

We thank the associate editor for checking again our manuscript and for her new constructive comments. Below are our responses in blue, with the original editor's comments in black. The new line numbers refer to the new version with modifications tracked.

*After reading and evaluating the manuscript for the second time, there are two things that are not clear to me yet. First, I am not sure whether the girdling and felling treatment was an explicit part of the experimental design. It appears that this was an intentional treatment that was replicated in the control and in the experimental plots. Also, it seems like the results of this treatment were presented in detail in another study, but here it appears that this experimental treatment still had a very significant effect on the acidity of not only soil, but also of runoff water. However, this does not become clear until the reader finishes the discussion. Therefore, it might be necessary to add a section in the introduction explaining how reducing basal area would have a feedback mechanism on soil and runoff properties derived from decreased N demand from plant uptake and decreased litter input. These mechanisms are invoked in the discussion without having been introduced properly. In addition, the authors would need to consider incorporating this treatment explicitly in their experimental design and statistical analyses.*

Girdling and felling were a supplementary treatment done on both the N-addition catchment and on the control (rainwater only) catchment. This supplementary treatment had a strong effect on nitrate leaching and this was the main result in Schleppe et al. (2017, cited in the manuscript). Girdling and felling may have had a transient effect on the pH of discharge water by reducing the acidification due the exchange with protons as roots take up base cations (L. 405).  $\text{Ca}^{2+}$  in discharge water is, however, much more concentrated than  $\text{H}^+$ , which means that changes due to reduced root uptake may affect  $\text{H}^+$  in runoff water without being visible for  $\text{Ca}^{2+}$ . It seems that there was a misunderstanding between this aspect and the difference in  $\text{Ca}^{2+}$  trends between the N-treated catchment and the control. The duration and the intensity of the N addition are well sufficient to induce such an effect. We thus write that the significantly different trend observed in the N-treated catchment compared to the control may be related to the acidifying effect of the N addition. Indeed, more acidity means more dissolution of limestone and more release of  $\text{Ca}^{2+}$ . As a buffering process it affects more  $\text{Ca}^{2+}$  fluxes than the resulting  $\text{H}^+$  concentrations. Buffering is described in the introduction and barely requires more explanations here in the discussion. Still, we need to make it clear that the discussion about  $\text{Ca}^{2+}$  in discharge water relates only to the N treatment and not to girdling and felling (in contrast with for nitrate leaching). This was done by recalling which treatment and which process is meant in each case (L. 357 and 405).

*Second, the experimental design and measurement methodologies are not explained clearly. It is not clear why while there were two catchments, the experimental plots where rainwater (with or without extra N) were outside these catchments. The spatial details and chronology of the study are not explained clearly. (...) Also, why are there experimental pairs of plots outside the monitored catchment? And why are there no experimental pairs of plots inside catchment two?*

It is true that we introduced the replicated plot design too late in the text. As suggested, we write now that we used a dual experimental design right at the beginning of the description of the N-addition experiment (section 2.2, L. 137-147). We now refer already here in the text to Fig. 1, which clearly shows how both designs lie together in the field. We clarified also the meaning of the symbols for the replicated plots in the legend of that Fig. 1. Only after this basic information about the dual design do we bring details about the treatment itself. We believe that this description from more general towards more detailed will indeed be preferable for the readers.

*Furthermore, there seem to be some inconsistencies in how water and soil samples were measured. The information regarding the analytical methodologies is hidden in different sections of the methods, but I think that in this study this is critical information and deserves a sub-section of its own. Also, in this sub-section, the authors should clarify potential changes in instruments or methodologies that were (understandably) implemented over the course of such a long-term experiment.*

There were indeed some inconsistencies in the soil analyses as we relied mostly on samples dried soon after collection, but in one case we used samples that had first been frozen. This is acknowledged in the text (L. 197). There was, however, no specific result that could be linked to this fact and would have needed discussion. Laboratory analyses were done with standardized methods regularly checked with international standards. Otherwise, there were only potential inconsistencies due to aging of the pH-meters and renewing of electrodes, in spite of calibrations before each measurement series. As explained in the text (L. 180 ff.), a long-term trend due to the instruments was checked by remeasuring together samples that had been frozen after collection along 21 years. The trends observed on fresh samples or on samples frozen and later measured together were not equal but not really significantly different either (L. 328-330). We show thus both regression slopes, and now also their standard errors (L. 256 and 260). We discuss also possible errors that may arise from the conservation of frozen samples (L. 332-336). It is true that uncertainties remain about the exact rate of change and we do not hide this at all. Still, the main fact is clear: a strong increase in pH over time.

Because (potential) inconsistencies are so different from each other, we do not consider it useful to gather them into a specific section but prefer to discuss them directly along with the corresponding methods and results.

*In line with these comments, one wonders whether the observed decrease in pH of runoff water in control plots would be simply due to the addition of extra rainwater, which already has a baseline concentration of N forms. How do these trends compare to similar measurements of pH in runoff water from similar catchments?*

Compared to the N addition of  $22 \text{ kg ha}^{-1} \text{ y}^{-1}$  in the treated catchment, adding ambient rain water to the control catchment corresponds to  $0.8 \text{ kg ha}^{-1} \text{ y}^{-1}$ . It is very unlikely that this small amount cannot induce almost as much pH decrease than the amount sprayed to the treated catchment. The difference in slope in Fig. 6 is indeed very small. However, we take the chance of this question to add in the text that the amount of precipitation water added represents 7% of the ambient precipitation, allowing the readers to immediately see the order of magnitude that it represents, also in term of acidifying inputs (L. 139).

In the discussion, we now compare our trend for pH in the forested catchments with a third nearby catchment (L. 414-416). This third catchment is fallow land, originally mostly abandoned grassland, now with more and more trees. Due to this slow change in vegetation, it is not really a good comparison for pH trends, but it is the only comparison that we have to answer this specific question. Therefore, we indicate the trend observed there without going into more details about the own dynamics of a fallow land, which would be a separate topic.

*In addition, I provide some more comments that I hope they will improve the clarity of the language, which can still be polished. Comments that I consider are more relevant are separated by an extra line space. Line numbers (L) refer to the version with track changes.*

*L11 N deposition is not exclusive to forests. Also “forest ecosystems” can be shortened to “forests”.*

Done (L. 11).

*L12-13 what do you mean by “sensitive receptors of the environment”.*

Critical loads (CL) are generally defined in two ways: empirical CL that are based on vegetation changes and mass balance CL that consider the budgets of chemical elements. We indicate now these two aspects in a broad sense: "biodiversity and ecosystem functioning" (L. 12).

*L20 name these “soil acidification processes”.*

We removed the word "processes", as we examine the soil acidification itself and not the processes (L. 19).

*L20-21 I do not understand the addition of the clause “while previous reports...” I think you could remove this from the abstract*

Done (L. 18).

*L23 the sentence “In spite of...” is repetitive, this same idea is given in the previous sentence.*

Done (L. 21).

*L23 Remove “First”.*

Done (L. 22).

*L25 what does “they” refer to here?*

We mean the mounds, not explicitly (L. 23).

*L31 Close your abstract with a brief conclusion or future perspective derived from your study.*

We included the main outlook from the conclusion into the abstract (L. 29-31).

*L33 remove “more than”.*

Done (L. 34).

*L36 more concise language: remove “are a” and change “of” by “from”*

Done (L. 37).

*L38 more concise language: “and deposited in dry or wet form”.*

Done (L. 39).

*L40 New paragraph starting with “As large regions...”*

Done (L. 42).

*L54 provide a citation for this.*

This sentence is probably correct considering all the decomposition processes, but it is a sum of processes consuming protons and other producing protons. We did not find any study that would be broad enough to really conclude that consumption generally dominates. Therefore, we decided to remove the sentence.

*L62-64 I am not sure this is very relevant for this study and you could consider removing it.*

Done.

L116 “at a site”

Done (L. 110).

L125 “a deeper water table” instead of a “lower laying water table” (unless you meant a “more shallow water table”, which is not entirely clear here).

Done (L. 119).

L130 specify the net rate of which process.

It's about nitrification, now explicitly (L. 124).

L136 remove “their”.

Done.

L146 & 179 provide manufacturer details.

The weirs are a self-construction and this is now written so (L. 154). The manufacturer of the pH-meter and its location are now both given (L. 177).

L141 Here is where you should present the experimental design in detail and how the treatments were applied (currently L180-191)

See the answer to the second general comment above.

L140 I assume these values of N deposition refer to the levels at the beginning of the experimental treatment, please clarify this.

Yes, now clarified (L. 133).

L142 According to the responses to the reviewers' comments, one of the catchments was not delimited by a trench, could you clarify this?

This was only clear in the legend of Fig. 1, now it is also written in the text (L. 151).

L166-167 Can you provide the maker, model and manufacturer details of the rain gauges. Also, how did these gauges collected and measured precipitation? Were the two rain gauges located in the same open area? Where was this area with respect to the experimental and control plot?

The rain gauges were self-made from three parts: a tube, a funnel and a collection bottle. These details are now indicated (L. 172 ff.).

L174-177 How consistent were measurements taken more than two decades apart? How can you relate potential drifts with changes in measurement instruments, technology and their precision?

See the answer to the second general comment above.

L190 what does “it” refer to here?

Clarified (L. 200).

L192 what do you mean here by “in general”?

Deleted (L. 202).

L214 provide the corresponding package citations for this function.

Function `acf` in the core of R and there is no need for more indications.

L215-216 This belongs to the results section, also, what are “periods” here?

The periods are always of two weeks. This is now explicit in the text (L. 304-305).

L232 I do not understand this, I thought the variable “treatment” was included as a fixed factor on the models testing for the trends of soil acidity.

There was a problem with the order of the explanations that indeed made this confusing. We reordered the text to make it clear (L. 238 ff.).

L245 remove “highly” and “atmospheric” (precipitation is by definition atmospheric).

Done (L. 254).

L246 what do you mean here by “at the beginning of the project”? Maybe use instead “at the start of the experimental treatment” or “pretreatment measurements”?

Done (L. 254).

L247 remove “almost”.

Done (L. 255).

L247 & 250 These rates of change should be accompanied by an error estimate.

Done (L. 256 and 260).

L248 what do you mean here by “weather events”?

This is now clarified (L. 258).

L251 where do you show this result? Please refer to the corresponding table or figure.

Reference to Tab. S1 and Fig. 2 added here (L. 261).

L263-269 replicates instead of replications. Also, consider writing your results in past tense.

Done (L. 272 ff.).

L269 “were the most acidic” instead of “have the highest acidity”.

Done (L. 277).

L307 add “of precipitation water” after  $\text{NO}_3^-$ .

Done (L. 317).

L319 pH instead of “it”.

Done (L. 327).

L325 remove “likely”

Done (L. 334).

L333-334 I am not sure I understand this, soil acidification has to be intrinsically linked to the concentration of  $\text{H}^+$ .

It is about the effect of the acidity in precipitation on soil acidification. This is now clarified (L. 341).

L339 remove “in greater detail”.

Sentence rewritten without these words (L. 348).

348-349 I am not sure I follow how decreased N demand by the trees has a feedback effect on the concentration of Ca<sup>+</sup>, could you explain this?

L349 But I thought the felling resulted in a decrease in N demand, this is contradictory.

See the answer to the second general comment above.

L350-351 Please provide a citation to support this claim of low nitrification rates.

A reference is now given (L. 360).

L364 what soil property is buffered?

Done (it's about cation exchange, L. 372).

L371 provide citations and examples. Also consider “lack” instead of “don't contain”.

We give now a reference about the history of research on this topic as well as a study centered on Switzerland, the country of our own experiment (L. 379).

L376 how would deeper rooting help buffer acidification due to increased N deposition?

Deeper rooting would facilitate the uptake of base cations from soil horizons, which sustains the buffering by cation exchange. This is now clarified in the text (L. 384 ff.).

L376-379 This line of argumentation linking plant litter, denitrification and cation transport between mounds and depressions is not clear.

Trees (both root systems and crowns) are larger than single mounds or depressions. Uptake of nutrients and their restitution to the soil as litter act thus as a redistribution of elements between the small topographical features. This is now clarified in the text (L. 384 ff.).

L383 “peaked”

Done (L. 396-397).

L416 “emissions of N and S oxides”

Done (L. 433).

Figure 1. Add a legend indicated what are the blue and red circles.

We added this in the legend of the figure (rather than a separate legend within the map because it would in our opinion add complexity).

Figure 2-6. Please, increase font size of all labels and axes.

Done.