

RC1:

We would like to thank Reviewer 1 for suggestions that helped to improve the current study. We agreed with all the suggestions and as such, made the necessary changes/additions. Please find our responses to comments in blue.

General Comments:

“The analysis of the data seems quite superficial, in that while Deming regression is an appropriate test, the intercept is fixed at zero and nonlinearities are largely ignored. This to me is an omission, because many elements of interest such as Pb, Cr and As very clearly demonstrate nonlinear relationships in the scatter plots. Given the remit of the paper, these must be discussed both in terms of the possible causes and implications. It may also be prudent to perform an additional numerical fit to quantify this effect, such as using a different fit function or testing a nonzero intercept from the Deming regressions. A cursory inspection would indicate to me that if the intercept on arsenic in particular were to be allowed to vary, a much more favourable slope would be yielded (for whatever reason), so these effects cannot be ignored.”

We agree that not including a secondary fit was an oversight. Therefore, a full re-analysis using Deming regressions without forcing through 0,0 was added to the supplement (Table S8 and Figure S6) and for two elements that indeed showed non-linear relationships, Pb and As, the intercept-fitted analyses were used in the main body as opposed to results forced through origin (Figure 8). Please see line 328 onwards in the track-changes document for the justification of retaining the origin-forced regressions as the main body (except for As and Pb), and line 388 onwards for the full analysis of non-origin forced regressions. Charts for the intercept-fitted regressions and a comparison of intercept vs non-intercept results are now found in the supplement.

“There needs to be more information regarding the offline sampling and analysis. While some of this may be considered obvious to those familiar with the reference methods, these must still be stated for the sake of those who are not. Specifically:

- More information on the Leckel samplers should be given (e.g. inlets, blank strategy)
- The manufacturer(s) and product codes of the filters used for offline analysis should be given, along with specification (in particular pore size) and any pre-treatment (e.g. washing, baking).
- The 'shaker' used for the aqueous extraction needs describing better and the product information on the syringe filter included.
- More information given on the extraction process used for ICP, in particular the use of a microwave digester (I think this is implied, but not stated) and the fact that the substrate is included in the digestion.
- While the digestion protocol is standard, the authors should still comment on the acid mixture used in the digestion and the likely recovery rate (particularly of silicates), noting that other mixtures based on aqua regia and hydrofluoric acid are used elsewhere.”

Based on the suggestions, we have added thorough detail to every stage of the methods process as requested (section 2.6).

“The conclusions does not feature nearly enough quantitative information, with the comparison using descriptive terms instead. Given the slopes are crucially important, these should be referred to

and summarised quantitatively with reference to a range that the authors would consider acceptable. Furthermore, quantitative comparisons of these compared to previous equivalent literature should also be included.”

We have added quantitative information to the conclusion, including explicit references to regression slopes and references to the previous literature (with the full comparison to previous literature found towards the end of section 3.6), strengthening this section.

Specific comments:

“Title: That the instruments are compared to ICP-MS as well is important because this represents an independent point of reference, so this should be incorporated into the title.”

We have added ‘vs ICP-MS’ to highlight this comparison.

“Line 51: The authors should refer to the most recent directive (2024/2881), even though the limits were maintained from previous.

Updated to current directive.

Line 107: A map should be included, showing the location of the sampling site and context such as the distance from major pollution sources.

Location of the measurement site added as suggested by the reviewer (Figure 1)

Sections 2.2.1/2.2.2: Because the manufacturers regularly modify instruments, more information on the specific physical units used should be given here (ideally hardware revision numbers). Furthermore, because some of the discrepancies found are later attributed to software, the version numbers of the inversion software should be given.

Hardware revision numbers, software versions (lines 139-140 and 153-154) and technical specifications (Table S1) are added for each instrument.

Figure 1: The means by which this data was calculated should be more explicit. 'Relative Accuracy' is not very specific.

Added in line 275.

Line 272: 'Gravimetric' is not an appropriate term because this refers to the weighing of filters. Beta attenuation is a 'gravimetric equivalent' measurement.

Corrected to gravimetric equivalent.

Figure 2 (and elsewhere): "Total Concentration" should not be used because it could easily be confused with PM. "Total Detectable Concentration" would be more appropriate.

Axis title has been changed to Total Detectable Concentration.

Line 293: Worth stating that Deming regression takes the uncertainty of both variables into account. It should also be stated that the intercepts were fixed at 0 (may seem obvious, but it is important to state)."

This has been clarified in lines 324 and 325 as well as under charts.

"Figure 3 (and other equivalent plots): The use of grey makes the plots hard to read, in particular the whiskers. Suggest using different colours. It would also be a useful guide to the eye to include the 1:1 line on these plots as well."

Charts now use red instead of grey as well as having a 1:1 line for clarity.

"Line 329: "Both departments" is a very strange turn of phrase. Revise."

We have changed 'departments' to 'metrics'.

"Line 468: The term "offset" I would take to mean a systematic additive error, but the analysis performed here would be insensitive to this because the intercepts were fixed at zero. The authors should be more specific about what they mean here. If they are saying there is such an error, then the fits should be performed again without the constraint."

The reviewer is correct in saying the use of the term 'offset' is wrong here. Instead, it has been changed to 'systematic differences' which better suits the phenomenon seen here. The text now includes reference to the intercept-fitted analysis to better explain these observations.

"Line 488: Given that the measurement site was part of various networks including EMEP and ACTRIS, surely some of the data is in the public domain?"

As the Xact625i is not a 'standard' instrument (as opposed to filter-based elemental analyses whose data are public) for these networks, data are not published. However, upon discussion, we have decided to publicly upload the data, and as such, have changed the data availability statement: "Data availability. The datasets generated and analysed in this study will be deposited in Zenodo and made publicly available upon publication of the article at <https://doi.org/10.5281/zenodo.17198675>".

RC2:

We would like to thank RC2 for their comments and recommended suggestions. We have made the necessary changes, strengthening the paper for publication. Please find our answers in blue.

Major comments:

“Methods, Section 2.3: The Xact 625i also has an internal standard every hour – does a similar check exist for the PX-375? Please include the internal standard in a paragraph”

We confirm that the PX-375 in fact does not include an hourly standard, unlike the Xact625i.

“Methods/General: It might be useful to have a table with the main parameters compared for the two instruments to have an easy overview – can be in supplementary.”

We agree with this suggestion and have added a comparative table of instrument specifications to the supplement (Table S1).

“Methods, Section 2.6: Sentence: More details are needed, such as filter manufacturer, blank levels, recovery etc. or paper reference to method if it is described with those details elsewhere or possibly reference method that is followed.”

We have added through details for this methods section including all suggested additions.

“Results and Discussion, Section 3.1, L237 onwards: Is the analysis area the same in the Xact 625i and the PX-375? If not, does that need to be considered when using the Xact standards?”

We thank the reviewer for raising this point. The analysis areas of the two instruments are indeed different: the Xact625i analyses a smaller fixed spot on the Teflon tape, while the PX-375 scans a larger portion of the filter tape. This difference could contribute to minor variations when using standards. However, since the Micromatter foils used here are uniform across their surface, the different analysis areas do not affect the validity of the test.

“Results and Discussion, Table 1 and associated figures: Please include a confidence interval for the slope. Also, it might be useful to not have the intercept fixed at zero to see if the offset (with confidence interval) is significant.”

We have added confidence to all slopes found in the paper. A full secondary analysis using Deming regressions with fitted intercepts has been added to investigate potential differences in the analyses. While the original analysis has been kept as the main analysis, the fitted intercept analysis was used for Pb and As that show nonlinear relationships. The fitted intercept charts and a table comparing results are found in the supplement.

“Results and Discussion, Section 3.6: I think this section would benefit if the comparison included 24hr means between the two ED-XRF methods, so it is more directly comparable to the filter comparison as well as the high time resolution comparison.”

We have added a table comparing results of the 2-hour and 24-hour averaged data in the Supplement. While we acknowledge that 24-hour averages provide direct comparability with offline filters, we chose to retain the 2-hour resolution as the main analysis, as this maximises sample size, avoids uncertainty from incomplete 24-hour coverage, and better represents the instruments' normal operating resolution (see line 457 in the tracked changes document).

“Introduction, L41: Sentence “Although low in weight contribution...” should be “Although low in mass contribution...”

Corrected ‘weight’ to ‘mass’.

Introduction, L43: I would change this to “..., most elements are only slightly affected by...”.

Corrected as suggested.

“Methods – please review method subchapter numbers as there seems to be a jump from 2.1 to 2.2.1”

A title “2.2 Online ED-XRF instruments” has been inserted.

“Methods, L115: Sentence “Online trace elements were performed...” should be “Online measurements of trace elements were performed...”

Thank you for pointing out this error, it has been amended.

“Results and Discussion, L263: Sentence “...data points below the LoD were kept in the analysis must be treated...” should be “...data points below the LoD were kept in the analysis but must be treated...”

We have changed this sentence for clarity: “Elements such as V and Cr that had 60-80% of data points below LoDs **were kept in the analysis, but their low data coverage was taken into account when attempting to draw conclusions from their results.**” (Line 292 onwards).

Here, the term ‘treated’ wrongly implied treatment of data – instead, these elements were treated the same as the rest of the data, but kept in the analysis despite low data coverage and as such were not used to make strong conclusions based on their results.

“Results and Discussion, L404 onwards: Sentence “Higher r^2 values in the PX-375 vs Xact625i comparison were found compared to ED-XRF vs ED-XRF across all elements” please revise as I think it should be ED-XRF vs ICP-MS.”

We understand that this sentence was not clear – we have clarified that this section is regarding the **direct internal comparison of ED-XRF vs ED-XRF using 2-hour data.**

“Results and Discussion, Figure 8: It might be worth showing the same energy range on the X-axis for easier comparison.”

Here, the difference in energy range originates from the excitation conditions of the two instruments. The Xact625i EC1 operates at 25 kV, allowing excitation and detection of fluorescence lines up to ~6 keV (e.g. Ti K α at 4.51 keV). In contrast, the PX-375 'EC1' operates at 15 kV, so effective excitation is limited to lighter elements, with measurable peaks only up to ~4.5 keV. Beyond this energy range, the PX-375 spectra contain no signal of interest.