Responses to Reviewer Comments

Dear reviewer:

Thank you very much for your time involved in reviewing the manuscript, and for providing helpful and specific feedback for how to improve this work. Below we have responded to all of your comments and have indicated how we will change the manuscript as a result of these suggestions.

- **Comment 1**

As outlined in the title, the main novelty of this manuscript would be the incorporation of straw return amount into the Digital Soil Mapping framework. But descriptions on how the straw return amount (the CRC factor) is quantified and mapped are lacking. The authors did give a reference (Huang et al., 2020; Liu et al. 2020 in Table 1 is not listed in the References) to the data source (I did not manage to get access to the paper), but a more detailed explanation should still be added into the paper. For example, what was the approach used for CRC mapping? What was the sampling design and sample size for CRC field quantification, at what time of year? Without this information, it is difficult to assess the data quality and thus the overall modeling approach.

- **Response 1**

Thank you very much for the comment.

1) The title of the paper is "Quantification of the effects of long-term straw return on soil organic matter spatiotemporal variation: A case study in typical black soil region". As outlined in the title, the novelty of the paper is the **quantification of the effects of long-term straw return on soil organic matter (SOM) spatiotemporal variation on a regional scale**, but not "the incorporation of straw return amount into the Digital Soil Mapping framework". **As a result, the confounding issues raised by the reviewer don't apply.** We appreciate the comment because our ambiguous writing may have led to the misunderstanding. We guess that the column in the Abstract is ambiguous: "due to lacking accurate spatial distribution of straw return, few studies took straw return as a variable to carry out rigorous research on the impact of straw return on SOM variation on a regional scale", so we corrected the sentences as "due to lacking accurate spatial distribution of straw return, few studies carried out rigorous research on the impact of long-term straw return on SOM spatiotemporal variation on a regional scale" to make the expression clearer and more accurate. Of course, the above is our guess. If there is anything else with a vague expression, please let us know so that we can continue to revise and present the best work to readers.

2) Straw returning is the focus of this paper all the time, so the straw return amount (the CRC factor) should be given a detailed description, as you suggested. We have given a detailed description in Section 2.3 along with the references (Liu, Z., Liu, Z., Wan, W., Huang, J., Wang, J., and Zheng, M.: Estimation of maize residue cover on the basis of SAR and optical remote sensing image, National Remote Sensing Bulletin, 25(06), 1308-1323, 2021. DOI：10.11834/jrs. 20210053):

"*Liu et al. (2020) provided a crop residue coverage at a 10 m resolution in 2018 by combining radar indices and optical remote sensing indices. 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 1 November 2018 to 10 November 2018 was used to estimate the CRC. The best model shows high accuracy. Lishu County has implemented the straw return policy since 2007. Therefore, the amount of straw return in 2006 can be regarded as 0. The straw returning amount in 2018 minus that in 2016 is numerically equivalent to the straw returning amount in 2018 (CRC), which is used as one of the variables for modelling SOM in 2018 and evaluating the effects of long-term straw return on SOM variation during 2006–2018.*"

3) At the same time, we would like to thank the reviewer again for this comment, which gave us some inspiration. We decided to conduct additional experiments in this study to evaluate the incorporation of straw return amount into the Digital Soil Mapping framework (please see Response 3).

- **Comment 2 (1)**

2. I think this is problematic because the effect of organic inputs on SOM dynamics is a mid-to-long term process, so linking spatial variability of SOM to CRC at one timepoint seems a bit farfetched to me. I would suggest the authors to look into the cumulative effect of straw return on SOM over a longer time period.

- **Response 2 (1)**

Thank you very much for the comment.

In this study, the CRC data is mainly used in two aspects: 1) to assess the impact of long-term straw return on soil organic matter variation; 2) to evaluate the importance of CRC in the spatial modeling of SOM in 2018. In both aspects, we used the value of **the straw returning amount in 2018 minus that in 2016** (Figure 5 (a)), which is the cumulative straw return amount. We appreciate the comment because although we provided the information in lines 88-89, it was not clearly stated: "Lishu County has implemented the straw return policy since 2007. Therefore, the amount of straw return in 2006 can be regarded as 0". So, we added more information in this section: "Lishu County has implemented the straw return policy since 2007. Therefore, the amount of straw return in 2006 can be regarded as 0. The straw returning amount in 2018 minus that in 2016 is numerically equivalent to the straw returning amount in 2018, which is used as one of the variables for modelling SOM in 2018 and evaluating the effects of long-term straw return on SOM variation." In addition, we've clarified our wording in the Notes of Table 1, Figure 3, Section 3.2 (291), and Section 3.3 (337).

- **Comment 2 (2)**

As far as I understood, the authors only used the CRC data in 2018 to evaluate the effect of straw return on SOM content in the same year.

- **Response 2 (2)**

Thank you very much for the useful comment. Your suggestion is valid. At present, there is no efficient method to investigate the cumulative effect of straw return on SOM spatiotemporal variation on a regional scale. As you pointed out, this study should utilize the cumulative straw returning amount several years ago (such as, 2015-2006). But we can only collect the spatial distribution data of straw returning in 2018. Based on the data available, this is the most efficient approach we can take.

- **Comment 3**

3. For the RF model, the authors included the CRC factor for the year 2018, but the relative importance of CRC appears to be low. This again questions the validity of the approach and the CRC data. At the very least, the authors should compare the predictive performances of models with and without CRC, so as to demonstrate whether incorporating CRC improves the model performance.

- **Response 3**

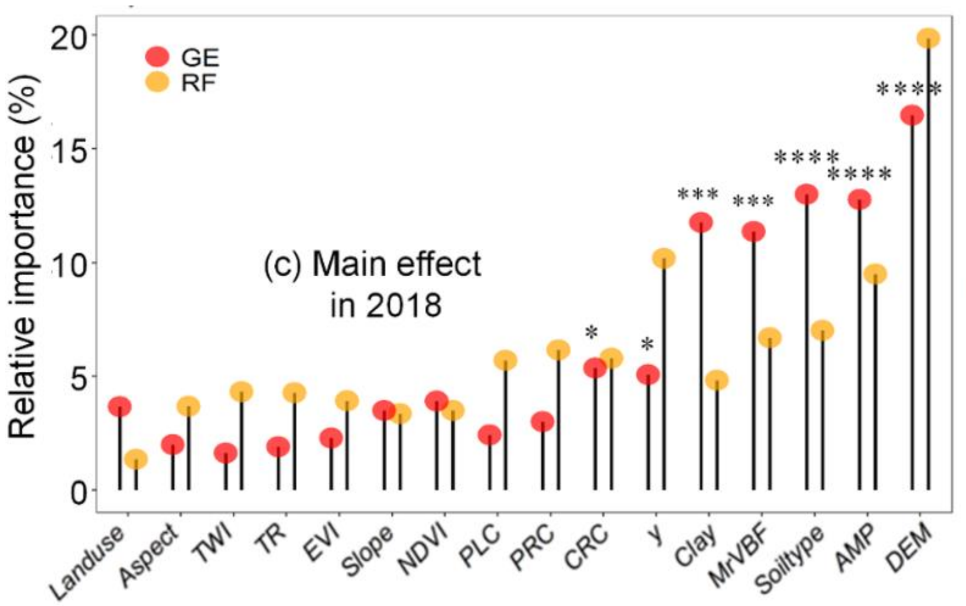
Thank you for the comment.

In order to avoid misunderstanding, we must emphasize that, as mentioned above, the focus of this paper is to evaluate the impact of long-term straw returning on soil organic matter change. The spatial modeling in 2018 was calibrated to obtain the spatial distribution of organic matter only, and it is not the focus. Although the relative importance of CRC in the RF nodel is not the focus of this article, as mentioned in comment 1, we thank you for the comment, because it provides us with new ideas, and this article can explore the significance of straw returning in the DSM. So, we will add the content about this idea according to your suggestion. As you suggested, we added a set of experiments. We established RF models with CRC and without CRC and found that, after removing CRC, CC was reduced by 14.9%, and RMSE was increased by 0.23 g kg^{-1} (Table 4). We have added the result and the related discussion in section 3.2: "*In areas dominated by agriculture, agricultural management practices (irrigation, tillage practices, farming systems, and residual management) increased the spatial heterogeneity of SOM, so the spatial distribution of SOM cannot be accurately reflected using individual vegetation indexes or topographic factors. However, few studies integrate these agricultural activities to explain the SOM spatial distribution due to the challenge of collecting them based on remote sensing technology. This study developed RF models with and without CRC. After removing CRC, CC was reduced by 14.9%, and RMSE was increased by 0.23 g kg^{-1} (Table 4). The results demonstrated that the RF model with the CRC factor achieved a higher accuracy than the model using common environment variables alone. Therefore, further research should consider the* *importance of geographic coordinates and long-term straw return in DSM.*"

Table 1 The performance of the random forest model.

Methods		2006		2018	
		CC	RMSE (g kg^{-1})	CC	RMSE (g kg^{-1})
RF-all	Calibration	0.55	5.04	0.44	6.99
	Validation	0.59	4.54	0.54	5.38
RF- (all-Y)	Calibration	0.50	5.22	0.38	7.18
	Validation	0.55	4.63	0.47	5.55
RF-(all-CRC)	Calibration	\	\	0.43	7.03
	Validation	\	\	0.47	5.61

To ensure the rigor of the research, RF and GE methods were used to analyze the importance factors, as shown in the figure, CRC ranked the seventh in this study. At the same time, it shows the CRC was significant using the significance test with GE. Straw returning is not the main cause of spatial variation of SOM (of course, topography, climate and soil type are the main contributions), but the CRC did contribute significantly in the spatial distribution of SOM.



- **Comment 4**

4. The predictive performances (CC and RMSE) for the 2006 and 2018 RF models were not exceptional (Figure 2) – the accuracy was actually worse in 2018 after the addition of CRC. This means that the predicted SOM values are associated with large prediction errors and uncertainties, thus weakening the obtained results from direct comparisons for the purpose of SOM monitoring.

- **Response 4**

Thank you for the comment.

As shown in the additional experiments (Table 1), when CRC was removed from the RF model in 2018, CC was reduced by 14.9%, and RMSE was increased by 0.23 g kg^{-1}. So the accuracy was higher in 2018 after the addition of CRC. For the comment "the accuracy was actually worse in 2018 after the addition of CRC", we're a little unclear what the reference is.

The modeling accuracy is the study's limitation. In general, the modeling accuracy of SOM or SOC in other similar studies (county-scale, and the study area characterized with plat farmland area) is low (Table 2), which may be because topographic and remote sensing factors are too homogeneous to effectively extract SOM information. The improvement of modeling accuracy for SOM or SOC in county-scale with similar characteristics is indeed a challenge. However, compared with similar studies on flat agricultural areas, the model of our study performed better. This content was added in Section 3.2: "*The accuracy of the study's model was not high, which is the study's limitation. By reviewing the SOM or SOC modelling in flat farmland areas, we found that the models of these similar studies all show poor performance (Table 5). This may be because common environmental variables, such as topographic and remote sensing factors are too homogeneous to effectively extract SOM information (Zeng et al., 2017). Compared with these studies in flat farmland areas, our prediction accuracy is comparable or even better (Table 5).*"

Table 2 Models' performance in SOM prediction in plain farmland areas.

Study area	Model with common environmental variables	Model performance	References	Characteristics
Cultivated land of Xuanzhou city and Langxi County	Random Forest	R^2=0.34	Yang et al., 2020	A typical plain rice production area
Chahe Town	Ordinary Kriging; Regression Kriging	CC=0.15-0.24	Wu et al., 2021	A typical plain farmland area
Agricultural soils in the north-eastern Iberian Peninsula	General Least Squares	R^2=0.20-0.27	Funes et al., 2019	Agricultural soils
Jianghan plain	Stepwise Regression; Partial Least Squares Regression; Extreme learning machine	R^2=0.14-0.53	Guo et al., 2021	Agricultural lands in low-relief areas
Miandoab county, West Azerbaijan, northern Iran	Random Forest; Cubist; Conditional Inference Forest; Conditional Inference Trees; Extreme Gradient Boosting; Classification and Regression Trees	CC=0.34-0.44	Goydaragh et al., 2021	The elevation varies from 1292 to 1342 m and the main land use is agriculture

Funes, I., Savé, R., Rovira, P., Molowny-Horas, R., Alcañiz, J. M., Ascaso, E., Herms, I., Herrero, C., Boixadera, J., and Vayreda, J.: Agricultural soil organic carbon stocks in the north-eastern Iberian Peninsula: Drivers and spatial variability, Science of the Total Environment, 668, 283-294, 2019.
Goydaragh, M. G., Taghizadeh-Mehrjardi, R., Jafarzadeh, A. A., Triantafilis, J., and Lado, M.: Using environmental variables and Fourier Transform Infrared Spectroscopy to predict soil organic carbon, Catena, 202, 105280, 2021.
Guo, L., Fu, P., Shi, T., Chen, Y., Zeng, C., Zhang, H., and Wang, S.: Exploring influence factors in mapping soil organic carbon on low-relief agricultural lands using time series of remote sensing data, Soil and Tillage Research, 210, 104982, 2021.
Wu, Z., Liu, Y., Han, Y., Zhou, J., Liu, J., and Wu, J.: Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data, Science of The Total Environment, 754, 142120, 2021.
Yang, L., He, X., Shen, F., Zhou, C., Zhu, A. X., Gao, B., Chen, Z., and Li, M.: Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data, Soil and Tillage Research, 196, 104465, 2020.

Specific comments:

- **Comment 5**

1. In the Abstract, one should briefly mention the size and characteristics of the study area.

- **Response 5**

Thank you for your useful recommendation. We have added these information in the Abstract: "*This study was carried out across an approximately 3000 km^2 area in Lishu County, Northeast China, a typical agricultural plain.*"

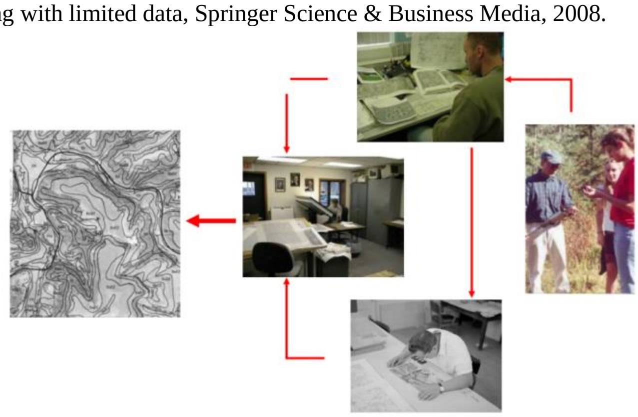
- **Comment 6**

2. Line 35-40, what is the difference between conventional mapping and DSM? Doesn't DSM also comprise the procedures you outlined in the first sentence of the paragraph?

- **Response 6**

Thank you for your comment. Comparing digital soil mapping to traditional soil mapping, the most notable difference is that digital soil mapping makes use of quantitative inference models to provide predictions of soil properties in a geographical database (raster) (https://www.nrcs.usda.gov/resources/data-and-reports/digital-soil-mapping-dsm). We originally intended to express the figure below, but the expression has a problem. We corrected the sentence "*Conventional mapping involves data collection, field investigation, interpretation, field inspection, calibration, and mapping.*" as: "*Conventional mapping involves laboriously constructing maps by planetable and alidade (Ahrens, 2008).*".

Ahrens, R. J.: Digital soil mapping with limited data, Springer Science & Business Media, 2008.



- **Comment 7**

3. Line 70-75, there is no mention of the sampling designs in 2006 and 2018. Also, what are the sample sizes?

- **Response 7**

Thank you for your useful recommendation. We have added these information in Section 2.3: "*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 and 319 sampling sites in 2018 were selected. 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)*"

- **Comment 8**

4. Line 85, I suggest the authors to specify how NDVI and EVI were calculated? Annual mean or based on images from a specific month?

- **Response 8**

Thank you for your useful recommendation. We have added these information in Section 2.3: "*The average reflectance of three image bands (B1 (Blue), B3 (Red) and B4 (Near-infrared)) of Landsat 5 SR and Landsat 8 SR products spanning May to September were processed to calculate the NDVI by (B4 – B3) / (B4 + B3) and EVI by 2.5 × (B4 – B3) / (B4 + 6 × B3 – 7.5 × B1 × 1). The band calculation, and image clipping were conducted in Google Earth Engine (GEE), and the images with less than 6% of cloud coverages were selected.*"

- **Comment 9**

5. The authors should specify the statistical method used for significance tests for all the boxplots. Otherwise, it is difficult to evaluate the appropriateness of the comparisons on changes in SOM with varying straw return amount.

- **Response 9**

Thank you for your commet. The statistical method used for significance tests for all the boxplots is wilcoxon test , which is realized in "stat_compare_means" in R 4.0.2. The information has been added in Section 2.4.3: "*The statistical method used for significance tests for all the boxplots is Wilcoxon test, which is realized in "stat_compare_means". The RF and GE models were implemented in the "caret" and "GD" libraries, respectively.*"

- **Comment 10**

6. The entire Results and discussion section was more focused on the interpretation of the results. An in-depth discussion on the strengthens and weaknesses of the methodology is missing.

- **Response 10**

Thank you for your useful recommendation. As your suggestion, we have added the information in Section 3.2: "*The accuracy of the study's model was not high, which is the study's limitation. By reviewing the SOM or SOC modelling in flat farmland areas, we found that the models of these similar studies all show a poor performance (Table 5). This may be because common environmental variables, such as topographic and remote sensing factors are too homogeneous to effectively extract SOM information (Zeng et al., 2017). Compared with these studies in flat farmland areas, our prediction accuracy is comparable or even better (Table 5).*"

- **Comment 11**

7. Overall, the writing of the manuscript should be improved.

- **Response 11**

Thank you for your useful recommendation. We have asked colleague who speacks English to carefully check, and we will improve the English writing in the revised manuscript.