Dear Reviewer 3:

Thank you very much for your time involved in reviewing the manuscript and your very useful comments. This feedback greatly improved the quality of our paper and made this article more rigorous.

Comment: The manuscript addresses an important topic: the effect of straw return on soil organic matter. However, the introduction is rather short and does not fully explain the spatio-temporal modelling of soil properties.

Response: Thank you for your comment. We appreciate your comments and agree that the introduction could be more detailed in explaining the spatio-temporal modelling of soil properties.

We have revised the introduction to provide more information on the spatio-temporal modelling of soil properties, including a brief explanation of the methods used to model soil properties:"Based on the soil-forming theory, digital soil mapping (DSM) uses statistical and geospatial techniques to model the relationship between soil properties and environmental covariates at a high spatial resolution. By analyzing the relationships between soil properties and environmental factors, DSM models can be developed to predict the soil properties of areas where no soil data exist. Therefore, it offers a promising solution for predicting soil properties with high precision and tremendous speed (Hengl et al., 2015; Dou et al., 2019; Liang et al., 2019; Schulze and Schütte, 2020). Moreover, DSM can also incorporate the temporal component in soil property mapping by taking time as an index and comparing soil maps at two moments to identify changes in soil properties over time. This is particularly useful in understanding the impact of land use and management practices on soil properties and identifying areas where remediation may be necessary." We have also included additional information on the mechanism of straw returning affecting SOM： "Straw return is beneficial for retaining soil moisture and preventing soil wind erosion, especially in arid and semi-arid regions. In addition, the decomposition process of straw promotes the activity of microorganisms and is conducive to SOM accumulation (Chang et al., 2014; Lu et al., 2009; Wang et al., 2015). Conversely, previous scholars have reported that the influence of straw return on SOM accumulation is non-significant (Pittelkow et al., 2015; Poeplau et al., 2015; Powlson et al., 2011). This may be because adding organic matter to the soil has no effect on its chemical, chemical, and biological properties, (Sosulski et al., 2011) or this practice may contribute to the SOM mineralization process and thus reduce SOM (Šimanský et al., 2019)."

Comment: Furthermore, the materials and methods lack precision and the protocols deviate from current practices. Why was the soil sieved at 0.25 mm, while the fine earth is generally defined as < 2 mm (line 78). There is also a confusion between SOM and SOC. The wet oxidation protocol should be explained more carefully, because these analyses determine the SOC and NOT the SOM content.

Response: Thank you for your comment. I understand your concerns regarding the precision of the materials and methods, as well as the confusion between SOM and SOC. Regarding the soil sieving, I agree that there may be some confusion in the manuscript. After air drying and grinding, the soil samples were thoroughly mixed and passed through a 2 mm mesh. The samples were ground and sieved to separate a particle size fraction (0.25 mm) to determine SOC concentration. I reviewed the materials and methods section to ensure that this was clearly explained.

As for the confusion between SOM and SOC, I understand that the wet oxidation method is used to determine SOC content. Specifically, it is the external heating potassium dichromate volumetric method, which can then be multiplied by a conversion factor of 1.724 to obtain the SOM amount. Because some studies have written that SOM was measured using external heating potassium dichromate volumetric method (Lu et al., 2022; Ma et al., 2022), we continued to write it this way. We have revised this in Section 2.2:

"Sun exposure, acid, alkali, and dust pollution were strictly prohibited. After air drying and grinding, the soil samples were thoroughly mixed and passed through a 2 mm mesh. The samples were ground and sieved to separate a particle size fraction (0.25 mm) to determine SOC concentration with the external heating potassium dichromate volumetric method, which can then be multiplied by a conversion factor of 1.724 to obtain the SOM amount (Liu et al., 1996)."

Lu, M. Y., Liu, Y., & Liu, G. J. (2022). Precise prediction of soil organic matter in soils planted with a variety of crops through hybrid methods. Computers and Electronics in Agriculture, 200, 107246.
Ma, R., Hu, F., Xu, C., Liu, J., & Zhao, S. (2022). Response of soil aggregate stability and splash erosion to different breakdown mechanisms along natural vegetation restoration. Catena, 208, 105775.

Comment: the effect of straw return (section 3.5) is difficult to evaluate. First of all, the term is not clearly defined. According to the materials and methods section it is the residue cover and not the percentage of the residue produced by the crop.

Response: Thank you very much for your feedback. I agree that the term "straw return" in our study is the crop residue cover. We have clearly defined it in the section 2.3. "Crop residue cover refers to the ratio of the vertical projected area of crop residue in a field per unit area to the total surface area of this unit area, with value ranging from 0 to 1." In addition, we have modified this term in the whole text, including the title, figures, and tables to ensure readers understand the intended meaning. At the same time, we have added more details about the CRC in Section 2.3, including how to measure straw coverage: "Liu et al. (2020) provided a crop residue coverage map at a 10 m resolution in 2018 by combining radar indices and optical remote sensing indices. Crop residue cover refers to the ratio of the vertical projected area of crop residue in a field per unit area to the total surface area of this unit area, with value ranging from 0 to 1. Firstly, the study divided the study area into a sandy soil area and a clay soil area to reduce the influence of soil properties on radar echo and spectral reflectance. Six radar indices and five optical remote sensing indices were then calculated from a Sentinel-1 SAR image and a Sentinel-2 optical image. Finally, the optimal subset regression based on these indices and 55 observations collected from November 1, 2018 to November 11, 2018 was used to estimate the crop residue cover. The 55 observations were measured using the Line-Transect method (Wollenhaupt and Pingry, 1991). The best model shows high accuracy."

Comment: Second, the statistical analysis of the effect (see Fig. 6) is poorly explained. If Fig. 6

displays the SOM content of the pixels in each class, these observations are not independent and therefore cannot be pair wise compared using a statistical test (ANOVA or t test). This remark also holds for figures 7-10.

Response: Thank you very much for your comment. To improve clarity and ensure consistency, we have revised figures 6, 7, 9, 10 to match the style of Fig. 8. Because some data did not meet the assumptions of normal distribution, we used the Wilcoxon signed-rank test (a non-parametric statistical hypothesis test) to compare the means of two groups. To satisfy the assumption of data independence, we randomly selected 1/200 of the total number of pixels using the *sample* function in R 4.0.2. Using the impact of straw incorporation on soil organic matter as an example, we categorized the crop residue cover into four levels (1: 0–0.15; 2: 0.15–0.30; 3: 0.30–0.60; 4: 0.60–1.00). We ordered these groups by the mean SOM values of each level and then conducted pairwise significant difference analyses using the "wilcox.test" method in the ggviolin function with the stat_compare_means setting for (1,2), (2,3), and (3,4) levels. The code is as follows:

```
my_comparisons <- list( c("1", "2"), c(2","3"),c("3","4"))
f%>%ggviolin(x="CRC",y="X2018.2006",fill="CRC",add="boxplot",add.params=list(fill="white"
)) +stat_compare_means(comparisons = my_comparisons,method="wilcox.test", label = "p.signif")
->p
```

We have added this information in section 2.4.3.

Lines 35-36 Would not it be better to express the functions of Jenny and SCORPAN with the dependent variable ‘soil property’ rather than ‘soil’. After all, ‘soil’ is a broad concept that cannot be quantified and you mention ‘soil properties ‘ in line 38.

Thank you for your comment. We have revised it as you suggested.

Lines 40-43 You have explained (not in great detail) the role of DSM for quantifying the spatial variation in soil properties. Here you also include the temporal component. This has to be explained in more detail.

Thank you for your comment. We appreciate your comments and agree that the introduction could be more detailed in explaining the spatio-temporal modelling of soil properties.

We have revised the introduction to provide more information on the spatio-temporal modelling of soil properties, including a brief explanation of the methods used to model soil properties: “Based on the soil-forming theory, digital soil mapping (DSM) uses statistical and geospatial techniques to model the relationship between soil properties and environmental covariates at a high spatial resolution. By analyzing the relationships between soil properties and environmental factors, DSM models can be developed to predict the soil properties of areas where no soil data exist. Therefore, it offers a promising solution for predicting soil properties with high precision and tremendous speed (Hengl et al., 2015; Dou et al., 2019; Liang et al., 2019; Schulze and Schütte, 2020). Moreover, DSM can also incorporate the temporal component in soil property mapping by taking time as an index and comparing soil maps at two moments to identify changes in soil properties over time. This is particularly useful in understanding the impact of land use and management practices on soil properties and identifying areas where remediation may be necessary.”

Lines 74 and 75 The sampling design and use of legacy data is not discussed, so it is difficult to interpret their effects on ‘prediction error’.

Thank you for your comment. We have provided more details on the sampling design: “**By taking into account the sample sites in the second national soil survey, local landform, and soil types, a total of 300 sampling sites in 2006 were selected. Except for considering these factors, grid sampling was combined to select 319 sampling sites in 2018.** The soil samples were collected on the surface (0–20 cm) from early October to mid-November in each year (from the harvest to the freezing). The corresponding longitude and latitude were also documented. The prediction error caused by the differences in sampling designs for the years 2006 and 2018 was not considered to make full use of legacy soil data” to improve the readers' understanding of the impact on prediction error. We use legacy data from both 2006 and 2018. The type and source of the legacy data are the same for both years.

Line 82 Please provide the reference for the ‘Resource and environment data cloud platform’

Thank you for your comment. We have provided a website for the ‘Resource and environment data cloud platform’ : “A 30 m resolution digital elevation model (DEM) was derived from the Resource and Environment Data Could Platform **(http://www.resdc.cn/)**.”

Line 87 Although spectral indices such as NDVBI and EVI are well-known, this is much less the case for the NDTI and STI. Please specify these indices.

Thank you for your comment. We have specified all the vegetation indices and described their calculation formulas in this paper.

Line 89 If I understand correctly, you use the CRC of 2018 for all fields between 2007 and 2018? This is a strong assumption as it does not take differences in crop performance or crop type into account. Please describe more clearly that the CRC is not used as a co-variate, but you compare two datasets (with and without residue).

Thank you for your comment. We have described more clearly the use of the CRC in Section 2.3: “Because Lishu County has implemented the straw return policy since 2007, the crop residue cover in 2006 can be regarded as 0. The difference between the crop residue cover in 2018 and 2006 (CRC) was used as one of the variables for modelling SOM in 2018 and to evaluate the effects of long-term straw return on SOM variation during 2006–2018. This study assumed that crop performance or crop types were the same except for CRC. The CRC was used to represent the straw return.”

Section 2.4.2 The technique of geographical detector is not as widely known as e.g. random forest.

The principles will have to explained in a couple of sentences.

Thank you for your comment. We have added a couple of sentences to explain the principles of GE: "GE (Wang and Xu, 2017) is a statistical method used in geographical analysis to identify the factors that contribute to spatial patterns. It is based on the idea that the variation in a dependent variable across a geographical area can be explained by a set of independent variables and their interactions."

Table 3 Please include a column with the number of samples.

Thank you for your comment. We have included a column with the number of samples.

Section 3.2 Please explain the abbreviations e.g. '(all-Y)'. As it stands the reader has to look them up in the figure caption. It is not clear either whether these statistics apply to the calibration or the validation data set.

Thank you for your comment. We have included a more comprehensive explanation of the abbreviations in the text of the manuscript: "Table 4 shows that **the validation results considering all the variables as predictors** (RF-all) (CC = 0.59, RMSE = 4.54 g kg^{-1}, taking 2006 as an example) (Fig. 2) performed better than **that considering the environment variables without latitude as predictors (RF- (all-Y))** did." The statistics presented in Section 3.2 refer to the performance of the model on the validation dataset. We have also added this information to make it clearer.

Line 180 and further on. Please define what you mean by 'straw return content'. As far as I can see it is the straw cover and not necessarily the percentage of residues produced.

Thank you very much for your feedback. I agree that the term "straw return" in our study is the crop residue cover. We have clearly defined it in Section 2.3. "Crop residue cover refers to the ratio of the vertical projected area of crop residue in a field per unit area to the total surface area of this unit area, with value ranging from 0 to 1.” In addition, we have modified this term in the whole text, including the title, figures, and tables, to ensure readers understand the intended meaning.

Figure 6 Please explain how the significance was calculated and what it means. Why did you not try to fit a regression and analyse the significance of the regression?

Using the impact of straw incorporation on soil organic matter as an example, we categorized the crop residue cover into four levels (1: 0–0.15; 2: 0.15–0.30; 3: 0.30–0.60; 4: 0.60–1.00). We ordered these groups by mean SOM values of each level and then conducted pairwise significant difference analyses using the "wilcox.test" method in the ggviolin function with the stat_compare_means setting for (1,2), (2,3), and (3,4) levels. The code is as follows:

```
my_comparisons <- list( c("1", "2"), c(2","3"),c("3","4"))
f%>%ggviolin(x="CRC",y="X2018.2006",fill="CRC",add="boxplot",add.params=list(fill="white"
)) +stat_compare_means(comparisons = my_comparisons,method="wilcox.test", label = "p.signif")
->p
```

We have added this information in Section 2.4.3.

We did not conduct regression analysis because each level of crop residue cover has been given a specific meaning according to the table as below, which will be utilized in future studies.

Crop residue cover	**Cover**
Conventional tillage	0-0.15
Low residue tillage	0.15-0.30
Conservation tillage	0.30-0.60
High residue tillage	0.60-1.00

(Chesapeake Bay Program: Annapolis, MD, USA, 2016)

Thomason W, Duiker S, Ganoe K, et al. Conservation tillage practices for use in Phase 6.0 of the Chesapeake Bay Program watershed model[R]. CBP/TRS-308-16. Chesapeake Bay Program.< https://www. chesapeakebay. net/documents/CT_6. 0_Conservation_Tillage_ EP_Revised_Full_Report_12-14-16.2 _FINAL_NEW_TEMPLATE. pdf, 2016.