

Spatial Multi-scale Relationships between Soil Organic Matter and Influencing Factors in Loess Plateau Basins: Postprint

Authors: Zhu Hongfen, Nan Feng, Xu Zhanjun, Jing Yaodong, Duan Yonghong, Bi Rutian

Date: 2018-01-09T00:00:00+00:00

Abstract

The effects of different influencing factors on soil organic matter content exhibit scale dependency. Taking soil organic matter in the Taiyuan Basin as the research object, sampling transects were established in the upper, middle, and lower parts of the basin, and multivariate empirical mode decomposition (MEMD) was applied to analyze the correlations between soil organic matter and influencing factors (elevation, slope, topographic wetness index, soil bulk density, sand, silt, clay, and spectral principal components, etc.) at characteristic scales, and to predict soil organic matter content at the sampling scale, aiming to investigate the spatial multi-scale relationship between soil organic matter and related factors in the basin region of the Loess Plateau. The research results indicate that: (1) The MEMD method can decompose the spatial sequences of soil organic matter at different locations within the basin into different characteristic scales, with the upper, middle, and lower parts of the basin having 6, 8, and 7 characteristic scales, respectively. Within the study area, the scale of approximately 1000 m represents the primary characteristic scale for soil organic matter, and the main characteristic scales of organic matter sequences perpendicular to the river direction in the basin exhibit dispersion along the river direction. (2) The spatial multi-scale relationship between soil organic matter and influencing factors demonstrates that the relationship between elevation and soil organic matter is primarily manifested at large scales, whereas the relationships between slope, topographic wetness index and soil organic matter in the middle and lower parts of the basin are more pronounced. Soil bulk density exhibits significant differences with organic matter at different characteristic scales across various locations. Among soil texture components, the silt content shows the most evident multi-scale relationship with organic matter. Spectral principal component 1 is significantly correlated with organic

matter at all characteristic scales across all sample transects. (3) The prediction accuracy for organic matter using the MEMD method is higher than that of stepwise multiple regression based on raw data. In summary, the research results can provide a theoretical basis for digital soil mapping, rational design of soil management units, and accurate prediction of organic matter in the basin region of the Loess Plateau.

Full Text

Preamble

Multiscale Spatial Relationships Between Soil Organic Matter and Influencing Factors in Basins of the Chinese Loess Plateau

ZHU Hongfen, NAN Feng, XU Zhanjun, JING Yaodong, DUAN Yonghong, BI Rutian*

College of Resources and Environment, Shanxi Agricultural University, Taigu 030801, China

Abstract

Soil organic matter (SOM) content is influenced by various factors operating at different scales. This study investigated the multiscale spatial relationships between SOM and influencing factors—including elevation, slope, topographic wetness index, soil bulk density, sand content, silt content, clay content, and soil spectral components—in the Taiyuan Basin, a typical region of the Chinese Loess Plateau. Soil samples were collected along transects established in the upper, middle, and lower parts of the basin. Multivariate empirical mode decomposition (MEMD) was applied to analyze the multiscale correlations between SOM and influencing factors at different basin locations, and to predict SOM content at the sampling scale.

The results showed that: (1) MEMD could decompose SOM transects into six, eight, and seven intrinsic mode functions (IMFs) for the upper, middle, and lower basin, respectively. A scale of 1000 m represented the main characteristic scale for SOM across the entire basin, with the number of representative scales increasing along the river direction. (2) Compared with correlations at the sampling scale, multiscale spatial correlations revealed that elevation-SOM relationships dominated at larger scales, slope-SOM correlations were significant in the middle basin, and topographic wetness index-SOM relationships were significant in both middle and lower parts. However, bulk density-SOM correlations were more complex, differing across scales and locations. In the upper basin, silt content showed stronger relationships with SOM than sand or clay content. Additionally, spectral component 1 was significantly correlated with SOM across all transects at all characteristic scales. (3) MEMD provided more accurate SOM predictions than stepwise multiple linear regression. These

findings provide a theoretical basis for digital soil mapping, rational design of farmland plots, and accurate SOM prediction in Loess Plateau basin regions.

Keywords: multivariate empirical mode decomposition; intrinsic mode function; multiscale; soil organic matter; influencing factors

Introduction

Soil organic matter content is not only an important indicator of soil fertility but also a significant source and sink in the global carbon cycle. Previous studies have extensively investigated the spatial variability of soil organic matter at different scales, demonstrating that soil heterogeneity and environmental factors jointly influence organic matter distribution at specific scales. Establishing accurate relationships between soil organic matter and influencing factors has become a method for indirectly understanding the spatial variability of organic matter. However, because spatial relationships between soil organic matter and influencing factors differ substantially across scales, when a factor simultaneously influences soil organic matter distribution at multiple scales, it may create nonlinear relationships that affect model accuracy and reliability.

In studies examining multiscale spatial relationships between environmental factors and soil properties, methods such as Pearson correlation analysis, wavelet analysis, and multidimensional analysis have been applied. These methods assume that environmental factors influence soil property distribution in a linear and stationary manner, which may not reflect actual spatial distributions. Multivariate empirical mode decomposition (MEMD), first proposed by Huang et al., is an alternative multiscale analysis method corresponding to wavelet analysis that avoids linear and stationary assumptions. MEMD can empirically decompose spatial data into multiple characteristic scales based on the data's intrinsic scale characteristics, offering clear advantages, though it has not been widely applied in soil property scale studies.

The Loess Plateau, formed by thick Quaternary loess deposits, is one of the world's most severely eroded regions. The basins within the Loess Plateau contain numerous plains and hills with relatively low elevation and suitable temperature-moisture conditions favorable for agricultural production. Due to diverse land use patterns and fragmented vegetation cover, soil organic matter distribution in this region exhibits disorder, and factor effects on organic matter are nonlinear. Investigating multiscale spatial relationships between soil organic matter and related factors can provide theoretical support for organic matter management. Since MEMD can be used for nonlinear and nonstationary multiscale analysis of spatial data, this study hypothesized that it could also be applied to examine multiscale relationships between soil organic matter and influencing factors. This research focused on the Taiyuan Basin in Shanxi Province to analyze characteristic scales of soil organic matter at different basin positions, examine multiscale spatial relationships between organic matter and influencing factors, and predict organic matter content at the sampling scale.

1. Study Area Overview

The Taiyuan Basin is located in central-eastern Loess Plateau, central Shanxi Province (36°55' -38°12' N, 111°42' -113°02' E), covering the middle reaches of the Fen River watershed. The basin extends approximately 150 km north-south and 12-60 km east-west, comprising nearly one-third of the province's area. The region has a warm temperate continental monsoon climate with annual sunshine duration of 2360-2796 hours, average annual precipitation of 420-457 mm (decreasing from south to north), and average annual temperature of 7.8-10.3°C. The terrain is surrounded by mountains on three sides (east, west, and north) with an average elevation of 800-900 m. The central plain consists primarily of alluvial-pluvial sandy loam and clayey loess, while peripheral hills are dominated by thick aeolian sandy loess, representing typical concentrated loess distribution. Main soil types include fluvo-aquic soil, lime concretion black soil, calcic cinnamon soil, and cinnamon soil-like soils. The primary crops are winter wheat and corn, making this a major agricultural production region.

2. Sample Collection and Processing

Based on elevation and Fen River flow direction (southwest), the basin was divided into upper, middle, and lower sections. After field investigation and map analysis, three sampling transects were established at different basin positions while avoiding large construction areas. Each transect was approximately 42 km long with sampling points at 330 m intervals. If a sampling point fell on construction land or roads, the nearest cultivated land was sampled instead, maintaining all points along a straight line as much as possible. Transect elevations ranged from 742-894 m (upper), 723-807 m (middle), and 707-1002 m (lower), with 121, 128, and 116 sampling points respectively.

[Figure 1: see original paper] Geographical location of study area and distribution of sampling points

At each sampling point, a spiral soil auger was used to collect 0-20 cm topsoil samples. Five subsamples were taken in a cross pattern and mixed to represent the sampling point. After air-drying, samples were passed through a 2 mm sieve and divided into two portions for soil spectral analysis and organic matter determination. Undisturbed surface soil samples were collected at cross-pattern centers using a 100 cm³ cutting ring to determine bulk density. Soil organic matter content was measured by the potassium dichromate method, soil texture by the sedimentation method, and bulk density by oven-drying at 105°C for 10 hours. Soil spectra were obtained using an ASD FieldSpec3 spectroradiometer (350-2500 nm range, resampled at 1 nm intervals).

3. Multivariate Empirical Mode Decomposition

Multivariate empirical mode decomposition is an extension of empirical mode decomposition (EMD) for spatial data. EMD decomposes data into multiple

spatial series based on the assumption that spatial data consist of simple superimposed oscillation patterns at different frequencies, and that overall variability results from multiple processes occurring at different scales with different intensities. Processes occurring at the same scale can be decomposed into identical intrinsic mode functions (IMFs), which must satisfy two conditions: (1) the number of local extrema and zero crossings must be equal or differ by at most one throughout the data range; (2) the oscillation is symmetric about the local mean, meaning the average envelope defined by local maxima and minima equals zero at any data point.

EMD uses a “sifting” process to extract IMFs from any function. The variance contribution percentage of each IMF can be calculated to determine its relative importance in explaining overall variance, thereby identifying the relative significance of processes occurring at different scales. The average oscillation frequency determines the corresponding IMF scale, while instantaneous scales provide the scale range represented by each IMF, obtained through Hilbert transform-derived instantaneous frequency. Hilbert spectrum analysis can calculate energy and frequency at each scale and location.

However, EMD requires calculating local means of spatial data sequences, which cannot be directly defined for multiple spatial data series. To address this, Rehman and Mandic proposed creating multiple direction vectors in multidimensional space through different projections. For multidimensional vector data $V(s) = \{v_1(s), v_2(s), \dots, v_n(s)\}$ representing soil organic matter and influencing factors, direction vectors $X_k = \{x_{k1}, x_{k2}, \dots, x_{kn}\}$ are generated at angles $\theta_k = \{ \theta_1, \theta_2, \dots, \theta_{n-1} \}$ ($k = 1, 2, \dots, K$, where K is the total number of directions). Projections $p_k(s)$ along given directions X_k are calculated, maxima positions s_k are identified, and multivariate envelope curves $e_k(s)$ are obtained. The mean envelope $M(s) = (1/K) \sum e_k(s)$ serves as the local mean, and residuals $D(s) = V(s) - M(s)$ indicate original data trends. This process repeats until residuals become monotonic functions.

4. Data Preprocessing

Soil spectral data noise in the 2451-2500 nm range was removed. Principal component analysis compressed the 401-2450 nm spectral data (2050 bands), with the first two principal components (accounting for most variance) selected as comprehensive influencing factors. Multivariate data sequences were constructed combining SOM with topographic factors (elevation, slope, topographic wetness index), soil physical properties (bulk density, sand, silt, clay content), and spectral principal components. MEMD decomposed each transect data series into different IMFs. Similar scales across different data sources were grouped to represent actual process scales.

Stepwise multiple regression models were developed for each IMF and residual to predict SOM at corresponding scales. Predicted SOM values from each IMF (S) and residual (R) were combined using: $SNP = S + R$, where SNP is the

predicted SOM at sampling scale. Prediction accuracy was evaluated using coefficient of determination (R^2), root mean square error (RMSE), normalized root mean square error (NRMSE), and residual prediction deviation (RPD). Wavelet energy spectra representing SOM spatial variability also assessed prediction capability. All processing was implemented in MATLAB.

1. Relationships Between SOM and Influencing Factors at Sampling Scale

Pearson correlation analysis examined linear relationships between SOM and topographic factors, soil physical properties, and spectral components. Results showed SOM was significantly correlated with elevation, bulk density, and spectral components across all transects, and with silt and spectral components in specific transects, indicating these five factors primarily controlled SOM at sampling scale. Local-scale processes (e.g., biological activity, tillage practices) disturbed other factor effects, resulting in nonsignificant correlations. Topographic factors and physical properties showed strongest influence in the middle basin, while soil texture most strongly affected the lower basin.

[Table 1] shows correlation coefficients between SOM and influencing factors at sampling scale. Stepwise multiple regression models using these factors explained 52%, 56%, and 54% of SOM variability in transects 1, 2, and 3 respectively.

2. Multiscale Spatial Relationships Between SOM and Influencing Factors

MEMD decomposed multivariate data sequences of SOM and influencing factors into different IMFs. SOM spatial sequences showed longer oscillation periods with increasing transect number. Residual sequences indicated original data trends, with large oscillation amplitudes at characteristic scales showing high spatial variability. Residual trends were most apparent in the middle basin.

Variance percentages showed SOM variability concentrated at specific scales: 30.61% at 8573 m in the upper basin, 17.62% and 12.53% at 1725 m in the middle basin, and 20.46% at <2000 m in the lower basin. The main characteristic scales were approximately 960 m and 6753 m in the upper basin, 1000 m and 8000 m in the middle basin, and 1000 m, 7000 m, and 12000 m in the lower basin. Vertical-to-river SOM characteristic scales showed dispersal patterns along the river direction.

[Table 2] shows stepwise regression models for SOM based on influencing factors. [Table 3] presents IMF scales for SOM across three transects, with coefficients of variation indicating similar characteristic scales between SOM and influencing factors at small scales. [Table 4] shows variance percentages explained by each IMF and residual.

Elevation variance primarily occurred in residuals (75.93%, 72.83%, 70.04% across transects), indicating its main characteristic scales exceeded extracted IMF scales. Other factors concentrated in IMF1 and IMF5 (upper basin), IMF1 and IMF3 (middle basin), and IMF1 (lower basin), showing their variability mainly at small scales (1530 m, 1725 m, 1504 m respectively).

Multiscale correlations differed from sampling-scale correlations (Table 5). Elevation-SOM correlations dominated at large scales (IMF5,6 in upper basin; IMF6,7,8 in middle; IMF6,7 in lower). Slope-SOM relationships were significant at large scales in transect 2. Topographic wetness index showed strongest relationships with SOM in the middle basin, contrasting with sampling-scale results. Bulk density-SOM relationships were complex, significant at middle basin small scales (~1500 m) and lower basin large scales (~10000-12000 m). Silt content showed the most significant multiscale relationship with SOM, significantly correlated at all characteristic scales. Spectral component 1 was significantly correlated with SOM across all transects and scales.

3. SOM Prediction Based on Multiscale Relationships

Stepwise regression models for each IMF showed all adjusted R^2 values between 0.44 and 1.00, with influence strength increasing with scale. [Table 6] presents predictive equations and regression statistics for each IMF and residual. Prediction accuracy parameters (Table 7) showed MEMD predictions achieved R^2 of 0.90, 0.86, and 0.91 for transects 1-3, significantly higher than direct stepwise regression (0.52, 0.56, 0.54). RMSE and NRMSE were substantially lower, while RPD was higher for MEMD.

Correlations between measured SOM at sampling scale and IMF/residual predictions (Figure 3) showed main contributors were IMF6 for transect 1, IMF5 for transect 2, and IMF6 for transect 3, with large-scale residuals also contributing substantially. Spectral component 1 was the most important predictor across all transects, followed by slope and bulk density. Terrain factors, bulk density, and texture most strongly influenced transect 2, with the order being middle > lower > upper basin.

Local wavelet spectra (Figure 4) revealed that stepwise regression errors mainly involved local overestimation at 19.8-36.3 km in transect 1 and 19.8-33.0 km in transect 2, and underestimation at 4-8 km in transect 3. MEMD better captured local variance patterns.

3. Discussion

MEMD decomposed SOM spatial sequences into characteristic scales, revealing 1000 m as the main scale. For grid-based soil property storage in this basin, a 1000 m grid width is recommended to capture major SOM spatial variability. The upper basin's large residual variance percentage (24.19%) suggests longer transects are needed to capture larger scales identified by MEMD. The

upper basin' s characteristic scales differed substantially from middle and lower basins, likely due to proximity to Taiyuan and Jinzhong cities causing land use fragmentation from human activities.

Vertical-to-river SOM characteristic scales showed dispersal along the river, indicating river-oriented factors increasingly dominated SOM distribution. At both sampling and multiscale levels, middle basin SOM showed strongest correlations with topographic factors, bulk density, and texture, possibly due to the area' s location at the intersection of erosion plains and subsidence basins. Weak correlations between middle basin SOM and spectral components may reflect low SOM variability reducing soil-spectral relationships.

While some factors showed nonsignificant correlations at sampling scale, their multiscale relationships were significant at certain characteristic scales, demonstrating that multiscale analysis captures more information. MEMD' s superior prediction accuracy confirms that single-scale analysis cannot reveal complex factor relationships. Though MEMD effectively analyzes nonlinear, nonstationary data, its recent development means limited application in soil property studies. Unlike wavelet analysis, MEMD lacks a defined basis function, producing different characteristic scales for different transects, which limits generalization but focuses on process analysis.

4. Conclusions

MEMD analysis of Taiyuan Basin SOM and influencing factors yielded these main conclusions:

- 1) MEMD effectively decomposed SOM spatial sequences into characteristic scales: upper basin (960 m, 6753 m), middle basin (1000 m, 8000 m), and lower basin (1000 m, 7000 m, 12000 m). The 1000 m scale was the main characteristic scale basin-wide, with vertical-to-river scales showing dispersal along the river.
- 2) Multiscale correlations differed from sampling-scale correlations. Elevation effects dominated at large scales (~8573 m basin-wide). Slope relationships were significant at large scales in the middle basin. Topographic wetness index effects were most apparent in the middle basin, though not significant at sampling scale in the lower basin. Bulk density relationships were complex and scale-dependent. Silt content showed the most significant multiscale relationships. Spectral component 1 was significant across all scales.
- 3) MEMD prediction accuracy was significantly higher than direct stepwise regression, demonstrating that multiscale analysis comprehensively reveals complex SOM-factor relationships across all spatial scales. The method effectively captures factor influences at specific scales, providing a robust tool for digital soil mapping and precision agriculture in Loess Plateau basins.

Acknowledgments

We thank Guo Ying, Wang Peng, and Xu Zhen for assistance with sample collection and testing.

References

- [1] Bauer A, Black A L. Quantification of the effect of soil organic matter content on soil productivity. *Soil Science Society of America Journal*, 1994, 58(1): 185-193. [2] Lal R. Soil carbon sequestration impacts on global climate change and food security. *Science*, 2004, 304(5677): 1623-1627. [3] Hu K L, Wang S Y, Li H, Huang F, Li B G. Spatial scaling effects on variability of soil organic matter and total nitrogen in suburban Beijing. *Geoderma*, 2014, 226-227: 54-63. [4] Li D F, Gao G Y, Lü Y H, Fu B J. Multiscale variability of soil carbon and nitrogen in the middle reaches of the Heihe River basin, northwestern China. *Catena*, 2016, 137: 328-339. [5] Zhou Y, Biswas A, Ma Z Q, Lu Y L, Chen Q X, Shi Z. Revealing the scale-specific controls of soil organic matter at large scale in Northeast and North China Plain. *Geoderma*, 2016, 271: 71-79. [6] Zhu H F, Bi R T, Duan Y H, Xu Z J. Scale-location specific relations between soil nutrients and topographic factors in the Fen River Basin, Chinese Loess Plateau. *Frontiers of Earth Science*, 2016: 1-10, doi: 10.1007/s11707-016-0587-y. [7] She D L, Shao M A, Hu W, Yu S E. Variability of soil water-physical properties in a small catchment of the Loess Plateau, China. *African Journal of Agricultural Research*, 2010, 5(22): 3041-3049. [8] Zhu H F, Hu W, Bi R T, Peak D, Si B C. Scale- and location-specific relationships between soil available micronutrients and environmental factors in the Fen River basin on the Chinese Loess Plateau. *Catena*, 2016, 147: 764-772. [9] Liu Q J, Shi Z H, Fang N F, Zhu H D, Ai L. Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet-ANN approach. *Geomorphology*, 2013, 186: 181-190. [10] Wang D, Fu B J, Zhao W W, Hu H F, Wang Y F. Multifractal characteristics of soil particle size distribution under different land-use types on the Loess Plateau, China. *Catena*, 2008, 72(1): 29-36. [11] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q A, Yen N C, Tung C C, Liu H H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1971): 903-995. [12] She D L, Shao M A, Timm L C, Reichardt K. Temporal changes of an alfalfa succession and related soil physical properties on the Loess Plateau, China. *Pesquisa Agropecuária Brasileira*, 2009, 44(2): 189-196. [13] [Reference incomplete in original] [14] Hu W, Shao M A, Si B C. Seasonal changes in surface bulk density and saturated hydraulic conductivity of natural landscapes. *European Journal of Soil Science*, 2012, 63(6): 820-830. [15] [Standard soil analysis method reference] [16] [Standard soil texture method reference] [17] [Spectral prediction reference] [18] Goovaerts P. Geostatistical tools for characterizing the spatial variability of microbiological and physicochemical soil properties. *Biology and Fertility of Soils*, 1998, 27(4): 315-

334. [19] Oladosu G. Identifying the oil price-macroeconomy relationship: An empirical mode decomposition analysis of US data. *Energy Policy*, 2009, 37(12): 5417-5426. [20] Rehman N, Mandic D P. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A*, 2010, 466(2117): 1291-1302. [21] Ur Rehman N, Mandic D P. Filter bank property of multivariate empirical mode decomposition. *IEEE Transactions on Signal Processing*, 2011, 59(5): 2421-2426. [22] Zhao X M, Patel T H, Zuo M J. Multivariate EMD and full spectrum based condition monitoring for rotating machinery. *Mechanical Systems and Signal Processing*, 2012, 27: 712-728. [23] [Holocene geomorphology reference] [24] She D L, Zheng J X, Shao M A, Timm L C, Xia Y Q. Multivariate empirical mode decomposition derived multiscale spatial relationships between saturated hydraulic conductivity and basic soil properties. *Clean-Soil, Air, Water*, 2015, 43(6): 910-918. [25] Hu W, Si B C. Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. *Geoderma*, 2013, 193-194: 180-188.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.