Reviewer 2 report:

Replies from authors will be in blue and italic.

This study addresses the crucial issue of cation retention in soils following rock dust treatments aiming to capture carbon via enhanced weathering and alkalinity export. Cation retention has the potential to delay carbon capture by years, decades or even centuries, with serious implications for monitoring, reporting and verification for ERW experiments. Using a range of soil types and feedstocks, most of these greenhouse-based mesocosm experiments described here ran for 650 days, longer than many previous laboratory studies. The mesocosms were all sown with *Lolium perenne* seeds with twenty earthworms. Leachate chemistry allowed quantification of CO2 consumption, while sequential extractions allowed determination of the likely fate of the cations released by the weathering feedstocks. Results indicate that cation retention and alkalinity export are very sensitive to the soil properties of cation exchange capacity, inorganic carbon content, and presence of swelling clays, as well as feedstock mineralogy and particle size distribution. The authors quite sensibly suggest that processes of cation exchange and carbon dynamics may lead to retention of feedstock cations and redistribution of cations associated with the inorganic carbon content of calcareous soils, for example. However, they consider that other processes could enhance weathering of native soil minerals, such as changes in hydrology due to swelling of clays coupled with greater pCO2 concentrations due to the Birch effect.

General comments

This study is timely and and the manuscript well written; the overall experiment is well replicated, and the qualitative discussion is sensible. However, there is a complete and perplexing lack of any statistical analysis. The figures often include uncertainty envelopes and/or error bars which are never explained, and there are no uncertainties of any kind (not even standard deviations or standard errors) in the tables of results. Where a single bar or value is given for a particular soil/feedstock combination, it is not clear what those represent; are they averages of the replicates? This study deserves a thorough statistical analysis before it is published. The statistical methods will need to be explained in a subsection of the Methods, and the meaning of all uncertainties, error bars or confidence envelopes clarified in the captions of the tables and figures.

We are sincerely grateful to the reviewer for the time and care invested in evaluating this manuscript. We appreciate the thoughtful and detailed assessment, which has identified important areas where the presentation and statistical treatment of the data can be strengthened. These comments are extremely valuable, and we are confident that addressing them will substantially improve the rigour, clarity, and overall quality of the manuscript.

Specific comments

Line 132: Please define Ø; is this radius or diameter?

We thank the reviewer for pointing this out, we changed it to 'diameter'.

Table S1 is comprehensive, but it does lack bulk density and mineralogy for these soils. Bulk density is undoubtedly available for these soils so this should be included; mineralogy should be included if available.

We agree that the inclusion of additional soil properties would improve the usefulness of Table S1. Mineralogical information is available for these soils, and we are happy to add it to the Supplementary Information. In addition, we will provide the corresponding bulk density values in the revised version.

Did the serpentinised peridotite contain any chrysotile, or other asbsestiform minerals?

The reviewer correctly refers to the potential health risks of the presence of asbestiform serpentine or amphibole minerals in an altered ultramafic rock dust. Quantitative XRD showed that this material contained no amphibole, and about 6.6 wt% of serpentine. No further analyses were undertaken to know exactly which type of serpentine minerals were present, or whether they were asbestiform. As this material is quarried in Norway and sold on the European market, the mining company is under legal obligation to regularly check its product for the presence of asbestiform minerals, and in case any are found, the corresponding batch of rock can not be sold. We therefore presume that there was no asbestiform chrysotile in this feedstock.

Were the sequential extractions (Table 2) also performed on these soils to determine the initial conditions before the experiment? As Table S1 gives effective CEC (NH_4Cl) and Table 2 determines the exchangeable fraction using NH_4Ac, the Table S1 data do not provide the initial conditions. It would be good to know whether CEC changed as a result of the treatment; it does not look like the controls are completely stable with time in Figure 8.

The initial mixtures were not analysed by sequential extraction prior to the start of the experiment, and therefore these data are not currently available to provide direct initial conditions for comparison with Table 2. We agree that such information would be valuable, particularly for assessing potential changes in exchangeable pools and CEC over the course of the treatment. In the revised manuscript, we will clarify this limitation more explicitly and note that the effective CEC values reported in Table S1 are not directly equivalent to the NH_4Ac-extractable exchangeable fraction shown in Table 2.

Lines 267, 300, 329, 402, 550, 572, 628, and 632: Does "significant" mean that some measure of statistical significance has been determined? If so, what statistical test was done?

Are any of the results shown in Figure 2 considered to be statistically significant? What statistical tests have been done?

Figures 3-7: Please explain how the variability was calculated; are the shaded envelopes 95% confidence intervals, for example?

We agree that the use of the term “significant” and the presentation of variability in the current version are not sufficiently clear. As also noted by Reviewer 1, the manuscript requires a more explicit description of uncertainty, variability, and statistical treatment.

In the revised manuscript, we will therefore revise the wording where needed to avoid ambiguity and will clearly indicate where statistical significance is meant and which statistical test was applied. We will also expand the figure descriptions to explain explicitly how variability was calculated and what the shaded envelopes represent.

More generally, as also outlined in our response to Reviewer 1, we will add a dedicated section describing the statistical approach used to process and interpret especially the leachate data. This will include an analysis of replicate variability, an assessment of whether differences among treatments and controls are statistically significant, and clearer reporting of uncertainty in the text and figures. For the leachate data, treatment responses will be summarised as mean ± one standard error, and inferential statistics will be based on a single endpoint value per replicate, calculated as the average over the final 45 days. Treatments will then be compared against the control using a two-sided Mann–Whitney U test, and Hedges’ g will be reported as the primary measure of effect size. These revisions will also clarify the interpretation of Figure 2, Figure 9 and Figures 3–7.

Lines 276, 295, 332, 333 and several of the figures: What is "mmeq"? On googling mmeq, Google AI came up with "morphine milligram equivalents" which is unlikely to be appropriate here! "meq" is sufficient for milliequivalents, if that is what was intended. Or is it microequivalents?

We thank the reviewer for identifying this error. The notation “mmeq” is incorrect; the intended unit is “meq” for milliequivalents. This will be corrected consistently throughout the manuscript and in the relevant figures.

Table 1 says (for example) that there were a total of 7 replicates per feedstock for the LUFA 2.2 B and 6S B soils. What is being shown in the panels of Figure 8? Are these averages over the replicates; in which case, where are the error bars for the various pools?

The reviewer raises an important point, and we agree that this aspect needs to be clarified more explicitly. The panels shown in Figure 8 represent results from individual pots rather than averages across the full set of replicates reported in Table 1. Consequently, replicate variability is not represented in this figure, and only measurement uncertainties can be shown.

We will revise the manuscript accordingly to make this explicit in both the figure caption and the main text, so that the relationship between Table 1 and Figure 8 is clear to the reader.

How was the effect size in Figure 9 calculated?

The effect size shown in Figure 9 was derived from the average values presented in Figure 2 of this study, using the difference between treatment and control as described in the figure caption. For the literature comparison, effect sizes were calculated from the corresponding treatment–control differences reported in the respective studies.

Does the CEC (Figure 9) reflect similar buffered or unbuffered extractions e.g. NH4Ac as in the Kelland et al experiment? It looks like the effective (unbuffered) CEC reported in Table S1 is being shown in Figure 9; these values may not be directly comparable to the literature values.

We agree that the terminology and comparability of the CEC data require clearer treatment. In the revised manuscript, we have replaced the general abbreviation "CEC" with the explicit term CEC_eff throughout the text where effective cation exchange capacity is meant. This clarifies that the values shown in Figure 9 and Table S1 refer to effective, unbuffered CEC.

In addition, we re-examined the methodological descriptions in the cited literature references and will clarify these in the manuscript where relevant. We also contacted the authors, as this detail is missing from a few of the published papers.

Application of Mg-rich, fast-weathering feedstocks could indeed result in displacement of Ca from exchange sites , followed by either Ca export or even carbonate precipitation if saturation indices permit. Looking at Figures S2 and S3, however, it is a little bit difficult to see the changes in the cation concentrations being discussed here. Perhaps the colour contrast between the Ca and Mg is not great enough; consider either making the blue of the Ca lighter, or the turquoise of the Mg darker.

We agree that the distinction between Ca and Mg in Figures S2 and S3 can be improved. We will revise the colour scheme of these plots to make the two cations more clearly distinguishable.

Line 506: Does "this study's experiments" refer to the study of Paessler et al? I assume "this study" refers to the study being presented in the current manuscript, but when reading this sentence it did seem at first to refer to Paessler et al's experiment discussed in the previous sentence. Please replace with "Paessler et al's experiments" if that is what is intended here.

We agree that the wording may lead to confusion between the experiments of Paessler et al. and those presented in the current manuscript. To make this distinction clearer, we

will revise this section by inserting a separate paragraph to make the difference between the outdoor lysimeter experiment and the greenhouse experiment clearer.

Paragraph beginning line 524 re swelling clays: No mineralogical data are presented for these soils, but if such data were available, it would be interesting to know whether there was a difference between smectite content in the 2023 and 2024 soil batches (assuming smectite, or other swelling phase, is present). Such a difference could also reflect heterogeneity of the site from which the contributing soil samples came, even though the LUFA Speyer staff undoubtedly made efforts to avoid such effects.

The reviewer raises a very interesting aspect. It would indeed be valuable to know whether differences in swelling clay content, such as smectite, existed between the 2023 and 2024 soil batches. Based on the available data, the difference in 2:1 layer silicate content between the batches is around 5.7%, which may indicate some variation in this mineral fraction. However, we do not have more specific mineralogical data, and in particular no quantitative information on smectite content, to assess this in greater detail. We will note this in the revised manuscript as an interesting aspect that could not be resolved with the available dataset.

Figure S5 (Soil pCO2 dynamics in LUFA 6S): What does the Experiment ID mean? Are the shaded areas around the curves confidence intervals? The depth(s) at which pCO2 was measured should be clarified in the caption.

We agree that Figure S5 could require a clearer explanation. In response, we will replace the experiment IDs with more descriptive names, clarify in the caption what the shaded areas represent, and specify that the pCO2 measurements were taken at half the depth of the bucket. This should make the figure easier to interpret.

Are the water-holding capacities (Figure 11) essentially field capacities, or do they reflect the porosities?

In Figure 11, 'water-holding capacity' was used to refer to field capacity rather than porosity. We agree that this terminology may be ambiguous and will therefore revise the manuscript to use the more precise term field capacity throughout.

Line 565, Section 4.4: It is a good idea to remind readers of the pH and CEC of the LUFA 2.2 B soil being discussed; however, this is an effective CEC and perhaps that too should be reflected here (e.g. effective CEC ~ 9.5 meq/100g)

We agree that the term should be stated more precisely here. This point has already been addressed in response to a previous comment, and we will revise the manuscript to use the explicit term effective CEC (CEC_eff) consistently throughout the text, including in this section where relevant.

Tables 3 and 4 would benefit from some uncertainties, given the number of replicates. It is not clear what is being shown in these tables; are they average values over the replicates? At the very least, standard deviations or standard errors should be shown

here (making it clear which one of course). These tables would also benefit from some indication of which soils these are (e.g. calcareous, acidic) so that readers do not have to refer to Table S1.

We agree that Tables 3 and 4 would benefit from clearer information on uncertainty and interpretation. The values shown in these tables are based on single-pot measurements taken at specific time points during the experiment, so they do not represent averages across replicates. For this reason, replicate variability cannot be reported directly in these tables. However, we will include the analytical error in the revised version to provide an indication of measurement uncertainty.

This also reflects the general challenge of working with small-pot experiments, where destructive sampling is required to obtain soil material. In other words, a soil sample can only be collected by destroying the entire pot at a given time point. Otherwise, a representative sample is not obtained and the hydrology is disturbed (i.e. preferential flows are created). Consequently, these values should not be interpreted as the fully defined, fixed properties of the soils at that exact stage, but rather as measurements from individual pots within the experimental system. At the same time, the broader range of behaviour is supported by the corresponding measurements available from the 6- and 12-month sampling points.

Additionally, we agree that clearer identification of the soil types in the tables would be beneficial, and we will revise them accordingly to make it easier for readers to distinguish between the calcareous and acidic soils without referring back to Table S1.

Line 593: The variability mentioned here needs to be quantified and presented.

We agree that the current wording is too vague and should be clarified. Because the solid-phase results discussed here are based on single-pot destructive sampling, variability cannot be quantified directly for these measurements in a statistically robust replicate-based manner. We will therefore revise this statement to avoid implying a level of quantification that the dataset does not support and instead state more clearly that, while the absolute magnitudes vary across the available individual pot measurements, the overall trends and mineral-specific cation signals are consistent across the datasets. This limitation will also be made more explicit in the revised manuscript.

Figure 12: Why are the panels labeled "Table 0" and "Table 1"? Please label them with the soil names instead, showing the pH, effective CEC and carbonate content as well so that readers can immediately understand the differences between these soils without having to go to Table S1.

We agree that the current panel labels are not sufficiently informative and confusing. In the revised figure, we will replace “Table 0” with LUFA 2.2 B and “Table 1” with LUFA 6S B. We will also add the relevant soil properties directly to the figure, including pH,

effective CEC, TOC, and TIC, so that readers can more readily understand the differences between the soils without referring back to Table S1.