We thank the two reviewers and Fabien Ferchaud for their useful comments that we believe have helped improve the manuscript.
The line numbers correspond to the track-change version of the manuscript.

Answer to reviewer 1

RC1: 'Comment on egusphere-2025-592', Moritz Laub, 23 Apr 2025 reply
In the following, we have numbered our answers as AARC1-n, where n is the answer number (Author Answers to RC1).

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). We additionally checked that there were 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). We also compared the only comparable parameter between the two sampling campaigns (the rock fraction), which showed no significant differences (**Figure R3**, Figure S1 in the revised MS supp mat, P12). 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 harmonisation 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 for 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).

3) *Regarding the discussion section, we propose to significantly revise this section to enhance clarity and coherence, and we propose to organise it into three chapters:*
   - ***a.*** *L570, P20-21: "**4.1 Effects of sampling depth and computation methods on organic carbon stock changes evaluation**": we will discuss the results obtained using 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 will also provide a mechanistic interpretation to explain the observed stock change patterns. See **answer AARC1-11**.*
   - *b. L598, P21-22: "**4.2. Possible causes of the observed SOC stock changes over 13.25 years**". This restructuring will address the reviewer's concerns about the section's length and allow for the removal of 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 the AMG model and the ICOS C balance approach, but add some AMG scenarios to help interpret the reasons for the observed SOC stock change. See **answer AARC1-10**.*
   - *c. L666, P22-23: **"4.3. Uncertainties in soil carbon stock changes"**: 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 our results as measured by the Hedges g effect and the minimum detectable difference (MDD) published by Schrumpf et al. (2011). We also added 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 the inconvenience. We did provide higher resolution images as an author comment in the EGUsphere discussion page. Anyway, we updated the figures with TIFF files with 800 DPI and now provide higher resolution images.*

- 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 only display figures with ESM in the main manuscript and discussing SOC changes based on ESM SOC stocks only. 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 Materials and Methods section by isolating a dedicated section 2.7 describing the ESM approach (L289-L321, P10):

"2.7 Soil carbon stocks calculation using the equivalent soil mass (ESM) and SOC stocks changes

To properly estimate SOC stocks evolution, one needs to consider changes in SOC content of the soil (SOC) but also the potential changes in BD_i due to compaction or decompaction, which may change the fine earth mass FE_i in each sampling depth (Lipiec and Hatano, 2003). 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 occur due to repeated surface traffic by heavy machinery (Liebhard et al., 2025; Lu et al., 2021). 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 (Ellert and Bettany, 1995; von Haden et al., 2020; Lee et al., 2009; Wendt and Hauser, 2013). This approach has the advantage of accounting for a common sampling bias with the hydraulic corer, which is soil compaction.
*The ESM-based SOC stock was computed using the R function "SimpleESM" (Ferchaud et al., 2023), which implements the classical ESM method (Ellert and Bettany, 1995) and ESM2, a model-based approach incorporating cubic splines (Wendt and Hauser, 2013). The reference fine earth mass (FE_{ref}) was derived from the median values in the 2005 dataset for the aggregated soil layers: 0–5 cm, 5–30 cm, and 30–60 cm (**Table 1**). The total FE in the 0–60 cm layer 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 fine earth mass (FE$_{ref}$) per layer used in the equivalent soil mass approach (ESM).

Layer	*Upper depth*	*Lower depth*	*FE$_{ref}$*
	cm		*kg m^{-2}*
L1	*0*	*5*	*63.2*
L2	*5*	*30*	*372.6*
L3	*30*	*60*	*453.1*

In the "classical" ESM approach (Ellert and Bettany, 1995), SOC stock is calculated by 1 mm increments (Autret et al., 2016; Mary et al., 2020). In brief, soil depth is discretised into elementary layers of 1 mm thickness, with FE density (g cm^{-3}) and carbon content (g kg^{-1}) assigned to each 1 mm layer. Since both FE 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, FE_i and SOC_{stock_i} are then computed cumulatively until the FE_{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" (Lee et al., 2009;

*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 fine earth mass (von Haden et al., 2020; Wendt and Hauser, 2013). As both methods yielded similar results (**Figure S7**), only ESM outcomes are reported in the following."*

- 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, in particular when there are more than 10 years between the two campaigns, and staff or protocols can change. Indeed, in the case of ICOS, many current sites had historical campaigns with a different sampling design, and only 10 years ago, ICOS defined a common standardised protocol to be used in all sites. This manuscript is therefore the occasion to evaluate whether we can use these historical data to compare with the ICOS sampling design, or in other cases when a different sampling scheme is adopted for any possible reasons.*

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 campaigns that would indicate a soil difference: the rock fraction. We chose strata in 2019 where the rock fraction was lower than or equal to 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 the 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, and 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**).*

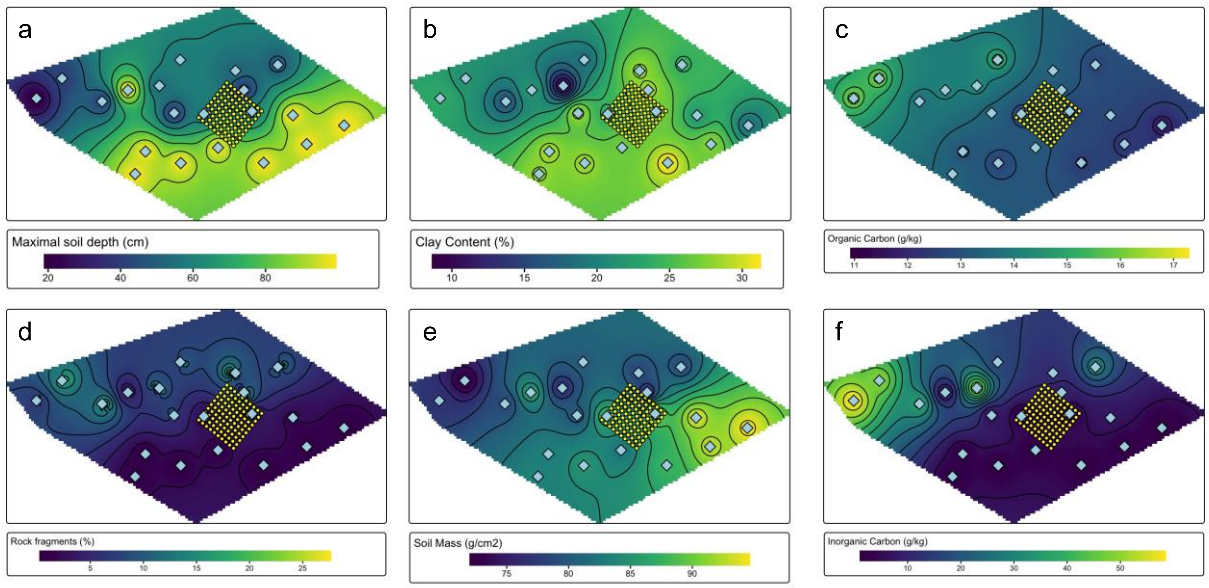

Figure R1 (Figure S2 in revised MS, L110, P13). *Interpolated 2019 soil properties used to construct the soil group map. Yellow circles indicate the 2005 sampling points, while blue diamonds represent the 2019 sampling locations. Maximum depth (cm) and clay content (%) were derived from the soil description report for the FR-Gri site according to WRB (2022). Clay content was measured in the diagnostic horizon and harmonised to the standard depth intervals (0–5, 5–15, 15–30, 30–60, and 60–100 cm) using cubic spline interpolation. All other variables were measured in the same five standardised layers.*

*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 characterised 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.*

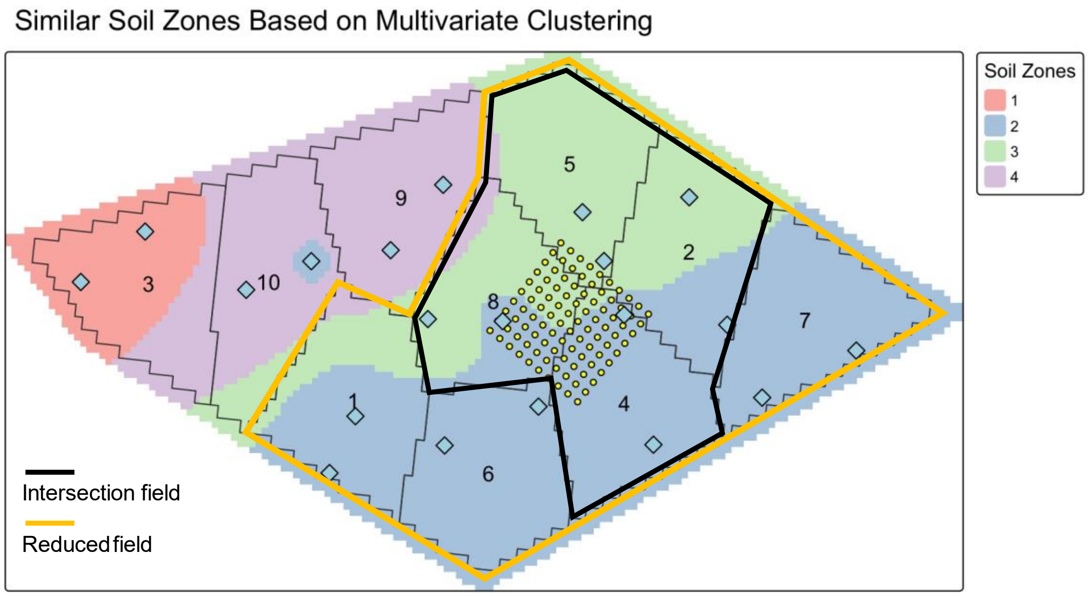

***Figure R2 (Figure 3 in revised MS, L340, P12)**. Map showing similar soil zones derived from multivariate clustering analysis. The study area is classified into four distinct soil zones (labelled 1–4), each representing regions with homogeneous soil properties based on statistical clustering of multiple variables (clay content, rock fragments, carbonates, organic carbon, fine earth mass, effective soil depth). The 2005 (yellow circles) and 2019 (blue diamonds) sampling points are superimposed together with the strata (numbered polygons) defined in the 2019 sampling. Yellow polygon represents the "Reduced field" and black polygon the* "Intersection field".

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 the 2019 samples, defined as the strata that contain at least one point of the 2005 sampling grid (**Figure R2**), 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**). 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**).

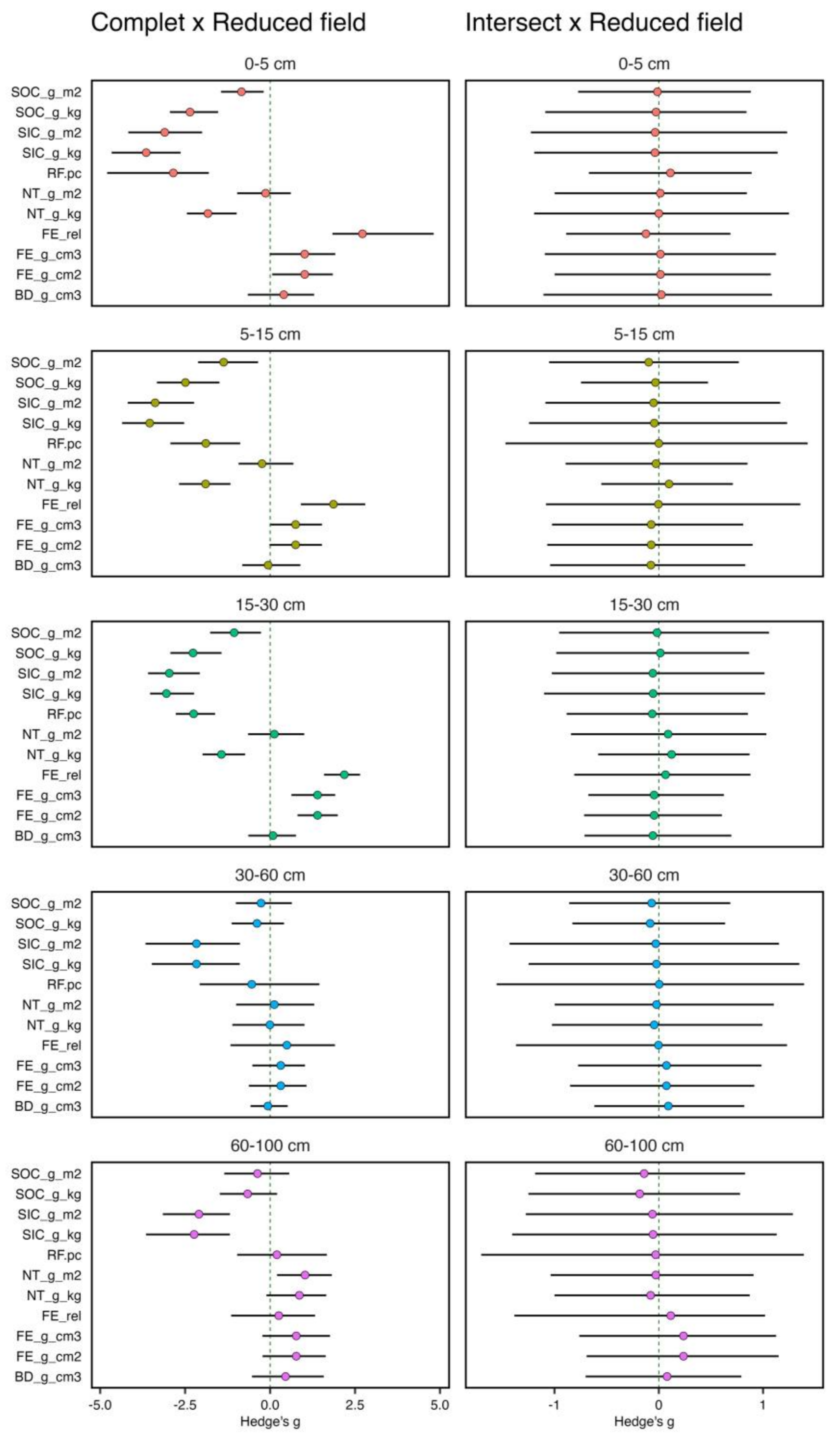

Figure R3 (Figure S1 in the revised MS supp mat, P12)*. Magnitude of differences in soil variables between different subsets of the 2019 dataset, measured using Hedges' g effect size method (Hedges, 1981). Circles represent Hedges' g values, and error bars denote bias-corrected and accelerated (BCa) 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 followed by their respective units: BD_g_cm3 – soil bulk density (g cm^{-3}); FE_g_cm2 – fine earth mass per area (g cm^{-2}); FE_g_cm3 – fine earth density (g cm^{-3}); FE_rel – proportion of fine earth mass (dimensionless, ratio); RF.pc – rock fragment content (%); SOC_g_kg – soil organic carbon concentration (g kg^{-1}); SIC_g_kg – soil inorganic carbon concentration (g kg^{-1}); NT_g_kg – total nitrogen* *concentration (g kg^{-1}); SOC_g_m2 – soil organic carbon stock (g m^{-2}); SIC_g_m2 – soil inorganic carbon stock (g m^{-2}); NT_g_m2 – total nitrogen stock (g m^{-2}).*

We propose to include two new sections, the first in the Materials and Methods 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 "Supplementary Material and Methods". We also propose to add Figures 3 (P12) in the revised manuscript and Figures R2 (Figure S1, P11) and R3 (Figure S3, P12) in the revised supplementary material

In the Materials and Methods, a new section 2.8 was added (L322-L368, P11-12) :

*"**2.8 Spatial comparison between 2005 and 2019.***

*To enable a reliable comparison between the 2005 and 2019 data and infer changes in SOC stocks, we minimised the effect of spatial heterogeneity in the 2019 sampling by subsetting the strata. To address this, we first employed a K-means-based clustering algorithm (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 characterised by its physical (clay and rock fragment contents, fine earth mass) and chemical properties (carbonates and SOC content). In addition, we extracted information from the soil descriptions of each 2019 SP-I **(Table S2)**, including effective soil depth (parent material layer not included) and clay content (%). 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), 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 were 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 clusters, hereafter, soil zones (**Figure 3**).*

[FIGURE 3]

Zone 2 is characterised 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. The Intersection and Reduced fields were chosen to ensure 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, 2018; Pebesma and Bivand, 2023), raster (Hijmans, 2010), and stars (Pebesma and Bivand, 2023). For inverse distance weighting (IDW) interpolation, we used gstat (Gräler et al., 2016; Pebesma, 2004). Plotting was carried out using tmap (Tennekes, 2018), and multivariate clustering was performed using stats (R Core Team, 2025). Effect-size analysis comparing soil variables in

*2019 across the original sampling layers 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 (**Figure S1**). 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 earth mass, 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 campaigns within the Reduced field to explore the SOC changes due to its consistent soil properties."*

An additional Supplementary material figure S1 and S2 (P12 and 13) and section (L19-46, P2):

"Effect-size using BCa bootstrap

Hedges' g was computed as:

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

(9)

$$d = \frac{\overline{X_1}-\overline{X_2}}{s} \qquad (s1)$$

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

Where d is the uncorrected effect size (Cohen's d), df represents the pooled degrees of freedom, Γ is the gamma function, s is the pooled standard deviation, X represents the individual observations in each group denoted by subscripts 1 and 2, $\bar{X}$ the sample mean in each group; n is the sample size in each group.

Confidence intervals for effect-size estimates were computed using 20000 nonparametric bootstraps with resampling and bias-corrected and accelerated (BCa) method (Canty et al., 2024; Efron, 1987; Kirby and Gerlanc, 2013). Negative values of Hedges' g indicate a reduction in SOC stocks from 2005 to 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).

The bias-corrected and accelerated (BCa) bootstrap confidence interval is computed as follow:

$$CI_{BCa} = \left[\widehat{\theta^{(\alpha_1)}},\ \widehat{\theta^{(\alpha_2)}}\right] \qquad (s2)$$

where the adjusted quantiles α_1 and α_2 are calculated as:

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0+z_{\alpha/2}}{1-\hat{a}\left(z_0+z_{\alpha/2}\right)}\right) \qquad (s3)$$

$$\alpha_2 = \Phi\left(z_0 + \frac{z_0+z_{1-\alpha/2}}{1-a\left(z_0+z_{1-\alpha/2}\right)}\right) \qquad (s4)$$

Here, Φ is the of standard normal cumulative distribution, and $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the corresponding quantiles of the standard normal distribution.

The bias correction factor is computed as the inverse of standard normal cumulative distribution:

$$\boldsymbol{z_0} = \boldsymbol{\Phi^{-1}}\left(\frac{\#\{\widehat{\boldsymbol{\theta}_b^*}<\widehat{\boldsymbol{\theta}}\}}{\boldsymbol{B}}\right) \qquad (s5)$$

Where $\widehat{\boldsymbol{\theta}}$ is the observed statistic (e.g., mean, median, effect size), $\boldsymbol{\theta}_b^*$ are the bootstrap replications, $\#\{\widehat{\boldsymbol{\theta}_b^*} < \widehat{\boldsymbol{\theta}}\}$ counts the number of bootstrap replication smaller than $\widehat{\boldsymbol{\theta}}$, and $\boldsymbol{B}$ is is the number of bootstrap samples."

- 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 approaches. But we did not compute 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 the 2005 and 2019 data may introduce uncertainties in our results. We, however, included detailed information about the statistical approaches in both the 2005 and 2019 datasets and reported confidence intervals in all mentions of soil stock changes. We propose adding the minimum detectable differences (MDD) as another metric to complement the p-values, based on a 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 metric (Hedges effect). Also, we provide a dedicated paragraph discussing the uncertainties related to the missing values of SIC stocks in 2005 (section 4.3, L666-L698, P22-23).*

Table R1 (Table 3 in revised MS, L532-L542, P 19). *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 annualised 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 show 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 gives 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's 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 emphasise 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 in 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 options 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. To do that, values of 0.63 and 0.68 were taken from Kanari et al. (2022), as an independent estimate made on similar soil sampled in a nearby field.*

*This leads to **Figure R4** below (Figure 7 in revised MS), which shows that (1) AMG in its standard setup overestimate the ESM soil sock change by a factor of 1.5, (2) the incorporation of 100% of the crop residues would lead to a 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 of 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.*

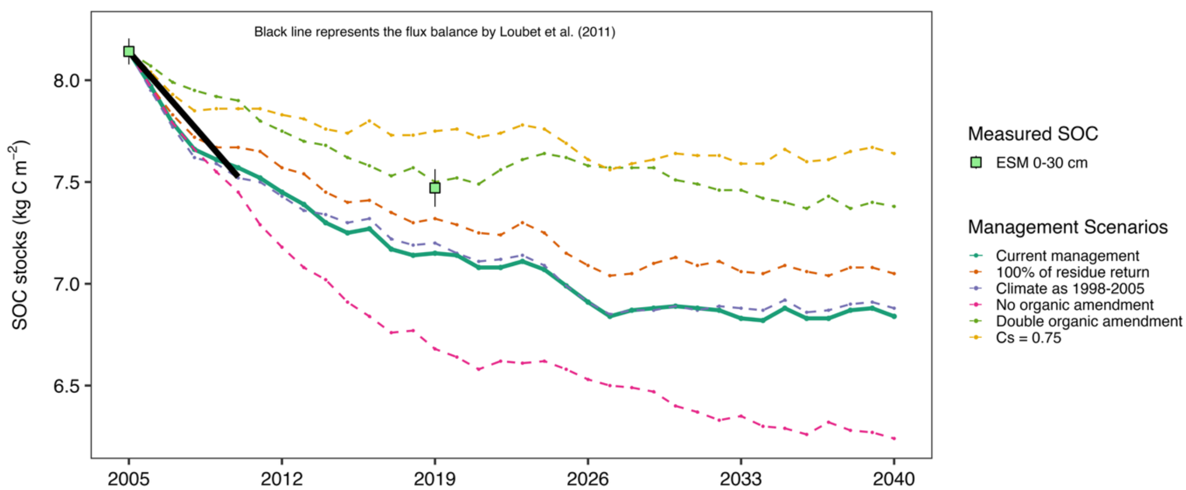

***Figure R4 (Figure 7 in revised MS, L565, P20)**. Soil organic carbon (SOC) stock in the 0-30 cm depth as simulated by the AMG model (plot lines), measured by soil sampling in 2005 and 2019 and computed using the equivalent soil mass (ESM; light-green square). The error bars show the sampling standard error. The 2019 SOC stock is given for the "Reduced field" only. Flux balance over the 2006-2010 period (black line) as published in Loubet et al. (2011), scaled to start at the measure SOC stock in 0-30 cm.*

We propose to include the following sentence at the end of the introduction to better explain the interest of using the models (L126-130, P4):

"Soil carbon cycling models such as DAYCENT (Parton et al., 1998), STICS(Brisson et al., 1998), RothC (Coleman and Jenkinson, 1996) or AMG (Clivot et al., 2019) are essential tools to further understand the observed SOC dynamics based on site-specific managements, and in particular exports, imports and residue returns. Models are also key in providing long-term simulation of SOC stock dynamics and scenario analysis. "

The modify the Material and method section related to AMG (L547-561, P 19):

“Under current cropland management practices, the model evidenced a declining trend of SOC stocks (**Figure 7**), which aligns with the decrease observed with the ESM approach in the 0-30 cm layer. Under current management, AMG models that SOC decreased from 8.24 kg C m^{-2} in 2005 to 7.25 kg C m^{-2} by 2019, reflecting a cumulative loss of approximately 0.99 kg C m^{-2} (−12%) over 13.25 years. SOC stocks appear to approach a quasi-steady-state from 2027 onwards, with fluctuations of ±0.02 to ±0.04 kg C m^{-2} yr^{-2}. By 2040, SOC stocks are projected to decrease to 6.94 kg C m^{-2}, representing an approximate 15% reduction from the 2005 baseline. Both the AMG model and measured SOC stocks were consistent with the flux balance approach reported by Loubet et al. (2011), during the early period from 2006 to 2010. The overall loss over a 22-year period (2005-2027) would then be of around 1.3 kg C m^{-2}, or 13 Mg C ha^{-1}, which amounts to 0.059 Mg ha^{-1} yr^{-1}. Overall, the sensitivity analysis across five scenarios shows the same declining patterns in SOC, with cumulative losses ranging from 5 to 18% by 2019, and from 6 to 23% by 2040. Increasing the residue return leads to a stabilisation of the SOC stock near 7.15 instead of 6.95 kg C m^{-2}, while doubling the organic carbon amendment would lead to an equilibrium of 7.48 kg C m^{-2}. On the contrary, suppressing the organic carbon amendment would lead to a stabilisation of 6.34 kg C m^{-2}.“

And the following sentences in the discussion sections (L616-620, P21):

“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 7)“

- 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 organised into the following topics.*

We now propose the following discussion section (L569-L698, p20-23):

“4. Discussion

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 have documented misleading interpretations of SOC stock increases with reduced or no-tillage when using the FD approach at shallow depths (≤ 30 cm) (Du et al., 2017; Xiao et al., 2020). Our results support this in the 0-5 cm layer, where FD indicates SOC stock losses while ESM shows gain (**Table S5, Figure S7**). 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 varies 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 in shallow layers (Wendt and Hauser, 2013; Xiao et al., 2020). 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 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 operations 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) (**Table S4**). Likewise, roots may contribute to subsoil SOC stocks through root growth, biomass accumulation, and rhizodeposition. The rhizodeposition process may account for up to 65% of root C and ~10% of total photosynthesised C, as shown for maize (Tardieu, 1988) and wheat (Zhang et al., 2020; Zou et al., 2022), 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, 20% of the SOC stock changes occurred in the 30-60 cm layer, confirming that sampling to at least 60 cm better captures root-related C inputs and reduces SOC bias estimate, as also emphasised by Baker et al. (2007) and Wendt and Hauser (2013). Furthermore, SOC stock estimates in deeper and multiple layers provide valuable insights into SOC dynamics across the profile, as mineralised carbon may percolate and accumulate in subsoil layers (Rumpel and Kögel-Knabner, 2011).*

4.2Possible causes of the observed SOC stock changes over 13.25 years

*SOC stock losses in cropland systems under various management practices have been widely reported in European studies (De Rosa et al., 2024). A major cause of carbon losses is the imbalance between carbon imports and exports, which progressively leads to a shift in the carbon stock from one state to a new one, higher if the imbalance is an excess of imports or lower in the opposite case (Ingwersen et al., 2024; Poyda et al., 2019). Over the 13.25-year period (2005–2019), the FR-Gri site has experienced a decrease in SOC stock of 0.95 kg C m^{-2} [95% CI: 0.51-1.4]. Our study evidenced that C losses in the intermediate soil layers (5–40 cm) are not offset by gains elsewhere down to 60 cm depth (~0-5 and 40-60 cm). On average, the system implemented at FR-Gri led to a carbon stock decrease of 72 ± 16 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 decline is primarily 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 at least over 100 years, with reduced tillage and crop rotation introduced in the past two decades. In the 1980s, the field received an unquantified but large amount of organic matter inputs from wastewater treatment plants. Moreover, since 2004, increased export of wheat straw for bioenergy has reduced crop residue return, while organic amendments were limited (Table 1). This shift in management practices may have contributed to a long-term imbalance between C imports and exports, leading to SOC stock 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 7**).*

The initial high SOC stocks may explain the observed declining trend at our site, as sustaining such high levels is difficult even with substantial organic inputs. In terms of soil processes, SOC stock declines during that period may reflect an imbalance between SOC mineralization and immobilization rates likely triggered by high fresh plant inputs with low C:N ratio, organic amendments, and nitrogen-rich fertilisation (193 kg

N ha^{-1}) in combination with climate drivers – notably elevated temperatures - that favour mineralization over immobilization (Bernard et al., 2022; Ceschia et al., 2010; Loubet et al., 2011). 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, and the initial soil C stock was also smaller. Dimassi et al. (2014) observed similar SOC stock 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 stock 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 the AMG model application, which accounts for carbon residue return, imports and exports, reproduce the trend of 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 7**). 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 stabilisation of the SOC stock to 7.2 kg C m^{-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 biogas plant on the farm, would lead to a stabilisation of 6.3 kg C m^{-2}. However, although not explicitly simulated in our study, digestate residues derived from biogas production could serve as an alternative organic amendment. While this residue typically contains lower content of labile organic carbon compared to fresh organic material, the remaining organic material tends to be more chemically recalcitrant and resistant to microbial decomposition. As a result, their incorporation in the soil may contribute to slight, but persistent, increases in SOC stocks over time (Keel et al., 2025; Thomsen et al., 2013).*

*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 7**). 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 fertilisations 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.*

4.3 Uncertainties in soil carbon stock changes

*We recognise the importance of distinguishing a true SOC stock change from artefacts introduced by differences in sampling designs in 2005 and 2019. To address this, the clustering of the soil based on 2019 soil properties (**Figure 3**) provides an objective way to subset the 2019 dataset to compare with the 2005 campaign over a similar soil condition. The data-driven area selection corroborates the farmer's expert knowledge of the field heterogeneity. The robustness, across both design- and model-based approaches, alongside the clustering of the soil properties to identify distinct soil groups, increases our confidence that the observed differences reflect real changes in SOC stocks over time.*

In the Reduced field, the observed SOC stock change between 2005 and 2019 in the 0-60 cm layer was -0.95 ± 0.22 kg C m^{-2}, exceeding the minimum detectable difference (MDD) of 0.73 kg C m^{-2} ($p < 0.01$), and this represents both significant and detectable changes given our sample size and design. In contrast, the Complete field did not show the same pattern, as the observed SOC changes fell below the MDD, indicating that the changes detected between 2005 and 2019 could be masked by spatial heterogeneity. Therefore, the results using the Complete field should be considered with caution due to the potential for Type II error (failing to detect a real effect). The larger MDD when all strata are included in our comparisons reflect increased soil heterogeneity, particularly related to the potential presence of Calcisol (shallow soil with high rock fragments and SIC content) on the north-western part of the field. These factors not only affect the soil bulk density and fine earth mass, but also the soil capacity in stabilising carbon through the positive interactions between Calcium (Ca) and soil organic matter (Kleber et al., 2021).

Additional uncertainty on the overall SOC 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, whose 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). 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 $CaCO_3$ (An et al., 2019; Liu et al., 2022).”

*In the supplementary section we added the Table S5 (P9) and Figure S7 (P17), **Figure R5 below**:*

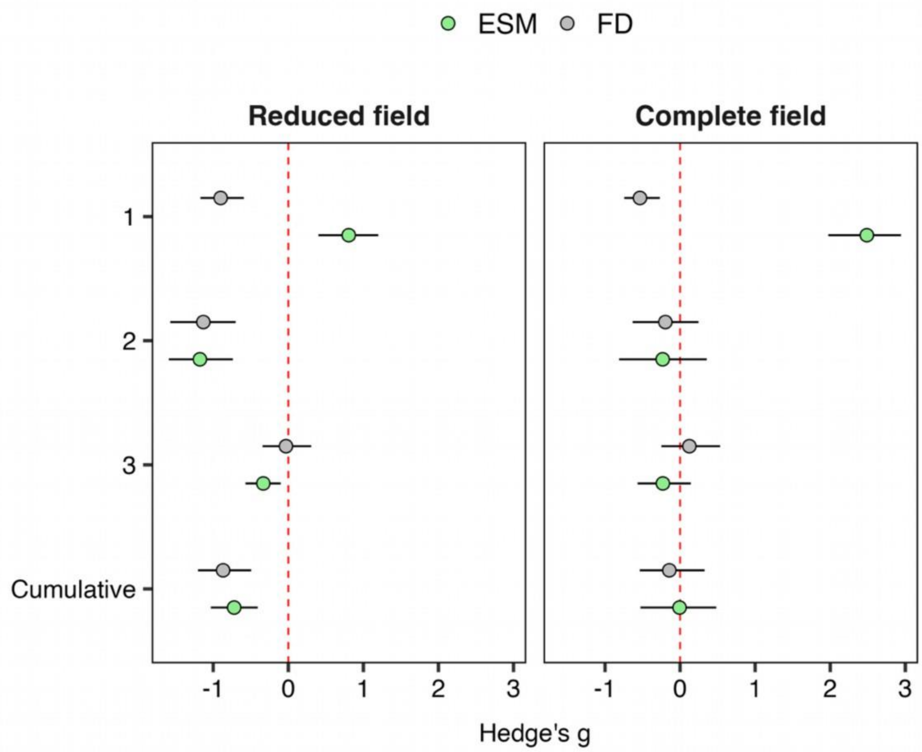

***Figure R5 (Figure S7 in revised MS, P17)**. Magnitude of differences in soil organic carbon (SOC) stock between 2005 and 2019 across soil layers, quantified using Hedges' g effect size (Hedges, 1981). Circles represent Hedges' g values, with their direction indicating the change in SOC stocks (kg C m^{-2}) in 2019 relative to 2005. Error bars show bias-corrected and accelerated (BCa) 95% confidence intervals, based on 20,000 bootstrap resamples. Confidence intervals that overlap zero indicate no statistically significant difference between the two campaigns for the corresponding variable. Soil layers 1, 2, and 3 correspond to sampling depth intervals of 0–5 cm, 5–30 cm, and 30–60 cm, respectively. The "Cumulative" category represents the full 0–60 cm soil profile for FD and 889 kg m^{-2} for ESM. The complete field includes 10 strata and the reduced field 7 strata.*

Specific comments:

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

We propose to change to (L26-29): "A significant soil decompaction was observed over the 13.5 years up to 30 cm. Bulk density decreased by 22% in the 0-5 cm layer (from 1.31 to 1.02 g cm^{-3}) and by 5% in the 5-30 cm layer (from 1.53 to 1.45 g cm^{-3}), likely due to the adoption of reduced tillage since 2004. "

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 the 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 to (L34-41):

"To further interpret this trend, we applied the soil carbon cycling model AMG to simulate SOC dynam-ics down to 30 cm depth from 2005 onwards. Based on site-specific exports and imports and estimated residue returns, the model predicted a continuous SOC stock decline under current management, stabi-lising around 2028. The SOC stock decrease aligns with the changes observed between 2005 and 2019 over the whole soil profile, but is larger than the stock change over the 0-30 cm depth. By 2040, SOC stocks are projected to decline to 6.9 kg C m^{-2}, representing an approximate 15% reduction from the 2005 baseline. Furthermore, the AMG simulation was also consistent with the carbon flux balance re-ported between 2006 and 2010 by Loubet et al. (2011). "

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 and have removed this sentence.

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

This is a good suggestion. We cited indeed the Sanderman paper here (L47 and L53).

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 (L69-72, P2): " 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 (L97-106, P3) : "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 are prone to systematic biases (Harbo et al., 2022; Knotters et al., 2022; Schrumpf et al., 2011). In some studies, the choices of PTF equations, along with the neglect of RF, lead to systematic un-certainties in SOC stock estimates, among which RF is the most critical (Beem-Miller et al., 2016; Poeplau et al., 2017; Wiesmeier et al., 2012). 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 circu-larity 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 changing to (L131-136): "Monitoring SOC stocks also requires an adequate sampling strategy that ensures unbiased and robust estimates with a limited number of samples (Arrouays et al., 2018; Don et al., 2007; Saby et al., 2008). Within the ICOS network, soils started to be sampled in 2017, adopting a Design-Based (DB) approach (Brown, 1992; Collins, 1992). This method relies on randomly chosen sampling points, which increases the precision of the mean or total estimates (Arrouays et al., 2018; Brus and deGruijter, 1997; de Gru-ijter et al., 2006; Loustau et al., 2017)."

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 site. 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 our response above 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 is:

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

Where DM_{AG} is the above-ground biomass, SRR is the shoot-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 the parameters in the material and methods. The AMG model was evaluated by Clivot et al. 2019 as already cited in the manuscript. We can also 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 (L423-L440, P14):

"Roots and rhizodeposition C inputs down to a considered depth i are computed as:

$$C_{BG(i)} = \frac{DM_{AG}}{SRR} * 0.4 * 1.65 * (1-\beta^{i}) \quad (9)$$

Where DM_{AG} is the aabove-groundbiomass, SRR is the shoot-to-root-ratio, 0.4 is the carbon content of the roots (40%), 1.65 is a factor accounting for the dead roots and rhizodeposition, assumed to be 65% of the living roots C, and $(1-\beta^{i})$ accounts for the roots' distribution in the soil, where β is a crop-dependent parameter.

The SOC stock changes were simulated on an annual timestep over the period 2005–2040, considering the 0-30 cm depth layer, which generally corresponds to the managed soil layer in cropland systems, where most crop roots and residue inputs occur. The baseline SOC stock was set to 8.25 kg C m^{-2}, based on measurements from 2005. The proportion of active C was set to 65%, as Clivot et al. (2019) proposed for agricultural fields with a long-term history of cultivation. To assess model sensitivity, we performed additional simulations varying key management and environmental factors: (1) residue returns to the soil *were increased to 100%, (2) organic amendments were either eliminated (set to zero) or doubled (multiplied by two), (3) meteorological conditions were held constant by repeating the 1987–2004 weather data for the 2005–2040 period, and (4) the ratio of active to stable carbon pools (Cs = 0.65) was varied using values of 0.63 and 0.75 taken from independent estimates on a nearby soil reported by Kanari et al (2022) to illustrate the model's response to this critical parameter."*

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 recognise the importance of distinguishing a true temporal change in SOC stocks from artefacts introduced by differences in sampling design. However, we cannot definitively attribute the entire effect to sampling differences. The robustness of our results across both design- and model-based approaches, 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 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 are the same, the 30 to 60 horizons in the 2005 samples can be combined by computing the mean, and 5-15/15-30 by a weighted average.

Also, the strategy to choose the "reduced" sampling area just by 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 2 and Figures R7, R8 and R9.

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 detail in the answer AARC1-1 and AARC1-3, and show the Bulk density, SOC content, and SOC stocks as a profile (Figure R6).

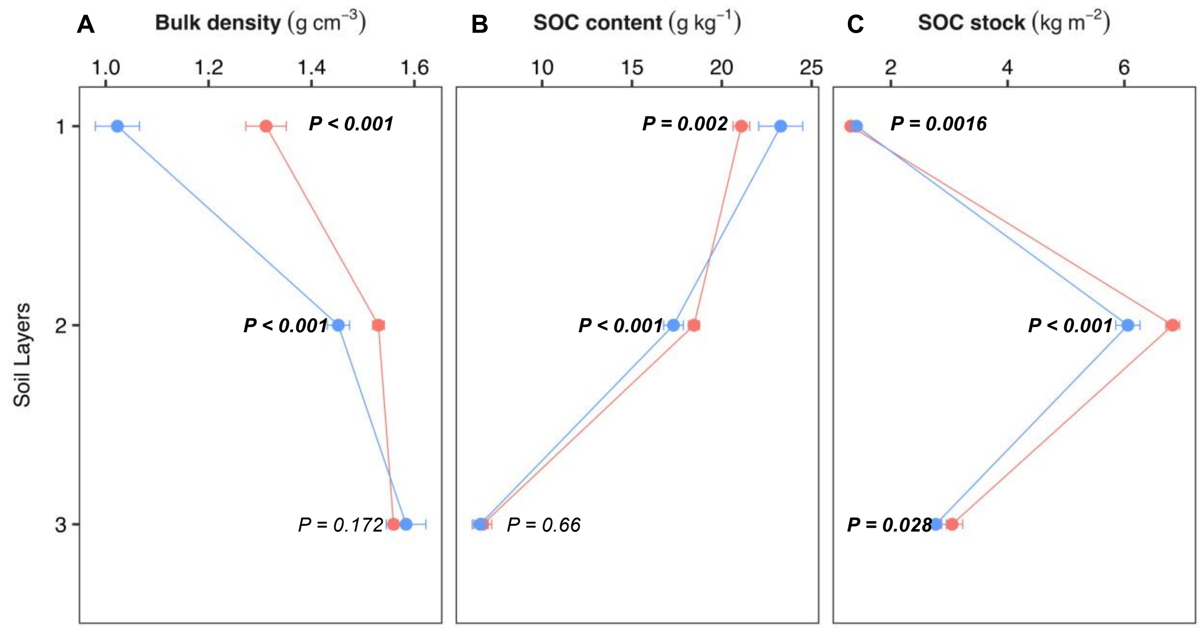

Figure R6 (Figure 4 in revised MS, L480, P16)*. Mean of bulk density and soil organic carbon (SOC) contents and stocks with their corresponding confidence intervals (CIs) in the 2005 and 2019 campaigns across three soil layers (0-5, 5-30, 30-60 cm). Only the "Reduced Field" is presented.*

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 displayed the ESM only in the revised manuscript (**Figure R7**).*

.

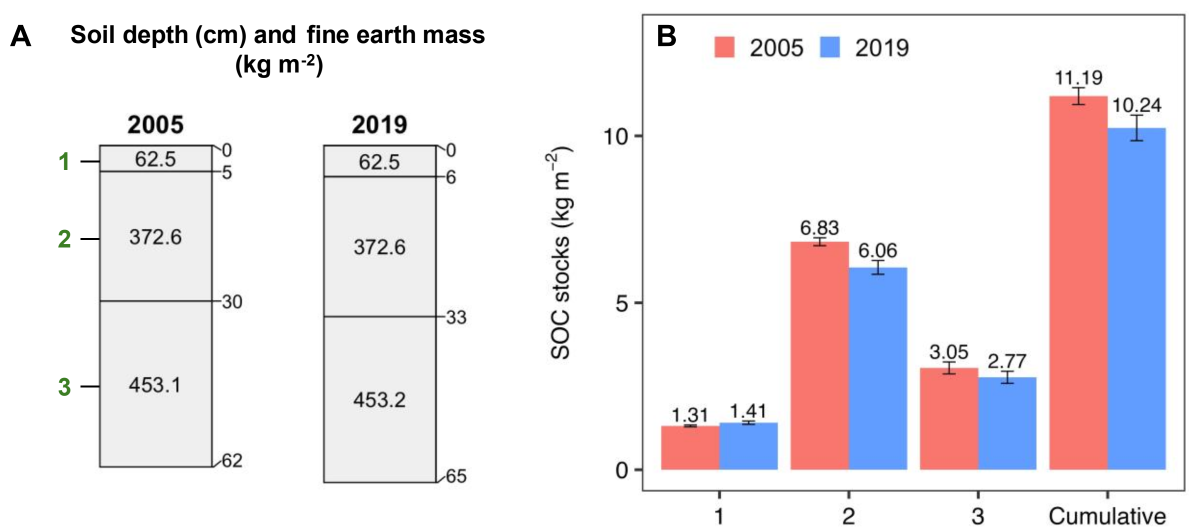

***Figure R7 (Figure 6 in revised MS, L515, P18)**. Mean soil organic carbon (SOC) stocks (kg C m^{-2}) estimated using the Equivalent Soil Mass (ESM), along with their corresponding confidence intervals (error bars), for the 2005 and 2019 campaigns. Adjusted soil depth (cm) and fine earth mass (kg m^{-2}) are also shown in Panel A. Commutation made for the "Reduced Field" where only seven strata from the 2019 dataset with pedological characteristics similar to the 2005 area were selected. Asterisks denote significant differences between both campaigns: $P < 0.0001$ (***), $P < 0.001$ (**), $P < 0.05$ (*).*

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 (L571-573, P20): "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 ($p > 0.80$)."

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 the revised manuscript, we focus on 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 percentages.

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 in that field. The discussion section was modified in depth and now the section reads (L609-819, P21): "The FR-Gri site has been under continuous cropland management for at least over 100 years, with reduced tillage and crop rotation introduced in the past two decades. In the 1980s, the field received an unquantified but large amount of organic matter inputs from wastewater treatment plants. Moreover, since 2004, increased export of wheat straw for bioenergy has reduced crop residue return, while or-ganic amendments were limited (Table 1). This shift in management practices may have contributed to a long-term imbalance between C imports and exports, leading to SOC stock declines since, on aver-age, 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 de-crease mainly explained by the low residue return and limited organic C application, while the mete-orology (+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 7)."

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

Indeed, their work was carried out in the 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 (section 4.3, L66-699, P22-23).

Conclusions : The uncertainties of the approach should be mentioned.

We have now made mention of it (L715-726, P24): "While our study detected SOC changes between 2005 and 2019, important uncertainties remain. Notably, the shift from a regular-grid design in 2005 (N = 100,

nested within the 2019 footprint) to a stratified random design in 2019 (N = 20 covering the entire C-flux footprint) may introduce artefacts related to soil heterogeneity and reduced statistical power. This calls for additional campaigns in the future with the same sampling design as in 2019. According to the AMG model runs, a change of around 0.3 kg C m^{-2} is expected between 2019 and 2028, which would be just above the standard error difference of 0.22 kg C m^{-2} found here, indicating that a sample in 2028 would be meaningful.
These uncertainties call for standardised, high-quality monitoring protocols such as those developed by the ICOS research infrastructure. Consistent sampling methodologies over time are needed to reliably assess the long-term impact of crop management on SOC stocks at sites like FR-Gri, and to improve our understanding of carbon dynamics in cropland systems. Integrating SOC stock data with CO_2 flux measurements will be crucial to exploring the underlying processes driving SOC changes."

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

We do agree on both comments.

Answer to reviewer 2

In the following, we have numbered our answers as AARC2-n, where n is the answer number (Author Answers to RC2). This allows referring to our answers in this document as well as in the answers to the other reviewers.

In this manuscript Loubet et al. present the first case study of SOC stock measurements with the ICOS protocol at the French crop site FR-Gri in 2019 and compute the soil stock evolution from 2005 to 2019. Overall, the paper is of a good scientific quality and stresses some important issues like digging deeper than 30cm when monitoring soil organic carbon stocks (unlike e.g., LUCAS or the LULUCF inventory which limit themselves to 30cm). There are however a few issues that need to be addressed before the manuscript can be accepted for publication.

A clear carbon balance should be included in this manuscript. Table 1 includes the imports and exports but no overall balance is calculated. Also is the mineralization rate known of this field? The outcome of the balance could be directly related to the SOC stock loss

AARC2-1. *We thank Reviewer 2 for his constructive comments. This first comment is very important indeed. Table 1 and the manuscript as a whole do not include a full carbon balance because this would imply computing the net ecosystem exchange (NEE) over the whole period 2005-2019. While in principle we have all the data to do such a carbon balance, we should stress that in 2018, there was a serious acquisition problem that invalidated almost the whole year of measurements, which we would like to gap fill with a crop model. Similarly, Kindler et al. (2011) have shown that lixiviation of carbon was an important term in the carbon balance from 2006 to 2010 in the FR-Gri field, which would require a modelling effort to evaluate the losses from 2010 to 2019. Finally, the spatial variability of the soil properties shown in this study also leads to spatial variability in the biomass production, as shown by Loubet et al. (2011), which will require additional footprint-based filtering on the NEE and biomass. All this work, in addition to an advanced uncertainty analysis on the carbon balance, is required to provide a sound comparison between the carbon balance and the carbon stock evolution. Unfortunately, this requires additional manpower, which leads to postponing the full carbon balance analysis for later publication, although we fully approve the interest of such a comparison.*

The differences between the sampling campaigns are very important and should be discussed into more detail. E.g., the difference in sampling depth intervals, the selection of the reduced sampling area.

AARC2-2. *We thank the author for this comment, which corresponds to RC1 comments. 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*** *and* ***Figure R2,*** *Figure 3 in revised MS, L340, P12). We additionally checked that there were 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 and Figure R2****). We also compared the only comparable parameter between the two sampling campaigns (the rock fraction), which showed no significant differences (****Figure R3****, Figure S1 in the revised MS supp mat, P12). This additional information provides additional confidence that the choice of the reduced sampling zone in 2019 corresponded to the soil characteristics of 2005.*

How was the measurement of the bulk density executed? Which protocol? Was the soil moisture content similar at the time of sampling in 2005 and 2019

AARC2-3. *The bulk density was measured by weighing the dry mass of the bulk sample obtained from a mechanical corer. The soil was first dried at room temperature prior to fractionating the fine soil, rock and roots fractions, then the rocks and roots were dried at 70°C prior to measuring the weight of rocks and roots. The remaining fine soil was subsampled and dried to 105°C to determine the remaining water content fraction, which was used to correct the dry mass of fine soil.*

The soil sampling was in December 2005 and in mid-March 2019. The surface moisture at these times was: 32.4% in 2005 and 39% in 2019, averaged over the 0-30 cm profile over the month.

The ICOS site focusses on arable land but in the introduction the effect of crops and crop type on carbon sequestration is hardly mentioned. The same for tillage type and the difference between inversion and non-inversion tillage which could lead to a redistribution of the carbon present in the soil.

AARC2-4. *We tried to give a short overview of these effects, which may not have been sufficiently put forward. We propose to rephrase the introduction to better emphasise these aspects (L58-L82, P3):*

"The impact of agricultural management on SOC stocks depends on the interaction between management practices, climate, and soil type (Paustian et al., 2016). Intensive farming in the 20th century, characterised by monoculture, soil tillage, and high fertiliser use, has aggravated climate change (Autret et al., 2016), notably through increased nitrous oxide (N_2O) emissions, which accounted for 74% of the total anthropogenic N_2O emissions in the last decade (Tian et al., 2024). Beyond N_2O, these practices also contribute to CO_2 emissions, primarily by releasing protected soil carbon into the atmosphere (Autret et al., 2016; Schmidt et al., 2011; Six et al., 2002). However, recent studies have shown that N fertilisation (Skadell et al., 2023) and occasional deep tillage can enhance SOC stocks by incorporating organic matter into deep layers (Krauss et al., 2022). Evidence also support the adoption of conservation agricultural practices, such as cover crop, high organic inputs, and diverse crop rotations, to restore SOC and mitigate climate change by enhancing C inputs and persistence in soil (Lal, 2004; Poeplau and Don, 2015, Schmidt et al., 2011), despite some carbon losses due to priming effects (Chen et al., 2019). Minimum and no-tillage have been promoted to increase SOC stocks, with a postulated sequestration of 2 to 3 Pg C yr^{-1} in the top meter of agricultural soils by increasing carbon stocks by 4 per 1000 annually. If achieved, this level of sequestration could offset approximately one-third of global anthropogenic greenhouse gas emissions (Minasny et al., 2017). However, these estimates may be overly optimistic, as C sequestration declines over time while the soil approaches a new equilibrium (Baveye et al., 2018; Franzluebbers et al., 2012). Moreover, growing evidence shows that their effectiveness may be limited unless they are combined with other conservation practices (Chenu et al., 2019). For instance, in France, Meersmans et al. (2016) suggested that significant SOC gains are only possible through the conversion of cropland to forest or grasslands, although expanding conservation practices and integrating temporary grasslands into crop rotations can also contribute meaningfully to SOC gains at the national scale (Launay et al., 2021). Although the primary objective today is climate mitigation through enhanced carbon stocks and diminished GHG emissions, including CO2, N2O and CH4, it is crucial that strategies designed to achieve this goal are balanced with other priorities, including ensuring global food security and minimising environmental impacts."

Figures are fuzzy and difficult to read

AARC2-3. *We apologise for the inconvenience that was due to a wrong setup of Adobe Acrobat on the very last version of the PDF submission with minor changes. All figures have been modified to improve* *the quality and were uploaded as an author response on 21 May 2025 "**AC1**: 'High quality figures to ease the reader', Benjamin Loubet, 21 Mar 2025".*

We have taken care of the figure quality in the revised version

Small remarks

L54: High organic matter inputs can lead to nutrient leaching. Important drawback.

We agree and appreciate your comment. We actually mentioned in the discussion section the evaluation of the organic and inorganic carbon leaching as measured by Kindler et al. (2011) (L542-544), which indeed showed that DOC leaching may be an important term in the overall carbon balance. "Dissolved organic carbon leaching was evaluated to contribute to the overall organic C loss from the field to around 10% (Loubet et al. 2011)."

We propose to specifically mention in the revised manuscript the potential negative feedback related to high organic matter inputs (e.g., priming effect → soluble C → leaching). We propose to add the following sentence in the discussion section (see detailed answer AARC1-11) (L621-625, P21):

"In terms of soil processes, SOC stock declines during that period may reflect an imbalance between SOC mineralization and immobilization rates likely triggered by high fresh plant inputs with low C:N ratio, organic amendments, and nitrogen-rich fertilisation (193 kg N ha^{-1}) in combination with climate drivers – notably elevated temperatures - that favour mineralization over immobilization (Bernard et al., 2022; Ceschia et al., 2010; Loubet et al., 2011)."

L175: Add the number of the strata to the figure. Why not limit the reduced sampling area to the strata that show overlap with the sampling area of 2005?

This is a very good comment which we have addressed in detail in our answer to the reviewer 1 comments (see answer AARC1-4). In short, we used as many strata as possible to increase the statistical significance. We showed in our answers to reviewer 1 that choosing the strata that overlapped 2005 did not change significantly the soil properties and SOC stock when compared to the reduced area that we chose in the original manuscript.

L185: Which soil characteristics are similar? Explain into more detail.

This is a very good comment which we have addressed at length in our answer to the reviewer 1 comments (see answer AARC1-4). The only similar characteristics that we can compare are the rock fraction, but using 2019 measured soil characteristics we identified groups of similar soil characteristics that correspond to the choice of reduced area retained in the analysis. And also, we showed that all the soil characteristics measured in 2019 were not significantly different in the reduced area and in the overlapped area (see previous answer).

L190: Root fraction >1mm in diameter?

The removal of roots larger than 1 mm in diameter aimed to facilitate the subsequent steps of soil preparation. After separating all soil fractions, the remaining roots were added to those previously measured (∅>1 mm).

L230: How deep was the reduced tillage practice applied? Was decompaction verified by e.g., penetration resistance measurements?

Thanks for this good question. Tillage was usually at most 15 cm after 2004, except in 2010 (25 cm) and 2012 (40 cm). However, soil decompaction was not verified by a direct technique, but was clearly shown *by observed changes in soil bulk density at the surface (see **Figure R6,** Figure 4 in revised MS, L480, P16, in answer to RC1).*

We propose to insert the following paragraph concerning soil management in the Materials and Methods section after Figure 1 (L169-173, P5):

"In 2004, as part of the implementation of reduced tillage and the crop rotation system, the soil was scarified to a depth of 50 cm to reduce compaction. Since then, most tillage operations have been restricted to the superficial layer (0–15 cm), using a stubble cultivator or clod crusher. Two additional scarification events were carried out: one in 2010 (to a depth of 25 cm) and another in 2012 (to a depth of 40 cm). Additionally, the soil is disturbed to a depth of 5 or 10 cm during seeding operations."

L345: How are the stocks calculated in this figure? ESM?

This question was also raised by Reviewer 1. Regarding the harmonisation 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 for 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 AARC1-3 and Figures associated in response to RC1).

L386: Which soil depth is used? 0-60cm?

The maximum soil depth in our study was 60 cm in 2005 and 100 cm in 2019. In the revised manuscript, we divided the soil profile into three aggregated layers: 0–5 cm, 5–30 cm, and 30–60 cm, and we removed the 60-100 cm layer when aggregating some soil variables and computing FD-based SOC stocks. Bulk density (BD), rock fragments, and carbon content were aggregated using a thickness-weighted mean, while SOC stocks and soil mass were calculated by summing values across layers. For the ESM calculation, we did not exclude the 60–100 cm layer from the 2019 data. Since the ESM approach adjusts both SOC stocks and layer depth limits, including the last layer is essential if the reference soil mass is not reached within the top 60 cm. The Figure R9 below (Figure 2 in revised MS) was also included at P8 to clarify the two sampling schemes.

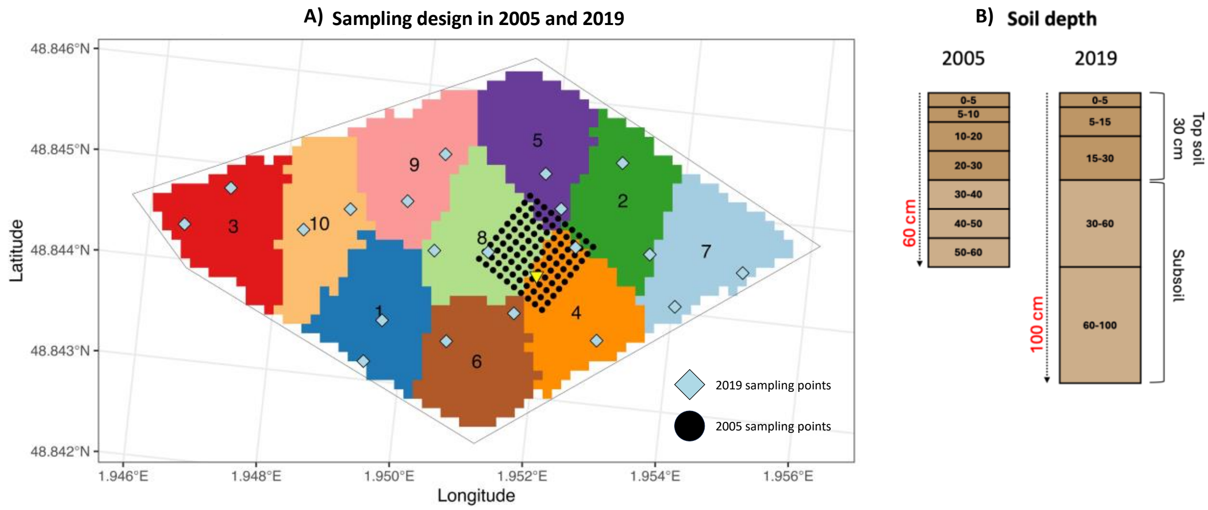

***Figure R9 (Figure 2 in revised MS, L215, P8)**. A) Map of the study area showing the spatial distribution of sampling zones and soil core depth segmentation. Soil sampling was conducted at two times: in 2005 (black circles) and 2019 (blue diamonds). For the 2019 sampling, the field was stratified into 10 strata (coloured polygons, labelled 1–10), and the sampling points were randomly located within each stratum. The 2005 sampling followed a grid-based sampling design partially covering strata 2, 4, 5 and 8, and mostly concentrated in strata 8 and 4. B) Segmentation of soil cores into depth intervals for two different sampling protocols: 60 cm cores (six layers: 0–5, 5–10, 10–20, 20–30, 30–40, and 50–60 cm) and 100 cm cores (five layers: 0–5, 5–15, 15–30, 30–60, and 60–100 cm). Latitude and longitude are shown in WGS 84 coordinates.*

L387: kg ha-1 is preferred over g m-2 in my opinion

We have harmonised the units in kg C m-2, with additional presentation in Mg C ha-1 in parentheses.

Answer to Fabien Ferchaud comments

***CC1**: 'Comment on egusphere-2025-592', Fabien FERCHAUD, 28 Apr 2025*

In the following, we have numbered our answers as AACC1-n, where n is the answer number (Author Answers to CC1). This allows referring to our answers in this document as well as in the answers to the other reviewers and vice versa.

This is an interesting paper that aims to compare SOC stock measurements, flux balance approach and modelling.
However, it seems quite surprising to use fixed depth (FD) calculations in your results and comparison of measured (carbon stocks, flux balance) vs simulated data (AMG), while it is stated in the introduction that "comparing SOC stocks on the same soil mass per unit area was recognised as a better practice than the fixed depth approach" (line 96). It is shown that the Equivalent Soil Mass (ESM) gave similar results than FD for the whole soil profile (~0-60 cm), but this is not true for the upper layers, as the bulk density strongly decreased in 0-5 and 5-10 cm (and increased in 30-60 cm). Furthermore, AMG simulations should not be compared to SOC stock calculated at fixed depth when bulk density changes with time. Indeed, such a simple SOC model does not take into account the change in bulk density, but consider a constant soil mass.
It would be therefore recommended to keep the FD calculations as supplementary results and mainly focus on ESM results. It would also help for the comparison with the literature (published results are often at ESM).
***AACC1-1.** We fully agree with Fabien Ferchaud's comment that ESM is the reference method, as he noticed we wrote this textually in the manuscript. We think we were rather clumsy in the way we presented the comparison between the FD and the ESM method. We, however, think it is a valuable result to show that FD may be a robust estimation as long as the sampling depth is large enough, and we propose to keep a figure in the supplementary material and in a Table in the main manuscript to show this and also show that the FD does not work for shallower depth (30 cm). This would promote sampling to > 60 cm depth.*

The initial choice of keeping the fixed depth method for comparison with the simulations could maybe be due to the fact that a lower decrease of SOC stocks (0-30 cm) was obtained from the ESM method calculation (decrease from 7.60 to 7.16 with ESM vs 8.25 to 7.28 with FD) and did not align completely well with flux balance and simulations with AMG model.
***AACC1-2.** This choice was made by a different reasoning: we observed that the soil stock changes were equal in the 30 cm and 60 cm layer as estimated with the FD method (all changes were attributed to the 0-30 cm layer in that case). Moreover, the FD and ESM methods in the 0-60 cm layer were equal. Hence, the 30 cm FD stock evolution was representative of the whole stock change over the entire soil profile. This was the reason for our choice, which was not indeed sufficiently explicated. In the revised MS we now aligned compare AMG to the ESM method in the 30 cm layer.*

However, for the ~0-30 cm calculation, it could maybe be tested if a slightly higher reference soil mass could be used to calculate the stocks at ESM or maybe consider a slightly deeper upper soil horizon (e.g. ~0-35 cm) to capture the main decrease in SOC stocks and for comparisons and modelling.
For example, the reference soil mass for the ~0-30 cm layer could be obtained from the mean soil mass measured in 2005: this would be more representative of the initial situation.

AACC2-3. *We fully agree with this analysis, which is the one that led to our choice of the 30 cm FD ~ 60 cm FD ~ 60 cm ESM. We have however stick now to the stricter 0-30 cm depth for the comparison with AMG and discuss this in the discussion section. (**Figure R4,** Figure 7 in revised MS, L565, P20).*

We further propose to add some sensitivity tests of the AMG model 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 those values of 0.63 and 0.68 as independent estimates on a soil nearby were taken from Kanari et al. (2022).

*This leads to **Figure R4**, which shows that (1) AMG in its standard setup overestimate the ESM soil sock change by a factor of 1.5, (2) the incorporation of 100% of the crop residues would lead to a 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 of 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.*

Concerning the comparison of SOC stocks dynamics, their simulations and the flux balance, it appears really frustrating to only have 2006 to 2010 data for the flux balance (data that are now 15 years old and published in 2011). Indeed, the comparison appears relatively limited because only restricted to the 5 first years. Data from recent years or at least until 2019 could be added to strengthen the comparison and improve the significance of this study.
AACC1-4. This very sound comment was also made by Reviewer 2 (see answer ***AARC2-1*** *reproduced here for quality*). *The manuscript does not include a full carbon balance because this would imply computing the net ecosystem exchange (NEE) over the whole period 2005-2019. While in principle we have all the data to do such a carbon balance, we should stress that in 2018, there was a serious acquisition problem that invalidated almost the whole year of measurements, which we would like to gap fill with a crop model. Similarly, Kindler et al. (2011) have shown that lixiviation of carbon was an important term in the carbon balance from 2006 to 2010 in the FR-Gri field, which would require a modelling effort to evaluate the losses from 2010 to 2019. Finally, the spatial variability of the soil properties shown in this study also leads to spatial variability in the biomass production, as shown by Loubet et al. (2011), which will require additional footprint-based filtering on the NEE and biomass. All this work, in addition to an advanced uncertainty analysis on the carbon balance, is required to provide a sound comparison between the carbon balance and the carbon stock evolution. Unfortunately, this requires additional manpower, which leads to postponing the full carbon balance analysis for later publication, although we fully approve the interest of such a comparison.*

The rationale on the reduced area should be further developed to better justify the choices made (notably by being more explicit on “only the strata having soil characteristics similar to the 2005 sampling area” line 185). As it stands, it is not easy to understand well while some 2019 sampling points close to the footprint from the Eddy covariance mat and to the 2005 campaigns were not integrated in the reduced area (e.g. sampling point in the grey area #9) while others that are quite far were considered.

AACC1-4. *This comment was also given by Reviewers 1 and 2. See detailed answers* ***AARC1-4*** *and* ***AARC2-2****.*

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://doi.org/10.32614/RJ-2016-014.

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.

Kindler, R., Siemens, J., Kaiser, K., Walmsley, D. C., Bernhofer, C., Buchmann, N., Cellier, P., Eugster, W., Gleixner, G., Grunwald, T., Heim, A., Ibrom, A., Jones, S. K., Jones, M., Klumpp, K., Kutsch, W., Larsen, K. S., Lehuger, S., Loubet, B., McKenzie, R., Moors, E., Osborne, B., Pilegaard, K., Rebmann, C., Saunders, M., Schmidt, M. W. I., Schrumpf, M., Seyfferth, J., Skiba, U., Soussana, J. F., Sutton, M. A., Tefs, C., Vowinckel, B., Zeeman, M. J., and Kaupenjohann, M.: Dissolved carbon leaching from soil is a crucial component of the net ecosystem carbon balance, Global Change Biology, 17, 1167-1185, 10.1111/j.1365-2486.2010.02282.x, 2011.

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.