Reviewer 1 report:

Replies from authors will be in blue and italic.

Hammes et al., reported results of a large greenhouse enhanced weathering (EW) mesocosm experiments to contribute to the understanding of the potential of carbon dioxide removal of EW. The experiments included a set of different combinations of EW feedstocks (13) and soil types (7). Over a period of 2 years, sample collection and analysis were conducted to study the export of Alkalinity in leachate, and the cations retained in different fractions of the soil. The authors observed that there was no measurable change between control and treated pot, or even a reduction of alkalinity export. This led them further investigate the potential traps for released cations within the soil system. The authors discussed the key controls of the exported Alkalinity and retained cations in the experiments, including mineralogy of the feedstock, soil characteristics, and hydrological properties. They concluded that in-soil retention dominated over the Alkalinity export in leachate, and that the partitioning of the cations in soil pools is a critical component for calculating CDR of EW. The experiments were well designed, and the number of experiments and replicates is the highlight of this study. This study has great potential to generate a data set to look into some of the areas that urgently require improved knowledge in in EW. However, due to a few concerns which I will provide more details below, I could not recommend publication of this manuscript in its current state.

In general:

We thank the reviewer for their careful and constructive comments. Overall, we appreciate their valuable suggestions and believe that these revisions will strengthen the manuscript and improve the transparency and interpretability of the dataset.

1. Data quality and statistics

This study generated a large dataset with the amount of experiments and replicates conducted. However, there lacks information to allow readers to assess the quality of the data. Firstly, in the sections describing the analytical methods, there lacks information about the evaluation of the quality of the instrumental analysis. Secondly, the uncertainties of the data were not reported in texts or all figures. Where there shows variability in data, there is no information about how such variability was assessed. Thirdly, the handling of such large datasets would require an approach to process the data statistically. Given the number of replicate and experiments, do the measurements show significant variation among replicates of the same experiment? I suggest that a section on the statistics of the data be reported in this study, with information about uncertainty.

Authors reply 1:

We recognise the importance of providing readers with sufficient information to enable them to assess the quality, variability and statistical robustness of the dataset.

In response to this comment, we will revise the manuscript to provide a clearer description of data quality and uncertainty. Firstly, we will expand the 'Analytical Methods' section to include further details on instrumental performance evaluation and quality control procedures, thereby clarifying the reliability of the measurements.

Secondly, we agree that uncertainties should be reported more explicitly. We will therefore revise the text and figures where appropriate to include information on data uncertainty and variability. We will also clarify how this variability was assessed across replicates and experiments.

Thirdly, we agree that the size of the dataset warrants clearer statistical treatment. Accordingly, we will add a section describing the statistical approach used to process and interpret especially the leachate data. This will include an analysis of replicate variability and an assessment of whether differences among treatments and controls are statistically significant.

To achieve this we will quantify leachate alkalinity across all treatment replicates over time, visualise cumulative trajectories using linear interpolation between measurement dates and summarise treatment responses as the mean ± one standard error. To avoid pseudoreplication from temporally autocorrelated cumulative data, inferential statistics will be based on a single endpoint value per replicate. This will be calculated as the average over the final 45 days and treatments will be compared against their respective control using a two-sided Mann–Whitney U test. In addition to significance testing, we will report Hedges' g as the primary measure of effect size, in order to quantify the magnitude of treatment effects relative to within-group variability.

2. The study used rainwater as irrigation water, and supplied the pots through a rainwater cistern (Line 150). Although this ensured that the contribution of cations to the leachate by the rainwater is the same between experiments, the influence of the temporal variation in this contribution on the experiments could not be ruled out as rainwater chemistry is not constant with time. This requires justification with analysis of rainwater chemistry, and the contribution of cations from the rainwater needs to be considered when discussing cations in the leachate.

Authors reply 2:

We agree that temporal variation in rainwater chemistry occurs and could be considered in general when interpreting cation composition of leachates. In our study, data on rainwater chemistry are available for all analysed metals and demonstrate variability over time. The time-weighted mean for the main base cations of all irrigation measurements results in the following: Na: 0.45 mg/l (SD 0.66 mg/l); Mg: 0.13 mg/l (SD 0.05 mg/l); K: 1.15 mg/l (SD 0.66 mg/l); Ca: 4.63 mg/l (SD 3.68 mg/l).
However, we do not expect this variation to have substantially influenced treatment effects, since all pots within each experiment were irrigated with the same rainwater source and received the same amount of water. Consequently, the input of rainwater-derived cations remained comparable across both the control and treatment groups in each experiment.

We also acknowledge that rainwater is a more realistic irrigation medium than ultrapure water, such as Milli-Q water. In other studies, this type of water may behave more acidic than natural rainwater and may therefore be less representative of environmental conditions. To address the reviewer's concern, we will revise the manuscript to include the available rainwater chemistry data showing the time-weighted mean and standard deviation in metal input, so that the reader can gain insight into this.

3. The influence of the biomass (the planted ryegrass) and biological activity (the added worms) needs more discussion than the very short mentioning in Line 485-486. Due to various types of treatment, the biomass and biological activity would vary, and that the feedstock and soil types are not the only variables in the experiments. Biomass the biological activity might affect the uptake of the cations by plants, the condition of the soil (e.g. organic matter degradation changing soil pH), and the physical properties of the soils as worms move. These will all have an impact on both the Alkalinity export in leachate and store in soil.

Authors reply 3:
The reviewer raises an important point regarding biological influences within the system. We agree that both ryegrass biomass and earthworm activity may have affected the soil–rock–leachate system and thus contributed to the observed responses.

That said, the present study was not designed to disentangle or quantify the individual effects of plants and earthworms. Rather, our objective was to assess the overall behaviour of the soil–feedstock system using the integrated response captured in the leachate and soil data. In this respect, the experiment takes a black-box approach, focusing on the response of the system as a whole rather than on the contribution of each separate biological component.

The inclusion of earthworms and plants was intended to maintain a more realistic and functional soil environment by supporting processes such as bioturbation and microbial activity, reducing the risk of clogging, and supplying continuous CO_2 input through plant activity. These components can reasonably be expected to affect cation uptake, organic matter turnover, soil pH, and soil physical structure, and thereby influence both alkalinity export in the leachate and alkalinity storage in the soil. However, with the current experimental design, the specific effect of earthworms cannot be isolated quantitatively.

A fuller discussion of ryegrass biomass would considerably expand the scope of the manuscript beyond its present focus. For this reason, we do not plan to develop this aspect further in the main text. Nevertheless, we agree that biomass and biological activity may influence the overall system response, and this should be acknowledged as a limitation of the study.

4. It was useful to know that the reported study was a part of a larger experiment. It is however not necessary, especially when the information about the larger experiment gets mixed up with that of the reported study. I suggest that the authors focus on providing information that is relevant to this study, and could include those from the wider experiment in the supplementary information.

Authors reply 4: This is a very important point. We agree that the manuscript should focus on information directly relevant to the present study. Details from the wider experimental framework may distract from the description of the reported work if they are too closely interwoven with it.

We will therefore revise the manuscript to more clearly separate these aspects and retain only the information necessary for understanding the present study in the main text. Additional details related to the broader experimental context will be moved to the Supplementary Information section.

Our intention in including this broader context was to improve transparency and make the dataset more accessible for future use in modelling or related studies, for example. However, we agree that this can be achieved more appropriately in the supplementary material without compromising the clarity or focus of the main manuscript.

Section 2.3: The description of the feedstock here is not clear. It seems that the authors were giving information about the wider experiments they were doing, which this study is part of, as well as the information for this specific study. The writing of it makes it very difficult to follow which was reported in this study, and generated some inconsistencies. For instance, Line 193-194: the authors described different application rates for basanite. However, only 2 rates could be found in Table 1. And in Line 195: the authors stated that the focus was 6 representative feedstocks. However, the experiments using the 7 feedstocks from EW suppliers were included in Table 1, reported and discussed in the manuscript.

As mentioned in 'Author's Reply 4', we will take a closer look at this issue and clarify it.

5. Discussion

The current discussion is insightful. But it could be discussed more with the data generated in this study.

Line 476: The authors reported the measurement of the cations in leachate. They could take the next step of looking into these data to test this mechanism.

The cation measurements in the leachates are indeed relevant for assessing this mechanism, and these data are presented in a dedicated figure in the Supplementary (Fig. S2), which is also referenced in a preceding sentence (line 471). To make this clearer for the reader, we will revise the text passage to strengthen the connection to that figure and to highlight more explicitly how these measurements support the discussion.

Section 4.3.2. The authors should provide some evidence to argue against or for this mechanism. Some analysis of the soil porosity would be useful.

We agree that additional evidence would be valuable when evaluating this mechanism, and that soil porosity data could in principle contribute to this discussion. However, in the present study, such measurements were not feasible within the chosen sampling *approach. Because the experiment was designed to obtain the most representative possible sample from each bucket, the entire soil–feedstock mixture from the upper, middle, and lower sections was homogenised before subsampling. This step was necessary to capture the heterogeneity of the system and to enable consistent downstream processing of the large number of samples. As a result, in situ measurements such as soil porosity or bulk density could not be obtained without compromising the representativeness of the final sample. We will try to find an alternative solution to this problem in future samples and trials.*

More comments (minor-major points) are given by section of the manuscript below:

Section 1.

Line 103: What did the authors mean by “comprehensive”? comprehensive geochemical analysis? Comprehensive experiment designs? The manuscript does not reflect either. The authors should be clear here.

By ‘comprehensive’, we intended to refer to the combination of measurements, controls, feedstocks, and soil types included in this study at a scale that, to our knowledge, has not yet been reported in this form. We did not mean to suggest that every possible geochemical parameter or experimental aspect was covered. To avoid ambiguity, we are happy to revise this wording in the manuscript and replace ‘comprehensive’ with a more precise term, such as ‘unprecedented’.

Line 109: The years should be spelled out here.

We agree and will correct this in the revised manuscript.

Line 123: The stated hypothesis is not what the work as described by the authors earlier in the same paragraph was testing, and focussed on discussing. A more clearly described hypothesis is needed here. It also lacks clear description of the objectives of the study. The writing of this paragraph could be improved by stating the hypothesis first, then describing the work done to test it.

In response to this helpful comment, we will revise the paragraph to state the hypothesis more clearly at the outset and to distinguish it from the study objectives. The revised text will first present the central hypothesis, namely that alkaline EW amendments increase alkalinity export in soil leachate, while also recognising that the strength of this response may vary with soil properties and internal alkalinity sinks. It then describes the experimental approach used to test this hypothesis and explicitly states the objectives of the study, including the evaluation of alkalinity production and export across contrasting soils and the identification of major soil alkalinity sinks.

Section 2.

Line 145: I could not see the cause-effect relationship here: all experiments experienced identical climatic conditions does not really allow for accelerated weathering conditions. It is not clear whether the authors meant to describe the comparability of the experiments without having climatic condition as a variable, or they meant to describe the conditions they set to allow for accelerate weathering. If the former, the sentence needs to be rewritten to reflect that. If latter, the conditions should be given.

We fully agree and thank the reviewer for pointing this out. The current wording can be indeed misleading.
‘This greenhouse setting ensured that all mesocosm experiments experienced identical climatic conditions. Moreover, having a temperature constantly above 19 °C (temperature range Fig. S1) and high leachate formation due to irrigation of at least 2,000 mm a-1 meant that the experimental period required to detect different biogeochemical trends could be shorter due to the accelerated weathering conditions.’

Line 147: The irrigation is up to 4000 mm a-1, instead of 2000 in the manuscript, based on Table S5.

We agree and will revise this, as the current wording may be misleading.
‘Moreover, having a temperature constantly above 19 °C (temperature range Fig. S1) and high leachate formation due to irrigation of ***at least up to 2,000 mm a^{-1}****… During the hot summer months, the irrigation rate needed to be increased up to 4,000 mm a-1 to ensure leachate formation due to the very high evapotranspiration rates (Tab. S5. Irrigation rates applied by period).’*

Table S1. The authors gave a very brief description of the analytical methods for each parameter. However, there is not any information given to assess the quality of the data. Such information includes, uncertainty, quality control, etc. If such information was published in separate literature, please provide reference.

As noted above in Author Reply 1, we will revise the tables and the presentation of the dataset accordingly to include clearer information on data quality, uncertainty, and quality control.

Line 182: CEC should be spelled out in full, cation exchange capacity, first before using abbreviation.

Thanks for pointing this out. The term was already written out in the preceding sentence, but the abbreviation CEC was not introduced there. This will be corrected in the revised manuscript.

Line 183: Inconsistent significant number for CEC.

We thank the reviewer for this comment. However, we are not fully certain that we understand the specific issue being referred to regarding the significant figures reported for CEC in Line 183. If it's a general comment to the significant figures of CEC values, we will change this accordingly to the same amount of significant figures.

Line 187: PSD was now clear. The abbreviation should be given following its full name first before using elsewhere.

We agree and will make this change.

Line 190: ‘representative materials’ is not very clear language. Please give information about what they are representative of.

We agree and will change it to ‘commonly used in the field of ERW’ to make it clearer.

Line 195: “ a short description of these materials is given below…” I could not see the information the authors said that was given below.

We apologise for the misunderstanding. It should refer to the appendix.

Line 199: More information should be given about the sampling containers , such as materials, whether it is cleaned with water or acid etc.

We agree that additional information on the sampling containers could be provided for clarity. In the revised manuscript, we will specify that the sampling containers consisted of borosilicate measuring cylinders and Erlenmeyer flasks. The cleaning procedure with the sample (water/leachate) has been described in this particular line.

Section 2.4.3.: Information should be given about the quality control of the analyses: repeated measurements of reference materials, for instance, and reporting this result is needed.

As noted above and in Author Reply 1, we will revise the tables and the presentation of the dataset accordingly to include clearer information on data quality, uncertainty, and quality control.

Section: 2.5.1: Information should be given about the quality control of the leaching procedures, ICP-OES and ICP-MS measurements.

Same here: we will revise the tables and the presentation of the dataset accordingly to include clearer information on data quality.

Figure 2: Application rate for EW 1-7 in legend is missing. Uncertainties should be plotted.

We thank the reviewer for noting this and will add it. Regarding the uncertainties, we will proceed as mentioned in Author Reply 1.

Figure 3-7: What are the numbers right after the symbol (beginning of the texts) in the legends? What are the dots (all the datapoints?) and lines (average of the replicates?)? How is the variability (shaded area) assessed? standard deviation? or? Such information should be described and given.

We fully agree. Thank you for pointing this out; the current version is indeed misleading. We will proceed as mentioned in Author Reply 1.

Line 299: 4 times should be spelled out instead of using 4x.

We agree and will revise this.

Line 333: The authors reported CO_2 sequestration rate, but did not give any information about how it was calculated.

We thank the reviewer for highlighting this. We agree that the calculation of the CO_2 sequestration rate should be explained more clearly in the manuscript. In the revised version, we will include a more detailed explanation of the calculation in the appendix to increase transparency.

Specifically, the estimate is based on measured alkalinity in the leachate, which is used as an approximation of bicarbonate under the experiment's pH conditions. We recognise that alkalinity is not identical to HCO_3^- concentration in a strict chemical sense. However, within the relevant pH range and under the assumptions applied here, alkalinity can be considered an approximate equivalent of the bicarbonate present in solution. The CO_2 sequestration rate was then estimated from the difference in measured alkalinity between the treatment and control groups, providing an indication of the additional bicarbonate generated by the amendment. This difference was then related to the experimental surface area and time to express the result as an equivalent CO_2 sequestration rate per unit area (ha) and time (year).

We will revise the manuscript to make both the assumptions and the calculation steps explicit.

Figure 8: Application rate should be given especially the limestone treatment was applied at a different rate from others.

We agree and will revise this for more transparency.

3.2 and figure 8: I do not think that the liquid and solid phase could be compared this way, i.e. by normalizing both to per pot. The amount of the Alkalinity in the leachate was measured in the unit of per volume of the leached liquid, while the cations in the solid phase were measured in the unit of per mass of the solid. Also, Alkalinity in water includes the cations and anions, and therefore is not really directly comparable with the sum of cations in solid phase. Plotting them on the same figure and make comparison are not really comparing the same thing.

We thank the reviewer for this thoughtful comment and acknowledge the concern regarding the comparison of liquid- and solid-phase data. We agree that these pools are not strictly equivalent from an analytical perspective. However, Figure 8 does not imply that leachate alkalinity and solid-phase cation concentrations are the same.

Instead, the comparison is made on a charge basis per pot in order to relate the cumulative alkalinity exported in the leachate to the pool of exchangeable and soil-associated cations, which represent the maximum potential alkalinity reservoir remaining in the system. The leached alkalinity reflects the actual charge exported in solution. In contrast, the solid-phase cations represent a potential source of future *alkalinity release, even if not all of these cations are necessarily mobilised into solution to balance bicarbonate or carbonate with other anions. Of course, this aspect may seem misleading, as it appears that we are making the soil pool appear larger than the actual resulting leachable alkalinity in the future. However, in the field of ERW, potentials have often been calculated using the absolute stoichiometric potential of feedstock XRF results. The comparison we make here allows us to compare the exaggerated promises or estimations with what is now located in the different soil pools.*

For this reason, we consider the current presentation to be justified in the context of the intended conceptual comparison and will retain it in the manuscript. However, to avoid any misunderstanding, we will revise the text and figure caption to make it clearer that this is a comparison of charge pools at the pot scale, and that it should be interpreted as a comparison between exported alkalinity and the remaining potential cation reservoir rather than as a direct comparison of identical geochemical quantities.

Section: 4.3.1

Line 466: it is discussed in more details as the 2nd point, why is it also mentioned here?

We agree that the current structure may appear somewhat repetitive at this point. In the revised version, we will differentiate these passages more clearly so that the discussion at Line 466 is better distinguished from the second point.

Line 585: I struggle to see why point 3 was put down here as one of the indices to support the argument originate form background soil carbonate.

We agree that the link between point 3 and the argument regarding background soil carbonate is not clear enough in the current version. However, point 3 was never intended as a direct argument for a carbonate origin, but rather as part of the broader interpretation of the LUFA 6S soil system. Points 1 and 2 address the dominant role of carbonates in this soil; point 3 was included to explain why weathering products may be released within the system but not necessarily appear in the leachate. As this section aims to discuss the contrasting behaviour of the two soil systems more broadly and not just the role of carbonates, we would retain point 3 in this context.