation is about 0.05 mic. However, spatial resolution of data should be mainly limited by the ca. 0.6 sec transfer function (washout time) of ablated material into the plasma (in contrast to line 465), which covers a range of about 15 mic of beam movement along a scan. The observed distribution of isotopic ratios in the direction of a scan should be the actual distribution convoluted with the transfer function of the instrument. This would cause data variations along a scan at a scale lower than about 15 mic to be smeared together by the response time of the instrument, which is much broader than the distance covered by a mass sweep (1.4 mic).

We agree that from an LA-ICP-MS imaging quality perspective, not accounting for the longer washout time when choosing the mass sweep duration (whether averaged to a higher value or not) can result in image smearing. However, Figure 2 shows that this smearing remains limited with a washout time of approximately 500 ms. Visual image resolution is not the primary objective in this case, although, of course, the highest possible resolution is desirable. The guiding principle here is to retain all measurements (each mass scan), even if signal mixing occurs, because the goal is to generate virtual spots, by analogy with static spot ablations. Signal mixing is common in static ablation, where a more homogeneous signal is sought to improve statistical reliability. In such cases, each ablation takes a different volume of the sample, but the overall signal is mixed/homogenized. A higher sampling time improves the statistical uncertainty on the mean, which decreases with the square root of N. In the case of the images produced here, the signal will not be as homogeneous as in static ablation, since the system's washout time remains relatively short (500 ms). This value represents a compromise between limiting image smearing along a line and obtaining a sufficiently reliable uncertainty on the mean. It should be noted that averaging the mass scans can have a beneficial effect on age accuracy for low-concentration samples, as illustrated by the example of C6-265-D5. Conversely, the example of BH14—based on both the results from Hoareau et al. (2021), who used an averaged mass scan of 540 ms, and the present study—shows that results are more satisfactory without averaging the mass scans. Therefore, for each sample, it is reasonable to test multiple configurations (with and without averaging) to select the most accurate and precise one.

Therefore, defining the mass sweep as a pixel seems a bit misleading. Pixels, in the sense of a fundamental area scale at which independent values can be measured, are really more like the size of the spots but they vary along a continuum so it might be better to avoid the term pixel altogether.

We initially used the term *pixel* in its general definition, which goes beyond isotopic imaging, namely 'the smallest addressable element in a raster image' (Wikipedia). This definition is independent of any potential signal correlation between adjacent pixels and is easily understandable, even for non-specialist readers. As an example, in the case of digital images, increasing the resolution from 100 px × 100 px to 200 px × 200 px multiplies the number of pixels by four based on the original image, with the new pixels necessarily being correlated (in terms of colour). Nevertheless, they are still referred to as pixels.

We prefer to retain the term *pixel* as a clear and accessible term, while adding a clarification in the Methods section of the manuscript to address the issues you raise from the perspective of LA-ICP-MS imaging.

The relatively slow instrument response time is largely the result of using a nebulizer chamber where carrier He and makeup Ar are mixed before injection into the plasma. Response time can be significantly lowered by the use of a

modified, commercially available, quartz injection tube where mixing between the He carrier gas and Ar occurs within the plasma. Further improvement would require the use of a multi-collector or time of flight mass spectrometer in which signals for different masses are effectively measured simultaneously.

We fully agree with these remarks, which go beyond the scope of the present study. The instrumentation used is not necessarily the most efficient for LA-ICP-MS imaging, and we hope to benefit from more advanced setups in the future.

The scale of observable compositional variations across (rather than along) scans will be limited by the beam size (25 mic).

Yes. To be more precise, the beam size is 15 μm but galvanometric scanners allow to rapidly move the beam to generate a 25 μm-width line, but the result is the same.

It should also be noted that although the scan line is claimed to have a depth of 40 mic at 500 Hz, the troughs should have a V-shaped profile for non-overlapping lines so the average depth would be 20 mic. One should therefore be mapping triangular segments beneath the surface. This should of course affect ablation bias so standards need to be analyzed the same way.

We fully agree. The troughs broadly have a V-shape, with the highest depth reaching about 40 μm as measured with a digital microscope. This maximum value is not an average. It is this maximum penetration depth that is critical for us when analysing thin samples, which must not be traversed. In any case, we always use strictly similar parameters to analyse the standards.

A more fundamental problem is the fact that regression of relatively imprecise ratios produces ages that can be significantly different using the Wetherhill plot and the more commonly used Tera-Wasserburg (T-W) plot.

We agree. This is clearly shown by the example of sample C6-265-D5.

I assume the reason is that one measures counts on the masses but one plots and calculates with ratios of these counts. Errors in the numerator mass count will propagate linearly whereas those in the denominator mass count will only propagate linearly to first order so second order effects become important when the relative error is large. This was the main reason that I developed the approach of regressing signals in a 3D space where the Wetherhill solution generally agrees most closely with the 3D solution for poorly radiogenic data sets and both disagree for highly radiogenic data sets (see Davis and Rochan-Banaga, 2021, Fig. 5).

Your comment is very interesting; indeed, we did not address this specific point, which you discuss in detail in Davis and Rochin-Banaga (2021). Clearly, this was not an aspect we intended to cover here, as we are not able to provide further insights beyond those you have already offered. We tested sample C6-265-D5 using the Brama2.0 software developed by Liu et al. (2023), which is a Python-based equivalent of UtilChron. We were also unable to obtain a satisfactory age. The initial regression in Wetherill space yields an age of around 115 Ma, whereas the Bayesian solution gives an age of approximately 63 Ma. This discrepancy is not resolved by averaging the mass sweeps in groups of 8, for instance (as done in Hoareau et al. 2021).

The reason can be visualized by considering that non-linear variations in the denominator isotope in T-W plots (^{206}Pb) will bias data along a diagonal, which usually is at a high angle to the mixing line (isochron), whereas in the Wetherhill plot the denominator isotopes are from U, which is usually the largest peak (lowest error) and the common Pb end member is at infinity so non-linear variations will tend to be sub-parallel to the mixing line unless the datum is very radiogenic. The manuscript demonstrates this clearly but does not offer any discussion on how to effectively deal with it. One way would be to do regressions in 3D but there may be ways of correcting for it in 2D.

In the manuscript, we propose resolving these age inconsistencies between TW and Wetherill regressions by averaging the number of mass scans (smoothing), even though this results in larger individual uncertainties for the virtual spots. For easily datable samples (e.g., ARB20-2D or BH14), TW and Wetherill ages agree within uncertainties, and smoothing is unnecessary: it does not change age values but increases their uncertainty. For low-concentration samples (e.g., C6-265-D5), smoothing helps eliminate erratic isotopic ratios caused by too many zero or negative counts for ^{207}Pb and ^{206}Pb, which usually result in overly high ^{238}U/^{206}Pb and ^{207}Pb/^{206}Pb ratios in TW diagrams. These ratios bias the regression towards a shallower slope and thus a younger age. We agree that this effect should be less pronounced in Wetherill regressions. However, the C6-265-D5 example shows that averaging the mass scans alters both the TW and Wetherill ages, which then converge towards a common value. We consider this convergent value more reliable, although it should ideally be tested on samples of known age.

To further expand on this very interesting issue, we present here a point we discussed in 2024 with Pieter Vermeesch, concerning both the age differences between TW and Wetherill regressions and the discrepancies in results between Isoplot and IsoplotR. Taking sample C6-265-D5 as an example, without averaging the mass scans and using virtual spots of 100 µm × 100 µm across the entire image (n = 42), the age obtained with Isoplot from ^{206}Pb/^{238}U and ^{207}Pb/^{235}U ratios in Wetherill space is 140 ± 25 Ma, identical to the age calculated using IsoplotR. In contrast, the age obtained from ^{238}U/^{206}Pb and ^{207}Pb/^{206}Pb ratios in TW space using Isoplot is 84 ± 16 Ma, which is higher than the age proposed by IsoplotR (76 ± 17 Ma). This discrepancy stems from differences in the regression methods used by the two software packages:

1. In Isoplot, regressions—whether in TW or Wetherill space—are performed using the York algorithm. The software can compute four different ages for a single sample depending on the chosen ratios and coordinate space.
2. In contrast, IsoplotR always performs regressions in Wetherill space, regardless of the input ratios (whether $^{207}Pb/^{235}U$ and $^{206}Pb/^{238}U$, or $^{238}U/^{206}Pb$ and $^{207}Pb/^{206}Pb$). The software, which uses the York and/or Ludwig (1998) algorithm (both yielding identical results), thus calculates only two ages. The results are then simply displayed in either the TW or Wetherill plot, depending on the user's choice.

Therefore, if one uses $^{238}U/^{206}Pb$ and $^{207}Pb/^{206}Pb$ ratios to perform a Wetherill regression in Isoplot, the resulting value will be identical to the one obtained in TW coordinates using IsoplotR, i.e., 76 ± 17 Ma. For easily datable samples such as BH14, both software tools yield identical or nearly identical results regardless of the chosen method. The discrepancies observed in bad samples such as C6-265-D5, as noted by Pieter Vermeesch, illustrate that "*both age estimates are likely biased and wrong*". In our view, comparing TW and Wetherill age estimates should become a systematic prerequisite to assess the robustness of U-Pb geochronology results in general.

The best application that I can think of for this method would be WC1. This is an excellent standard because of its relatively old age and high U concentration but shows evidence of not being homogeneous in age based on the high ID-TIMS age error (2.7%) and the work of Guilong et al (2020, https://doi.org/10.5194/gchron-2019-20). If it were possible to isolate the predominate phase, this would be a much more useful standard.

This is an interesting comment, although we believe that our approach may also prove useful for other types of samples.

Technical corrections:

If one choses not to indent paragraphs there should be a space left between them, as well as between references.

OK. We will check this in the new version.

In some places I found the phasing unclear or ambiguous. Suggestions for improvement are made on an annotated copy.

Thank you for the time spent for the phrasing. We will try to follow all suggestions made.

The authors refer to Wetherhill Concordia plots as 'concordia' and the inverse (but more commonly used in this application) Tera-Wasserburg concordia plot

as T-W. This is confusing because both are concordia plots. It would be better to refer to T-W and W plots.

We agree that it is unclear as presented. We will make the appropriate changes.

Line 62:

Presumably the authors mean 207Pb/235U, but why would one want to use this ratio, rather than the more precise 206Pb/238U ratio as a criterion for sorting? They both encode the same information (age and radiogenicity).

We suggest to directly contact Kirsten Drost to discuss this, as it is not our proper work. The problem with $^{206}Pb/^{238}U$ sorting is that it tends to overweight outliers with high $^{238}U/^{206}Pb$ ratios in the TW space, resulting in results biased towards too young ages.

Line 73:

The 3D regression also allows for editing outliers. This aspect is a separate problem from the best regression approach as discussed above.

We agree. However, we wanted to emphasize that “bad quality samples” may also give biased results with the Bayesian approach: the prior estimate is obtained with a York regression on individual pixel values to which uncertainties are added from Poisson statistics. If the Isoplot estimate is biased for the reasons exposed above, the Bayesian result might be biased too, as we understand from our experience with this approach.

Lines 125, 129, Table 1:

A better use of English would be to refer to line width (rather than height) as the diameter of the laser beam and line length (rather than width) as the scanning distance. Otherwise it can be confusing to the reader.

OK. We will change the terms in the next version of the manuscript.