Reviewer 1

We thank the referee for the detailed comments which helped a lot to improve the manuscript. In this manuscript, we stated a new hypothesis, revised some explanations and conclusion as follows:

Line 18: We changed the sentence “Stagnant water conditions may release phosphorus (P) in soil solution that was formerly bound to Fe oxides” into “Phosphorus (P) species in colloidal and “dissolved” soil fractions may have different distributions”.

Lines 31-34: We changed the last sentence into “We conclude that P species composition varies among colloidal and “dissolved” soil fractions after characterization using advanced techniques, i.e. AF4 and NMR. Furthermore, stagnic properties affect P speciation and availability by potentially releasing dissolved inorganic and ester-bound P forms as well as nano-sized organic matter-Fe/Al-P colloids.”.

Q: The extract used (MQ water) is quite harsh compared to natural waters such as rain water or pore water and would result in significantly greater release of P than that possible by contact with water in a natural environment, due to desorption and dissolution of poorly crystalline authigenic mineral phases. Living cells within the soil would also certainly undergo significant osmotic stress likely resulting in osmotic rupture and release organic and inorganic P found in intracellular components. The potential ramifications of these effects on the results should be clearly stated and discussed, as there are clearly implications as to the origin and mobility of identified P species in a natural context. Are the species identified in the size fractions indicated present in the natural soil or a result of alterations during the extraction procedure?

A: We agree that a contact of soil to rain and pore water would provide a more realistic scenario; yet, rain and pore water chemistry is variable and thus hard to standardize. As a result, the release of natural nanoparticles from soil could also be variable. Using MQ water for extraction instead of aqueous solutions with higher ionic strength, however, has also two advantages. On the one hand, as also stated by the reviewer, it increases sample dispersion in that we get access to potentially dispersible colloids. We stated this more clearly now. On the other hand, there are analytical advantages, because we avoid interference of additional ions with the retention of particles on the membrane in the channel of FFF, and because MQ water better allows to freeze-dry large amounts of soil solution. Natural water would increase viscosity for the re-dissolved solution, which could increase line broadening and thus decrease the ability to differentiate peak resonances from one another (Cade-Menun and Liu, 2014). We agree that MQ water has potential ramifications of the effects on the results as the reviewer mentioned; as such we added related discussion in the main manuscript as follows:

Lines 123-130: It is worth noting that Mill-Q water was used here to extract soil colloids instead of rain water or pore water, since total amounts of WDFCs will likely be larger when using Mill-Q water, i.e., we consider these WDFCs as potentially water-dispersible colloids. In addition, the use of Mill-Q water facilitates subsequent sample processing with AF4 and NMR. It is inevitable that Mill-Q water would result in the release of P due to desorption

and dissolution of poorly crystalline authigenic mineral phases. Additionally，living cells within the soil would also certainly undergo significant osmotic stress, likely resulting in osmotic rupture and releasing organic and inorganic P found in intracellular components.

Q: I have concerns with the way in which the results are framed within the context of oxygen availability and iron redox cycling. The first sentence of the abstract "Stagnant water conditions may release phosphorus (P) in soil solution that was formerly bound to Fe oxides" implies that the P release investigated is due to reductive dissolution of ferric oxides in the absence of oxygen. Undoubtedly, oxygen availability differences between the soil samples selected resulted in differences to iron speciation, particle size, organic carbon content and P speciation. The handling of the soil samples in the laboratory does not appear to have preserved the field redox conditions and likely resulted in considerable oxidation of reduced iron species during processing, particularly in the sampled Stagnosols. Oxidation of aqueous Fe^{2+} and colloidal ferrous particles can be very fast (seconds to minutes) therefore the extraction in presumably oxic MQ water for 18 hours almost certainly changed the composition and speciation of the colloids, which were later characterized. Although the importance of Fe oxidation and reduction processes on P speciation generally is highlighted in the manuscript, the impact of these processes during sample processing and on the final dataset is not discussed. The differences between the three soil types are convincing but I question whether the analyzed species and size fractions are representative of the soils themselves or of differences in response to the extraction procedure based on different initial soil redox conditions. Extracting soils in MQ water, under oxic conditions, is not representative of P released during reductive dissolution, as implied in the abstract and in fact would result in the opposite process (oxidative precipitation of Fe hydroxides).

A: We agree. However, we also have to annotate hear that stagnant water conditions do not mean that there was stagnant water to the very top of the land surface at time of sampling. By definition, stagnant water dominates for most time of the year and most parts of the soil profile, but it must not (and was) not present in the very surface soil at each time of sampling. When we sampled, the soils were not saturated, i.e., they must have been aerobic already (as common in these surface soils, also in Stagnosols). Hence, the experiment process with Mill-Q water under oxic conditions has potential impact on oxidation of aqueous Fe^{2+} and colloidal ferrous particles, but we do not see this risk as very severe, because we sampled (and stored) the soils in aerobic conditions. We mentioned it in the manuscript as follows:
Line 130-136: It is worth noting that the experimental procedure with Mill-Q water under oxic conditions may have an impact on oxidation of aqueous iron (Fe^{2+}) and colloidal ferrous particles. However, at time of sampling, the very surface soils were not fully water saturated as allowed even for Stagnosols for time of the year. As such, the analyzed species and size fractions are representative of differences in response to the extraction procedure based on different soil redox conditions that reflect a kind of legacy of former redox cycle, but at time of sampling and analyses the soils were aerobic.

Q: 89. Inclusion of a site map would be useful here in the main manuscript rather than in the supporting information. A scale should also be included to establish the distance between the sampling sites.

A: We added the map with a scale in the main manuscript as follows:

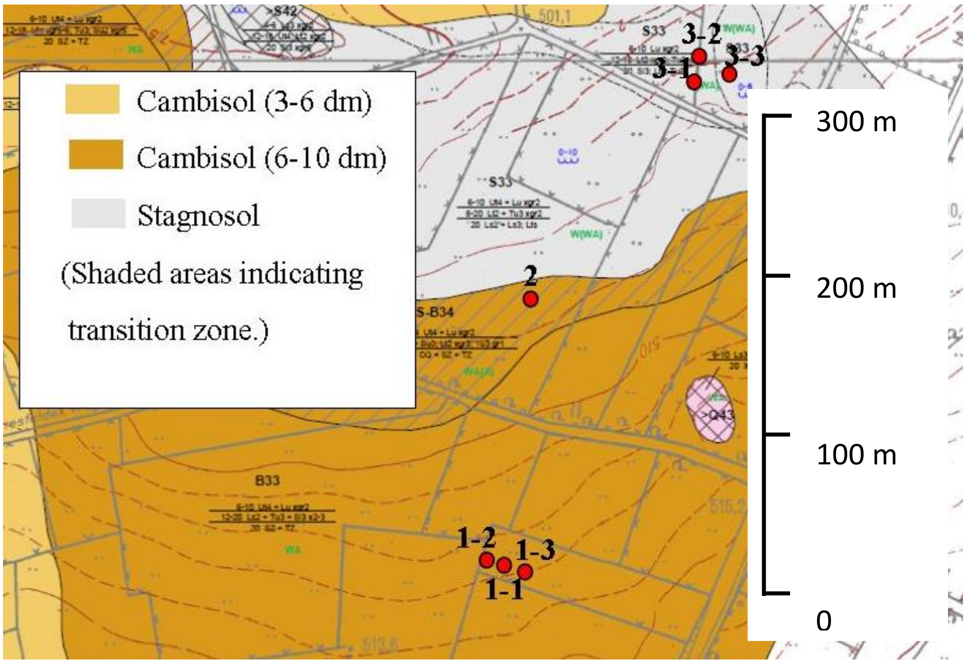

Fig. 1 Excerpt from the soil map of the test site at Rollesbroich (*modified from Geologischer Dienst Nordrhein-Westfalen, 2008*). Numbered red dots indicate location of plots.

Q: 101. How long were the samples stored at 5°C? Long storage times prior to extraction and preservation could result in significant speciation changes. It is impossible to evaluate the importance of these changes if the storage time is not provided.

A: The samples were sieved immediately to < 5 mm and stored at 5°C for less than 6 months before the extraction. All samples were stored in similar manner. The FFF characteristics of WDFCs did not change significantly in the 6 months period of the investigation. We added this information in the manuscript (lines 158-159).

Long storage time under oxic condition have potential impact on the forms of Fe-minerals in soil. However, it is also worth noting that we sampled topsoil (2-15 cm) from Stagnosol which is not the horizon where water is actually stagnating. Additionally, stagnic water conditions do not mean that the soils are under reduced conditions for the whole year – only for some significant time of the year. Although all samples were treated the same way, differences among the samples were consistent with soil characteristics at each site. This suggests that the influences of treatment and storage were minimal.

We gave related discussion as follows:

Lines 103-106: It is worth noting that Stagnic water conditions do not mean that the soils are under reduced conditions for the whole year – only for some significant time of the year. We sampled a Stagnosol, but only the topsoil (2-15 cm) which was not under perching water, i.e., it was aerobic at time of sampling.

Lines 255-259: We cannot rule out any effects from sample storage or from the use of Mill-Q

water, as discussed in the Methods section, However, although all samples were treated the same way, differences among the samples were consistent with soil characteristics at each site. This suggests that the influences of treatment and storage were minimal, but further investigation is warranted in future studies.

Q: 112. Please list the material of the 0.45 micron membranes.

A: The material of membrane was cellulose mixing ester and we added it in the manuscript (line 122).

Q: 119. There is no justification for the choice of analytes - Fe, Al, Si and Ca? The rationale for this may not be clear to some readers.

A: These elements containing minerals (e.g. clay minerals and Fe oxides) were main soil minerals which can be associated with P. We added information about this to the text (lines 145-146).

118. What were the limits of detection and precision for the analytes measured by ICP-MS?

A: The limits of detection (LOD) depend highly on the element, matrix, possible interferences and last but not least the daily performance. The precision, on the other hand, depends mostly on the concentration but also on the element and matrix. Analytes with a concentration close to the LOD have a rather poor precision, whereas higher concentrated analytes achieve a precision of typically 3-10% (relative standard Deviation) depending on the matrix and homogeneity of the samples. The LOD of measured elements in this manuscript is typically around 0.1 to 1 ug/L. We added information about this to the text (lines 157-158).

146. Was neutralization of the NaOH-Na_2EDTA extracts with HCl performed to avoid break down of polyphosphate species? Perhaps the rationale for not doing so could be included here?

A: We did not neutralize the NaOH-Na_2EDTA extracts with HCl prior to lyophilization, although this was recommended by Cade-Menun et al. (2006, EST 40:7874-7880). Neutralization of samples has not been widely adopted, and was not used by Liu et al. 2014 in their study of WDCs. However, it is something to consider for future studies.

169 – “for identify” I believe this should read “to test for significant differences” or “to identify significant differences”.

A: Yes, we changed it into “to test for significant differences among soil fractions” in the manuscript (line 201).

170 – Which tests were employed to determine distribution normality?

A: We used a Shapiro-Wilks test for normality. This is now indicated in the text (line 202).

178 – Analysis of Ca is not previously mentioned. Either calcium analysis should be included in the methods section, and the statement clarified i.e. what constitutes a low concentration? Or this statement could be removed here.

A: We removed this statement.

202 – The effect of pH, and differences between the sampled soils should probably be discussed here.

A: Please see our response to Reviewer 2 on this topic.

216 – This discussion needs to take into account the effect of the reactions that likely occurred during the oxic extraction procedure.

A: We added the following comment on lines 255-259: We cannot rule out any effects from sample storage or from the use of Mill-Q water, as discussed in the Methods section, However, although all samples were treated the same way, differences among the samples were consistent with soil characteristics at each site. This suggests that the influences of treatment and storage were minimal, but further investigation is warranted in future studies.

280 – clay-Fe oxides is an interpretation based on elemental analyses, it is not certain that the colloids identified contain clay minerals from the analyses conducted.

A: We cannot identify clay minerals according to FFF and element analyses. However, clay minerals with Si and Al elements and Fe oxides are common minerals for soils. We did TEM experiments for arable soils in a prior study (Jiang et al., 2015) and found clay minerals and Fe oxides in soil colloids.
Lines 145-146: These elements were analyzed as part of the main soil minerals (e.g. clay minerals and Fe oxides) that can be associated with P (Jiang et al., 2015a).

Table 1 – Dissolved or total organic carbon? If this refers to the bulk soil it is not dissolved organic carbon but total organic carbon? The table caption refers to uppercase letters but the letters indicating significant differences are lower case.

A: We changed the table caption as suggested.

Table 2 – To help distinguish between bulk solid analyses and analysis of water extracts I suggest using mg kg-1 for bulk soil analysis and mg L^{-1} for water analysis. Also –TOC for bulk solid analysis and DOC for aqueous and colloidal analyses.

A: We changed TOC into DOC as suggested. With respect to the unit of water analysis, we

still prefer mg kg^{-1} because mg L^{-1} cannot directly tell readers the concentrations of colloidal and dissolved elements compared to those in bulk soil.

Table 3 – The formatting and alignment issues make this quite hard to read. “below detection limit” is noted but the detection limit is not defined?

A: Below detection limit <0.05%. We added it in the Table 3.