Anonymous Referee 1

This is a well-designed and potentially interesting study to test the hypothesis that the combination of an organic amendment with lime may improve the efficiency of liming at depth. Three soils have been studied in column experiments. A single organic amendment and a single rate of both lime and organic amendment to the soils have been considered. At the end of the experiments soil columns were sliced and analyzed. For one of the soils sieved and repacked soil and undisturbed soil columns were compared.

Although I appreciate the volume of work involved and the careful planning in the experimental set-up, I do have serious concerns.

Comment 1: Firstly there were few replicates. No repetitions for the control soil and organic amended soil, and only duplicates for the lime amended soil (with and without organic amendment). Soil columns were sliced after the end of the experiments, but apparently slices were analysed without dividing into subsamples. Surely there was enough material to do so, thereby allowing repetitions and a better appreciation of variability.

Reply 1: Point taken and text added in revised document. We agree that more column replicates per treatment would have enhanced the statistical power of the experiment. Unfortunately, the column set-up did not allow for this in practice because this set-up is a rather unique one for unsaturated flow using the suction plates. The standard column set-up with free drainage is unsuitable, leading to water-saturated conditions at the outlet and local pH increases due to anaerobic conditions. Even though that artefact is straightforward, it is still overly used, and the suction plate method is a much better alternative, with the drawback of limited replication and statistical power. Despite this, we proved to find significant differences between the lime & lime+OA treatments (both in duplicate). Apart from that, there were indeed no analytical replicates within one slice of a single column. Since we were mainly interested in the depth profile of the soil pH and DOC concentration, we chose to sample one representative composite sample rather than a few point samples.

To address the comments, a justification for the current setup compared to the free drainage setup is included in the Materials & Methods section of the revised manuscript (lines 137-140).

Comment 2: Since there were no comparisons of the amount of lime, or organic amendment nor comparison of organic materials, I suggest a better explanation of the choice of amount of lime and amendment should be given.

Reply 2: **Point taken and table+text added in revised document**. The selection was done as follows:

- Type of organic amendment: 7 different organic amendments (three types of green compost, a pig manure, a mix of green compost+pig manure, and two types of farmyard manure) were analyzed for pH, DOC concentration, and aromaticity (SUVA method). A titration curve was also constructed for each type of amendment to determine its buffer capacity. A high DOC concentration was desired to maximize leaching of organic anions, while a low aromaticity was desired to minimize sorption to iron and aluminum oxides in the soil; based on this, the “ILVO green compost” was selected.
- Dose of lime & organic amendment: typical application rates in field experiments on highly weathered soils are 2 tons/ha lime and 10 tons/ha farmyard manure. The application rates in this leaching experiment were chosen to be somewhat higher as a

proof of concept, in an attempt to ensure a leaching effect. If the doses of 5 g/kg lime and 10 g/kg compost are converted to field scale, assuming a plough layer of 10 cm depth and a soil bulk density of 1400 kg/m^3, the resulting field application rates are 7 ton/ha lime and 14 ton/ha compost.

To address the comments, (i) a table with the tested organic amendments and their characteristics is included in the Supplementary Information of the revised manuscript (Table S2), and (ii) a justification of the doses of lime & OA is added in the Materials & Methods section of the revised manuscript (lines 143-144).

Comment 3: Secondly the leachate could be collected, but there is no information on the composition of the leachate.

Reply 3: **Text not changed.** In this study, we did not collect the leachates. The primary research objective was to understand how alkalinity (either or not in combination with DOC) is retained and distributed within the soil matrix, rather than how much was lost via leaching. As alkalinity leaching is a notoriously slow process, it was expected that leaching losses, if any, would be minimal. The data (Figure 1 & S3, Table S2 & S3 in draft manuscript; Figure 1 & S3, Table S3 & S4 in revised manuscript) also show that pH and DOC concentration changes are restricted to the upper 5 cm of the soil columns.

Comment 4: Third the leaching experiments were relatively short, three weeks for the comparison of soil with different adsorption affinities for Dissolved organic carbon, and eleven weeks for the comparison of repacked and intact soils for a third soil. Why were these periods chosen? I understand that the experiments may have been limited by equipment availability.

Reply 4: **Point taken and text added in revised manuscript.** These periods were chosen based on the number of leached pore volumes, rather than the duration needed to achieve this. The number of leached pore volumes is calculated as the total volume of rainwater applied to the column during the experiment divided by the pore volume of the column (which depends on the soil bulk density). In the first experiment, we aimed for at least one leached pore volume. At the highest rainfall intensity that was possible without overflowing the columns, this took three weeks. Since the results of the first experiment indicated a very limited alkalinity leaching, we aimed for two leached pore volumes in the second experiment. However, the rainfall intensity had to be lowered since the undisturbed columns were more prone to overflowing. The duration needed to achieve two leached pore volumes was eleven weeks in this case.

To address the comments, we included a short explanation on the choice of the duration of the experiment in the Materials & Methods section of the revised manuscript (lines 154-157).

Comment 5: I found it difficult to follow the adsorption/release of DOC in organic amended soil. Kd are given as soil and pH dependent (Table 1), and the adsorption isotherms in Figure S2 are nonlinear and show classical form. However in the main text, the authors explain difficulties with desorption of soil DOC when suspended.

Reply 5: **Concerns addressed and adsorption isotherms revised.** Indeed, in the draft manuscript, the isotherms in Figure S2 were derived by first adding the negative intercept of the y-axis of net DOC sorption to the measured values for the Podzol and Retisol, resulting in the total DOC sorbed. Then, a Freundlich isotherm was fitted to this data (Figure S2 of the draft manuscript). This was done because there was a net release of DOC during batch sorption experiment at every concentration of added DOC for these two soils. The K_D in Table 1 was then derived by fitting a linear line through the lower DOC concentration range data points, of which the slope represents the K_D. This was described in line 118-133 of the draft manuscript. However, following comment 2 by Prof. Jan Siemens, a better method for fitting adsorption isotherms in case of both desorption and adsorption was used in the revised manuscript.

To address the comments, the original Freundlich adsorption isotherms were changed to modified Langmuir adsorption isotherms that included a parameter C accounting for the desorbable amount of DOC in soil. Figure S2 was adapted accordingly, with better fits as a result. Details are described in lines 119-134 of the revised manuscript.

Comment 6: The authors attempt to use optode images to identify heterogeneity with preferential movement of amendment. Why was this method only used for intact soil? It is thus not correct to claim that there was different movement in repacked and intact soil columns. The authors explain that poor contact may introduce artefacts in the pH measurement using these planar optodes. I was certainly perplexed by the images. pH at depth appeared to be very acid, much more acid than the soil. Could a better calibration not have been found?

Reply 6: **Text not changed.** The method was used for all 12 columns of the second experiment (both intact and sieved), however, only three images representative of the leaching process were selected for the manuscript (Figure 2). For the sieved soils, pH imaging of the lime+OA treatment failed due to insufficient wetting of the sensor foil and soil before sensor foil deployment. The pH sensor foils are designed for use in aqueous solutions; the use in unsubmerged soils is possible if the sensor is sufficiently pre-soaked, but this was not always equally effective. Regarding the calibration of the optode, we fully agree that better calibration should be established in future experiments. The technique is quite new and the foils currently do not cover the full pH range. There are two types of foils available: one for pH range 2.5-4.5 and one for pH range 5.5-7.5. This is the "main response region", meaning that the sensors still respond outside of this range but with lower sensitivity. Both sensors need to be calibrated separately and preferably at the ionic strength of the soil solution. In our study, we faced following problems that probably caused the shift in calibration:

- Application of the sensor foils to unsubmerged soils (insufficient wetting)
- The soil pH range in the columns superseded that of each sensor foil separately (e.g. pH 3.3-6.5 (0.01 M $CaCl_2$, 1:5) in undisturbed limed columns). Therefore, the two sensor foils were applied next to each other. In theory, the images of both foils should be combined to achieve the best result, but this was not possible in our case due to the complex pattern of preferential flow pathways. Therefore, we chose to only show the result of the lower pH foil, but since the pH of the upper soil layers falls outside of the main response region, the calibration is shifted.
- The sensor foil calibration buffers should be prepared at the same ionic strength as the sample. However, the ionic strength of the soil solution within one column ranged from 14 mM (control soil, deeper layers) to 41 mM (lime+OA, upper layers). Therefore, an average value of 25 mM was chosen for the ionic strength of the calibration buffers. This is quite low for the planar optode (use in samples with ionic strength of 200 mM was recommended) and can cause a shift in calibration.

In our opinion, however, the optode images still have an added value since they show preferential flow patterns that can not be detected by the conventional soil pH measurements.

Comment 7: Amendments were added to the top of the undisturbed soil column mixed with sieved soil, to a depth of 2 cm. Was sieved soil similarly added to the top of un amended undisturbed soil?

Reply 7: **Text added to revised document.** Yes, indeed, the top layer of the undisturbed control column is unamended sieved soil. In Table S3 (S4 in the revised manuscript), it can be seen that the soil pH in the upper two cm of the undisturbed control column is about 0.2 units higher than that of the layers below, due to air-drying and rewetting of the sieved soil.

To address the comment, this was clarified in the Materials & Methods section of the revised document (lines 146-147).

Comment 8: I find the authors too affirmative and positive in their conclusions. Sieved Retisol actually shows less pH increase at depth with addition of organic amendment in addition to lime, the effect is not significant for ferralsol, and small, even if significant for the podzol.

Reply 8: **Text not changed**. Attempts were made to correctly reflect these observations in the conclusions sections, which indeed states that the Ferralsol shows only limited alkalinity transport (lines 325-327), and that alkalinity leaching was not more pronounced in structured soils compared to sieved soils (lines 329-330).

Minor points

Comment 9: The soils are identified by geographical origin and not by soil type in Table 1 and the Supplementary information.

Reply 9: **Accepted and revised** (Table 1 and SI).

Comment 10: The difference in pH and DOC with respect to the control soil is given as delta, not D in axis labels.

Reply 10: **Accepted and revised.** We made sure "delta" and not "D" was given in all axis labels.

Comment 11: I suggest colour and symbols should be used to distinguish treatments instead of shades of grey.

Reply 11:. **Partially revised.** Due to the author guidelines of this journal concerning the use of colors ("make sure that the colour schemes used in your maps and charts allow readers with colour vision deficiencies to correctly interpret your findings") we eventually decided to keep the shades of grey for the revised manuscript, but different symbols were used per treatment.

To address the comment, Figure 1 was revised, using different symbols to distinguish among treatments.

Comment 12: Findings should not be described as troublesome, line 259, I suggest disappointing or limits the relevance to …

Reply 12: **Accepted and revised** (line 262).

Comment 13: Did the authors consider attempting to quantify the movement of Ca? It is possible that any change would have been too small to be detected.

Reply 13: **Text not changed.** Unfortunately, the Ca concentration in the soil slices was not measured in this study.

Comment 14: The sigmoidal fit used is not appropriate for DOC profiles (Figure S3).

Reply 14: **Text not changed.** Indeed, the DOC concentrations in unamended soils decrease with depth and then, a sigmoidal fit is not appropriate. However, our main objective here was to compare the depth of the inflection point of the curves (so the depth of enrichment in DOC) between different treatments, which we believed could be best achieved with a sigmoidal fit.

Anonymous Referee 2

Surface application of lime often fails to amend subsoil acidification because of limited vertical transport of alkalinity. The downwards migration of alkalinity might be restricted due to the strong retention of dissolved carbonate forms by the soil matrix. The presence of more mobile anions can support deep leaching of alkali and earth alkali cations, which can displace subsoil exchangeable acidity. One way suggested to provide mobile anions to accompany lime-derived cations is co-application of lime and organic material. Given that the released predominately anionic dissolved organic matter is not retained strongly by the mineral matrix, such as under preferential flow conditions, it likely will facilitate cation leaching into subsoil.

The present study tested this idea in percolation experiments with columns either filled with homogenised (sieved) and undistributed acid soil. The top parts of columns received either no amendment, lime, compost, or lime + composts, and were irrigated with artificial acidic rainwater. The results revealed significant downwards transport of alkalinity only for one soil receiving lime + composts. Seemingly, the mobility of the dissolved organic matter was crucial for the mobility alkalinity. The authors concluded that surface co-application of lime and organic materials could indeed facilitate the amendment of subsoil acidity in case of little retention of dissolved organic matter. This approach might work best for sandy soil allowing for rapid percolation of soil water and having little capacities for organic matter sorption. The results on the soil column with naturally structured soil gave no clear hints to the possible deep percolation of alkalinity along preferential flow paths.

Overall, the manuscript reads. The study is well justified and addresses a well-defined and important research question. The experimental setup is well suited to address the research question and the experiments were carried out with care. Using a planar optode to in-situ track changes in acidity is innovative. The obtained results are mostly well presented and their interpretation is largely sound and considers the relevant solution and soil properties.

Comment 1: The results, although not truly novel, are important for improving and further developing strategies to mend subsoil acidification occurring in various soils around the world. The probably weakest point of the study is the little number of soils being tested. Especially, the effect of soil structure has been tested only for one soil and therefore warrants more in-detail studies. Given that small number of soils tested, I am reluctant to fully support the immediate publication of the manuscript. I therefore suggest re-working the manuscript for pointing out even more clearly the preliminary nature of the study.

Reply 1: **Point taken and text added to the revised document.** Thank you for your feedback. The goal of our study was to identify the underlying mechanisms of organic matter-mediated alkalinity leaching, not to prove effects on a bigger scale or to make claims about all Ferralsols, Retisols or Podzols. For our mechanistic study, we selected a few soils that were representative examples of soils exhibiting low, average and high DOC sorption, with the intention of creating a gradient in the success of organic matter-mediated leaching. To prove the concept of organic matter-mediated alkalinity leaching in soils all around the world, a higher number of soil types and replicates per soil type would indeed be needed; unfortunately, this was not logistically possible in this study. I refer back to our response to the first comment of Referee #1 and the changes that we made to the text: Unfortunately, the column set-up did not allow for more replicates in practice because this set-up is a rather unique one for unsaturated flow using the suction plates. The standard column set-up with free drainage is unsuitable, leading to water-saturated conditions at the outlet and local pH increases due to anaerobic conditions. Even though that artefact is straightforward, it is still overly used, and the suction plate method is a much better alternative, with the drawback of limited replication and statistical power.

To address the comments, a justification for the current setup compared to the free drainage setup is included in the Materials & Methods section of the revised manuscript (lines 137-140).

Comment 2: Also, I am puzzled by description of the statistical evaluation of results since the number of replicates per soil and treatment was hard to guess. The tests used would require three replicates per treatment at minimum, resulting in each 12 columns used for the three homogenised soil. Is that correct? The authors may consider to provide a sketch depicting to entire experiment setup for.

Reply 2: **Accepted and revised.** The experimental set-up allowed for twelve columns to be leached at the same time. In Experiment 1, six columns were filled with the Podzol and six columns with the Ferralsol. For each of these soils, there was one control column, one compost-amended column, two lime-amended columns and two lime+compost amended columns. In Experiment 2, six columns were filled with sieved Retisol and six columns with intact Retisol. The treatments were the same as for Experiment 1. This information had been given in line 139-141 of the draft manuscript (144-146 in the revised manuscript). The statistical analysis comparing pH or DOC concentration values between the different soil treatments within one soil and at each depth was done with an ANOVA followed by a Tukey test, but as Prof. Jan Siemens also noted in the Community Comment below, this does not allow for the derivation of robust conclusions. The statistical program JMP uses the Tukey-Kramer adjustment to allow for unbalanced designs but the validity of the test is compromised. Therefore, we adjusted this analysis for the revised manuscript in the way proposed by Prof. Jan Siemens. However, the statistical comparison of leaching depths in lime versus lime+OA treatments for each soil was possible since these treatments were in duplicate per soil. For each soil, data of all four columns were fitted in one set with an assumed difference in parameter values of the sigmoidal fits for each of the four parameters (ΔpH_{max}, ΔpH_{back}, slope and b). The statistical difference in parameter values were tested using dummy variables to identify significant treatment effects on the extent and depth of penetration of the alkalinity.

To address the comments, the ANOVA followed by a Tukey test for the differences in pH among treatments was left out of the revised manuscript and focus was given only to the functional approach (i.e. non-linear regression), which did allow for comparing the limed treatment (n=2) with the lime+OA treatment (n=2).

Comment 3: The authors assume that the results of the optode imaging are probably hampered by incomplete contact to soil. I therefore suggest to consider transferring the result to supplement section since presumably not reliable.

Reply 3: **Text not changed.** In our opinion, the optode images still have an added value since they show preferential flow patterns that can not be detected by the conventional soil pH measurements. For the lime+OA treatment in the sieved soil, the imaging failed due to insufficient contact with between the soil and the sensor foil, therefore we only showed images for the intact soil in Figure 2. The sentence "The absolute values of the soil pH are solely indicative since incomplete contact between the sample and sensor foil may occur in unsaturated samples" was added more out of caution than that we actually think this happened.

Comment 4: The derivation of partition coefficients from the sorption data is hampered by the mostly less than ideal fits of the applied sorption model. The authors may consider using other approaches. Several have suggested and tested by Kothawala et al. (2008, Geoderma 148, 43–50. doi: 10.1016/j.geoderma.2008.09.004).

Reply 4: **Accepted and revised.**

To address the comments, the revised manuscript uses a modified Langmuir equation to describe the DOC adsorption isotherms, as suggested by Prof. Jan Siemens. Details are given in lines 119-129 of the revised manuscript.

Comment 5: Finally, please refer to the soils by type and not regional origin.

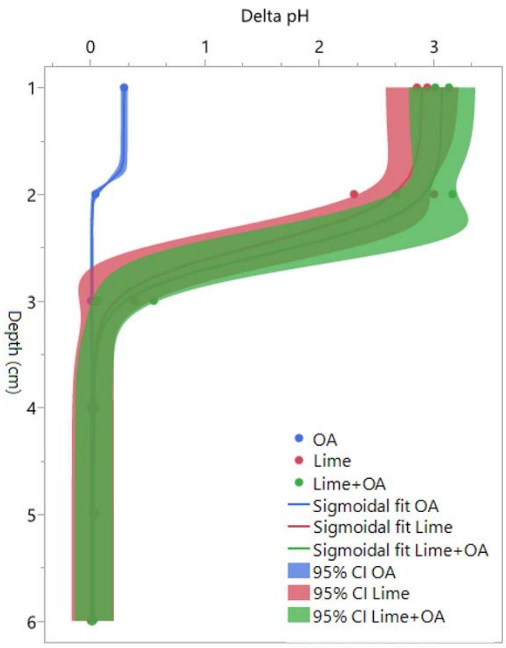
Reply 5: **Accepted and revised** (Table 1 and SI).

Jan Siemens (Community Comment)

The manuscript at hand addresses a highly relevant, but understudied topic, the amelioration of acid soils for increasing the productivity of soils. The study is based on well-formulated hypotheses and carefully conducted experiments that allow testing these hypotheses. I have two main recommendations.

Comment 1: The first relates to the statistical evaluation of differences between the treatments in soil column experiment 1. The soil column experiment 1 comprised one control column (n = 1), one column with organic amendment (n =1), and two soil columns (n = 2) for each of the lime and lime plus organic amendment treatment. Differences in pH values of the treatments were tested with an ANOVA with a Tukey post hoc test (line 175-176). I don´t understand how an ANOVA with n = 1 for part of the treatments works, and even if it works, I am skeptical that the results of an ANOVA with n = 1 or 2 allow the derivation of robust conclusions. As an alternative to the ANOVA with Tukey test, I suggest to plot all data points in Figure 1 (i.e. two data points per depth interval for the lime treatment and lime plus OA treatment) and to add 95% confidence intervals of the fitted sigmoidal functions. In my view, this would increase the transparency of the data set and its interpretation. Data points would represent single measurements for all treatments and not single measurements for the OA treatment and means of duplicates for the lime and lime + OA treatments, as it is the case in current version of the figure. The confidence intervals of the sigmoidal functions and their overlap would provide the reader a rapid overview of the differences between treatments.

Reply 1: **Accepted and partially revised.** Thank you for your comment, you are absolutely right. The statistical program JMP does give results for the Tukey test in this case, it uses a pooled variance estimate from ANOVA and adjusts for unequal group sizes by modifying the standard error in each comparison. However, it still assumes equal variances across group and requires enough degrees of freedom in the residuals to be reliable, and as you noted, this is not the case here. The results of the Tukey test were therefore excluded from the revised manuscript. Figure 1 was adapted to include all data points, i.e. data points now represent single measurements for all treatments. However, we finally decided not to add the 95% confidence intervals to the figure for following reasons: (i) the added value of the confidence intervals was only to show a deeper alkalinity leaching in the lime+OA treatment compared to the lime only treatment in the Dutch soil, which was already proven by a significantly larger b parameter, which is the depth of the inflection point of the sigmoidal curve (lines 211-213 of the revised manuscript), and (ii) the resulting figure became too complex and not aesthetically pleasing, as illustrated for the Vietnamese soil below:



To address the comments, the statistical analysis was reinforced by (i) excluding the Tukey test for differences in pH among treatments, (ii) revising Figure 1 to include all data points.

<u>Comment 2:</u> My second recommendation relates to the sorption isotherms for the Podzol and the Retisol. The net release of DOM during batch sorption experiments is a common phenomenon that has often been described by using the initial mass isotherm (e.g. also in the study of Kaiser et al 1996 that is cited in the manuscript). Siemens et al. (2004, DOI 10.1046/j.1365-2389.2004.00596.x) added an additional parameter to the Langmuir sorption isotherm to describe the net desorption of P in batch sorption experiments. The resulting isotherm allows the estimation of the P (or DOC) concentration in equilibrium with the sorbed P (DOC) concentration from the intercept with the x-axis. Both approaches could also be used to describe the DOC (de)sorption data of the study at hand. The approaches do not allow the straightforward "back of the envelope" calculation of DOC transport that is now included in the discussion section. However, I don't think that the use of a KD derived by adding the negative intercept of the y-axis of net DOM sorption to the measured values in the current analysis of the authors (lines 124ff) is useful for calculating DOM retardation on the Dutch and Belgian soils. After all, for the Dutch and Belgian soils this calculation indicates a retardation of DOM while in fact the data of the batch experiment indicate that the topsoil leachate with raised pH might cause a net mobilization of DOM instead of a retardation. Therefore, I suggest to plot the DOC (de)sorption isotherms of Figure S2 using a modified Freundlich isotherm in the form $s = K_F * c^n - D$ with D describing the desorbable DOC concentration (y-axis intercept) or as modified linear isotherms in the form $s = K_D * c - D$ for the lower concentration range of the isotherms and to calculate DOC retardation and transport only for the linear isotherms for which D is not significantly different from zero (the Ferralsol isotherm).

<u>Reply 2:</u> **Accepted and revised.** Thank you for your recommendation, we implemented the proposed changes following the method used in your paper (DOI 10.1046/j.1365-2389.2004.00596.x), with better fits as a result.

To address the comments, modified Langmuir equations were used to describe the DOC adsorption isotherms (lines 119-129 + Figure S2 of the revised manuscript). The "back of envelope" calculations in the discussion section were adapted to only include the soil with net DOC adsorption, i.e., the Ferralsol (lines 265-295 of the revised manuscript).