Answers to referee 2:

We thank the reviewer for thoughtful comments which helped to clarify the manuscript.

General comments:

The manuscript may benefit from specific objectives and hypothesis. Data analysis seems to be "fishing expedition" and the fact that some significant correlations were found is not sufficient evidence to prove that they are ecologically interesting. Correlation is not causation.

We agree that simple correlation is not causation and the ecological relevance of the results would indeed be questionable if not linked to familiar, well-established concepts of soil organic matter cycling/stabilization, as acknowledged by the referee later in the review. The study was not designed as an experiment that allows for formulation and testing of hypothesis for unambiguous cause and effect relations. This is not possible based on just regressions, especially when considering the variations in site properties. Our approach was rather to start from existing concepts on soil OC storage and turnover. Most of them have been developed based on a small number of sites or for topsoil layers, or laboratory studies. Testing their suitability for a range of sites under different management and with different properties offers the opportunity of rating them in a larger context. We will state that more clearly in the revised version and focus the manuscript more on specific objectives and hypothesis.

Main goal: "to test whether general controls emerge even for soils that vary in vegetation, soil types, parent material, and land use" or to test for common controls that are exhibited in a large range of soils. Their analysis looks for explanatory variables that are significant across broad environmental gradients, but not within individual vegetation or land use types. If they wanted to establish whether explanatory variables are important in many soil types, they have to us for example mixed-effect models to test for correlations between 14C or pool size and potential explanatory variables, while modeling land use or vegetation type as random effect. Currently they lump together all the soil types into simple regressions without accounting for soil type in their statistical models. This makes it impossible to tease apart the effect of land type as separate from the explanatory variables of interest.

We see the point the referee wants to make. As land use, management, soil type, climate … all vary at the same time at our study sites it is not possible to sort out individual factors apart. This is also the reason why we are only able to test for correlations across sites, but not within environmental gradients. Linear mixing models are quite useful but require sufficient replicates, e.g., of certain soil types and a sufficient number of different soil types to be reasonably applied. As stated before, our dataset is not sufficiently clustered to allow for determination of random effects. For proper variance partitioning, we do not have enough sites indeed. Therefore, our focus was mostly on studying effects across sites for which the lack of clustering of the data is useful. We used different symbols for different land-use types in each of the graphs, so the reader can see for example if land use causes clustering of the data along the total gradient or if sites within land-use classes follow other patterns than the total dataset. The small number of replicates impedes proper statistical tests of these trends. However, the used approach allows for identifying common mechanisms and factors.

This authors state that “roots rather than aboveground litter are the main source for LF-OC”. I would contend that this pattern could be due to the fact that the authors examine vegetation types with widely differing ratios of above- and belowground biomasses (e.g. aboveground litter input should be more important in forests with more biomass aboveground than in grasslands). Differences in foliage chemistry between vegetation types mute “universal” relationships. The greater importance of roots may occur across vegetation types but not within vegetation types.

We never claimed that root litter was more important than aboveground litter within vegetation types, as the number of replicates (maximal 3 sites per land use type) does not allow for such statements. In general I would rather assume greater differences between above- and belowground litter input between vegetation types than within them due to differences in plant allocation mentioned by the referee. Still, we found a close agreement in the depth distributions of roots and LF-OC despite differences between forests and grasslands. This straight forward approach supports the idea of roots being important sources of LF-OC across vegetation types. After all, I do not understand why roots should be less important as sources for LF-OC within vegetation types than across vegetation types.

The discussion appears somewhat unfocused, and though a great number of interesting concepts are discussed, it isn’t always convincing that the data presented in the paper actually significantly supports the concepts discussed. The manuscript may benefit from an overall reduction in the length of the text and the number of figures. It’s more repetitive than necessary, and 17 figures is excessive to illustrate key points. The organization could be improved and streamlined by combining the results and discussion, with subsections for each of the topics listed in their summary. As an organizing principal they could aim to describe the key findings summarized in their conceptual figure (Fig. 16) as efficiently and simply as possible. All of the concepts discussed in this manuscript are well-established theories of soil organic matter cycling/stabilization, though the results of the current manuscript seem to confirm these previous findings, the fact that these are familiar concepts makes it possible to describe the current investigation's findings more efficiently.

We restructure the discussion to focus it on specific objectives and, eliminate unnecessary repetitions.

We shorten the manuscript by removing former figures 6, 10, and 12. Figure 2 is reduced to the relevant information on OC-to-TN ratios and ^{14}C values of sites affected by fossil C.

Additionally, many times the data is presented in the form of somewhat unintuitive indices (e.g. percentage contribution of 10 cm depth increments to total stocks in 0-60 cm, contribution of fraction in 0-10 cm to total OC in fraction, contribution of roots in 0-10 cm to roots in 0-60 cm soil depth, etc.). It takes quite a bit of effort to tease out what these derived indices are supposed to represent conceptually, then an additional period of time to tease out what the correlations among indices are supposed to prove. Complexity is of course not a bad thing, but perhaps derived parameters should be used for very specific reasons to support a specific course of inquiry. The connection between these complicated correlational analyses and the initial research questions isn't stated in an explicit way anywhere in the manuscript at this point. I would suggest that the authors stick with straight-up pool size if possible.

We add simple pool sizes to show depth declines in LF- and HF-OC to Figure 5, and explain the motivation and calculation of indices in the methods section. Despite absolute amounts of OC in density fractions and of root masses varied greatly between sites, they still had the same depth distributions. One way to show that is by calculating the relative portion each depth layer contributes to total masses of roots or to OC in fractions, and to compare those among sites. Figure 6, which showed the correlations between indices, has been removed as it adds no additional information to the main objectives.

I would also suggest that qualitative description of patterns may be fine in some instances and more intuitive than the regressions the authors have calculated. They do not necessarily need to quantify the strength of every relationship to make their points.

While intuition can be quite useful in science, results and conclusions should nevertheless be properly backed up by the data and data analyses. The referee complains on the one hand in an earlier part of the review that the presented data and the applied statistics do not convincingly support the concepts discussed, while encouraging us now to work with less statistics and more intuitive. As no examples are given for each case it is difficult to reply to these comments.

Specific comments:

It was unclear to me what variables the authors equated with SOM "stability." For example, HF pool size? The relative proportion of total C in the HF? Pool 14C abundance? The authors should be more explicit and consistent about how they define stability. The word “age” is used throughout to describe differences in radiocarbon values among different soils. Since soil organic matter is composed of a variety of materials with widely varying radiocarbon values, the term “mean residence time” may be more appropriate. Depth is an extremely (and most of the time overriding) important factor in determining the bioavailability and mean residence time of organic matter in soils, yet depth is often ignored as a confounding variable in the correlational analyses.

Stability of OC was assumed when samples had small specific mineralization rates but radiocarbon data indicated long average turnover times. This is explicitly stated in the revised version.

We agree that OC is composed of a mixture with varying radiocarbon values so that the term “mean” or “average” is more appropriate. However, residence time is only appropriate if referring to residence time in the ecosystem and not in a specific fraction or soil layer. Therefore, we think that age defined as time since C fixation is a more intuitive and shorter term.

We are aware that depth has a strong impact on biological activity and the mean residence time of OC and therefore we analyzed different depth increments separately. Using soil depth as a covariate is only necessary when samples from different soil depths are bulked into correlation analyses.

Why were soils sampled by depth instead of genetic horizon? Many of the indices of soil physiochemical character used here vary more profoundly with genetic horizon than depth (e.g. clay content, Al and Fe oxide content, root abundance). Though this obviously cannot be changed at this point, the motivation behind sampling by depth might be a good point to discuss in the methods.

The soil sampling design was developed and optimized for soil carbon monitoring. Sampling by fixed depth intervals will have less bias from processors and sampling by depth instead of horizon reduces the variation of OC stocks and enhances accordingly the chance of detecting changes. Sampling by depth increments also facilitates easy comparisons between sites. While topsoil horizons might be used for comparisons between sites, it can be difficult to judge, which of several subsoil horizons of one soil type should be used for comparison with one large subsoil horizon at another site or the completely different subsoil horizons of a third site. Identification of different horizons at replicate soil cores would already complicate averaging within sites. We agree, the approach is less suitable for studying soil genesis effects within sites.

When mineralization rates were measured, were the samples periodically vented to prevent CO2 concentrations from inhibiting respiration rates? Incubations under laboratory induced conditions (moderate temperature, high oxygen availability) do not mimic the conditions present in natural environments, therefore the results of laboratory incubations may have very limited applicability to interpreting organic matter trends observed in natural soils.

Yes. The containers were closed only during the period of CO_2 accumulation (2-5 hours) and aerated during most of the incubation time, however, only with a 5-mm diameter aperture to prevent strong desiccation. We are aware that laboratory incubations cannot be directly transferred to field conditions; this issue is raised in the discussion section.

Section 13101-13102: The paper states "Increasing LF-OC with decreasing clay contents, and increasing LF-OC at higher OC loadings of clay particles or pedogenic oxides in the uppermost soil layers indicate a greater importance of LF-OC for OC at sites with limited sorption capacity of the HF". This is a very complicated statement, and it may be confusing two very different phenomenon that actually have very different causes. It is true that the amount of total C allocated to the light fraction is almost always higher in sandy soils than in clayey soils since mineral surface area is small in the former case. It is also often the case that soils with higher organic inputs have high "C loading" values. However, these are two different scenarios and have little bearing on each other. The former is caused by a lack of available surface area for sorption of organic inputs. The latter is due to high organic matter inputs. In both cases, it could be argued that an increase in free/light fraction does not really represent any type of "importance in storage" since free/light fractions (especially in surface horizons) have short turnover times.

We agree that the larger amounts of LF-OC are probably caused by a lack of mineral surface area. Nevertheless, there is no plausible reason to believe that soils with high OC loading of mineral surfaces receive larger OC input; it might be even the opposite. Sites with high amounts of OC in the HF despite small amounts of clay and pedogenic oxides are the dry, sandy grassland site Bugac in Hungary, the sandy coniferous forest in Bordeaux in France, and the coniferous forest Norunda in Sweden. These sites rather share small contents of clay-sized material than large productivity. Therefore we hypothesize that potential binding sites at mineral surfaces are mostly occupied in clay-poor topsoil layers leading to large OC loadings of the HF, while that is not necessarily true in topsoil layers rich in binding sites. The question arising is when and why LF-OC accumulates in excess to other sites in sandy soils as the LF typically turns over fast. It is either fed by large input rates, keeping average LF-OC stocks up as long as the input is constantly high but renders the OC accumulation more sensitive to disturbance than OC accumulating in the HF. Or LF-OC accumulation is due to environmental constraints reducing decomposition, such as poor litter quality, acid soil conditions (e.g., under coniferous forests) and summer-drought at the grassland site. We cannot solve the question with the available data set. Since it is not the focus of the manuscript, and to avoid misunderstandings, we decided to remove Figure 10 and the discussion on links between OC saturation of the HF and LF-OC accumulation.

Section 13101-13102: "Our study confirmed previous observations that secondary hydrous Fe and Al phases are generally more important to OC accumulation in the HF than clay-sized particles." Fe and Al oxides often are clay-sized particles. It may be more clear to say "secondary phyllosilicates" or "total clay content"

"clay sized particles" are replaced with "total clay content".

In addition to the points made in their summary, it would be very interesting to see an additional section dealing with transport mechanisms. It would be good to see this topic brought from the supplement to the main paper (e.g. It's discussed briefly on p. 13103). Regulation of SOM inputs and transport could have very important influence on 14C profiles, and this is treated too cursorily.

We agree lateral and vertical transport is important for shaping ^{14}C profiles. As we have no other information than DOC fluxes for a smaller subset of the study sites available, we can only speculate about other transport like bioturbation or erosion. However, grasslands and forest are typically not extensively prone to erosion, and the arable sites were all in rather leveled settings (relevant for the eddy covariance measurements performed at the sites). Erosion is therefore only a minor factor at the study sites. The close depth distributions of LF material and roots suggests little input of aboveground litter at the study sites, thus, bioturbation is seemingly a minor factor at our study sites, too. As the manuscript is already quite lengthy, we prefer not discuss these mechanisms in great detail, but add bioturbation explicitly to the summary figure.

The figures and tables in the supplementary materials are mislabeled in the text (i.e. numbers are mismatched between text and actual figures/tables).

Correct application of numbers of figures and tables from the supplement were crosschecked and revised.

Figure 4: This figure is missing a key for the symbols

A key for the symbols is added in the revised version.

Figure 7: One cropland is included in this figure, but it’s difficult to interpret the influence of root abundance on OC stocks given that crops may vary from year to year and root abundance will certainly change drastically with season/harvesting.

The cropland sites Gebesee and Grignon were excluded because they were tilled prior to sampling, confounding any relation with roots. Spring barley was cultivated at the third cropland site Carlow each year since 2000 and samples were taken prior to ploughing. Nevertheless, tillage will typically have an impact on litter input and distribution despite rooting patterns.

Figure 8: Why is the “Ca” site labeled in this figure? If it is being considered an outlier, was it included in the correlational analyses or not?

Carlow is labeled to show that this cropland site has lower LF-OC amounts than expectable based on other soil properties. It was not excluded from the correlation analyses. We remove the label as it is confusing.

Figure 9: Outliers in the dataset are giving falsely high r values. The data needs to be transformed and the linear regression reapplied.

It is true that the Andosol at Laqueuille has exceptionally large amounts of Alo and Feo. But, e.g., logarithmic data transformation does not change that. Excluding the Andosol we still get significant correlations for Feo and Alo (for Alo, the site Laqueuille was already excluded).