In the following, we have numbered our answers as AARC1-n, where n is the answer number (Author Answers to RC1). This allows referring to our answers in this document as well as in the answers to the other reviewers and vice versa.

***RC1**: 'Comment on egusphere-2025-592', Moritz Laub, 23 Apr 2025 reply*
In this study, Loubet et al., describe two measurement campaigns in the FR-Gri site, which studies SOC stocks in 2005 and 2019. The also report differences in SOC between both time points, and simulate the evolution of SOC in 0-30 cm, using the AMG model.

Overall, the study is an interesting description of how SOC stock changes could be assessed by repeated samplings and the results are worth publishing. The overall context is well described, the authors describe clearly what they did (except the AMG model) and I think the results hold value. However, in the present form there are several major issues that the authors neglect. The first is the differences in sampling design between 2005 and 2019. This should be addressed explicitly. Using only a subset of 2019 samples might be the way to deal with it, but it must be ensured that the subset is representative of the 2005 locations, which is not demonstrated in the current article. The uncertainty arising should be addressed in the discussion. Additionally, the authors should try their best to harmonize the 2005 and 2019 sampling campaigns. This means that the horizons should be adjusted to represent the same depths/soil masses, for example by weighted means and by using an ESM quadratic spline for the stocks (Wendt and Hauser, 2013). Finally, the discussion should try integrating the results with other studies on a deeper level than what is currently done. Just stating they found X, we found Y, without much discussion why this could be, is a bit shallow.

I think if these issues are addressed properly, the study will make an interesting article.

***AARC1-1.** we thank the reviewer for his constructive review and for raising important issues that we will try to clarify in this answer and incorporate the key elements in the manuscript:*

1) *Regarding the differences in sampling design between 2005 and 2019, we provide additional analysis of the spatial variations of soil characteristics in the field that shows that the 2005 campaign and the 2019 "reduced area" sampling points correspond to the same soil type (see detailed answer **AARC1-4** below and Figure R2). We additionally checked that there was no significant differences between the soil characteristics of the samples (e.g., carbon content, density, rock fraction) from the 2019 reduced area and the 2019 sampling points that were in the 2005 sampling area (see detailed answer **AARC1-4** below and Figure R2). We also compared the only comparable parameter between the two-sampling campaign (the rock fraction), which showed no significant differences (Figure R3). This additional information provides additional confidence that the choice of the reduced sampling zone in 2019 corresponded to the soil characteristics of 2005.*
2) *Regarding the harmonization of the two sampling campaigns, we fully agree that the ESM method is the reference method. This was maybe not sufficiently clear in the original manuscript, and we will clarify it in the revised manuscript. In the original manuscript we focused on the overall carbon stock estimated by the ESM technique over the entire profile 0-60 cm. In the revised version, we will provide additional information in the soil profile to the three coarsest layers of 2019: 0–5 cm, 5–30 cm, and 30–60 cm. The bulk density (BD), rock fragments, and carbon content were aggregated using a thickness-weighted mean, while SOC stocks and soil mass were calculated as cumulative sums across the respective layers. We will further include equivalent soil mass (ESM)-based SOC stocks calculated using two approaches from the SimpleESM function (Ferchaud et al.,*

*2023). These methods converged to a C soil stock losses of 0.95 kg C m^{-2} in the 0-60 cm layer over time (see detailed answer below and **Figures R6, R7 and R8**).*

3) *Regarding the discussion section, we propose to significantly revise this section to enhance clarity and coherence and we propose to organize it into three chapters:*
    a. *"**4.1 Effects of sampling depth and computation methods on organic carbon stock changes evaluation**": we will present the results obtained using both FD and ESM approaches, explicitly discussing the limitations of FD when comparing multiple layers, but recognising the agreement between the two approaches for SOC stocks at 60 cm depth. We also will provide mechanistic interpretation to explain the observed stock change patterns. See **answer AARC1-11**.*
    b. *4.2. SOC stock changes are significantly detected over the 13.25 years: organic carbon stock equilibrium". This restructuring will address the reviewer's concerns about the section's length and allow to remove redundancies. The revised chapter will compare ESM-based SOC stocks evolution with other studies using the same approach. We further explore the mechanisms driving the observed changes, with particular focus on the C imports and exports associated with the management implemented at the FR-Gri site. We keep the comparison with AMG model and the ICOS C balance approach but add some AMG scenario to help interpreting the reasons of the observed SOC stock change. See **answer AARC1-10**.*
    c. *"4.3 Uncertainties in soil carbon stock changes in FR-Gri": We will keep this section but will add a dedicated section addressing the uncertainties associated with our results. We will explicitly discuss potential biases arising from differences in sampling design and include a discussion based on the new soil map produced and the significance of ours results as measured by the Hedges g effect and the minimum detectable difference (MDD) published by Schrumpf et al. (2011). We kept a small discussion on uncertainties on total C stock due to potential losses of inorganic C. See answer **AARC1-11**.*

Major concerns:

- *The image quality of all figures is very bad. A higher resolution is needed to make them acceptable and readable!*

***AARC1-2.** We are sorry for this inconvenience. We did provide higher resolution images as author comments in the EGUsphere discussion page. Anyway, we updated the figures with TIFF files with 800 DPI and provide higher resolution images in this document.*

- *While the study acknowledges that the equivalent soil mass (ESM) approach is the much better one, compared to fixed depth, this does not show in the results. The fixed depth results are always reported first and discussed in more detail. If anything, the ESM results should be reported first, and the fixed depth approach only briefly mentioned, knowing that BD changed and it is therefore biased.*

***AARC1-3.** We agree that the original manuscript was confusing regarding the reference method. This comment was also made by Fabien Ferchaud (CC1: 'Comment on egusphere-2025-592', Fabien FERCHAUD, 28 Apr 2025). In the revised manuscript we still present both FD and ESM, but the discussion about SOC* *changes is only based in ESM SOC stocks. These methods converge to a C soil stock losses of 0.95 [-0.87 1.03 CI] kg C m^{-2} in the 0-60 cm layer over time.*

To clarify the role of ESM, we propose to modify the Material and method section by isolating a dedicated section describing the ESM approach:

*"**Soil carbon stocks calculation using equivalent soil mass (ESM) approach and SOC stocks changes.***
*To estimate SOC stocks evolution change in reduced and no-tillage system, one needs to consider changes in SOC content of the soil (SOC_{FSS}) but also the potential changes in BD_i due to compaction or decompaction, which may change the mass of fine soil stock FSS_i in each sampling depth (Powlson et al, 2014). Additionally, soil erosion driven by rainfall or wind can export soil particles – mainly silt and clay - out of the field. Erosion is thought to be negligible at the FR-Gri site due to a slight slope and systematic winter inter-cropping. Decompaction may have happened since the site was converted to reduced tillage from 2000 onwards (Loubet et al., 2011), but compaction in subsoil may also occurs due to repeated trafficking of the soil surface with large vehicle loads (Liebhard et al., 2025). To consider possible changes in BD_i, the SOC stock evolution was estimated using the equivalent soil mass method (ESM), where the SOC stock is integrated down to a varying depth corresponding to a reference soil mass that is set equal for each campaign (Wendt and Hauser, 2013; Ellert and Bettany, 1995; Lee et al., 2009). This approach has the advantage of accounting for a common sampling bias with hydraulic corer, which is soil compaction. The ESM-based SOC stock was computed using the R function "SimpleESM" (Ferchaud et al., 2023), which provides two calculation method: the classical ESM method originally proposed by Ellert & Bettany (1995), and ESM2, a model-based approach incorporating cubic splines (Wendt & Hauser, 2013). Firstly, we computed the reference soil mass (that is, reference FSS_{ref}) across three aggregated soil layers (0–5 cm, 5–30 cm, and 30–60 cm) using the median FSS per layer from the 2005 data (**Table 2**). The total fine soil stock in the 0–60 cm layer (Eq. 4) ranged from 852 to 967 kg m^{-2} in 2005, and from 831 to 953 kg m^{-2} in 2019 (Table S4).*

***Table 2.** Reference soil mass (kg m^{-2}) from the fine soil stock fraction for each soil depth.*

Layer	*Upper depth*	*Lower depth*	FSS_{ref}
	cm		*kg m^{-2}*
L1	*0*	*5*	*63.19*
L2	*5*	*30*	*372.62*
L3	*30*	*60*	*453.08*

In the "classical" ESM approach (Ellert & Bettany, 1995; Lee et al., 2009), SOC stock is calculated by 1 mm increments. According to Autret et al. (2016) and Mary et al. (2020), soil depth is discretized into elementary layers of 1 mm thickness, with FSS density (g cm^{-3}) and carbon content (g kg^{-1}) assigned to each 1 mm layer. Since both FSS density and the SOC content are typically reported as average values over macro-layers (e.g., 0–5 cm), these values are assumed to be constant within each 1 mm sublayer. Subsequently, FSS_i and SOC_{stock_i} are then computed cumulatively until the FSS_{ref} is reached. This approach is referred to as "ESM non model" by Peng et al. (2024). The ESM2 approach is based on the "material coordinate system" (McBratney and Minasny, 2010) or the "cumulative coordinates approach" (Rovira et al., 2015). This method uses a post-hoc model - a cubic spline interpolation - to mathematically adjust SOC measurements to a common mineral soil mass (von Haden et al., 2020; Wendt and Hauser, 2013). Unlike the direct 1 mm increment method used in ESM, which computes stocks layer-by-layer, ESM2 *produces a continuous function that interpolates smoothly between measured data points. The spline type was set to Hyman, ensuring monotonicity (i.e., a non-decreasing function). SOC stocks based on ESM2 methods are then interpolated at the reference soil mass."*

- *The areas are so different that it is hard to state whether the difference between SOC in 2005 and 2019 is due to sampling design or real change in SOC!*

AARC1-4. *We agree that this question is key regarding the conclusion of this study. We should also stress that it is an important issue for many observation sites an especially ICOS sites. Indeed, many current ICOS sites had historical campaigns with a different sampling design, because ICOS is imposing a standardised protocol to all sites. This manuscript is therefore the occasion to evaluate whether we can use these historical data to compare with the ICOS sampling design.*

In the original manuscript we sub-sampled the 2019 samples into a "reduced area" based on expert knowledge of the site by the farmer, soil properties, and yield maps all indicating that the north-western band of the field was shallower. We further selected the 2019 strata so that the only measured quantity in both sampling campaign that would indicate a soil difference: the rock fraction. We chose strata in 2019 where the rock fraction was lower or equal than that measured in all samples in 2005. These criteria altogether led to the chosen 2019 "reduced area".

*In the revised manuscript we further propose to use a map of the soil characteristics to determine if 2005 and 2019 sampling zones were sampling the same soil**.** The map was constructed based on the following 2019 soil core properties from 0 to 60 cm depth: maximal A+B horizon depth, clay content by hand-textured, rock fragment, fine earth mass, inorganic carbon content, organic carbon content. These variables were selected not only for their impact on SOC stocks but also because they are commonly used in standard field-based soil descriptions (e.g., WRB, 2022). Moreover, some of those attributes are often associated with the degree of soil development, making them suitable indicators for identifying comparable pedological characteristics across years. To construct the map, we first interpolated each of the soil variables using the Inverse Distance Weighting (IDW) method with a power parameter of 2 [Shepard, 1968] (Figure R1).*

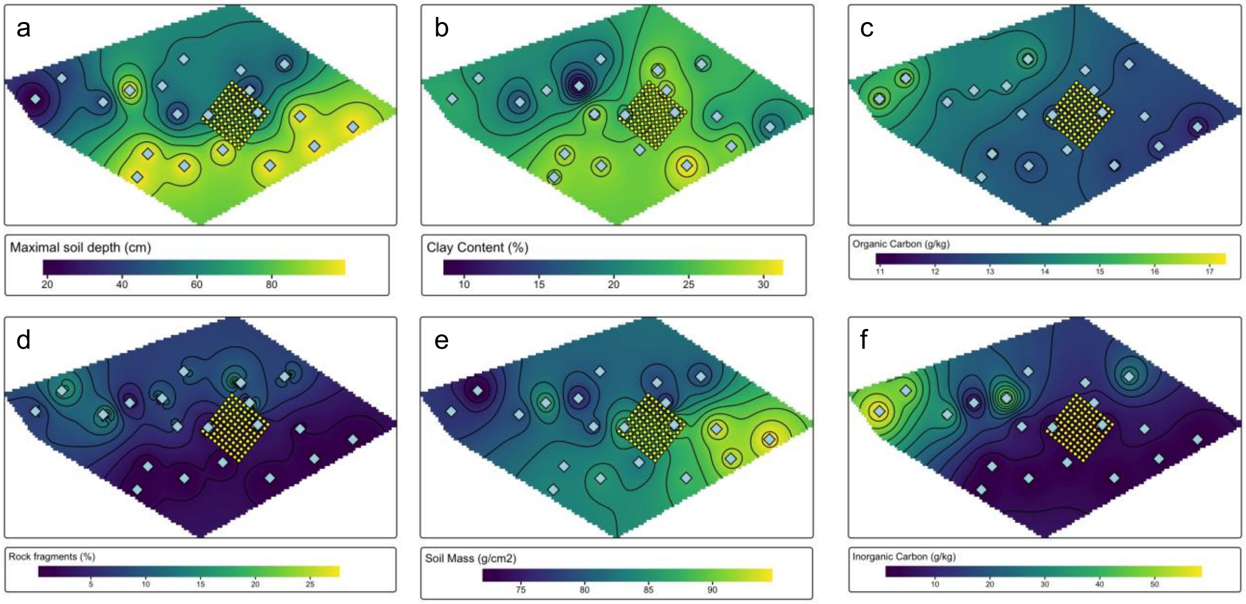

Figure R1. *Interpolated 2019 soil properties used to construct a soil characteristic map. The yellow dots represent the 2005 sampling points and the blue losanges are the 2019 sampling points.*

Subsequently, a clustering analysis was performed on the interpolated surfaces to identify zones with similar pedological characteristics [WRB, 2022]. This analysis resulted in the identification of four distinct zones (clusters) (Figure R2). Zone 2 is characterized by deep soils (>100 cm) with low contents of rock fragments and inorganic carbon. Zones 1 and 4 are composed of shallow soils (20–50 cm) with high concentrations of rock fragments and carbonates, and with less than 20% clay content. Zone 3 presents intermediate soil properties: soils are moderately shallow (40–60 cm) with medium levels of rock fragments and carbonates. The 2005 sampling points are located within Zones 2 and 3, therefore confirming that the chosen 2019 "reduced area" had similar soil pedogenetic and morphological characteristics as in 2005 which should ensure a reliable comparison between the 2005 and 2019 carbon stocks.

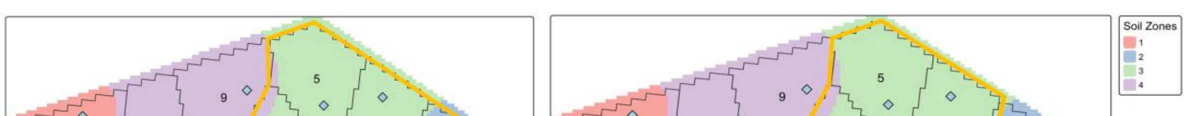

Figure R2. *Soil zones identified by clustering analysis based on the maps of Figure R1. The yellow dots represent the 2005 sampling points, and the blue diamond are the 2019 sampling points. The yellow lines represent the reduced fields selected in the revised manuscript based on the soil characteristics (left) and based on 2019 strata that were sampled in 2005 (right).*

To complement the design-based results and better account for the variation due to soil heterogeneity, we propose an additional model-based effect-size analysis with bootstrapping to compute the confidence interval. We apply this method to evaluate if a stricter selection in 2019 samples, defined as the strata that *contains at least one point of the 2005 sampling grid (Figure R2 right), would lead to differences in mean soil properties. We observe no significant differences between the two subsets of data in 2019 based on the model-based approach (Figure R3 left). We therefore conclude that the choice of the 2019 "reduced area" used in the original manuscript was the best choice to compare to 2005 as it includes a higher number of sampling points (N = 14), which helps reduce statistical variance, while having mean soil properties not significantly different from the stricter choice Figure R3.*

We propose to include two new sections, the first in the material and method section describing the maps and metrics used to complement the uncertainty analysis due to the different sampling designs in 2005 and 2019, and the second in the result section. We also propose to add Figures R2 and R3 in the supplementary material

<u>In the Material and Method section:</u>

*"**Spatial comparison between 2005 and 2019.** To enable a reliable comparison between the 2005 and 2019 and infer changes in SOC stocks, we minimized the effect of spatial heterogeneity in the 2019 sampling by subsetting the strata. To address this, we first employed a K-means-based clustering algorithms (MacQueen, 1967) to identify pedologically homogeneous zones. We used the 2019 dataset, which offered well-distributed coverage across the study area, with each sampling point characterized by its morphological (colour, redox features), physical (clay and rock fragment contents, fine earth mass) and chemical properties (Carbonates and SOC content) according to the World Reference Base for Soil Resources (WRB, 2022). These variables were aggregated to a depth of 60 cm using weighted mean (e.g., for carbonates) or summation (e.g., for soil mass) and then interpolated using Inverse Distance Weighting (IDW) of power 2 (Shepard, 1968). IDW was selected as the most suitable method for interpolation given the limited number of samples (n < 30). The resulting raster maps were stacked in a single raster and resampled to a common resolution. Subsequently, K-means-based clustering algorithms was applied (with K = 4) using the "stats::kmeans" (R core team, 2005) to delineate zones with comparable pedological characteristics. The number of clusters (K) was defined using the elbow within-cluster sum of squares (Thorndike. 1953). This analysis resulted in the identification of four distinct cluster, hereafter called soil zones (Figure R3). Zone 2 is characterized by deep soils (>100 cm) with low contents of rock fragments and inorganic carbon. Zones 1 and 4 are composed of shallow soils (20–50 cm) with high concentrations of rock fragments and carbonates, and with less than 20% clay content. Zone 3 presents intermediate soil properties: soils are moderately shallow (40–60 cm) with medium levels of rock fragments and carbonates. The 2005 sampling points are located within Zones 2 and 3. Based on the soil property map, the 2019 data was subset in three different ways:*

1. *Complete field = Including all the strata and corresponding SP-I plots.*
2. *Intersection field = Including the four strata of the 2019 campaign (Strata 8, 4, 2, and 5) that intersected the 2005 sampling area.*
3. *Reduced field = Including the strata within the same soil zones within the 2005 sampling area, that is, the strata 1, 2, 4, 5, 6, 7, and 8. These zones ensure more similar pedological properties in both campaigns. The SP-I_19 plot intersected Zone 2, while SP-20 did not fall within any of the target zones. Since both samples are part of Stratum 10, they were excluded from the analysis.*

All these analyses were performed using a set of R packages. For vector and raster manipulation, we used sf (Pebesma and Bivand, 2023; Pebesma, 2018), raster (Hijmans, 2025), and stars (Pebesma and Bivand, 2023). For inverse distance weighting (IDW) interpolation,we used gstat (Pebesma, 2004; Gräeler et al., *2016). Plotting was carried out using tmap (Tennekes, 2018), and multivariate clustering was performed using stats (R Core Team, 2025).”*

“[…] We performed an additional statistical analysis to quantity the magnitude of SOC stocks changes between 2005 and 2019 by calculating effect sizes using Hedges’ g (eq. 9-10). This metric is a standardized mean difference method that includes a correction for small sample sizes. For instance, N_{2019} reduced to 14 and 8 when comparing 2005 with 2019 in Reduced field and Intersection field, respectively. Here the two sampling designs are treated as a simple random sampling, which the replicates are obtained through the bootstrapping.

$$s = \sqrt{\frac{\sum(X_1-\overline{X_1})^2+\sum(X_2-\overline{X_2})^2}{n_1+n_2-2}} \quad (9)$$

$$d = \frac{\overline{X_1}-\overline{X_2}}{s} \quad (10)$$

$$g = d\,\frac{\Gamma\left(\frac{df}{2}\right)}{\sqrt{\frac{df}{2}}\Gamma\left(\frac{df-1}{2}\right)} \quad (11)$$

Where s is the pooled standard deviation, X represent the individual observations in groups 1 (2005) and 2 (2019), respectively; $\bar{X}$ is the sample means of each group; n is the sample size in each group, d is the uncorrected effect size (Cohen’s d), df represents the pooled degrees of freedom, and Γ denotes the gamma function, which generalizes the factorial function and is used in the correction factor to reduce bias in small samplesConfidence intervals for effect-size estimates (eq. 12) are computed using 20,000 nonparametric bootstraps with resampling and bias-corrected and accelerated (BCa) method (Efron (1987; Canty and Ripley, 2012). BCa method correct the CIs for bias, that is, when the bootstrap distributions are different from the observed statistic, and for skewness in the bootstrap distribution. In this analysis, the 2005 campaign was treated as the control, therefore, negative values of Hedges’ g indicate a reduction in SOC stocks in 2019, while positive values indicate an increase. If the confidence intervals (CIs) include zero, it suggests that there is no significant difference in SOC stocks between the two sampling years. These analyses were performed using the R package “bootES” (Kirby and Gerlanc, 2013).

$$CI_{BCa} = \left[\widehat{\theta^{(\alpha_1)}},\ \widehat{\theta^{(\alpha_2)}}\right] = \left[\widehat{\theta_{lo}},\ \widehat{\widehat{\theta_{up}}}\right] \quad (12)$$

Where $\theta^{(\alpha)}$ are the bootstrap percentiles corresponding to the adjusted lower (lo) and upper (up) CIs. The adjusted quantiles are computed as:

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0+z_{\alpha/2}}{1-\hat{a}\left(z_0+z_{\alpha/2}\right)}\right) \quad (13)$$

$$\alpha_2 = \Phi\left(z_0 + \frac{z_0+z_{1-\alpha/2}}{1-a\left(z_0+z_{1-\alpha/2}\right)}\right) \quad (14)$$

*Where Φ is the standard normal CDF and z_0 is the **bias correction factor, computed as the eq (15),** $\hat{a}$ is the **acceleration parameter**, measuring skewness of the bootstrap distribution, $\hat{\theta}$is the observed statistic (e.g., mean, median, effect size), θ_b^* are the bootstrap replications, and B is is the number of bootstrap samples.*

$$z_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\theta_b^*}<\hat{\theta}\}}{B}\right) \quad (15)”$$

In the result section:

*“**Comparative analysis of soil properties in complete and reduced fields.** Effect-size analysis comparing soil variables in 2019 across the original sampling layers (Figure R3) revealed no significant differences between the Intersection field and Reduced field, as indicated by Hedges’ g values close to zero and confidence intervals spanning zero. In contrast, the Complete field and Reduced field exhibited significant differences in several soil variables – such as SOC and SIC content, SOC and SIC stocks, rock fraction percentage, total nitrogen content and stocks, and fine soil stocks – with different differences pronounced up to 30 cm. Notably, SIC content and stocks differ consistently across all layers. Given these discrepancies in soil properties, the following sections focus on comparisons between the 2005 and 2019 Complete and Reduced fields (Figure R3).”*

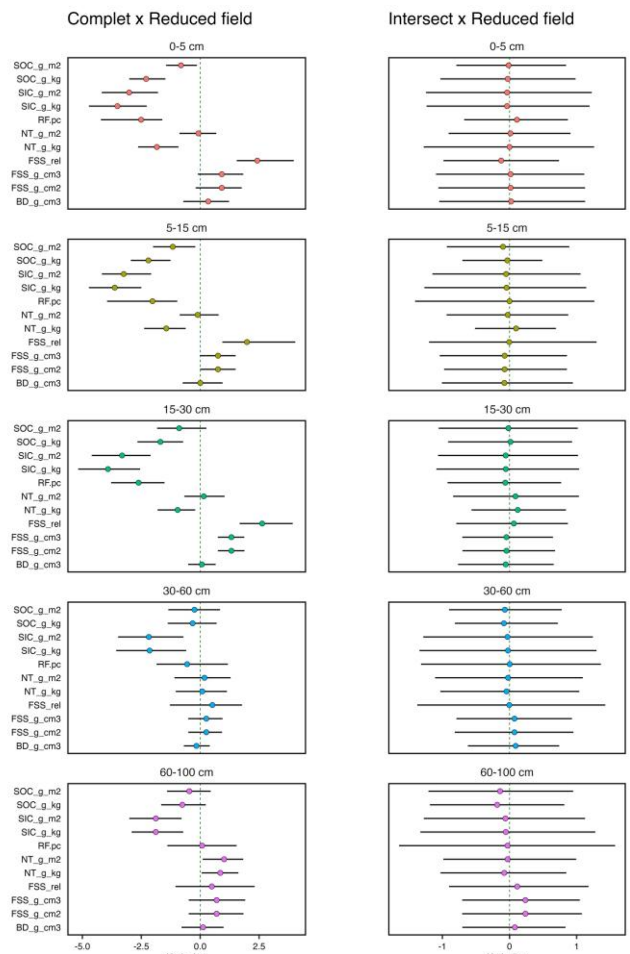

***Figure R3.** Magnitude of differences in soil variables between different subsets of the 2019 dataset, measured using Hedges’ g effect size method (Hedges, 1981). Circles represent Hedges’ g values, and error bars denote bias-corrected and accelerated (BCa) 95% confidence intervals based on 20,000 bootstrap resamples. The complete field includes 10 strata, the reduced field comprises 7 strata, and the intersection field comprises 5 strata. Variable codes are fol-lowed by their respective units. BD: bulk density; FSS: fine soil stock; FSS_rel: proportion of fine soil stock; RF.pc: rock fragment content (%); SOC: soil organic carbon; SIC: soil inorganic* *carbon; NT: total nitrogen. Confidence intervals overlapping zero indicate no statistically meaningful difference between datasets for the corresponding variable.*

- *The same is for the different depths and soil layers sampled. In a way this article compares apples to pears without trying to make adjustments (e.g., by harmonizing the data to have the same depth intervals).*

AARC1-5. *We actually did compare the soil carbon stock over aligned layers in both the FD and ESM approach. But we did not compute indeed the soil properties in equivalent layers. We hence propose to add soil layer aggregation to 0-5, 5-30 and 30-60 cm as explained in AARC1-1.*

- *The uncertainties that arise from this should at least be highlighted in the results (proper wording, lower and higher instead of decrease and increase) and the discussion (state the uncertainties explicitly).*

AARC1-6. *We acknowledge that the difference in sampling designs in 2005 and 2019 data may introduce uncertainties in our results. We however included detailed information about the statistical approaches in both 2005 and 2019 datasets and reported confidence intervals in all mention to soil stock changes. We propose adding the minimum detectable differences (MDD) as another metric to complement the p-values, based on t-test with 95% confidence and 90% power (α = 0.05, β = 0.10), as Schrumpf et al. (2011). We propose to include MDD in the discussion on the uncertainty in our comparative analysis: the sampling design differences (grid-based vs. stratified sampling), the effect of spatial coverage of sampling points (reduced field vs intersection between 2005 and 2019), and an additional statistical metrics (Hedges effect). Also, we provide a dedicated paragraph discussing the uncertainties related to the missing values of SIC stocks in 2005.*

Table 3. *Summary of soil organic carbon (SOC) stock changes between 2005 and 2019 in the reduced field at the FR-Gri site, assessed for the 0–30 cm and 0–60 cm soil layers using both the Equivalent Soil Mass (ESM) and Fixed-Depth (FD) approaches. SOC changes are reported in absolute terms (kg C m^{-2}), relative change (% of initial stock), and as annualized rates. The Minimum Detectable Difference (MDD) represents the smallest true difference that can be statistically detected given the observed variability and sample size. If the observed ΔSOC exceeds the MDD and $p < 0.05$, the change is considered detectable. If ΔSOC is less than the MDD, the change is not statistically distinguishable. A large MDD reflects high variability or limited sensitivity, whereas a small MDD indicates high precision in detecting SOC changes. These estimates were also used as input parameters for the AMG model simulations.*

Metric	Equivalent Soil Mass		Fixed depth	
	~ 0-30 cm † 435.1 kg m^{-2}	~ 0-60 cm 887.6 kg m^{-2}	0-30 cm ‡†	Fixed depth 0-60 cm
2005 SOC stocks (kg C m^{-2})	8.14	11.19	8.25	11.12
2019 SOC stocks (kg C m^{-2})	7.47	10.24	7.28	10.17
ΔSOC (kg C m^{-2})	-0.67	-0.95	-0.97	-0.96
Standard Error difference (kg C m^{-2})	0.11	0.04	0.12	0.04
Lower CI difference (kg C m^{-2})	-0.71	-1.03	-1.02	-1.03
Upper CI difference (kg C m^{-2})	-0.62	-0.87	-0.92	-0.80
P values (two-sided)	< 0.001	< 0.001	< 0.001	< 0.001
Minimum Detectable Difference (kg C m^{-2})	0.39	0.71	0.43	0.76
SOC stock change (% of initial stock)	-8.2%	-8.5%	-11.8%	-8.6%
SOC stock change (% initial Stock yr^{-1})	-0.62% yr^{-1}	-0.65% yr^{-1}	-0.89% yr^{-1}	-0.64% yr^{-1}
SOC stock change (per mil initial Stock yr^{-1})	-6.2‰ yr^{-1}	-6.5‰ yr^{-1}	-8.9‰ yr^{-1}	-6.4‰ yr^{-1}

- *It is not clear how the "reduced" sampling area was chosen. In Figure 4, it says it was chosen to have the same fine earth fraction as the 2005 sampling. More details are needed in the methods, as this is a potentially influential choice for the rest of the study. And why was it not chosen based on spatial proximity?*

AARC1-7. *As detailed in our previous response the sub-setting choice was made based on the fine earth fraction, and expert knowledge from the farmer and the lab on this field. A more detailed analysis based on an interpolation and a spatial clustering of the soil properties showed that the initial choice was robust in identifying similar soil properties between 2005 and 2019 (see answer AARC1-4).*

- *Any results of SOC stocks with the fixed depth approach should be reported with a high level of caution! By stating that the BD was significantly different between the samplings, you practically invalidate them and I would almost suggest that you completely eliminate them.*

AARC1-8. *We agree that FD SOC stock should be taken with caution, and in particular not used between 0 and 30 cm. But the results from this study shows that the FD SOC stock changes between 0 and 60 cm were not significantly different from the ESM approach, which is an interesting result that give more ground to existing FD data down to 60 cm depth. We have however added both ESM estimations at 5 and 30 cm and added soil layer aggregated soil variables in these layers following the reviewer advice (see answer AARC1-3).*

- *If you want to report SOC stocks for different layers, then use the approach of Wendt and Hauser, fitting a quadratic function of cumulative SOC stocks to cumulative soil mass (a+ bx+cx^2) with 0 intercept per each auger point, make sure that the R^2 is close to 0.99 for each, and then predict the SOC stocks for the layers that you want. Then do statistics on these, which ensures that you are now comparing data that is normalized to the same equivalent soil mass and you "compare apples to apples".*

AARC1-9. *We agree with this approach, but we should anyway emphasize that the ESM approach we used to 60 cm was comparing apples to apples indeed as they compared soil stocks. We have however now included estimation at further depths with the ESM method (see answer AARC1-3).*

- *The use of the AMG model is an interesting addition to the study. However, it is not well integrated with the rest of the text. It should be clarified in the introduction why it was used (I guess to compare input/output based SOC stock changes to measurements). Also, much more detail, such as parameters, inputs, etc. are needed in the methods. And why not use the AMG model to explore other management options that export less of the straw?*

AARC1-10. *The AMG model was used to test if the observed decrease of the SOC stock over time was supported by the known crop management at the site. An important result of this study is that AMG indeed confirms that the observed decrease in SOC is explained by the management. A detailed description of the model was already mostly given in the material and methods section, giving all inputs and parameters (Lines 288-302 and Table 1). We have added a description of the computation of the root's exports to the soil. We also propose to include a sensitivity to the most sensitive parameter of AMG: the ratio of active* *to stable organic C pool (Cs) which was set to the average value for this site as previously estimated by Kanari et al. (2022).*

The idea of using the method to test management option is a very good one. We propose to add a simple sensitivity study to organic imports and residue return effect as well as meteorological conditions similar to the 30 previous years: (1) the residues return to the soil were increased to 100% of the agronomically sound possibility, (2) the organic amendments were cut to zero or (3) were multiplied by two, and (4) the meteorology was set to a repeated 1987-2004 meteorology for the 2005-2050 period. Additionally, a sensitivity to the ratio of the active to stable C pool (Cs = 0.65) was made to illustrate the sensitivity of the model to this crucial parameter. For that values of 0.63 and 0.68 as independent estimate on a soil nearby were taken from Kanari et al. (2022).

This leads to the Figure R4 below which shows that (1) AMG in its standard setup overestimate the ESM soil sock change by a factor 1.5, (2) the incorporation of 100% of the crop residues would lead to 20% less decrease, while doubling the organic amendments would lead to 40% less decrease. It is also clear that the meteorology cannot explain the observed soil-stock difference between, 2005 and 2019, while the model is highly sensitive (as expected) to the Cs parameter. When changing this parameter to the range of evaluated values for this soil (0.63-0.68, from Kanari et al. (2022), we find a change of SOC stock 0.3 kg C m^{-2} in 2019. The Cs ~0.68 would lead to a better agreement with the observed stock evolution from 0 to 30 cm.

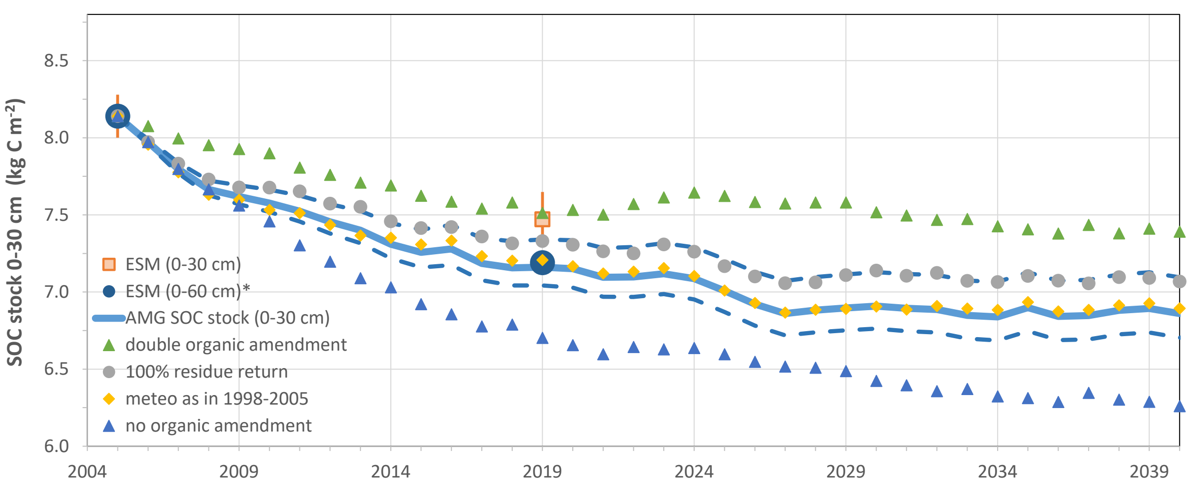

***Figure R4.** Comparison of the AMG simulations with the ESM stock evolution over the 2005-2019 period in the 0-30 cm and in the 0-60 cm. The blue line is the reference simulation (Cs = 0.65), and blue dashed lines correspond to a range of plausible values for Cs (Cs = 0.63 and Cs = 0.68). The dots correspond to 100% residue return, a doubling of organic amendment or no organic amendment, and a meteorology identical to the 30 years period before 2005. * To allow comparison of the 0-60 cm layer with the 0-30 cm layer the 2005 0-60 cm soil stock was shifted to that of the 0-30 cm layer.*

We propose to include the following sentence at the end of the introduction after Line 120 to better explain the interest of using the AMG model:

"The use of the AMG model allows analysing the yearly dynamics of the carbon stock over the period, and examine the sensitivity to residue return to the ground"

The following sentence in the result section after the new Figure 8 (Figure R4):

"This leads to the Figure R4 below which shows that (1) AMG in its standard setup overestimate the ESM soil sock change by a factor 1.5, (2) the incorporation of 100% of the crops residues would lead to 20% less decrease, while doubling the organic amendments would lead to 40% less decrease. It is also clear that the meteorology cannot explain the observed soil-stock difference between, 2005 and 2019, while the model is highly sensitive (as expected) to the Cs parameter. When changing this parameter to the range of evaluated values for this soil (0.63-0.68, from Kanari et al. (2022), we find a change of SOC stock 0.3 kg C m^{-2} in 2019. The Cs ~0.68 would lead to a better agreement with the observed stock evolution from 0 to 30 cm. "

And the following sentences in the discussion sections:

"The AMG simulations corroborate this hypothesis showing a decrease mainly explained by the low residue return and limited organic C application, while the meteorology (+0.3°C when comparing the period of 30 years before 2005 and the period 2005-2019) does not have a significant effect on the soil C stock (Figure R4)."

And

*"The results from our AMG model, which accounts for carbon residue return, imports and exports, reproduces the observed SOC stock declines, though with a higher decline than the observed one, providing evidence that the system is not in carbon equilibrium and that this imbalance is the most plausible cause of the observed changes (**Figure R4 see answer AARC1-10**). Furthermore, the AMG simulation suggests that the SOC stock should diminish at the same rate until 2027 and then stabilise. A sensitivity analysis shows that increasing the residue return would lead to a stabilization of the SOC stock to 7.2 instead of 6.95 kg C m^{-2}, while doubling the organic carbon amendment would lead to an equilibrium of 7.5 kg C m^{-2}. On the contrary suppressing the organic carbon amendment which may be a reality with the installation of a methaniser on the farm would lead to a stabilization to 6.3 kg C m^{-2}. "*

- *The discussion only loosely connects the results of this study to those of others (we found X, they found Y). A more integrative discussion is needed, discussing what the results of this study mean in the context of the others, and what the novelty of this study is.*

AARC1-11. *We thank the reviewer for this valuable suggestion. We have significantly modified the discussion section by deepening the integration of our findings with previous studies. We now explore potential mechanisms underlying the observed SOC stock patterns in our study, as well as factors that may explain divergent results reported in the literature. Our discussion is organized in the following topics.*

We propose now the following discussion section:

*"**4. Discussion***

Our data suggest that the management system implemented in the FR-Gri site may reduce soil carbon stocks levels, thereby increasing CO_2 emissions to the atmosphere. Our study highlights one key finding: The cropland at the Grignon station, although managed under reduced tillage, seems to lose approximately 0.07 kg m^{-2} yr^{-1} (or 7 Mg ha^{-1} yr^{-1}) between 2005 and 2019. Additionally, our study shows that cumulative SOC stocks at 60 cm depth computed using FD and ESM approaches were comparable at *our site, indicating that both methods may be suitable for monitoring total carbon if sampling down to this depth at this site.*

4.1. Effects of sampling depth and computation methods on organic carbon stock changes evaluation

Our results reveal that cumulative SOC stock changes between 2005 and 2019 under reduced tillage management were similar when using either the FD or ESM approach, differing by only 3% in the 0–60 cm sampling layer ($p > 0.80$). Previous studies (Du et al., 2017; Xiao et al., 2020) have documented misleading interpretations of SOC stock increases in conservation systems (e.g., reduced tillage, no-tillage) when using the FD approach at shallow depths (≤ 30 cm). Our results support this: FD at 30 cm indicates SOC stock losses in 2019, while ESM in a similar layer showed gains. Indeed, FD approaches are prone to bias when soil bulk density or SOC content changes, irrespective of the soil management (von Haden et al., 2020). Because BD often vary with management in agricultural soils, especially at shallow depths (≤ 30 cm), multilayer sampling and equivalent soil mass approaches are essential to capture the temporal response of SOC stock (Xiao et al., 2020; Wendt and Hauser, 2013). At the FR-Gri site, the topsoil (0–15 cm) is frequently disturbed by shallow tillage using a stubble cultivator or clod crusher, and deep tillage operations have occasionally been applied to depths of up to 40 cm. In addition to residue return, these practices likely influence BD and soil mass distribution, particularly within the upper 40 cm of the profile. Additionally, the potential compaction caused by repeated machinery traffic cannot be excluded (Hamza and Anderson, 2005), since the compaction tends to accumulate over time below 40 cm due to limited tillage operation of the subsoil (Zhang et al., 2024). Roots may also alter BD, including in subsurface layers, by modifying the physical properties (e.g., aggregation, porosity) as roots efficiently explore deeper layers. In the FR-Gri site, we find a significant decrease of BD in the 0-5 cm and 5-30 cm layers and no significant change in the lower layer (30-60 cm) (Figure R6, see answer to Figure 5 comment). Likewise, roots may contribute to subsoil SOC stocks through root growth, biomass accumulation, and rhizodeposition. The rhizodeposition process may account for ~30% of root C and ~10% of total photosynthesized C, as shown for maize (Tardieu, 1988) and wheat (Zhang et al., 2020; Zuo et al., 2006), the main crops at the FR-Gri site. Fan et al. (2016) reported that approximately 95% of root biomass lies above 100 cm. In our field, 80% of the SOC stock changes occurred in the 5–30 cm layer, confirming that sampling to at least 60 cm better captures root-related C inputs and reduces SOC estimate bias, as also emphasized by Baker et al. (2007) and Wendt and Hauser (2013). Furthermore, SOC stock estimates in deeper and multiple layers provides valuable insights into SOC dynamics across the profile, as mineralized carbon may percolate and accumulate in subsoil layers (Rumpel and Kögel-Knabner, 2011).

4.2. SOC stock difference between 2005 and 2019 explained by changes in carbon exports and imports

SOC stock losses in cropland systems under various management practices have been widely reported in European studies (De Rosa et al., 2024) and often arise from soil and crop management factors such as soil tillage and cover crop (Breil et al., 2023), and the unbalance of carbon imports and exports (Poyda et al., 2019; Ingwersen et al., 2024). Over the 13.25-year period (2005–2019), the FR-Gri site has also experienced a decrease in SOC stock of 0.95 [0.87-1.03] kg C m^{-2} (95% CIs in brackets). Our study evidenced that C losses in the intermediate soil layers (5–40 cm) are not offset by gains elsewhere (~0-5 and 40-60 cm). In average, the reduced tillage system implemented at FR-Gri led to carbon loss of 72 ± 3 g C m^{-2} yr^{-1} over the 13.25 years in the 0-60 cm soil layer, irrespective of the SOC estimation method. Our hypothesis is that SOC declines is primary related to a long-term imbalance between carbon imports - limited by reduced crop residue return - and high biomass exports. The FR-Gri site has been under continuous cropland management for over 100 years, with reduced tillage and crop rotation introduced in the past *two decades. In the 1980s, the field received unquantified but large amount of organic matter inputs from waste treatment plants, however, since 2004, increased export of wheat straw for bioenergy has reduced crop residues return, while organic amendments were limited (Table 1). This shift in management practices may have contributed to a long-term unbalance between C imports and exports, leading to SOC stocks declines since on average, the field exports were around threefold higher than imports and twice higher than the import and aerial residue return combined. The AMG simulations corroborate this hypothesis showing a decrease mainly explained by the low residue return and limited organic C application, while the meteorology (+0.3°C when comparing the period of 30 years before 2005 and the period 2005-2019) does not have a significant effect on the soil C stock (Figure R4).*

In terms of soil processes, SOC stock declines during that period may reflect an unbalance between SOC mineralization and immobilization rates likely trigged by a priming-effect through fresh organic matter inputs - mainly from legumes, plants with low C:N ratio, organic amendments, and nitrogen-rich fertilization (193 kg N ha^{-1}) (Ceschia et al., 2010; Loubet et al., 2011; Bernard et al., 2022). Mary et al. (2020) reported stable SOC stocks over 45 years in a winter wheat–maize rotation, with SOC gains in the surface layer (0–10 cm) offset by losses in deeper layers (10–30 cm), regardless of tillage intensity. However, unlike our study, their maize was not harvested for silage, which likely resulted in higher residue returns. Dimassi et al. (2014) similarly observed SOC declines (-0.018 to -0.76 kg C m^{-2} yr^{-1}) in the 0–30 cm layer under reduced residue returns and when maize was removed from the rotation. Both studies support our findings that reduced residue return and specific crop choices, such as silage maize, can significantly lower SOC stocks in cropland systems. Keel et al. (2019) reported ESM-based SOC losses ranging from 0.01 to 0.135 kg C m^{-2} yr^{-1} across various crop systems in Switzerland, with an average loss of 0.034 kg C m^{-2} yr^{-1} in the topsoil (~0–20 cm). Their highest SOC stock losses were observed under a crop rotation similar to that of FR-Gri, with a comparable initial stock (~7 kg C m^{-2} in the 0–20 cm layer), but implemented on an Orthic Luvisol. We notice that their C inputs from residue return and organic fertilisation (0.090–0.32 kg C m^{-2} yr^{-1}) are comparable to ours (0.265 ± 0.030 kg C m^{-2}), but they attributed the C losses to the recent grassland (with high SOC stock) to cropland (with low SOC stock) conversion, which may explain the doubled carbon stock change compared to this study.

*The results from our AMG model, which accounts for carbon residue return, imports and exports, reproduces the observed SOC stock declines, though with a higher decline than the observed one, providing evidence that the system is not in carbon equilibrium and that this imbalance is the most plausible cause of the observed changes (**Figure R4 see answer AARC1-10**). Furthermore, the AMG simulation suggests that the SOC stock should diminish at the same rate until 2027 and then stabilise. A sensitivity analysis shows that increasing the residue return would lead to a stabilization of the SOC stock to 7.2 instead of 6.95 kg C m^{-2}, while doubling the organic carbon amendment would lead to an equilibrium of 7.5 kg C m^{-2}. On the contrary suppressing the organic carbon amendment which may be a reality with the installation of a methaniser on the farm would lead to a stabilization to 6.3 kg C m^{-2}. Our simulation agrees with those reported by Meersmans et. Al. (2016). Likewise, the integrated carbon fluxes from 2006 to 2010 (Loubet et al., 2011), confirm a carbon loss from the soil similar to the AMG simulations during that period (**Figure R4**). Although the uncertainties on the integrated carbon fluxes are very large, the convergence between the two approaches corroborates a large soil carbon loss in the years 2005-2010 which is consistent with the small organic carbon fertilisation and residue return during that period (Table 1). We also note that the yearly carbon loss from Loubet et al. (2011) is not significantly different from the yearly carbon soil destocking found in the present study. In the north-western part of Switzerland, in a Cambisol soil, Leifeld et al. (2011) also compared the integrated carbon fluxes and soil sampling methods over 5 years on an* *intensive and an extensive grassland, both recently converted from intensive cropland. They concluded that the large uncertainties in both methods prevented detecting a significant change over 5 years in the intensive field. On the contrary, in the extensive field, they found a significant decrease of the SOC stock of -0.217 ± 0.143 kg C m^{-2} yr^{-1} by soil sampling, but a lower loss of –0.065 ± 0.092 g C m^{-2} yr^{-1} based on the integrated carbon fluxes method. We note that in Leifeld et al. (2011) and in our study, the two methods provide yearly losses that are not significantly different from each other.*

4.3. Uncertainties in soil carbon stocks changes in FR-Gri

*We recognize the importance of distinguishing a true SOC change from artifacts introduced by differences sampling designs in 2005 and 2019. However, the mapping of the soil based on 2019 sampling clustering (Figure R2 in answer **AARC1-4**) provides an objective way to select a reduced area in 2019 to compare with the 2005 campaign over a similar soil type, which corroborates the farmer expert knowledge on the field heterogeneity. The robustness or our results across both design- and model-based approach, alongside the clustering the identified distinct soil groups, increases our confidence that the observed differences reflect real changes in SOC stocks over time (Figure R3 in answer **AARC1-4**).*

In the reduced field, where soil properties were assumed to be comparable between both campaigns (Figure 4), the observed SOC stock changes between 2005 and 2019 in the 0-60 cm layer -0.95±0.21 kg C m^{-2} (showed a p-value < 0.001 and minimum detectable difference (MDD) of 0.71 kg C m^{-2} (Table 2, see answer AARC1-6). As both the p-value < 0.01 and the observed SOC changes exceeded the MDD, this indicates that SOC changes in FR-Gri was not only statistically significant but also detectable given our sample size and variability. In contrast, the complete field did not show the same pattern, as the observed SOC changes was lower than MDD, indicating that the changes detected between 2005 and 2019 falls below the detection thresholds and may not be confidently interpreted as a real difference. Therefore, while the trend may indicate a possible decline in SOC in Complete field, the result must be interpreted with caution due to the potential for Type II error (failing to detect a real effect). Our effect-size analysis (Figure R5 below) confirmed that the SOC changes was significant within the Reduced field, where CIs did not include zero. The higher MDD when all strata are included in our comparisons reflect the field soil heterogeneity, related to the presence of Calcisols (shallow soil with high rock fragments and SIC content) on the north-western part of the field (Figure R3 soils 1 and 4). The difference in soils not only affect the soil bulk density and fine soil stocks, but also the soil's capacity to stabilize carbon through the positive interactions between Calcium (Ca) and soil organic matter (Kleber et al., 2021).

Additional uncertainty on the overall carbon stock change at the site may come from inorganic carbon losses. Indeed, previous measurements of carbon leaching at the FR-Gri site indicated that inorganic carbon, which stock change could not be evaluated with the 2005 sampling data, may also contribute to significant soil carbon losses. Kindler et al. (2011) showed that, in 2010, the site was losing 28 g C m^{-2} yr^{-1} through leaching with a contribution of 21 g C m^{-2} yr^{-1} as dissolved inorganic carbon (DIC). Inorganic carbon leaching hence dominates at the site, with 75% of the leached C being inorganic, indicating a clear carbonate dissociation to DIC leaching, due to H^+. Although not measured directly as a soil stock change, we can therefore evaluate that carbonate leaching would lead to an additional inorganic soil carbon loss of 21 g C m^{-2} yr^{-1}, leading to a total of 72 + 21 = 93 g C m^{-2} yr^{-1} carbon loss. The inorganic carbon loss would therefore represent a very significant amount of 22% of the total carbon lost from the field, which could be induced by high nitrogen fertilisation (193 kg N ha^{-1} as half organic, half mineral, Table 1) and base cations exports by harvest (Raza et al., 2021; Song et al., 2022; Zamanian et al., 2021; Hao et al., Guo et al., 2010). We should however bear in mind that even if C is lost by DIC-DOC leaching from the 0-60 cm *layer, it may lead to a deep C sequestration by formation of secondary CaCO3 (Liu et al., 2022; An et al., 2019).*

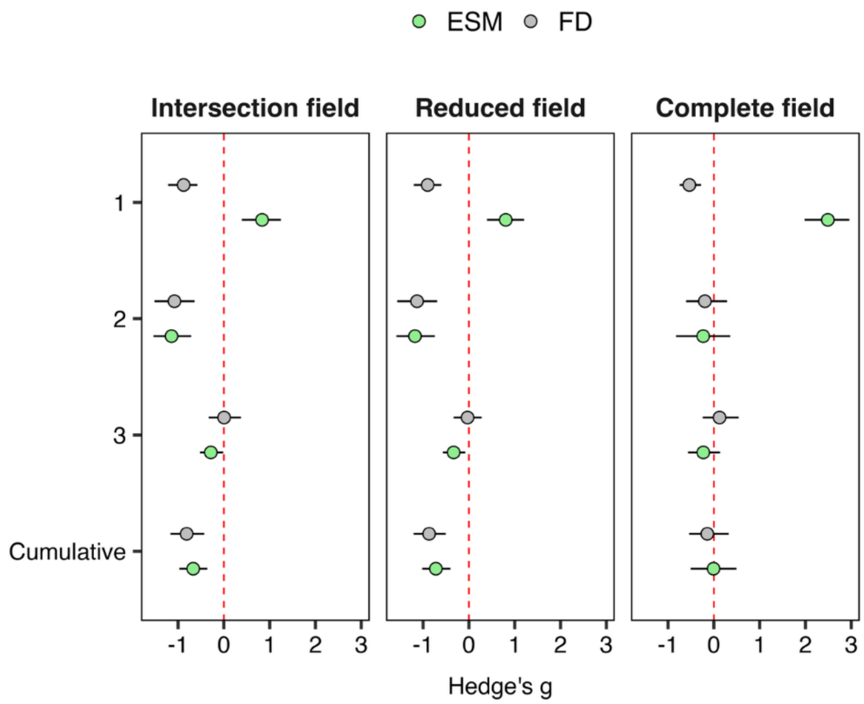

Figure R5. *Magnitude of differences in soil layers between 2005 and 2019, measured using Hedges' g effect size method (Hedges, 1981). Circles represent Hedges' g values, and error bars denote bias-corrected and accelerated (BCa) 95% confidence intervals based on 20,000 bootstrap resamples. Confidence intervals overlapping zero indicate no statistically meaningful difference between datasets for the corresponding variable. Layers 1, 2, and 3 correspond to the 0–5 cm, 5–30 cm, and 30–60 cm depth intervals, respectively. The "Cumulative" category refers to the full 0–60 cm soil profile. The complete field includes 10 strata, the reduced field comprises 7 strata, and the intersection field comprises 5 strata.*

Specific comments:

L25+27 Would be good to state initial values of bulk density and of SOC

We propose to change to: *"A significant decompaction was observed over the 13.25 years up to 30 cm. Bulk density decreased by 22% in the 0-5 cm layer (2005 = 1.31, 2019 = 1.02 g cm^{-3}) and by 5% in the 5-30 cm layer (2005 = 1.53, 2019 = 1.45 g cm^{-3}). These reductions are likely attributable to the adoption of reduced tillage practices performed since 2005."*

L31 Since you indicate a change in in bulk density, I think you should only present the results based on equivalent soil mass! The other ones are not right, that is why you used the equivalent soil mass approach.

See previous answer to L25+L27.

L32 It must be 0.6% of the initial carbon not % of soil carbon. Please clarify.

It is indeed 0.6% / year of the initial carbon. This is therefore assuming a linear decrease over year to give a rough idea of the rate of change per year.

L35 Should you not evaluate the change in SOC stock since 2005 rather than absolute SOC stock?

This is indeed what we did but it may have been ambiguous. We therefore propose to modify this sentence. Instead of:

"We utilised the soil carbon cycling model AMG to simulate the soil carbon dynamics over the period and beyond. Based on recorded exports and imports and estimated residue return to soil, the model was in perfect agreement with the soil stock measurements."

We propose:

"The AMG model simulating SOC stock changes between 2005 and 2040 in the top 30 cm revealed a declining trend under current management practices, consistent with the decreases observed using both ESM and FD approaches for the same depth. By 2040, SOC stocks are projected to decline to 6.94 kg C m^{-2}, representing an approximate 16% reduction from the 2005 baseline."

L40 While I agree that 4pm is probably unrealistic, you should not claim that the results from a single field are a good disproof.

We fully agree with this comment. We however feel that our phrasing did not say it was a good disproof but only that it "questioned" its feasibility. We propose anyway to temper the sentence as: *"If confirmed by further studies, our results question the feasibility of the 4per1000 aspirational target [...]"*

L50 Should you not cite the Sanderman paper here as a more recent estimate of soil C loss?

This is a good suggestion. We will cite indeed the Sanderman paper here.

L58 Evaluated? Where is the proof? Or do you mean postulated?

This number was evaluated with the so called 4per1000 approach. We agree that this evaluation is rather a postulation. We propose to change the sentence to : *" It was, however, <u>postulated</u> that 2 to 3 Pg C yr^{-1} could be sequestered in the top meter of agricultural soils by increasing carbon stocks by 4 per 1000 annually through changes in land use and the adoption of more conservation-oriented agricultural practices. <u>If achieved, this level of sequestration could</u> offset approximately one-third of global anthropogenic greenhouse gas emissions, as estimated in 2016 (Minasny et al., 2017)"*

L69 Large uncertainty is an understatement if the confidence interval includes the 0 within less than one standard deviation.

Yes indeed. But since "large" is subjective, the reader can make himself an opinion with the number given.

L82 And, even worse, if the PTF includes SOC, as it often does, SOC and BD measurements are not independent. This can then lead to systematic errors.

Thanks for this comment. We propose to modify the paragraph to : *"Measuring soil fine earth fraction relies on bulk density (BD) and RF measurements, which are time-consuming. Bulk density is often estimated by pedo-transfer functions (PTF), which are known to produce random errors and be prone to systematic biases (Knotters et al., 2022; Harbo et al., 2022; Schrumpf et al., 2011). In some studies, the choices of PTF equations, along with the neglect RF, lead to systematic uncertainties in SOC stock estimates, among which RF is the most critical (Poeplau et al., 2017; Beem-Miller et al., 2016; Wiesmeier et al., 2012; Ipcc, 2003). Likewise, many PTFs use SOC content as an input variable to predict BD which is subsequently used to estimate SOC stocks. This interdependence introduces a potential circularity in the estimation process. If the uncertainty associated with the PTF is not properly accounted for, this dependency can introduce systematic bias and increase the variance and overall uncertainty of SOC stock estimates (Schrumpf et. al. 2011; Xu et al., 2015)."*

L107-109 Since you will only evaluate one site, I think the two sentences about sampling design are a bit off-topic. To be removed, potentially.

This is a sound comment. However, this sentence was used to give a bit of context for the next sentence which presents the choice made in ICOS. Therefore, we suggest to change to: "Monitoring SOC stocks also requires an adequate sampling strategy that insure unbiased and robust estimates with a limited number of samples (Arrouays et al., 2012; Schrumpf et al., 2011; Saby et al., 2008; Don et al., 2007). Within ICOS, soils started to be sampled in 2017 adopting a design-based (DB) approach. This meth-od relies on a randomly chosen sampling points, which increase the precision of the mean or total estimates (Loustau et al., 2017; Arrouays et al., 2018; Brus and Degruijter, 1997; Laslett et al., 1997; Jaap J. et al., 2006)."

L123 It is the long-term monitoring site in Grignon, right? I would at least spell out the name once, as it is widely known.

Yes, Grignon is the name of the city. This is a good suggestion which we will follow.

Figure 2 Why is strata 9 excluded in 2019? And why are strata 7 and 10 included? They are much further from the 2005 samplings. If anything, you should eliminate 7 and 10 and include 9.

Please see the answer AARC1-4 that gives a detailed explanation.

L210 Would be better to show all factors that you multiply with to do unit conversion than a single value that you collapse them to.

Thanks for the sound comment. We will explain the factor 10 in equation (1).

L250 A slightly better approach would be to follow the approach of Wendt and Hauser (2013), fitting a cubic regression of cumulative SOC stocks vs cumulative soil mass. As I understand it, your approach assumes a linear change of SOC stocks with soil mass, which may not be the best representation. Given that most soil stocks are in the topsoil, I guess it is not a huge error that you get by this, though.

Please see above our response regarding the ESM method computation (answer AARC1-3).

L268 Did you include all strata? Or just the ones that overlap with the 2005 sampling campaign? The latter would be preferable since the 2005 area covered is much smaller.

Please see the response above regarding the sub-setting of the 2019 dataset (AARC1-4).

L300 A lot of details missing on the AMG model. How was management simulated? Were plants simulated? What allometric equations were used to estimate root inputs (i.e., what percentage of NPP went into roots)? Where and how can the AMG model be assessed? Which parameterization was used?

The plants are not simulated in AMG but only the soil with an annual time-step. The management (residue return to the field) is an input and is not simulated. It was either measured or estimated using allometric coefficients and is provided in Table 1 of the manuscript. The management is not simulated. It is only checked that the depth of the soil management is not below 30 cm. The allometric equations for the root C inputs to the soil were indeed not provided in the original manuscript as they are provided in Clivot et al., 2019, 2023 as already cited in the manuscript. The C input to the roots are:

$$C_{BG(i)} = \frac{DM_{AG}}{SRR} * 0.4 * 1.65 * (1-)\beta^{i}$$

Where DM_{AG} is the above ground biomass, SRR is the shot-to-root-ratio, 0.4 is the carbon content of the roots, 1.65 is a factor accounting for the dead roots and rhizodeposition that are assumed to represent 65% of the living roots C, and $(1-\beta^{i})$ accounts for the roots density in the soil, where β is a crop-dependent parameter.

We already mentioned all parameter in the material and methods. The AMG model was evalued by Clivot et al. 2019 as already cited in the manuscript. We also can mention the work of Levavasseur et al. (2020). Clivot et al. 2019 showed that AMG had an RMSE of 0.3 kg C m-2 when evaluated against 60 long term experiments with a prediction error similar to the soil stock measurement error.

We propose to add these elements in the revised manuscript.

L315 The sample name SP-15 is not meaningful for external readers. I suggest you define it.

Thanks for the suggestion that we will follow. This is indeed the ICOS nomenclature.

L317-319 With such a different sampling design I would be careful to infer temporal changes such as "decrease of bulk density" and just state that it was lower in the 2019 campaign.

We recognize the importance of distinguishing a true temporal change in SOC stocks from artifacts introduced by differences in sampling design. However, we cannot definitively attribute the entire effect to sampling differences. The robustness or our results across both design- and model-based approach, as well as the clustering identifying the different soil groups, increases our confidence that the observed differences may reflect, at least partially, real changes in SOC stocks over time. See answers AARC1-4.

Figure 4 It is not a meaningful displaying of the data if you mix all soil horizons into one graph, especially, given that they had different layers and sampling depths (60 vs 100 cm). I think to make any meaningful comparison, you HAVE TO display per depth horizon. E.g., the 0-5 cm horizons is the same, the 30 to 60 *horizons in 2005 samples can be combined by computing the mean, and 5-15/15-30 by a weighted average.*

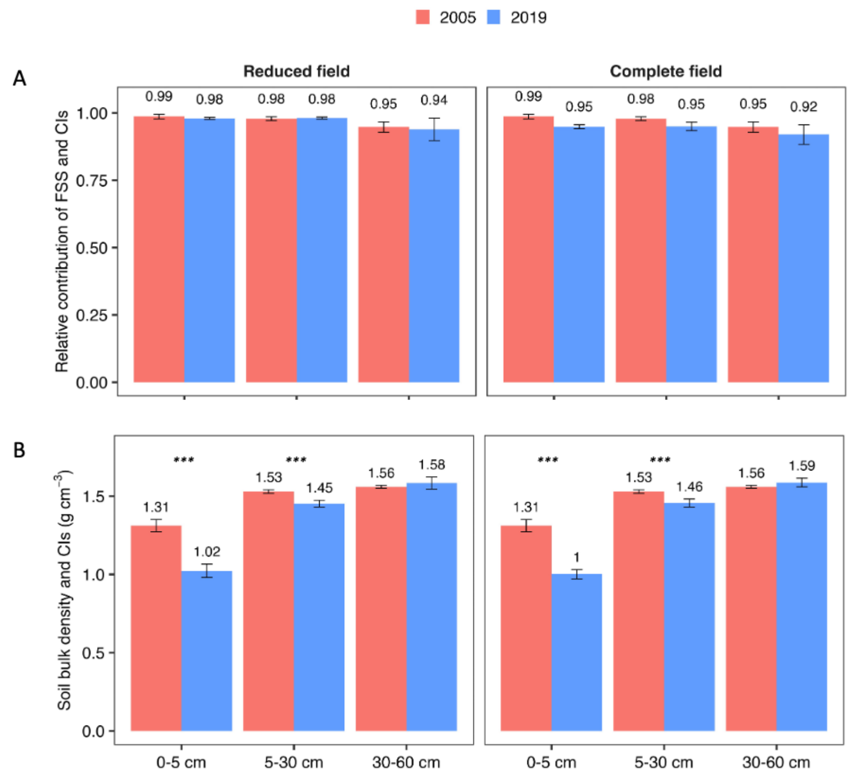

Figure R6. Means of relative contribution of fine soil stock and soil bulk density of the whole soil and their corresponding confidence intervals (CIs) in the 2005 and 2019 campaigns, computed using a design-based approach. In the "Complete field" panel, all strata from 2019 were considered. In the "Reduced Field panel", seven strata from the 2019 dataset with similar pedological characteristics to the 2005 area were included. Asterisks denote significant differences between both campaigns: P < 0.0001 (***), P < 0.001 (**), P < 0.05 (*).

Also, the strategy to choose the "reduced" sampling area just on having the same fine earth fraction as the 2005 sampling seems suboptimal. Why not choose based on spatial proximity?

Please see the detailed answer AARC1-4.

L338 Again, "decreased" should be "was lower" since the sampling positions were not really the same. This wrong wording exists through the rest of the text, but I will not mention it each time. Please try to correct throughout.

See previous answer on the same comment (*L317-319*).

L 339 Please add p values to your statements of significance

We reported the p values in the main text of the Results section, while asterisks were used to indicate significance in the figures. The P values (two.sided alternative) were computed using Welch's t-test. (***) P < 0.001; (**) P < 0.01; (*) P < 0.05. See Table R2 and Figures R6-R8.

Figure 5 I think it would be best to display this data as dots with error bars for your confidence intervals. The mean depth of each horizon could be used to display the dot and dots should be connected. At the moment it is really hard to read. Also, can you use annotations to show where there were significant differences? Also figure 5 Why are the SOC stocks of 2005 missing in 30-60?

We have modified the figure to show common horizons after aggregating the data as explained in detailed in the answer AARC1-1 and AARC1-3, and show the Bulk density, SOC content, and SOC stocks as a profile (Figure R7).

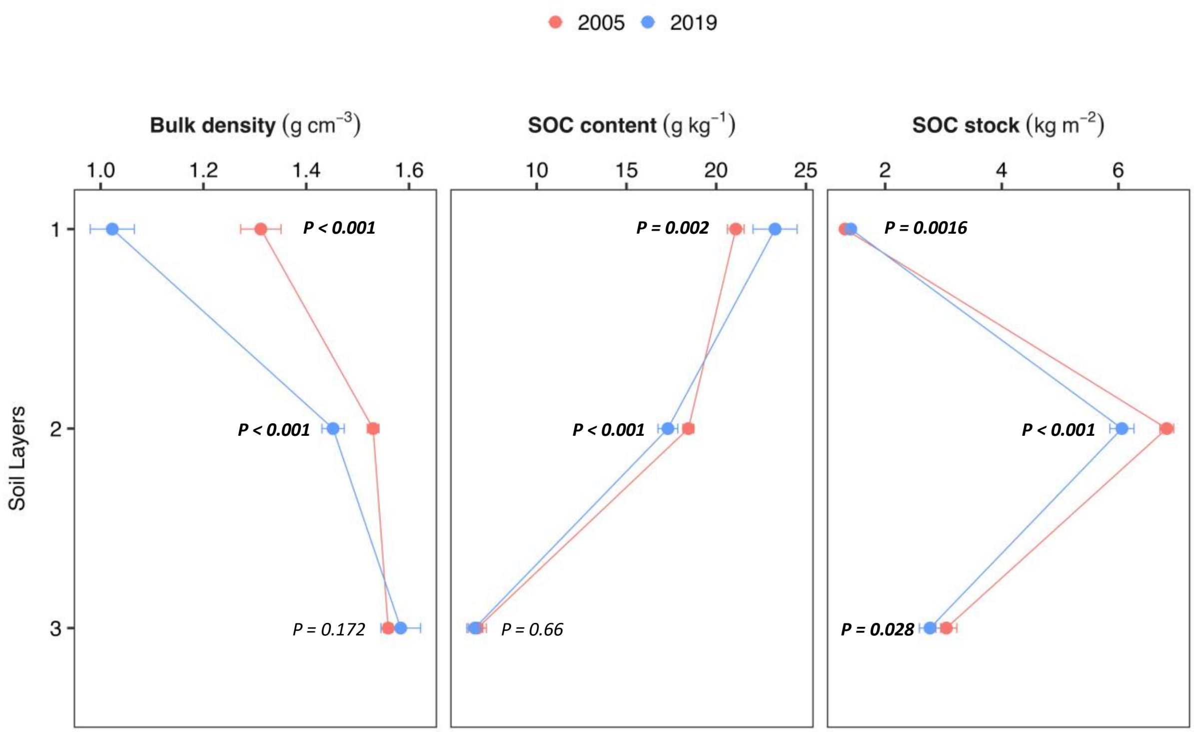

Figure R7. *Mean of bulk density and soil organic carbon (SOC) contents and stocks with their corresponding confidence intervals (CIs) in the 2005 and 2019 campaigns across three soil layer (0-5, 5-30, 30-60 cm), computed using a design-based approach. Only the "Reduced Field" is presented.*

L348 For the SOC content you use the word "higher" to compare, which is the right way to address it.

Thanks for this comment.

L352 What is slightly significant? $p < 0.1$? Please add p values.

See previous answers (L 339).

L353 What "correlation" do you mean? If you specify correlation, please give the r, also say if it is Pearson or Spearman. From the Figure S2, it does in fact not look like a strong correlation at all.

We propose to remove this figure that was confusing and does not provide much new information.

L357-364 Please do not use words such as "diminished" which imply temporal connection. Especially, when you talk about equal depth intervals. Focus on your ESM results, which is the only correct way to view it, given the BD has changed.

We do agree. See previous answers.

L375 This sentence is discussion. Also, what makes you so sure that the difference is not due to the different sampling design?+

See our detailed answer L317-319 and AARC1-4.

Figure 6 I suggest showing both 0-30 and 30-60 in this figure but ONLY with the ESM approach. You may display the fixed depth approach in the supplement.

Thanks for the suggestion. We however think it is a valuable information to show both in a single figure which also merged Figure 6 and Figure 7 (see Figure R8 below).

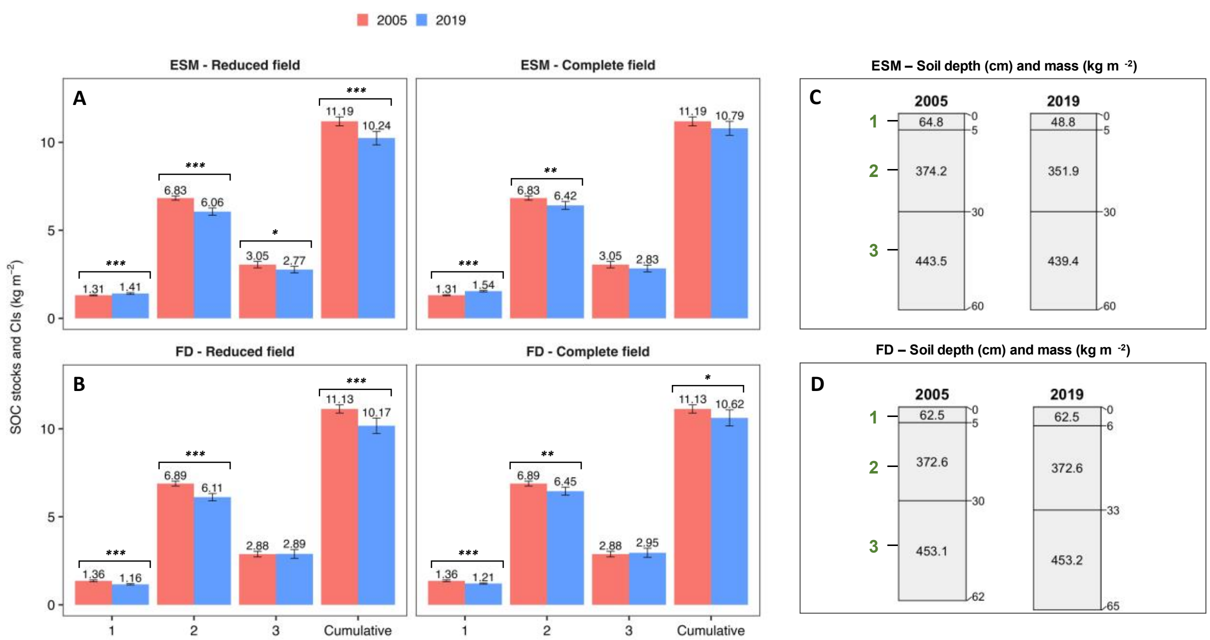

Figure R8. Mean soil organic carbon (SOC) stocks (kg m^{-2}) estimated using the Fixed Depth (FD, Panel A) and Equivalent Soil Mass (ESM, Panel B) approaches, along with their corresponding confidence intervals (CIs), for the 2005 and 2019 campaigns. Estimates were computed using a design-based approach. Soil depth (cm) and fine soil mass (kg m^{-2}) are also shown in Panels C and D. In the "Reduced Field" panels, seven strata from the 2019 dataset with pedological characteristics similar to the 2005 area were included. *, **, and *** stands for significance (p< 0.1, p < 0.01 and p< 0.001).

Figure 7 I suggest you replace this figure by ESM based SOC stocks, based on soil masses (that you report together with the cm) that represent roughly the depths you want to show. Simplest way to do this would be Wendt and Hauser, that you also cite.

See previous answer to Figure 6.

L431 I would start the discussion with a statement based on your data, and only then follow with the other studies in the following sentences.

Thanks for the suggestion. We propose to start the discussion section with this added sentence as follows: *"Our data suggest that the management system implemented in the FR-Gri site may reduce soil carbon stocks levels, thereby increasing CO_2 emissions to the atmosphere. Our study highlights one key finding: The cropland at the Grignon station, although managed under reduced tillage, seems to lose approximately 0.07 kg m^{-2} yr^{-1} (or 7 Mg ha^{-1} yr^{-1}) between 2005 and 2019. Additionally, our study shows that cumulative SOC stocks at 60 cm depth computed using FD and ESM approaches were comparable at our site, indicating that both methods may be suitable for monitoring total carbon if sampling down to this depth at this site."*

L436 Again, if you apply the Wend and Hauser approach, you could easily estimate SOC stock differences for an ESM representing 0-15cm.

This sentence has been removed

L441 Not necessarily. How much more robust 0-60 is really depends on how much BD changes and on how strong your SOC gradient is with depth (as you correctly argue in L455). ESM is the standard!

We fully agree that BD changes is a key driver affecting SOC stock estimation. The paragraph has been modified and is presented in the response related to the "chapter 4.1"

L450 As long as you do not use equivalent soil masses, your comparisons of 0-5cm etc. layers are really kind of meaningless!

The ESM approach was chosen to compare stocks across the soil layers.

Chapter 4.1 could be significantly shortened. A lot of the discussion is around what we already know – ESM is the better approach. Using the ESM approach for all depth/soil mass layers would really help to improve the discussion. E.g., are the differences in the upper horizons still there, if ESM is used.

Please see detailed answer AARC1-11.

Chapter 4.2 Can you really conclude that you have a change in SOC stock with the difference in sampling design? Is the MDD of Schrumpf et al. (2011) already considering a different sampling design, or is it meant only for the exact same sampling design?

Please see detailed answer AARC1-4.

Chapter 4.3 It would be important to state for any study that you report, whether is was ESM or fixed depth based. Overall, too long as well. The results of all the other studies could be summarized much more concisely.

In our revised manuscript, we propose to focus the comparisons with studies using the ESM approach.

L487 Why the switch to % losses from absolute values?

We switched to % because the study from Van Wesemael et al. (2010) reported percentage.

Chapter 4.4 is the best discussion chapter. It really tries to dig deeper and explain. The other chapters also provide room for that. For example, it could be discussed in more detail why there is a need to use ESM based on your and other data, and what the error is using a fixed depth approach.

Thanks for this comment. See answer AARC1-11.

L520 What do you mean by "it is excluded that the change comes from a change in crop cultivation". Please clarify.

This sentence was not well formulated. What we wanted to mean is that since the site was cultivated for over 100 years, the decrease in soil organic carbon observed between 2005 and 2019 could not be due to a recent land use change of that field. We propose to modify the sentence for clarification to: "Since the site was cultivated for over 100 years, the decrease in soil organic carbon observed between 2005 and 2019 could not be due to a recent land use change of that field"

L532 Please mention where was the Leifeld study was conducted. Switzerland?

Indeed, their work was carried out in north-western part of Switzerland under a Stagnic Cambisol. See answer AARC1-11.

L553 N per ha and YEAR?

Yes, thanks for spotting the typo. We will correct the units.

Chapter 4.5 While interesting, the part on inorganic carbon leaching is mostly speculation. Your study adds little here so this part should be kept to a minimal. In contrast, you MUST include a paragraph about the uncertainty that arises in your study due to the 2005 and 2019 sampling campaigns using different sampling designs! This uncertainty may reduce if you use ESM and discuss the error reduction properly.

We now have a dedicated chapter discussing the uncertainties related to our results, including not only those arising from sampling design and statistical approaches but also those due to incomplete measurement of carbon stocks—specifically, the missing information about SIC changes.

Conclusions The uncertainties of the approach should be mentioned.

L578 Rather demonstrating the need to use ESM! (But increasing sampling depth is also good)

We do agree on both comments.

References

Bernard, L., Basile-Doelsch, I., Derrien, D., Fanin, N., Fontaine, S., Guenet, B., Karimi, B., Marsden, C., and Maron, P.-A.: Advancing the mechanistic understanding of the priming effect on soil organic matter mineralisation, Funct. Ecol., 36, 1355–1377, https://doi.org/10.1111/1365-2435.14038, 2022.

Canty, A. J., and Ripley, B.: Package 'boot' (Version 1.3-4), R Project for Statistical Computing, 2012. http://cran.r-project.org/web/packages/boot/

De Rosa, D., Ballabio, C., Lugato, E., Fasiolo, M., Jones, A., and Panagos, P.: Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade?, Glob. Change Biol., 30, e16992, https://doi.org/10.1111/gcb.16992, 2024.

Efron, B.: Better bootstrap confidence intervals, J. Am. Stat. Assoc., 82, 171–185, https://doi.org/10.1080/01621459.1987.10478410, 1987.

Ferchaud, F., Chlebowski, F., and Mary, B.: SimpleESM: R script to calculate soil organic carbon and nitrogen stocks at Equivalent Soil Mass, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), 2023.

Gräler, B., Pebesma, E., and Heuvelink, G.: Spatio-temporal interpolation using gstat, R J., 8, 204–218, 2016. https://journal.r-project.org/archive/2016/RJ-2016-014/RJ-2016-014.pdf

Hamza, M. A., and Anderson, W. K.: Soil compaction in cropping systems: A review of the nature, causes and possible solutions, Soil Tillage Res., 82, 121–145, https://doi.org/10.1016/j.still.2004.08.009, 2005.

Hijmans, R.: raster: Geographic Data Analysis and Modeling, R package version 3.6-32, https://doi.org/10.32614/CRAN.package.raster, 2025.

Ingwersen, J., Poyda, A., Kremer, P., and Streck, T.: Harvest residues: A relevant term in the carbon balance of croplands?, Agric. For. Meteorol., 349, 109935, https://doi.org/10.1016/j.agrformet.2024.109935, 2024.

IUSS Working Group WRB: World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps, 4th Edn., International Union of Soil Sciences (IUSS), Vienna, Austria, 2022.

Kanari, E., Cécillon, L., Baudin, F., Clivot, H., Ferchaud, F., Houot, S., Levavasseur, F., Mary, B., Soucémarianadin, L., Chenu, C., and Barré, P.: A robust initialization method for accurate soil organic carbon simulations, Biogeosciences, 19, 375–387, https://doi.org/10.5194/bg-19-375-2022, 2022.

Kirby, K. N., and Gerlanc, D.: BootES: An R package for bootstrap confidence intervals on effect sizes, Behav. Res., 45, 905–927, https://doi.org/10.3758/s13428-013-0330-5, 2013.

Kleber, M., Bourg, I. C., Coward, E. K., Hansel, C. M., Myneni, S. C. B., and Nunan, N.: Dynamic interactions at the mineral–organic matter interface, Nat. Rev. Earth Environ., 2, 402–421, https://doi.org/10.1038/s43017-021-00162-y, 2021.

Liebhard, G., Toth, M., Stumpp, C., Bodner, G., Klik, A., Zhang, X., Strohmeier, S., and Strauss, P.: Developing topsoil structure through conservation management to protect subsoil from compaction, Soil Tillage Res., 253, 106669, https://doi.org/10.1016/j.still.2025.106669, 2025.

MacQueen, J.: Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 281–297, 1967.

McBratney, A. B., and Minasny, B.: Comment on "Determining soil carbon stock changes: Simple bulk density corrections fail" [Agric. Ecosyst. Environ. 134 (2009) 251–256], Agric. Ecosyst. Environ., 136, 185–186, https://doi.org/10.1016/j.agee.2009.12.010, 2010.

Pebesma, E., and Bivand, R.: Spatial Data Science: With Applications in R, Chapman and Hall/CRC, 2023. https://doi.org/10.1201/9780429459016

Pebesma, E.: Simple features for R: Standardized support for spatial vector data, R J., 10, 439–446, https://doi.org/10.32614/RJ-2018-009, 2018.

Peng, Y., Chahal, I., Hooker, D. C., and Van Eerd, L. L.: Comparison of equivalent soil mass approaches to estimate soil organic carbon stocks under long-term tillage, Soil Tillage Res., 238, 106021, https://doi.org/10.1016/j.still.2024.106021, 2024.

Poyda, A., Wizemann, H.-D., Ingwersen, J., Eshonkulov, R., Högy, P., Demyan, M. S., Kremer, P., Wulfmeyer, V., and Streck, T.: Carbon fluxes and budgets of intensive crop rotations in two regional climates of southwest Germany, Agric. Ecosyst. Environ., 276, 31–46, https://doi.org/10.1016/j.agee.2019.02.011, 2019.

Powlson, D., Stirling, C., Jat, M., Gerard, B., Palm, C., Sanchez, P., and Cassman, K. G.: Limited potential of no-till agriculture for climate change mitigation, Nat. Clim. Change, 4, 678–683, https://doi.org/10.1038/nclimate2292, 2014.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2025. https://www.R-project.org/

Rovira, P., Sauras, T., Salgado, J., and Merino, A.: Towards sound comparisons of soil carbon stocks: A proposal based on the cumulative coordinates approach, Catena, 133, 420–431, https://doi.org/10.1016/j.catena.2015.05.020, 2015.

Rumpel, C., and Kögel-Knabner, I.: Deep soil organic matter—a key but poorly understood component of terrestrial C cycle, Plant Soil, 338, 143–158, https://doi.org/10.1007/s11104-010-0391-5, 2011.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM National Conference, 517–524, https://doi.org/10.1145/800186.810616, 1968.

Tennekes, M.: tmap: Thematic maps in R, J. Stat. Softw., 84, 1–39, https://doi.org/10.18637/jss.v084.i06, 2018.

Thorndike, R. L.: Who belongs in the family?, Psychometrika, 18, 267–276, https://doi.org/10.1007/BF02289263, 1953.

von Haden, A. C., Yang, W. H., and DeLucia, E. H.: Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties, Glob. Change Biol., 26, 3759–3770, https://doi.org/10.1111/gcb.15124, 2020.

Xu, L., He, N. P., Yu, G. R., Wen, D., Gao, Y., and He, H. L.: Differences in pedotransfer functions of bulk density lead to high uncertainty in soil organic carbon estimation at regional scales: Evidence from Chinese terrestrial ecosystems, J. Geophys. Res.-Biogeosci., 120, 1567–1575, https://doi.org/10.1002/2015JG002929, 2015.

Zhang, B., Jia, Y., Fan, H., Guo, C., Fu, J., Li, S., Li, M., Liu, B., and Ma, R.: Soil compaction due to agricultural machinery impact: A systematic review, Land Degrad. Dev., 35, 3256–3273, https://doi.org/10.1002/ldr.5144, 2024.