Response to reviewer 1

Many thanks for your careful reading of the manuscript and your valuable comments and suggestions, which have led to significant improvement of this paper (expressed in **Acknowledgements**). We took it seriously, and we have carefully revised the paper over the past two weeks. In the following, we would detail the changes we have made on the paper.

Specific comments

1. Comments: *Lines 2: The title is long, you should shortened it.*

Answers: Many thanks for your suggestion. We have revised the title as follows (L2-3):

"Estimating Soil Copper Content for Urban Land in a Megacity Using Piecewise Spectral Pretreatment"

2. Comments: *Line 22: Abstract contain some information about Which prediction model (PLSR), study area, number of soıl samples, and the range of Vis-NIR (350-2500 nm) used in this study.*

Answers: Many thanks for your suggestion. We have revised the abstract as follows (L17-32):

"Heavy mental contamination in urban land is a serious environmental issue for large cities. Visible and near-infrared spectroscopy has been rapidly developed as a new method for estimating copper (Cu) levels, which is one of the heavy metal. Spectral pretreatment is essential for reducing noise and enhancing analysis. In the traditional method, the entire spectrum is uniformly pretreated. However, in reality, the influence of pretreatment on the spectrum may vary depending on the wavelengths. Limited research has been conducted on breaking down the en-tire spectrum into distinct parts for individualized pretreatment, an innovative method called piecewise pretreatment. This study gathered 250 topsoil samples (0-20cm) in Shengzhen City, southwest China, and obtained their vis-NIR spectra (350-2500nm) in the laboratory. This study divided the spectrum into three parts, each processed by six commonly used spectral pretreatments. The number of pretreated parts varied from 1 to 3, resulting in 342 PLSR models being built. Comparison to the traditional method, piecewise pretreatment showed an increase in mean residual predictive deviation (RPD) increased from 1.55 to 1.71 and in the percentage of positive outcomes in ΔRPD increased from 33.33% to 55.56%. Thus, we concluded that piecewise pretreatment generally outperforms the traditional method and aims to choose the most effective pretreatment method for each part to optimize the Cu estimation model."

3. Comments: *Line 24: which is one of the heavy metal.*

Answers: Many thanks for your suggestion. We have revised the sentence as follows (L18-19):

"Visible and near-infrared spectroscopy has been rapidly developed as a new method for estimating copper (Cu) levels, which is one of the heavy metal."

4. Comments: *Line 55: Delete it, because these sentences detailed ınformation about methods. Therefore, delete or romeve to the relevant section.*

Answers: Many thanks for your suggestion. We have removed the method details and revised the paragraph as follows (L51-55):

"Spectral pretreatment is important when developing vis-NIR spectroscopy models to estimate

Cu in soil [14]. Raw spectra can be influenced by instrumental status, experimental conditions, soil particle size, and surface roughness [15-17]. Spectral pretreatment can reduce noise and enhance the accurate of analysis and interpretation [18]. Therefore, the spectral pretreatment method must be carefully considered."

5. Comments: *Line 99: you should add more information about mean elevatian and slope.*
Answers: Many thanks for your suggestion. We have added the information about elevation and slop as follows (L105-107):
"The elevation ranges from 0 to 943.7m with a mean value of 82m, and the average slop is 7° [34]."

6. Comments: *Line 119: How many soil samples?what is the exact number?250 or around 250 ?*
Answers: Many thanks for your suggestion. We have revised the sentence as follows (L123-124):
"At each site, approximately 1.5 kg of topsoil (0–20cm) was collected, resulting in total of 250 samples gathered in November 2016 (Figure 1)."

7. Comments: *Line 121: Did yoo calibrate the ASD FieldSpec before scanning?If yes, you should add detailed information,If not, you should scan all soıl samples considering calibration procuderes!?*
Answers: Many thanks for your suggestion. We calibrated the ASD FieldSpec® 3 portable spectro-radiometer using a Spectralon® panel with 99% reflectance before measurement. We has revised the sentence as follows (L135-137):
"A Spectralon® panel with 99% reflectance was utilized to calibrate the spectrometer before measurement. The scans were performed 10 times, and the results were subsequently averaged [31]."

8. Comments: *Line 126: You should add more information about resolution of spectral range of Vis-NIR.*
Answers: Many thanks for your suggestion. We have added information about spectral resolution. We have revised the sentence as follows (L131-132):
"Spectra were recorded with a sampling resolution of 1 nm so that each spectrum comprised reflectance at 2151 wavelengths [36]."

9. Comments: *Line 134: Which software and package did you use to performe this preprocessing methods?*
Answers: Many thanks for your suggestion. We have added more information about the software as follows (L153-155):
"These six traditional spectral pretreatments were carried out using the PLS_toolbox (Eigenvector Research, Inc.) in the Matlab environment (The MathWorks, Inc.)."

10. Comments: *Line 136: you should add detailed information about window size.*
Answers: Many thanks for your suggestion. We have revised the sentence as follows (L145-147):

"(2) SG smoothing a technique that uses a low-pass filter to smooth spectra by getting rid of high-frequency noise and allowing low-frequency signals to pass through [42]. In this study, the size of the window (filter width) is 15 nm and the order of the polynomials is 3."

11. Comments: *Line 147: Which software and package did you use to performe this preprocessing methods.*
Answers: Many thanks for your suggestion. We have revised the sentence as follows (L183-194):
"The three strategies were carried out using the PLS_toolbox (Eigen-vector Research, Inc.) in the Matlab environment (The MathWorks, Inc.)."

12. Comments: *Line 176: Which software and package did you use to performe this prediction models.*
Answers: Many thanks for your suggestion. We have revised the sentence as follows(L190-191):
"The PLSR was carried out using the PLS_toolbox (Eigen-vector Research, Inc.) in the Matlab environment (The MathWorks, Inc.)."

13. Comments: *Line 182: Did you applied 10-fold cross validation or boostrap for calibration dataset to prevent overfitting in the train function. You should ad mode infoation above. You should add cross validation, parameters for PLSR.*
Answers: Many thanks for your suggestion. We have revised the sentence as follows (L201-202):
"The number of the latent variables was determined using the leave-one-out cross-validation (LOOCV)."

14. Comments: *Line 183: Why did choose this rate, mainly the calibration and validation rate %70 and %30 in soil science papers.*
Answers: Many thanks for your suggestion. The proportion of validation samples could range from 20% to 50%, as shown in Table 1. While it is common for many research studies to choose 30% of samples for validation, in this study, we decided on a 20% validation set—a choice that has also been made by previous researchers. However, if you think 30% is necessary, we would change it in the next round of review.

Table 1. A review of division of validation and calibration

	Calibration	Validation	Reference
1	70%	30%	Wang, *et al.* (2024) [1]
2	80%	20%	Mancini, *et al.* (2024) [2]
3	75%	25%	Yang, *et al.* (2022) [3]
4	67%	33%	Xiao, *et al.* (2022) [4]
5	80%	20%	Li, *et al.* (2022) [5]
6	75%	25%	Mousavi, *et al.* (2020) [6]
7	67%	33%	Hong, *et al.* (2019) [7]
8	80%	20%	Liu, *et al.* (2019) [8]
9	75%	25%	Gholizadeh, *et al.* (2018) [9]

10	70%	30%	Dotto, *et al.* (2018) [10]
11	50%	50%	Shi, *et al.* (2017) [11]
12	70%	30%	Dotto, *et al.* (2017) [12]

We have revised the sentence as follows (L193-195):

"Previous studies showed that the portion of the validation samples ranged from 20% to 50%, and we decided on a 20% validation set [9,21,26,30,32]."

15. Comments: *Line 203: not similar, you can use " almost similar".*

Answers: Many thanks for your suggestion. We have revised the sentence as follows (L215): "The statistical indicators for the calibration and validation sets were almost similar."

16. Comments: *Line 205: move them under the line.*

Answers: Many thanks for your suggestion. We have revised the Table as follows (L216-219):

"**Table 1.** The descriptive statistics of 250 soil samples for the calibration and validation sets.

Sample	Number	Cu (mg·kg⁻¹)								
		Range [1]	Min	Max	Median	Mean	Std [2]	CV [3]	Skewness	Kurtosis
Total	250	82.79	20.45	103.24	59.44	58.29	15.57	0.27	0.13	0.12
Calibration	200	82.79	20.45	103.24	59.44	58.29	15.60	0.27	0.13	0.15
Validation	50	71.85	25.21	97.06	59.18	58.30	15.63	0.27	0.13	0.12

[1] Range denotes the difference between the maximum and minimum observations.
[2] Std denotes standard deviation.
[3] CV denotes coefficient of variation."

17. Comments: *Line 205: You should check all CV and Skewness values, there are similar !.*

Answers: Many thanks for your suggestion. We have reviewed the data and recalculated it, confirming that the values of CV and skewness are accurate.

18. Comments: *Line 210: Which data set?All, Calibration validation.*

Answers: Many thanks for your suggestion. The figure displays all the samples, including both validation and calibration samples. We have revised the tile of the figure as follows (L228-230): "Figure 2. Boxplot and histogram of Cu content of all samples. Repoint (•) denotes the mean value. Blue line (|) denotes the median value. Hollow circle (○) denotes the outliers. The black box de-notes the interquartile range."

19. Comments: *Line 219: not relevent, you should delete it.*

Answers: Many thanks for your suggestion. We have deleted it.

20. Comments: *Line 450: You should add more discussion about the prediction performance of Cu with other studies? Which spectral range can be important to predict soil Cu? why your spreadiction performance is low or high when compared to other studies results.*

Answers: Many thanks for your suggestion. We have add two new paragraphs as follows (L452-473):

"In this study, the Cu estimation model produced acceptable results (Figure 3) with the R_p^2 of 0.75 and the RMSEP of 8.86 mg kg^{-1}. In comparison, Liu et al. (2011) achieved a R_p^2 of 0.37 for soil Cu estimation, indicating relatively lower predictability [50]. Riedel, et al. (2018) only obtained a R_p^2 of 0.01 using vis-NIR spectroscopy for soil Cu estimation [51]. Cheng et

al. (2019) achieved a R_p^2 of 0.26 for Cu estimation in suburban soils [52]. Some researcher have reported higher accuracy, with a R_p^2 of 0.67 [53], 0.91 [54] and 0.92 [55]. Therefore, in comparison with previous studies, our results appear to be acceptable.

The important wavelengths utilized in the PLSR models for estimating soil Cu were identified using variable importance in projection (VIP) scores [56]. Wavelengths with VIP scores exceeding 1 were regarded as highly correlated and significant for estimating soil Cu [57]. In this study, the critical wavelengths for Cu estimation were identified within the ranges of 1031-1890nm, 1973-2193nm, 2213-2266nm, and 2474-2497nm (Figure 13). The region of 2213-2266nm is attributed to overtones and combinations of fundamental vibrations of organic molecules, such as C–H, N–H, S–H, C=O, and O–H [18]. Some research has reported significant wavelengths at 700nm, 1000nm, 1400nm, 1900nm, and 2200nm [25,50]. The variation in important wavelengths may be attributed to different soil parent materials and environmental conditions.

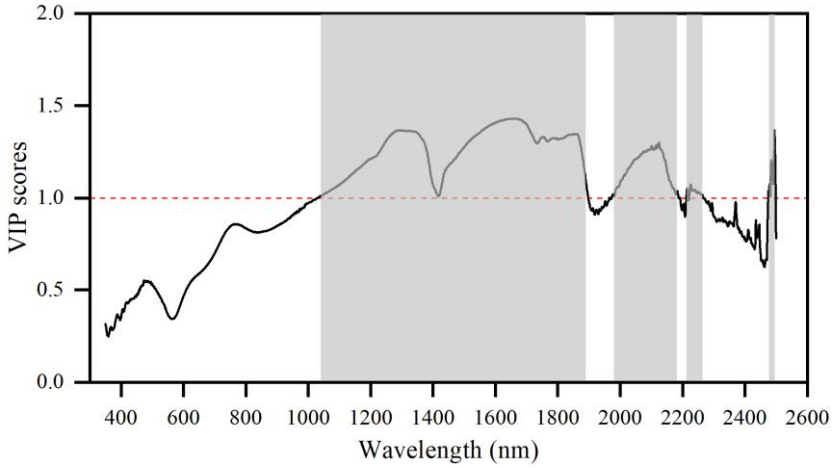


Figure 13. Variable importance projection (VIP) scores associated with the cross-validation of partial least squares regression model for soil Cu estimation. The threshold of VIP was set to 1 (red line)."

21. Comments: *Line 572: delete it, you should provide general outcome!*

Answers: Many thanks for your suggestion. We have deleted it.

Reference

[1].Wang, Z.; Chen, S.C.; Lu, R.; Zhang, X.L.; Ma, Y.X.; Shi, Z. Non-linear memory-based learning for predicting soil properties using a regional vis-NIR spectral library. Geoderma 2024, 441.

[2].Mancini, M.; Andrade, R.; Silva, S.H.G.; Rafael, R.B.A.; Mukhopadhyay, S.; Li, B.; Chakraborty, S.; Guilherme, L.R.G.; Acree, A.; Weindorf, D.C.; et al. Multinational prediction of soil organic carbon and texture via proximal sensors. Soil Sci. Soc. Am. J. 2024, 88, 8-26.

[3].Yang, W.; Xiong, Y.; Xu, Z.; Li, L.; Du, Y. Piecewise preprocessing of near-infrared spectra for improving prediction ability of a PLS model. Infrared Physics & Technology 2022, 126.

[4].Xiao, Q.; Tang, W.; Zhang, C.; Zhou, L.; Feng, L.; Shen, J.; Yan, T.; Gao, P.; He, Y.; Wu, N. Spectral preprocessing combined with deep transfer learning to evaluate chlorophyll content in cotton leaves. Plant Phenomics 2022.

[5].Li, S.; Rossel, R.A.V.; Webster, R. The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon. Eur. J. Soil Sci. 2022, 73.
[6].Mousavi, F.; Abdi, E.; Ghalandarzadeh, A.; Bahrami, H.A.; Majnounian, B.; Ziadi, N. Diffuse reflectance spectroscopy for rapid estimation of soil Atterberg limits. Geoderma 2020, 361.
[7].Hong, Y.; Shen, R.; Cheng, H.; Chen, S.; Chen, Y.; Guo, L.; He, J.; Liu, Y.; Yu, L.; Liu, Y. Cadmium concentration estimation in peri-urban agricultural soils: Using reflectance spectroscopy, soil auxiliary information, or a combination of both? Geoderma 2019, 354, 113875.
[8].Liu, Y.; Liu, Y.; Chen, Y.; Zhang, Y.; Shi, T.; Wang, J.; Hong, Y.; Fei, T. The Influence of Spectral Pretreatment on the Selection of Representative Calibration Samples for Soil Organic Matter Estimation Using Vis-NIR Reflectance Spectroscopy. Remote Sens. 2019, 11, 450.
[9].Gholizadeh, A.; Saberioon, M.; Carmon, N.; Boruvka, L.; Ben-Dor, E. Examining the Performance of PARACUDA-II Data-Mining Engine versus Selected Techniques to Model Soil Carbon from Reflectance Spectra. Remote Sens. 2018, 10.
[10].Dotto, A.C.; Diniz Dalmolin, R.S.; ten Caten, A.; Grunwald, S. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma 2018, 314, 262-274.
[11].Shi, T.; Liu, H.; Chen, Y.; Fei, T.; Wang, J.; Wu, G. Spectroscopic Diagnosis of Arsenic Contamination in Agricultural Soils. Sensors 2017, 17, 1036.
[12].Dotto, A.C.; Diniz Dalmolin, R.S.; Grunwald, S.; ten Caten, A.; Pereira Filho, W. Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. Soil Tillage Res. 2017, 172, 59-68.