Revision report soil-2020-21

Modelling of long term Zn, Cu, Cd, Pb dynamics from soils fertilized with organic amendments.

The manuscript presented here deals with the model prediction/description of measured data that reflect the Zn, Cu, Cd and Pb dynamics in a long-term field trial amended with different organic amendments. Then, after evaluation, the used model is extrapolated to the future, to evaluate the possible risks of TEs by long-term application of organic amendments on agricultural fields. The advantage of this model is, according to the authors, that it has a restricted amount of input parameters. However, there are several issues with the manuscript in its current state.

1) The model description is not sufficient and scattered over different sections, which makes it difficult to link the different parts in the model. In addition, it should be much more clear which input parameters are needed for each part of the model (i.e., SOM and pH are likely used to calculate Kf's? What are input parameters for the WHAM + which complexes are considered?) + references for ageing parameters are not presented. At least an overview of considered reactions should be presented in the SI, to allow the reader to evaluate the restrictions of the model.

2) The measurements of the TEs lack quality control, the limit of quantification of each element and the relation with the measured data should be presented and, maybe most importantly, the Cd concentrations in the extracts are measured with ICP-OES, what could be troublesome regarding the known interferences during ICP-OES measurements with As in soil extracts.

3) The implementation of DOC in the model. Because no measurements were available, the DOC is fixed at 7 mg/L for all treatments (as in the beginning of the experiment) and remained constant, despite the application of different organic amendments. I am critical to this approach, due to the important effect of DOC to metal leaching. I would expect increasing (or changing between treatments) DOC concentrations over the years of the different organic amendment applications or at least increased DOC fluxes right after organic amendments that could increase metal leaching.

4) The artefacts associated with the experimental design of the ZOFE, namely plots touching each other + mixing of plots edges by ploughing. In the model evaluation, it appeared to be critical to introduce lateral mixing. This does not allow to evaluate the model without mixing, which is later used to extrapolate into the future. The lateral spread of TE concentration should be validated in a transect across plots in the ZOFE experiment by some new measurements, to underpin this model approach.

5) In the results, the treatment effects should be evaluated relative to the control data, to evaluate and contribute observed trends to the organic amendments solely, which is the scope of this study.

In addition, both the abstract and introduction lack quantitative data, the English writing could be improved and the final discussion of the results becomes difficult to follow starting from lines 408 to the end.

Further point-by-point comments are presented below.

Abstract

36: abbreviation of model

38: soil plots, different. Are there more organic amendments than the ones summed up here (particularly is not the best link word here)

39-40: link with previous sentence is missing, maybe this sentence can be declined here + don't start sentence with an abbreviation.

41: better provide quantitative measure of model performance.

41: abbreviation ZOFE

Wouldn't it be interesting to add the range of EDTA-extractable concentrations here in the abstract?

46: labile = EDTA-extractable? + provide projections, i.e. after XX years, concentrations could increase to YY.

Introduction

57-58: this sentence is too vague and is not really necessary here, can be skipped.

59-64: it would be interesting if you would add the concentration ranges at which the essential and the non-essential TEs become of a concern.

68: an

68-71: please rephrase to make the message more clear.

71-75: the link with the previous section is not clear. In addition, what do you mean with "limited natural availability of nutrient elements such as P"? Preferably start your sentence with the main message, for example "Organic amendments are considered to be more sustainable then inorganic mineral fertilizers, due to XX and YY".

77: isn't it just the transformation of the organic amendment to SOM that contributes to carbon sequestration as such, not additional C sequestration from atmosphere? If not, please explain more, but only if relevant for this study!

Actually, line 71-77 could be skipped from this introduction and you could go right at the possible introduction of TEs into soil by organic amendments, to keep the introduction to the point and relevant for this study.

77-81: please provide concentration ranges for the TEs in the different organic amendments.

83: fate? + please rephrase second part of sentence

86: you mean mobility (for solubility)?

87-89: please provide examples or more explanation of importance of TE speciation vs toxicity

93: direct reactions?

90: I think you can even say that it is not only useful, but even necessary

97-98: you already stated this in line 90-91

99-100: Please rephrase: I would not say that mechanistic models are site specific. Indeed, models that predict TE mobility and transfers over time by using as much as possible underlying physical and chemical mechanisms are likely only useful on a limited size scale, due to the high input needed, but they can be used at every site (when data are available), so they are not site specific. And what about empirical models?

102: what do you mean with behaviour?

102-106: Ah, here you talk about empirical models. Please merge with text above and try to be more concise.

112-115: this is vague, could you specify more what variables you are talking about, and which mechanistic level of understanding is wanted.

Has this model been used already (and at what scale)? If yes, please provide the current state-of-the art of the performance on this model. What is the knowledge gap here for this study now?

121: IDMM-**ag**?

122: What is "larger scale" here?

123-124: "The hypothesis was that, if the model is successfully applied at field scale with no need of calibration, it might be used at larger scale as well, provided adequate inputs." -> can you test the second part of your hypothesis with this study? If not, please rephrase the hypothesis.

125: ZOFE?

131-133: please clarify sentence: "large scale", "broad trends", TE concentrations in soil? + rephrase final part of sentence.

M&M

150-152: for TE accumulation, the total applied amendment will likely be important, depending on the data collected on the TE content (per kg of material, per kg of OM,… ?).

160: 1M HNO_3 extractable metals are not total TE concentrations, please just write 1M HNO_3 extractable metals. And please also provide more experimental details, L:S ratio, extraction time. In addition, it would be interesting if this 1M HNO_3 extracted TEs could be related somehow to "total" element concentrations, measured by aqua regia or XRF or other, more standardized extraction protocols for total soil metal concentration.

162-163: please shortly describe, conc. of EDTA, L/S ratio, extraction time.

164: try to avoid the use of "we"

164-166: on what is the use of EDTA:1MHNO_3 extractable metals as a measure of lability based? Please clarify. Is this already used/tested? If so, please provide references.

What about quality control of these measurements, what is the limit of quantification of these methods and how do the measured soil samples relate to this? This could be described already here, or in the results section. In addition, the determination of Cd with ICP-OES in extracts from soil samples is troublesome due to the interferences with As, even at relative low As soil concentrations. For example see: "A comparison of reliability of soil Cd determination by standard spectrometric methods, M. McBride, JEQ 2011 (40, 1863-1865, doi: 10.2134/jeq2011.0096) and likely many other publications. Did you take this into account? If not, the reliability of the Cd measurements from this study can be severely questioned.

178: preferably write : free and adsorbed TE ions, in contrast to "free TEs"

176-179: please provide an overview of the Freundlich isotherms and TE complexes considered during this study in the supporting information. What are the input parameters for the Freundlich model (i.e. which extraction did you choose to represent the adsorbed fraction, and how does this relate to the adsorbed fraction represented in the initial models of Groenenberg (I think they used 0.43M HNO_3 acid extractable metals as "reactive" soil metals). In addition, what other soil properties were measured to calculate the K_F by the transfer functions of Groenenberg and how are these soil

properties measured? What are the input parameters of the WHAM model and how are they measured?

180: please provide the first-order rate constants used for each element and explain wherefrom they are derived (references).

184: please provide the start year used in the calculations

186-187: please rephrase this sentence, it is not clear what is stated here

204-206: please clarify, not clear.

207-208: please be consistent in choice of unit

208-211: please provide the data of these fitted mineral weathering fluxes and compare with literature data, if possible.

217: on what are these transfer functions based? Please shortly describe.

220: P loading from the manure? You mean addition of P to the fields by manure application? You could quickly give the data here.

222: not clear, did you take the X:P ratio's from literature (=1100 data points) or from own measurements (= 2 data points)? Was P also measured in the FYM? That is not been described previously. In addition, how does the measured X:P ratio related to the literature reported?

227: derivation of these factors? One time decrease or decrease linearly with time? Not clear.

231: the detection limit in the caption of table 1 is expressed in mg/L while the concentration data are expressed in mg/kg. Please provide detection limit in mg/kg, to provide a clear idea of the lowest measureable concentration in the FYM. See comment above on the analytics of soils and organic amendments.

232: written like this, Figure S2 is about detection limits, which is not. Please rephrase.

238: **t**TEs + avoid "we"

245: which peaks? In the soils?

246: soil metal concentrations negligible?

Figure S3: in the swiss sludge trend, the Cd fluxes trend is deviating from the other metals. Why?

253-255: Ok at the start of the experiment, but I would expect increasing or changing DOC concentrations over the years of organic amendment applications (or between treatments) or at least DOC fluxes right after organic amendments. I think this approach (constant & low DOC over time) is not a good approach to simulate metal leaching over time in organic amendment treated plots (I assume this constant DOC is then the WHAM input?).

256: on what is this plausible range based?

257: minor increase for fitted additional input flux? Not clear. + what is minor?

261: not clear

261: measurements of plant material has not been presented + data for Pb? In addition, changes of plant TE concentrations with changes in labile TE concentration in soils were not considered?

267-268: It is not clearly explained how SOM and pH affect soild/solution partitioning, aging and speciation, because the input data/model description for the Freundlich, WHAM model and aging model are not well specified.

284: pecks?

290-291: why?

305-306: Please state how the soil total concentrations are/should be measured in this Swiss Ordinance and comment on own measurement data.

311-313: not clear + why 0.1?

Results and discussion

324: the trends in the organic amendment plots should be investigated relative to the trends in the control plots, to exclude all other enrichment/losses other than use of organic amendments, which is the core of this study. Then, the statistical analysis should be repeated on these relative data.

334-336: compare the measured Cu loss with the literature values + on what is this "expected" Cu leaching based.

348-364: same comment, compare treatment effects relative to control.

350-357: you should test correlations between the data to underpin these suggestions.

370: P-overfertilization? Based on what?

378-388: this was already (partly) described in the M&M section (and provides answer to above comments), please move this section to the M&M.

390: in the figures (also in SI), the ZOFE trend is mentioned. Is this the "Idealized trend"?

399-401: I have severe doubts of the applicability of the modelling results of TE dynamics to a realistic scale, as the experimental conditions of the field experiment are so specific, i.e., high TE concentration plots "contaminate" low TE concentration plots, so all the treatment effects are obscured by an experimental artefact (i.e. the plowing, the plots being so close to each other…). To be more clear -> the model was not capable to predict the measured concentrations, because the measured concentrations are affected due to the specific design and maintenance of the experimental plots, but such experimental plot are not relevant for real agricultural fields (i.e. narrow soil strips with different amendments that are influenced by lateral mixing), which makes the "fixing" of the model with the lateral mixing not really important for real situations. In addition, due to the fixing of the model by lateral mixing, the true performance of the model cannot be evaluated, and the extrapolation (done in figure 9) to real fields is questionable. However, I understand that this is related to the specific nature of this experimental design and that it is nevertheless worth to investigate the data available, due to the valuable information present from these long-term experiments. However, to verify the overall modelling approach (including lateral mixing and excluding it again to extrapolate the model), I think a simulated transect from figure 6 should be validated by measurements-> i.e. sample along a transect in the ZOFE experiment, measure labile concentrations and remodel for the sampling year.

402: I guess only the r-value of the Cd is significant? Provide significance of r-values.

408 and further: not clear anymore. Initial measurements underestimated? I've understand that these were fitted?

471-472: but the EDTA-extractable concentrations were measured? Couldn’t this provide information of the “lability” of the TE input by the organic amendments…