

## Author Response to Anonymous Referee #2

Review comment	Author response
<p>All LTFE are situated in flat areas (a data evaluation in this respect would be nice and not too difficult to do). This means that they exclude major lateral processes (interflow, surface runoff) and differ largely from typical agricultural fields.</p> <p>This deficit may be especially pronounced for grassland experiments because grassland either occupies lowland areas that are too wet for arable use or areas that are too steep.</p>	<p>Indeed, lateral processes are typically not analysed in LTFE and they are not designed for such questions. Different design such as the 'Wishmeyer plots' are implemented for erosion studies. We are going to write a section explicitly about deficits in the setup of LTFE and about related experimental setups such as Wishmeyer plots. We also like to write something about the differentiation from LTFE and soil monitoring sites (BDF).</p>
<p>For grassland experiment, which in fact are meadow experiments (grazings seems to be missing; also a major deficit). Such critical assessment would be extremely helpful to guide the installation of future LTFEs and to show the limitations in the conclusions that can be drawn from the existing LTFEs.</p>	<p>We will include grazing as example in the discussion of limitations of existing LTFEs.</p>
<p>Were lysimeter experiments included, which would allow assessing at least vertical water fluxes? Do long-term experiments with lysimeter exist at all in Germany?</p>	<p>No, lysimeter experiments were excluded because they were considered as an own category. Some reasons are that soils are often transferred and not undisturbed in lysimeter experiments and tillage has to be conducted by hand instead of machines, which can bias some results. Indeed, longterm lysimeter experiments exist in Germany as part of the Tereno network. We will clarify this.</p>
<p>Were experiments included that allow quantification of lateral processes (runoff, soil loss)? I could imagine that the measurements in Trier (Stehling and Schmidt 2017) or those by Jung and Brechtel (1980) qualify for LTFE. If they don't qualify, this would again illustrate a major deficit of present LTFEs.</p>	<p>Our response to the first review comments also holds here. In addition, we check again the mentioned resources to evaluate that.</p>
<p>In the discussion I missed a wider view. Do similar compilations also exist in other countries? Are the German LTFE experiments similar to what was done and is done in other countries?</p>	<p>We will include a section about the international situation.</p>
<p>Furthermore, the authors give the impression that they still focus on the old questions of LTFEs (mainly yield) that became boring. I had this impression for two reasons. First, little examples are given how LTFEs can be used in fascinating modern research on urgent questions. Second, using LTFEs in modern research applying new techniques requires access to the experiments. Hence it makes a big difference whether an experiment is still ongoing or not. However, this information is given nowhere. Second, it often requires</p>	<p>Information on whether an LTFE still exists or not can be found in the extensive data set, which can be found under the following DOI: <a href="http://doi.org/10.20387/BonaRes-3tr6-mg8r">http://doi.org/10.20387/BonaRes-3tr6-mg8r</a>, 2019</p> <p>We asked in a questionnaire whether there were any retain samples or not. This information is available for 40 LTFE. The relatively small number compared to the total number stopped us from integrating this information into the data set. But we can do that in a next version of the data set.</p>

<p>archived samples (as an example what can be done with modern techniques and archived samples, Köhler et al. 2012 comes to my mind but there are certainly more examples). This information, whether archived samples are available, should be included. Generally, I missed information about which data could be obtained from the LTFEs.</p>	<p>Information about which data have been collected in the experiments can also be found in the dataset (“research parameters”).</p>
<p>Most of my other remarks are mainly editorial issues. The weakest part in this respect is the table in the Appendix, which is most important because it resolves the LTFEs and thus allows access (see below).</p>	<p>We could imagine to provide the whole dataset, which is very extensive, as supplemental material instead of the Appendix.</p>
<p>12: add "during the growing season"; I would even change the abbreviation to CWBg because usually an entire year is considered in a CWB. I was very surprised when suddenly somewhere in the manuscript the information 'growing season' popped up</p>	<p>We will do that.</p>
<p>13: Müncheberger Soil Quality Rating seems to be a combination of German and English. Shouldn't it be 'Müncheberg'?</p>	<p>You are right, we will change accordingly.</p>
<p>35: I welcome this definition of the control that is certainly better than the often used but wrong assignment of the strongest and most unrealistic intervention as control, namely the long-term nutrient removal. However, I did not find this definition to be used later in the manuscript.</p>	<p>Yes, we used this term only to give an example on how LTFE could be analysed collectively. We are going to write this part more detailed, also due to the comments of Referee #3.</p>
<p>46: Bai et al.</p>	<p>We will change accordingly.</p>
<p>116: Not clear how PET was derived. Was it taken from DWD? Is it Haude?</p>	<p>The PET was already included in the DWD data of CWB.</p>
<p>126: This is strange. Later only 6 classes of the MSQR are used, not 102. I wonder whether different properties like soil structure, wetness, relief, contaminations can be combined in one indicator of six classes. This may be possible for one specific target like yield but will fail for most other targets or require other classes. Is a better resolution than these six classes possible?</p>	<p>The soil qualita rating is is performed on an ordinal scale of 0-102 and clustered into six quality classes. We will add this information to clarify.</p>
<p>128: I guess this should read 'available water capacity'</p>	<p>The source says 'profile available water', just as Mueller 2010</p>
<p>130: What is unsuitable? This always requires the definition of a target.</p>	<p>We cited the source correctly, but we can add "for crop production" here.</p>
<p>139: This leads to the question: Were lysimeter experiments included? If not, why not?</p>	<p>See above</p>
<p>155: The title does not have this restriction; also the Abstract does not. I wonder why it suddenly pops up in the results. I also wonder how this is defined (what is bioeconomy?) and whether these experiments really aim at sustainable soil</p>	<p>Most LTFE were originally implemented for agronomic purposes. Accordingly and particularly for grassland LTFE, most research questions are agronomic in nature and not closely related to the soil. In this paper, we</p>

use. They exclude many things that make soil use unsustainable (erosion, compaction) and hence are unsuitable to test sustainability (in this general sense). I also wonder even more why the criterion sustainability excludes some grassland experiments. This is contrary to what I would expect.	intended to reveal the value of LTFE for soil related questions. We therefore only included those LTFE in our study, for which soil data are existing. We will state this more clearly. Besides that we will point to the deficits in LTFE setup as mentioned above (erosion, compaction, grazing).
160: Establishment was in the past. Hence past tense would be appropriate. The question of correct tense is rather difficult to answer given that 30% of the experiments have come to an end already and others will come to an end in the future, I wonder whether the mostly used present tense is justified.	Ok, we consider past tense.
171-172: One sentence is usually not a paragraph. Furthermore, temporal aspects were treated in the first paragraph of the results. I suggest moving this sentence.	Ok, good idea.
173: sentences usually do not start with a number; this also applies in other cases (e.g. L. 181, 184).	Ok, we will write out the numbers with letters; I think that will be correct.
178 : Move opening parenthesis	Yes, thank you
208-209: This should be moved to the M & M section; this is the first time that growing period is mentioned although CWB appeared already several times. Furthermore, it would be good to explain the rationale behind this decision than let the reader speculate	Ok
266-269: I would reverse the argument. In my view the critique by Franko is well justified and shows that 6 classes of the MSQR are insufficient. I do not suggest to include an assessment of the complexity of soil parameters but it is also not justified to say that the LTFEs are representative regarding soils just because they match the rather coarse and restricted (to yield) MSQR criterion.	We agree. We intended to say which CWB/MSQR combinations are less well represented in the existing LTFE having biomass production suitability in mind. For specific questions such as the representation of C-dynamics in simulation models other requirements to long term information exist. We will clarify the part. Furthermore most likely we are going to include in addition to MSQR and CWB an assessment of the distribution of LTFE according to clay content with clay data from JRC.
References: The format varies among references. Please homogenize	Ok
Fig. 2: The pie charts are an attempt to illustrate the manuscript. However, they do a poor job. They require a legend, which is difficult to read (because font size is smaller than that of ordinary text) and contain information that is better suited for a table or even could be given as plain text. For Fig. 2 a, a density graph would be more appropriate	Ok, we put this information into a table respectively a density graph.
Fig. 3: A graph usually has not a title but a caption. The colors are impossible to distinguish	We are going to change the colours respectively to change the whole figure.

<p>Are they necessary? Can they be simplified? Wouldn't the year when an LTFE was closed be equally interesting?</p>	
<p>Table 1: It is not clear whether 'organic fertilization' also includes straw and compost (there is not an equivalent 'Mineral fertilization'). Furthermore, why are green manure, compost and sludge mentioned, but not the main type of organic manure? This classification appears inconsistent. It surprises me that only two of the grassland experiments have organic fertilizer although grassland use unavoidably produces manure. Have all except for two experiments used an unrealistic design that does not allow application of the results to typical situations? Better call 'plant protection' 'crop protection'</p>	<p>We will improve the table accordingly.</p>
<p>Fig. 4: same remark as Fig. 2</p>	<p>Ok</p>
<p>Table 2 + 3: 'vegetation period' should not be in the column head but in the caption. Also the lines separating groups of variables are not consistent (why are CWB class and range separated by a line? Isn't the unit for CWB mm/yr?</p>	<p>We will change the tables accordingly.</p>
<p>Fig. 5: Here four classes of LTFE are sufficient. Why does Fig. 3 require eight classes (that cannot be read anyhow)? LTFE should not be repeated five times in the legend. It is not necessary at all. CWB is in mm/yr</p>	<p>For the map we simplified the classes to avoid complexity. We are going to simplify figure 3 also. We skip LTFE from the legend. We change the unit of CWB.</p>
<p>Fig. 6: Delete LTFE</p>	<p>Ok</p>
<p>Fig. 7: column widths could be much smaller while larger row heights would allow a larger font size. Presently the numbers hardly can be read. It is not necessary repeating 'MSQR class' six times. Better use a larger font size. The colors of the legend should agree with the colors in the graph.</p>	<p>We will change the figure. Referee #1 also commented on this figure and suggested points or lines.</p>
<p>Table A 1: This is likely the most important table because it allows access to the LTFEs. However, it is rather inconsistent and difficult to read. E.g., the IDs cannot be read; some institutions got abbreviations (why?) others not; some places are mentioned, others not (why?). Mentioning the main institution may be fine in hierarchical organizations but this is clearly insufficient for big universities. Whom should one ask there? I suggest replacing the information in column 3 by a number and the place and resolving the number below the table by reporting the full addresses. This would also create room for the other columns. Furthermore, I see no reason why umlauts are</p>	<p>We will change the table accordingly.</p>

replaced. This is poor technology of the past century and again a waste of space.