Answers to reviewer comments BGD:

We are grateful for the valuable comments, questions and suggestions provided by both anonymous reviewers. They helped us to substantially improve the manuscript. Please find attached the final author comments.

Anonymous Referee #1

This manuscript describes a comparative experiment on soil organic carbon stocks. They compare automatic and soil sampling results in order to know if they are equivalent to define the small spatial and temporal variation. The concept and results are novel. This study is interesting and must have involved a great deal of work. While the results merit publication, with respect, I think the paper requires major revision. This study pointed the need to more information about the different type/technic of soil C storage in order to standardize the results and pointed the advantage and inconvenient of them. Nevertheless, added a table will be a good idea to underline these differences and the need to more advances, and the necessity of this study.

Please see answer to comment 1.

The methods and statistical analyses seem not totally appropriate.

Please see answers to comments regarding the method section below. No comment regarding our statistical analysis was raised, wherefore we do not exactly know to which statistical analyses this comment refers.

With respect, your discussion need more attention in order to forward your innovative results. Currently, your paper tend to look like a technical report but without enough "technical information", and I think that it is more than that.

Please see answers to comments regarding the Result and Discussion section below.

General comments:

1. Firstly, more information are needed in order to see the real advantage of your methodology. Presently, we are quite lost and the benefit of AC-based C budgets is not enough forward.
   We rewrote the Abstract to more directly address the advantages of using the AC-based approach (mainly the small-scale spatio-temporal resolution of gained ΔSOC values).

   "Tendencies and magnitude of ΔSOC values derived by AC-measurements and repeated soil inventories corresponded well. The period of maximum plant growth was identified as being most important for the development of spatial differences in annual ΔSOC. Hence, we were able to confirm that AC-based C budgets are able to reveal small-scale spatial and short-term temporal dynamics of ΔSOC. "

   In addition we now describe the benefits of the AC-based approach within the Introduction in more detail by directly comparing it to:

   1. soil sampling based approaches
   2. other measurement systems used for estimating the gaseous C exchange

   "Compared to mentioned approaches for detecting ΔSOC by either repeated soil sampling or observations of the gaseous C exchange, automatic chamber (AC) systems combine several advantages. On the one hand flux measurements of the same spatial entity avoid the mixing of spatial and temporal variability, as done in case of point measurements by repeated soil inventories. On the other hand, AC measurements combine advantages of EC and manual chamber systems because they not only increase the temporal resolution compared to manual chambers but also allow for the detection of small-scale spatial variability and treatment comparisons regarding the gaseous C exchange (Koskinen et al., 2014)."

2. Need more details on the soil information and their effects on the soil C storage process: different soil layers taking into account; what about the roots, which are the main C input in the soil.

   We agree that process studies are needed, however, they are not within the scope of the presented MS, which compares two methods to show the accuracy and precision of AC derived ΔSOC values for on an exemplary field site. Hence, we discuss soil related processes and soil C storage processes with respect to the plausibility of observed ΔSOC using both methods (see section 4.2). None of both methods are usually able to differentiate observed ΔSOC between soil layers. As stated, most repeated soil inventories are based on topsoil soil sampling, which disables a distinct investigation and interpretation of different soil layers (except if samples are taken for the different layers). Opposing to that, the presented AC-based approach, integrates ΔSOC over the entire soil column (thus including processes in all soil layers), which however in return hampers a soil layer specific investigation.

   In case we misunderstood this comment, and it refers to more details about soil sampling given in section 2.3; we want to refer to changes made in section 2.3.

   "After soil manipulation, a 5-m raster sampling of topsoils (Ap horizons) was performed during April 2011. Each Ap horizon was separated into an upper (0-15 cm) and lower segment (15-25 cm), which were analyzed separately for bulk density, SOC, Nt and coarse fraction (< 2 mm) (data not shown). From these data, SOC and Nt mass densities were calculated separately for each segment and finally summed up for the entire Ap-horizon (0-25 cm). The mean SOC and Nt content for the Ap horizon of each raster point was calculated by dividing SOC or Nt mass densities (0-25 cm) through the fine-earth mass (0-25 cm). In December 2014, composite soil samples of the Ap horizon were collected. Composite samples consist of samples from four sampling points in a close proximity around each chamber."

   Concerning the root issue, we refer to a statement in the MS *"Usually, coarse organic material is discarded prior to analysis (Schlichting et al., 1995) and therefore, total SOC is not assessed (e.g., roots, harvest residues, etc.)."*. This is in line with standardized routines of sample preparation in soil sciences.

3. Why do you clearly underestimate the deep soil in your C budget?

   It is not state in the MS, that we underestimate deep soil layers in our C budget. As mentioned above, AC-measurement derived C budgets account for the entire soil column underneath the chamber, since the observed net flux is a result of the soil and crops components underneath the chamber. This is not always the case when detecting ΔSOC by repeated soil inventories which often sample the upper soil horizon only. Since we show that both methods are able to detect almost the same ΔSOC over the study period, C allocation to deeper soil layers is most likely not relevant in this study (please see answer to comment 24).

   "In contrast to the soil resampling method, we postulate a higher accuracy and a lower precision in the case of the AC-based C budget method. The reasons for this include a number of potential errors affecting especially the measurement precision of the AC system, whereas over a constant area and maximum soil depth, integrated AC measurements increase measurement accuracy."

4. Did you have more information about the seasonal variation of the soil chemistry, soil density in link or not with the different plant species?

   As a comparison of two different methods to detect ΔSOC, the seasonal course of soil chemistry and density was not within the focus of this MS. Repeated soil inventories are usually based on soil samples taken at a frequency of one to five years (e.g. Van Wesemael et al., 2011), and studies investigating the seasonal course are scares (as stated in L 79-84 within the Introduction).

However, seasonal variations of different soil chemistry parameters were measured throughout the entire study period at two profiles of the depression (details are given in Rieckh et al. 2013).

5. In the abstract, line 43 page 2, you talk about soil properties but nothing after.
We changed the sentences.

"The measurement site is characterized by a variable groundwater level (GWL) and pronounced small-scale spatial heterogeneity regarding SOC and nitrogen (Nt) stocks.

6. The temporal variation were nicely represented with 4 years of measurement, but concerning the spatial variation, I think that there are some overestimation because of you are only one chamber by topographic step, so no replication by topographic step; and on the other hand, this topographic gradient seems to be too little, with only "difference in altitude 1m within in a distance of approx. 5 m of each other. Page 7 line 140. So for me there are not enough difference to "called" spatial variation.
We agree on the fact that the topographic gradient is rather small. As mentioned in the title and abstract, and specified in the Introduction, we aimed at showing that AC-measurements are in principle suitable to detect small-scale or in-field spatial differences (10-30 m), since their ΔSOC values fit well to those derived by repeated soil inventories. This is a prerequisite to detect spatial heterogeneity, which is not only common for the study area but also wider areas of the northern hemisphere. Whether or not these differences are too small depends however on the precision of the approach used to detect these differences (in the case of this study, ΔSOC). We added results of a Wilcoxon-rank sum test, to indicate the difference (in monthly NEE, NPP and, ΔSOC) between the chamber positions. This test showed significant differences between the chambers, and thus indicated the presence of spatial variation within a transect length of <30 m.

7. It's not clear your hypothesis about the potential difference between the four topographic steps. Could you add some information about that, and confirm it in the discussion?
We added a schematic representation of the topographic gradient to Fig. 2. Along the topographic gradient we hypothesized an increase in wetness downslope due to a groundwater level closer to the surface as well as a related trend of decreasing redox potentials. As these gradients are strongly related to the annual weather conditions, esp. rainfall dynamics, we avoid an a-priori hypothesis on their (net) effects to carbon balance or NEE.

8. Estimation about the ecosystem compartment effect? For Reco, which part of soil and aboveground compartment?
We do not fully understand this comment. R_{eco} refers to the (total) ecosystem respiration as the sum of autotrophic and heterotrophic respiration. Thus it includes root, shoot and soil respiration. We added a short description of R_{eco} to the MS.

"The atmospheric sign convention was used for the components of gaseous C exchange (ecosystem respiration (R_{eco}; sum of autotrophic and heterotrophic respiration), gross primary production (GPP) and NEE), whereas positive values for ΔSOC indicate a gain and negative values a loss in SOC."

Specific comments:

9. Maybe the abstract need to more concise.
We shortened and specified the Abstract to make it more concise (please see answer to comment 1).

10. P 4, L 67, what kind of land-use?

We stated in the L 67 of the MS, that "Erosion and land use change" (such as ploughing of grassland or peatland drainage for agricultural purposes) are reinforcing "natural spatial and temporal variability". Wherefore we are actually referring to the change in land use itself, irrespective of certain kinds of land use. However, some kinds of land use such as agriculture are known to reinforce erosion and thus also reinforce small scale spatial heterogeneity through tillage and bare soil periods.

11. P4 L 71: I am not sure to understand the third point. For me it is also time dependent.
    We agree that the magnitude of ΔSOC compared to total SOC stocks is dependent on the respective time horizons of the observations. However, the *"rather small magnitude of ΔSOC compared to total SOC stocks"* complicate the detection of ΔSOC in a short or medium term time horizon, which is usually a requirement during scientific studies, that aim to compare e.g. fertilization treatments or different crop rotations. We therefore state, that a method is advantageous, when it is able to detect ΔSOC in a short- to medium term (3-5 years).

12. P6, L 110: with the same land use?
    No, Leifeld et al. (2011) showed temporal dynamics of a degraded intensively and extensively used grassland on drained peat. We accordingly specified the sentence.

    "Even though temporal dynamics in ΔSOC were shown e.g. for grazed pastures and intensively used grasslands (Skinner and Dell 2015; Leifeld et al., 2011), no attempt was made to additionally detect small-scale differences in ΔSOC."

13. P6, L 110, could you add also the study of Skinner & Dell 2015.
    We already referred to Skinner and Dell 2015 in L 93. In addition we now added Skinner & Dell 2015 to L 110 of the MS.

14. P6, L 127: soil or air temperature?
    We clarified this sentence by adding "air".

15. P8, L 157: So 5 different crops during your study?
    As stated in the MS a crop rotation of maize – winter fodder rye – sorghum sudan grass hybrid – winter triticale – alfalfa was measured, resulting in a total of 5 different crops. To better address this we changed this sentence.

    "The measurement site was cultivated with five different crops during the study period, following a practice-orientated and erosion-expedited farming procedure. The crop rotation was silage maize (Zea mays) - winter fodder rye (Secale cereale) - sorghum-Sudan grass hybrid (Sorghum bicolor x sudanese) - winter triticale (Triticosecale) - alfalfa (Medicago sativa)."

16. P8, L 174: it's a closed system?
    This sentence refers to the "closed chamber measurement system" as described in detail by e.g. Livingston and Hutchinson, 1995) but not the measured ecosystem. As mentioned by Livingston and Hutchinson (1995) "The terms "dynamic" or "open" are sometimes used synonymously to describe steady-state systems and "static" or "closed" are often applied to non-steady-state systems". Since we define the measurements as flow-through non-steady-state measurements right before, we understand the confusion resulting from the use of "closed" in this sentence. We therefore deleted the word "closed".

17. P9, L 189: could you use the same unit for volume liter or m3
    We changed the unit of the flow rate of the pump (1 l min^{-1}) into 0.001 m^{-3} min^{-1}.

18. P10, L206: I surprised because the C sink is a negative value and a C source a positive value, basically.

As stated in the MS, the atmospheric sign convention was used for the components of gaseous C exchange (ecosystem respiration (R_{eco}), gross primary production (GPP) and NEE)" whereas the soil perspective was used for ΔSOC, thus indicating a gain with positive values and a loss in SOC with negative values (as used in soil sciences). We corrected signs given in Tab. 1 accordingly.

19. P10, L 212: Could you add the equation of your fluxes? and more information about your choice: time length for the measurements, number of measurement by day

We added the Eq. 1 to the MS as suggested. Length of measurements and number of measurements per chamber per day are already addressed in section 2.2.1 of the MS:

"The chambers closed in parallel at an hourly frequency, providing one flux measurement per chamber and hour. The measurement duration was 5-20 minutes, depending on season and time of day. Nighttime measurements usually lasted 10 min during the growing season and 20 min during the non-growing season (due to lower concentration increments). The length of the daytime measurements was up to 10 min, depending on low PAR fluctuations (< 20 %)."

Nighttime and non-growing season measurements were in general longer, due to a lower concentration change (more details are given in Hoffmann et al. 2015).

20. P11, L 241: unit: µmol-1 m-2 s-1 => µmol CO2 m-2 s-1, right?

We changed the unit.

"R_{eco} is the measured ecosystem respiration rate [µmol C m^{-2} s^{-1}]"

21. L 243 : temperature of "air" ?

Not exactly, according to Hoffmann et al. 2015, Arrhenius-type temperature-dependency models are derived by using different temperature sets measured during flux measurements. Similar to Hoffmann et al. (2015), we used soil temperatures in 2 cm, 5cm and 10 cm soil depth as well as the air temperature. Thus T_{ref} can be referring to soil or air temperature, depending on the chosen fit. To better address this important issue, we rewrote this paragraph.

"To account for measurement gaps and to obtain cumulative NEE values, empirical models were derived based on nighttime R_{eco} and daytime NEE measurements following Hoffmann et al. (2015). For R_{eco}, temperature-dependent Arrhenius-type models were used and fitted for recorded air as well as soil temperatures in different depths (Lloyd and Taylor 1994; Eq. 2).

$$R_{eco} = R_{ref} * e^{E_0\left(\frac{1}{T_{ref}-T_0}-\frac{1}{T-T_0}\right)} \qquad [Eq. 2]$$

where R_{eco} is the measured ecosystem respiration rate [µmol m^{-2} s^{-1}], R_{ref} is the respiration rate at the reference temperature (283.15 K; T_{ref}); E_0 is an activation energy like parameter; T_0 is the starting temperature constant (227.13 K) and T is the mean air or soil temperature during the flux measurement. Out of the four R_{eco} models (one model for air temperature, soil temperature in 2 cm, 5 cm and 10 cm depth) obtained for nighttime R_{eco} measurements of a certain period, the model with the lowest Akaike Information Criterion (AIC) was used"

22. P12, L 263: could you explain the range 2-21 consecutive days?

As mentioned in the MS *"Due to plant growth and season, parameters of derived R_{eco} and GPP models may vary with time."*, wherefore model parameters obtained based on a seasonal data set are not suitable (smoothed model) to gap-fill measured CO_2 dynamics. However, the actual length of a period, showing similar temperature and PAR dependencies and thus model parameter sets might vary with time as well. During the non-growing season even longer periods may not show a change, whereas during the short period of crop growth and senescence a change in temperature or PAR dependency of the flux components may occur within a couple of days. We therefore decided to use a variable moving window (2-21 days; user defined), which searches for the appropriate length of a data set taken into account to obtain parameters for R_{eco} and GPP used for subsequent gap-filling. The minimum length of the variable moving window was set to two, since short term measurement gaps as well as short nights during summer might reduce the used data set to <5 measurements. The maximum length of 21 days was primarily set to avoid extensive calculation time. In reality, most data subsets used for parametrization were longer than 2 days but shorter than 21 days, and in principle longer during the non-growing season and shorter during the growing season.

23. P14, L 298: Could you give us a mean of your CH4 measurements? and what about the N2O ? If you want to discuss about the C budget, you need to add information about the two others greenhouse gases.
    We added the average annual CH_4-emission (-0.01 g m^{-2} y^{-1}; small uptake during drier years and small CH_4 release during the rather wet summer 2011) to the MS. We disagree on the statement that N2O measurements (0.34 g m^{-2} y^{-1}) are needed for estimates of the C budget. Since we do not discuss the GHG budgets, we decided to not include N_2O emission measurements within the MS.

24. P15, L 324: only topsoil? what about the subsoil ? We know that the subsoil have a high contribution to the soil C stock.
    Indeed, subsoils at depositional sites are important segments for total SOC stocks, e.g., down to 1m depth. The colluvial subsoils at the CarboZALF-D study site store 60% of total SOC stock whereas the plough layer (Ap) contains 40%. This is mainly due to the larger thickness of the colluvial horizons. However, as the manipulation affected only the plough layer, we expected/assumed detectable and significant SOC changes during our 4 years of observation only for the plough layer. Here, strong transient states were induced by the manipulation, because the soil material from eroded upslope soils is under saturated in respect to its C-sequestration potential. Furthermore topsoils, like plough layers, represent the soil compartment of highest SOC turnover (C-input; O_2 supply). Of course, we will include subsoils (down to 1.5 m) in a resampling campaign 10 years after manipulation (2020). This is the expected time scale on which we might detect (probably small) SOC changes in subsoils by using the soil resampling approach. In addition, the good fit of changes in SOC between the two methodological approaches support this assumption since it shows that changes in subsoil SOC are minor during the four years of our study.

25. P15, L 335: unit: I think that it will be better: gC m-2 y-1 , right ?
    We changed the Latin abbreviation for year (“a” for “annus”) for “y” throughout the entire MS.

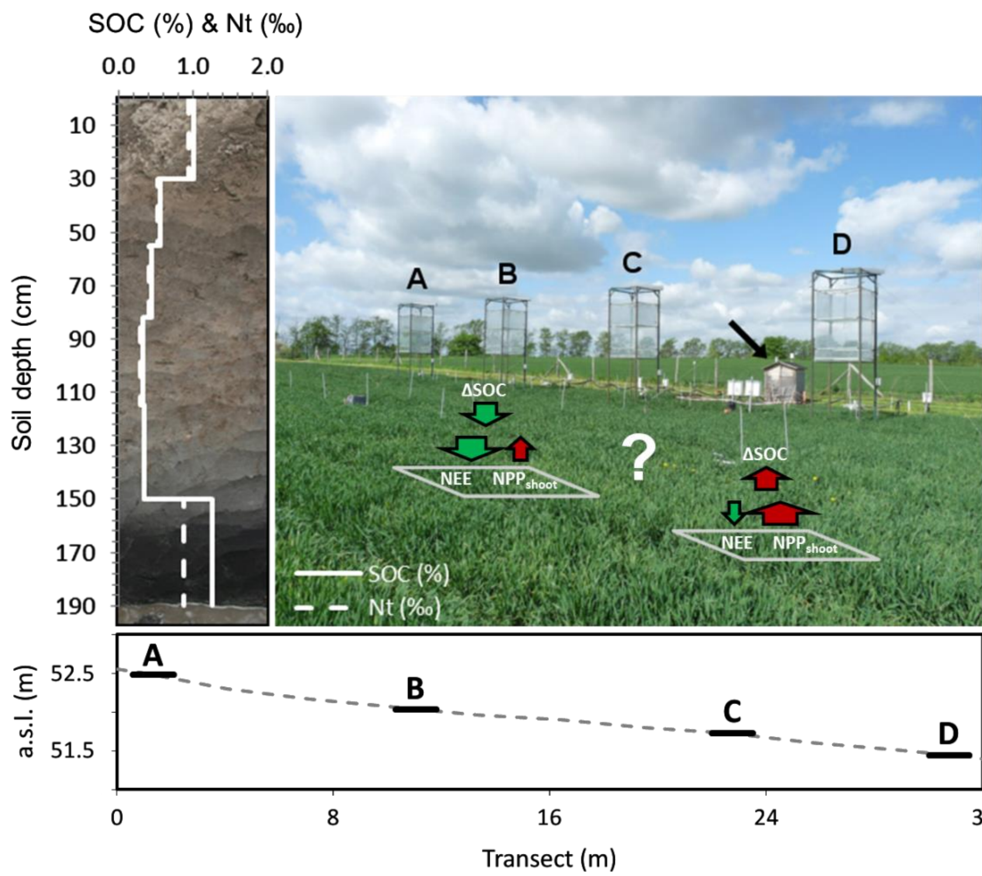
26. P19, L 431: add reference about the gap filling? your own method or adapted to already published methodology ?
    We added Hoffmann et al. (2015) as reference to this sentence. The method used for gap filling, is also explained within the method section of the MS (please see 2.2.2).

27. P21, L 465: what about the daily pattern of NEE, R_{eco} and GPP, and so soil C storage?
    We added a short description of monitored daily patterns. C storage (ΔSOC), was only calculated on a daily frequency (due to daily frequency for NPP_{shoot} estimates). Hence no daily patterns for ΔSOC can be given using the presented AC-based approach.

28. P21, L 477: Could you add more references? they are lots of studies on the soil C sequestration in pastures in different biomes, inverse to crop land.
Since this sentence compares obtained results for alfalfa on an agricultural used landscape with balances given in literature we are referring to references for perennial crops/grasses instead of permanent grassland like pastures, which are most-likely not comparable due to different C dynamics.

29. Figures 1: The temporal measurement was nicely represented. But, did you add a schematic representation of the spatial aspect?
The spatial aspect (that different Pedon's might show a different development in ΔSOC) is referred to in Fig. 2. We decided to not add it to Fig. 1 (in terms of multiple lines with a different development), in order to keep the concept of temporal changes in SOC easily understandable for the reader.

30. Figures 2: I prefer to see SOC in stock (g C m-2) rather than %. Could you add a scale in your picture? and also a cross-section of your site in order to see the different altitude and distance among chamber.
We added the scale (length of transect) and cross section of our site to Fig 2. We decided to give SOC in "%" rather than "g C m^{-2}", because we wanted to give N_t in the same figure ("‰"). If needed we will change "%" to "g C m-2" for SOC. However, in this case we will have to delete N_t.



31. Figure 5: Could you add in the caption the signification of the 4 symbols, and add in the graph a dotted horizontal line for zero.
    We added the significance of (spatial) differences between the chamber positions to tab. 1. Zero is already marked by a solid horizontal gray line within the figure. Used symbols are explained by the figure legend. In addition we added a short explanation to the figure caption.

32. Figures 6: problem with unit: gC m-2 y-1 , right ?
    We changed the Latin abbreviation for year ("a" for "annus") for "y" throughout the entire MS.
    Could you add a test (a t-test ?) in order to know if they are difference between the two estimations of C budget in each chamber?
    Even though there is no minimum sample size for a t test to be valid, we decided to test for significant (spatial) differences between chambers based on monthly instead of annual NEE, NPP and ΔSOC (pleases see Tab. 1). Results of the test between the results obtained by the two methods for the entire period used to detect ΔSOC (2011-2014) are added to the figure caption. Since the *t* test requires normal distribution, we used the Wilcoxon rank-sum test instead.
    Currently, we need more information about the added value of your C budget.
    We are not sure whether this refers to the initial manipulation in 2010 or not. If it refers to the added C due to soil manipulation it does not matter for this figure, since this figure only refer to the period after manipulation (as shown in Fig. 5 and mentioned in section 3.3). Since this seems to be misleading, we added the period taken into account to the figure caption. Moreover, we now mention it more precisely in section 2.1.

    "The change in SOC for each chamber was monitored by three topsoil inventories, carried out (I) prior to soil manipulation during April 2009, (II) after soil manipulation during April 2011, and (III) during December 2014. ΔSOC derived through soil resampling and AC-based C budgets, was compared for the period between April 2011 and December 2014 (Fig. 1)."

    Same as for figure 3, 4, 5. It will be nice to know if the chambers are significantly different.
    Please see answer to comment 32 above.