COMPARISON OF METHODS FOR PREDICTING SOIL PROPERTIES IN DONGHAI COUNTY OF CHINA

WANG, J. X.[1,2] – ZHOU, S. L.[1,*] – LIU, L.[1] – WANG, X. R.[1] – LIU, B.[3] – SHI, Z. B.[3] – SUI, X. Y.[3]

[1]Department of Geographic and Oceanographic Sciences, Nanjing University
163 Xianlindadao, 210023 Nanjing, Jiangsu, China
(e-mails: njuwjx@gmail.com; wuzhiliu@foxmail.com; njuwxr@163.com)

[2]Key Laboratory of the Coastal Zone Exploitation and Protection Ministry of Land and
Resource, 43-8 Beijingxilu, 210024 Nanjing, Jiangsu, China

[3]Jiangsu Land Consolidation and Rehabilitation Center
58 Shuiximendajie, 210009 Nanjing, Jiangsu, China
(e-mails: 13951661269@139.com; shizhb@126.com; xueyan322@sina.com)

*Corresponding author
e-mail: zhousl@nju.edu.cn; phone: +86-138-0517-1474

Abstract. Different methods have been used to predict soil properties and obtain continuous spatial distributions. However, studies that compare the accuracy of these methods are scarce. In addition, few studies considered human activity variables as well as landscape ecology variables to predict soil properties. The objective of this research was to determine the optimal method for predicting soil properties using human activity and landscape ecology variables in regions with complex topographies. The study results are as follows. (1) Introducing land-use type as human activity variable to the soil-landscape model can improve the accuracy. (2) In the study region, the mean absolute error (MAE) and mean relative error (MRE) of the soil organic matter, pH, and soil particle size predicted by the soil-landscape model were lower than those predicted by ordinary Kriging method. (3) Specifically, the regions in which the soil-landscape prediction model was more accurate were all concentrated in the central and west hilly regions in the study region, whereas the regions in which the geostatistical interpolation was more accurate were mainly in the east plain region. The differences in the results are based on the topography of the study area.
Keywords: soil-landscape, geostatistical interpolation, Donghai County, hilly region, precision comparison

Introduction and review of literature

The purpose of soil sampling is to acquire the spatial distribution characteristics of target variables and study the soil characteristics across an entire study region. Soil surveys can achieve optimal land management by exploiting the available information in the study region (Dent and Young, 1981; Carter, 1993; Minasny and McBratney, 2006). Because current land management is becoming more refined and soil sampling is limited by human, material, and financial resources, methods that reduce sampling costs through rational sampling yet can still satisfy soil survey accuracy requirements have received increasing attention. Current scholars have conducted plenty of in-depth research on sampling design optimization methods. Some scholars (Hengl and Rossiter, 2003) have constructed a prediction model of soil properties using auxiliary variables and simultaneously optimized the distribution of sampling points. In addition, some researchers (van Groenigen et al., 2000; Ferreyra et al., 2002; Simbahan and Dobermann, 2006; Brus and Heuvelink, 2007) optimized the number and spatial

distribution of sampling points using a simulated annealing algorithm and obtained the optimal distribution of soil sampling points. However, optimizing the methods of collecting soil samples may promote a heterogeneous spatial distribution of soil sampling points and soil properties, which can affect the accuracy of soil quality assessments. Under the assumption that additional soil sampling points cannot be added, methods that infer soil property information in unknown regions must be used to obtain a continuous spatial distribution of soil properties. Current soil studies primarily apply soil-landscape models and geostatistical methods for predicting different soil properties.

Soil-landscape models are derived from Jenny's equation (Jenny, 1941). This equation describes soil as a function of climate, potential biota, relief, parent material, and time (Jenny, 1941; Zhou and Wang, 2004). Numerous studies have used this model to predict soil properties. Jones (1973) studied the West African Savannah and found that the soil carbon, nitrogen and soil particle size were related to annual precipitation. Some research (Jones, 1973) found a piecewise linear relationship between the soil organic carbon, nitrogen, pH value and the gradient angle. Using soil-landscape models, Grunwald (2016) predicted the soil organic carbon content under different land-use types using a neural network with climate and topography factors as the input variables. A geostatistical approach was first applied by the South African mining engineer Krige (1951) for mineral exploration calculations used in the mining industry in the early 1950s. In soil property studies, the spatial variability of soil physical properties became a focus starting in the 1970s in North America and certain Western European countries (Krige, 1951; Minasny and McBratney, 2006).

In summary, numerous studies have investigated issues related to soil property predictions performed using soil-landscape models as well as geostatistical methods. However, the study regions in which these two methods were applied to predict soil properties were mostly concentrated in plains regions where the soil properties were relatively homogenous. Studies on regions with large topographic variations (such as mountains and hilly regions) are scarce, and comparative studies on the prediction results and accuracy of the two methods are limited. Therefore, in the present study, (1) the Huanghuaihai hilly region of Donghai County in Jiangsu Province was selected as the study region, and based on the large amount of data available for this study region, a soil-landscape model and a geostatistical model was used to predict the soil properties and then the prediction results and accuracy of the two methods were compared to provide information for selecting methods of predicting soil properties in regions with large topography variations; and (2) based on the frequent human activities in the study region, land-use types were added to represent of the effects of human activities as landscape factors to improve the prediction accuracy of the soil-landscape model and then analysed and verified the prediction results.

Materials and methods

Study area

Donghai County (*Fig. 1*) is located in the northeast of Jiangsu Province, China at $34°11'\sim34°44'$N, $118°23'\sim119°10'$E. Donghai County encompasses a plains hilly region at the southeast margin of the Huanghuaihai region. Its area is 2073 km^2. The elevation is high in the west and low in the east. The east plains region is flat and presents numerous lakes and reservoirs, the west region is a hilly region with an undulating relief, and the centre region presents gentle slopes as the plains transition to hills. The

Maling Mountains, on the western boundary, extend from north to east. The substrate consists of purple sandshale, and purple soil is distributed as well. The parent material of the purple soil is the weathered remnants of the purple sandshale. In the centre and west, numerous hills and gullies are distributed and cross each other, and the terrain is primarily hilly. The parent material of the brown soil distributed throughout these regions is the weathered remnant of acidic metamorphic granitic gneisses, representing an ancient alluvial deposit covered by slope wash. The elevation of the lacustrine plain in the east is only 2-5 m, and the parent material of the soil is secondary ancient fluvial-lacustrine loessial deposits formed in the Quaternary.
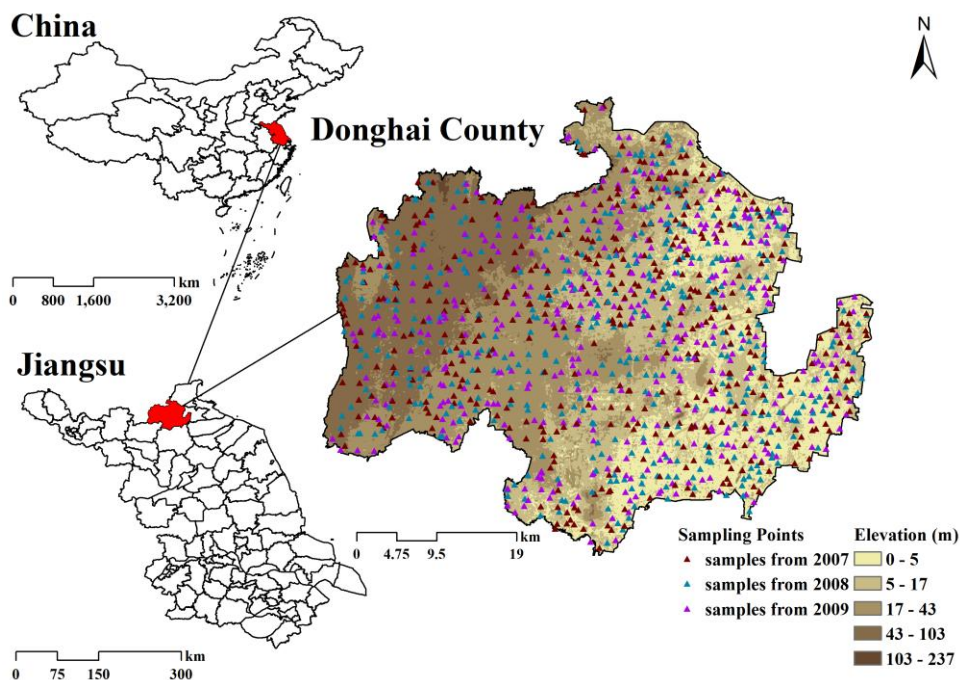

Figure 1. *Location map of the study region and design of the sampling points*

Soil sampling

The 1268 sample points used in this study were collected in November 2007, November 2008, and November 2009. Each soil sample has been collected only one time in three-year period (detailed design shown in *Fig. 1*). The sampling locations were all on farmlands, and the collection times were all after the rice harvesting period and before the winter wheat sowing period. The sampling locations were determined using geographic coordinates obtained with Garmin-76 GPS receivers to reduce the accuracy errors associated with global positioning systems. The soil samples were all collected using stainless steel soil samplers, which collected soil from the 0-20 cm topsoil layer. At each sampling location, 5 samples were collected along 10-m diagonals and then mixed homogenously. After that, 1 kg of soil was retrieved by quartering, and the samples were sealed in plastic bags and brought back to the laboratory. Plant roots and rocks in the collected soil samples were removed, the soil samples were spread out on plastic disks, and naturally dried in ventilated locations. The dried samples were crushed and ground using wood rods or mortars, passed through 20-mesh, 60-mesh, and 100-mesh nylon sieves, and separately sealed in plastic bags until the physiochemical

analyses. The physiochemical analyses of the samples were conducted at the Institute of Soil Science, Chinese Academy of Sciences, in Nanjing. Each sampling point included measurements of the pH value, organic matter content, and soil particle size. During the laboratory evaluation, each sample was divided into four parts and tested parallelly to prevent any errors that may occur during test to ensure the accuracy of soil properties. The three soil indices used in this study and their measurement methods were as follows: (1) soil pH was determined using the potential method; (2) soil organic matter was determined using the potassium dichromate volumetric method; and (3) soil particle size was determined using a laser particle size analyser and the Malvern method following the US grade system. 70% of the sample points were randomly selected (918) as the training set to build the soil-landscape prediction model and geostatistical model (ordinary Kriging interpolation). The remaining 30% of the sample points (350) were used as the test dataset for the models.

The statistical features of the three soil properties measured in areas with different land-use types are shown in *Table 1*. The organic matter in the soil from dry lands was significantly less than that from paddy fields. The soil moisture in the paddy fields was high, and organic matter decomposed slowly and accumulated more easily. The soil from the dry land was more acidic than that from the paddy fields. The particles of the soil from the paddy field were significantly larger than those of the soil from dry land, and the paddy field soil was mainly loamy and clayey, whereas the dry land soil was mainly sandy. The three soil properties of the paddy field showed a more significant normal distribution pattern, whereas the soil properties of the dry land showed a skewed distribution, with high kurtosis and concentrated data.

Table 1. *Statistical features of the soil properties of the different land-use types in intensively sampled regions*

		Minimum	Maximum	Mean	Variation coefficient	Skewness	Kurtosis
Paddy field	Organic matter (g/kg)	3.5	41.0	20.7	30.68%	0.42	-0.24
	pH value	5.0	8.2	6.8	8.85%	0.17	-0.31
	Soil particle size (%)	4.0	89.7	41.1	58.35%	0.15	-1.30
Dry land	Organic matter (g/kg)	1.2	28.7	12.5	29.63%	0.70	2.11
	pH value	5.1	9.9	6.4	8.45%	1.29	8.77
	Soil particle size (%)	3.5	78.3	17.8	65.73%	1.68	4.15

The above analysis indicates that the differences in the three soil properties from the paddy field and dry land were significant. The data statistics showed that the land-use types significantly affected the soil properties and that land-use types should be used as a factor in soil-landscape models.

Soil-landscape multivariate stepwise regression method

Soil-landscape models establish quantitative relationships between soil properties and different environmental factors (topography, landscape, hydrology, and human activities) by studying the correlation features between soil and environmental factors, and variation patterns of the spatial and temporal distribution of the soil properties are

predicted based on quantitative relationships (Gessler et al., 2000; Kerry and Oliver, 2004; Grunwald, 2016).

This study determined the environmental factors that affect the soil in the study region based on the soil-forming factor theory and soil-landscape multivariate stepwise regression method. The soil-forming factor theory determined five categories of soil formation factors: climate, parent material, relief, potential biota, and time. The soil-landscape model theory proposes that different soil environmental conditions can form different soil types. When specific combinations of soil-landscape factors are determined, the soil type that is formed is also uniquely determined (Grunwald, 2016). Therefore, this study selected landscape factors that are relevant to soil forming and predictive for topography, morphology, climate, hydrology, vegetation, and human factors. For the soil factors, the soil organic matter and pH in the study region are important soil properties. Furthermore, because the study region is hilly, the soil particle size varies significantly. Additionally, the organic matter, pH, and soil particle size represent the comprehensive index, chemical index, and physical index of the farmland soil quality, respectively, and can be used to represent farmland soil quality to a great extent (Xia et al., 2000; Yuan et al., 2008; Zhu, 2008). Therefore, this study selected soil organic matter, pH, and soil particle size as the soil property indices. The definitions, equations, and meanings of the landscape factors in this study are provided in *Table 2*. To acquire specific data acquisition, a 30 m × 30 m digital elevation model (DEM) of the study region was used to select the relevant topographic factors in ArcGIS 9.3, including the elevation, gradient angle, plane curvature, profile curvature, topographic wetness index, and stream power index. The 5-year mean normalized difference vegetation index (NDVI) value was calculated in the region based on 30 m Landsat Thematic Mapper (TM) remote sensing data for the five year period 2005-2010. The land-use types were identified based on the current land-use map of Donghai County for 2010 (1:5000 scale). This study used the soil properties of the 918 sample points in the training set and a multivariate stepwise regression method to build the soil-landscape prediction model.

Geostatistical interpolation method

To predict the target variables (soil properties) of many independent and common variables in space, the commonly used geostatistical interpolation method, the ordinary Kriging (OK) method, was used. Based on the 918 sample points, the ordinary Kriging interpolation was used to predict the soil organic matter, pH value, and soil particle size in the study region. The Kriging interpolation is also called the spatial local interpolation method, and it is based on variogram theory and structural analyses and yields the best unbiased estimations of regionalized variables in limited regions. The Kriging method is one of the most important methods in geostatistics and provides the best linear unbiased estimations of elevation based on the spatial location of the points to be interpolated and the measured elevation of neighbouring points. The original topography in the study region is represented by a Kriging interpolation map of the elevation. The overall equation is as follows (*Eq. 1*):

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \qquad (Eq.1)$$

where $Z(x_0)$ represents the value of the unknown sample points; $Z(x_i)$ represents the value of the known sample points surrounding the unknown sample points; N is the

number of known sample points; and λ_i is the weight of the ith sample point. This equation is determined by a semivariogram analysis, and based on the statistical requirements of unbiased and optimal, the equation between the weights and semivariance can be derived using Lagrange's minimization principle. For a more detailed and comprehensive description of the method, please refer to Cressie (1988), Odeh et al. (1995), and Mishra et al. (2009).

Table 2. *Definitions, equations, and meanings of the landscape factors*

Environmental factor	Definition and equation	Representation
Elevation (m)	Vertical distance from a point to the absolute reference surface	Changes in vegetation and hydrothermal conditions, affects soil development
Gradient angle	Angle between the normal and vertical directions at a ground point $$\alpha = \arctan\sqrt{f_x^2 + f_y^2}$$ (Gunstensen et al., 1991)	Reflects the degree of tilt of the local surface, which affects the topsoil stability and surface runoff discharge and convergence
Plane curvature (m-1)	At any given surface point, the curvature of the curve defined by the intersection between the horizontal plane passing through this point and the ground surface $$H_c = -\frac{f_y^2 f_{xx} - 2f_x f_y f_{xy} + f_x^2 f_{yy}}{\left(f_x^2 + f_y^2\right)\left(1 + f_x^2 + f_y^2\right)^{3/2}}$$ (Roddier, 1988)	Affects the convergence and divergence of the movement of surface material
Profile curvature (m-1)	At any given surface point, the curvature of the curve defined by the intersection of the plane passing through the normal direction and the direction with the maximum elevation gradient at this point and the ground surface $$P_c = -\frac{f_y^2 f_{xx} + 2f_x f_y f_{xy} + f_x^2 f_{yy}}{\left(f_x^2 + f_y^2\right)\left(1 + f_x^2 + f_y^2\right)^{3/2}}$$ (Zevenbergen and Thorne, 1987)	Affects the acceleration and deceleration of the movement of surface material
Topographic wetness index	$$TWI = \ln\left(\frac{A}{L \times \tan\alpha}\right)$$ (Barling et al., 1994)	Quantitatively reflects the comprehensive condition of soil water storage and discharge
Stream power index	$$SPI = \frac{A \times \alpha}{L}$$ (Hack, 1973)	Quantitatively reflects the potential erosion capability of runoff
Normalized difference vegetation index	$$NDVI_i = \frac{band4 - band3}{band4 + band3}$$ $$\overline{NDVI} = \sum_i^n NDVI_i$$ (Townshend and Justice, 1986)	Quantitatively reflects the vegetation distribution and coverage in the region
Land use type	Paddy field, dry land	Different land use types reflect the degree of human disturbance and change to soil

Note: f_x and f_y are the first derivatives of the elevation surface in the x and y directions, respectively; f_{xx} and f_{yy} are the second partial derivatives of the elevation surface in the x and y directions, respectively; f_{xy} is the second mixed partial derivative of the elevation surface in the x and y directions; A is the confluence area; and L is the length of the contour lines

Method of analysing accuracy errors

The method of determining the accuracy errors in this study includes the mean absolute error (MAE) and mean relative error (MRE) indices to describe the differences between the model predictions and the measured values. Smaller values indicate a higher accuracy of the model predictions. $\tilde{Z}_i$ is the predicted value, and Z_i is the measured value (*Eqs. 2* and *3*).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\tilde{Z}_i - Z_i\right| \qquad \text{(Eq.2)}$$

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\tilde{Z}_i - Z_i\right|/Z_i \qquad \text{(Eq.3)}$$

Results

Prediction results of the two prediction methods

Soil-landscape models

The study object in this study is farmland; thus, all of the sample points were on farmland. Prediction models were built separately for paddy fields and dry land, which are shown in *Table 3*. All of the prediction models passed the F test at the 95% confidence level. The test analysis results showed that the model prediction results were good. The R^2 values of the soil property prediction models showed that the goodness of fit of the prediction model for the three soil properties in the paddy fields was higher than that in the dry land. The landscape factors in the paddy fields could explain 65.9%, 51.7%, and 63.5% of the spatial variance of the soil organic matter, pH value, and soil particle size, respectively. The soil-landscape factors in the dry land could explain 56.1%, 46.4%, and 48.3% of the spatial variance of soil organic matter, pH value, and soil particle size, respectively.

Table 3. Soil-landscape prediction models for the different soil properties

Land use	Soil property	Regression equation	R²	F
Paddy field	Organic matter	$y_1 = 27.647 - 4.489\ln x_1 + 0.245x_2 + 4.832\,x_3 + 2.713x_7$	0.659	183.49
	pH	$y_2 = 6.753 - 0.119\ln x_1 + 0.009x_5 + 0.244\,x_7$	0.517	108.05
	Soil particle size	$y_3 = 0.443 - 0.122\ln x_1 + 0.016\,x_2 + 0.005x_5 + 0.194x_7$	0.635	128.88
Dry land	Organic matter	$\ln y_1 = 3.382 - 0.234\ln x_1 - 0.0433l\ \ln x_4$	0.561	77.02
	pH	$y_2 = 6.572 + 0.153\ln x_1 - 0.062x_4 - 0.0622l\ \ln x_5$	0.464	44.62
	Soil particle size	$y_3 = 0.371 - 0.043\ln x_1 - 0.0077l\ \ln x_6$	0.483	46.84

Note: x_1: elevation; x_2: gradient angle; x_3: plane curvature; x_4: profile curvature; x_5: topographic wetness index; x_6: stream power index; and x_7: NDVI

For the three soil properties, the goodness of fit of the prediction models was highest for organic matter. The soil-landscape model provided accurate predictions of the condition of the soil organic matter. Among the eight soil-landscape factors in the paddy fields and dry land, elevation was always a factor for the prediction models, indicating that elevation is one of the most important factors that affects soil properties.

Using the above soil property prediction models, The organic matter content, pH value, and soil particle size was separately predicted in the paddy fields and dry land. The prediction results are shown in *Figure 2*.
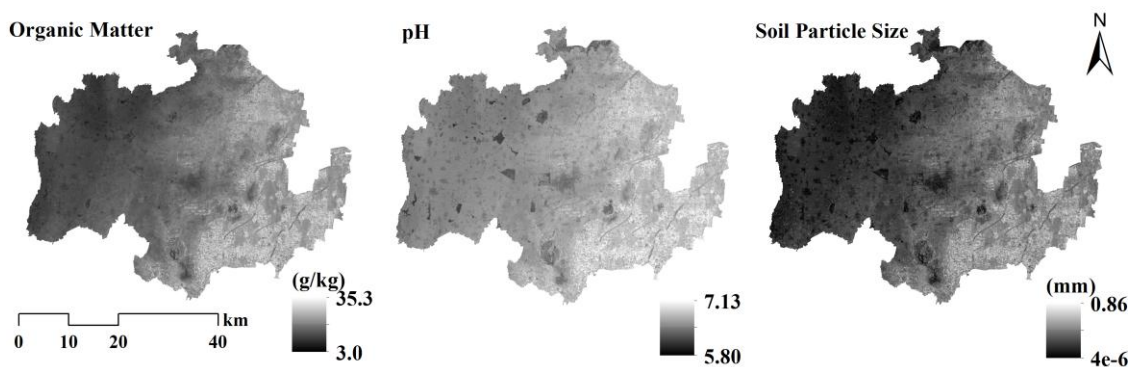

Figure 2. Soil property prediction results of the soil-landscape model

The results indicate that the soil organic matter content gradually decreased spatially from east to west. The pH values were high in the east and west of the study region but low in the centre. The soil particle size showed a similar spatial variation as the organic matter. The soil in the east region was made up of large/small particles, whereas the soil in the west region showed more sandy content.

Geostatistical interpolation

Ordinary Kriging interpolation was used to predict the soil organic matter content, pH, and soil particle size for the 918 sampling locations in the study region, and the results are shown in *Figure 3*.
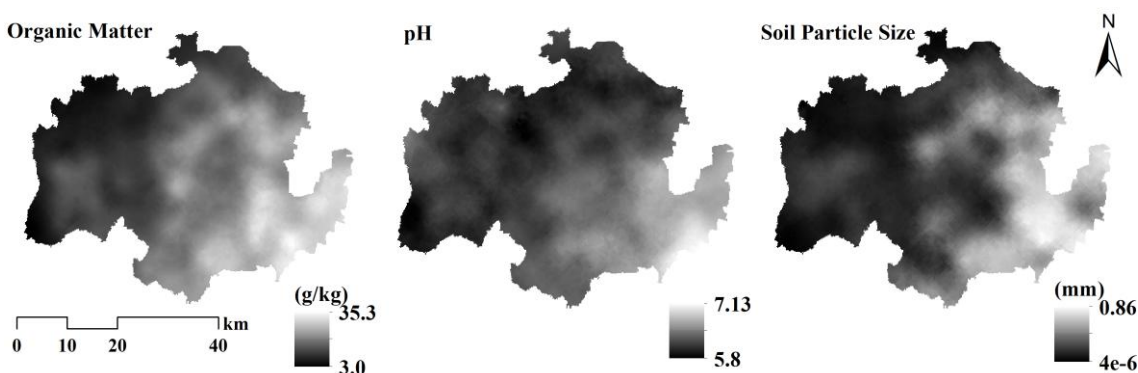

Figure 3. Soil property prediction results of the Kriging interpolation

Based on the spatial distribution of the interpolation results, the overall spatial distribution pattern of the three soil properties were similar between the Kriging interpolation results and soil-landscape model prediction results. However, certain local regions showed significant differences. The Kriging interpolation prediction was based on the characteristics of the data statistics and spatial distribution of the interpolation points. The spatial variation of the soil properties in the interpolation results was smooth, although several regions showed a significant 'bull's eye' effect. Several extremely high or low property data values strongly affected the interpolation results, which is expected when using this geostatistical interpolation method.

Accuracy errors of the two prediction methods

Soil-landscape model

350 sampling points were used to construct a test set and then quantitatively analysed the accuracy of the soil property prediction results in relation to the measured values.

The error analysis of the soil property prediction results is shown in *Table 4*. The analysis results show that the difference between the predicted results and measured values of the three soil properties in the paddy fields and dry land was extremely small; therefore, the three soil properties predicted by the model and interpolation method were accurate based on the prediction results. The mean absolute errors of the soil organic matter content, pH, and particle size in the dry land were smaller than those in the paddy fields by 41.18%, 20.00%, and 121.28%, respectively, although the relative error of the prediction results for the organic matter content in the paddy fields was smaller than that in the dry land by 13.33%, indicating that the model prediction results for the organic matter in the paddy field were more accurate than those in the dry land. The relative errors of the pH, soil particle size, and topsoil thickness in the paddy fields were extremely close to those of the dry land. All of the differences presented a significance of 0.01, which indicated that the prediction accuracy of the three soil properties by the prediction models was similar between the paddy fields and dry land.

Table 4. *Error analysis of the prediction results by the soil-landscape models*

Soil property	Paddy field (n = 220)				Dry land (n = 130)			
	Measured value	Predicted value	MAE	MRE	Measured value	Predicted value	MAE	MRE
Organic matter (g/kg)	20.6	21.2	0.24	0.13	12.1	12.6	0.17	0.15
pH	6.7	6.7	0.24	0.04	6.4	6.4	0.2	0.03
Soil particle size (%)	38.9	41.8	10.71	0.48	16.8	18.9	4.84	0.47

The relative errors of the 350 test points were interpolated to further study the spatial distribution of the prediction errors for the three soil properties predicted by the soil-landscape models in the whole study region. The results are shown in *Figure 4*.

The spatial distribution of the relative errors of the prediction results indicated that the relative error of the predicted soil organic matter content was high in the dry land in the west; therefore, the model showed low accuracy in this region. The relative error of the prediction results for soil pH value was high in the paddy fields in the centre and east of the study region but low in dry land, and the relative error of the prediction

results for the soil particle size was high in the central study region. Overall, the accuracy of the soil-landscape model predictions for the three soil properties was spatially heterogeneous and showed high values in most regions and low values in certain regions.
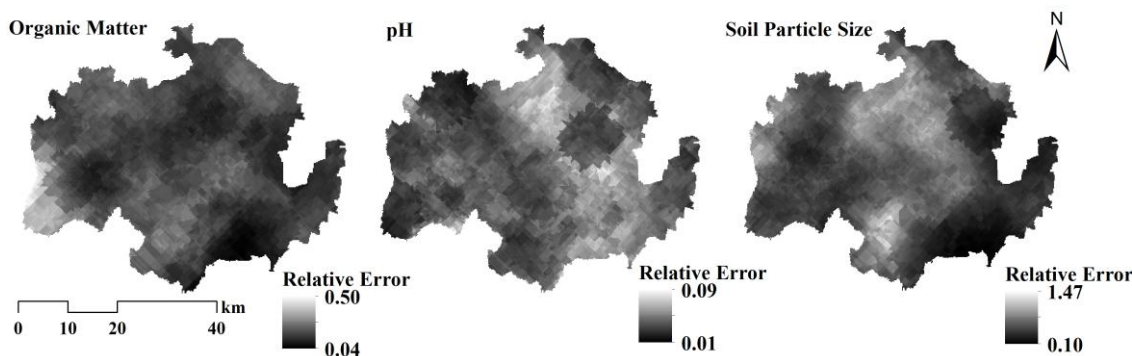

***Figure 4.** Spatial distribution map of the relative error of the soil property predictions by the soil-landscape models*

Geostatistical interpolation

The accuracy of the soil properties predicted by the statistical interpolation was calculated, and the error analysis of the interpolation results is shown in *Table 5*. Among the three soil properties, the MAE and MRE of the interpolation results for the organic matter content and pH value in the paddy fields were slightly smaller than those in the dry land, with the MAE's smaller by 3.57% and 6.90%, respectively, and the MRE's smaller by 40.00% and 20.00%, respectively. These results indicate that the Kriging interpolation results for the organic matter content and pH value in the paddy fields were more accurate than those in the dry land. The MAE and MRE of the interpolation results for the soil particle size in the paddy fields were larger than those in the dry land by 29.17% and 8.47%, respectively, which indicated that the accuracy of the Kriging interpolation results for this soil property was higher in the dry land area.

***Table 5.** Error analysis of the prediction results by the Kriging interpolation*

Soil property	Paddy field (n = 220)				Dry land (n = 130)			
	Measured value	Predicted value	MAE	MRE	Measured value	Predicted value	MAE	MRE
Organic matter (g/kg)	20.6	21.1	0.27	0.15	12.1	13.1	0.28	0.25
pH	6.7	6.7	0.27	0.04	6.4	6.5	0.29	0.05
Soil particle size (%)	38.9	42.2	10.76	0.64	16.8	21.0	8.33	0.59

A kriging interpolation was conducted on the relative errors of the 350 test points to further study the spatial distribution of the prediction errors of the three soil properties predicted by the Kriging interpolation for the entire study region. The results are shown in *Figure 5*.

Based on the spatial distribution of the relative error of the prediction results, the relative error of the predicted soil organic matter content and pH value was high in the west dry land region. The accuracy of the results predicted by the Kriging interpolation for the organic matter content and pH value in this region was low. The relative error of the interpolation results for the soil particle size was low in the central study region, whereas the accuracy of the predicted soil particle size was high in this region.
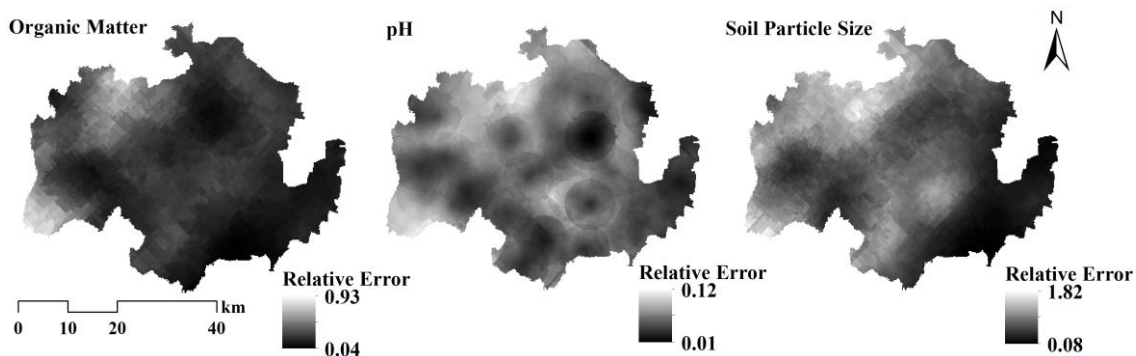

Figure 5. Spatial distribution map of the relative errors of the soil properties predicted by the Kriging interpolation

Comparison between the accuracy of the two prediction results

The comparison between the errors of the predicted soil properties by the two prediction methods is shown in *Table 6*. The results indicate that the MAE and MRE of the soil-landscape model predictions for the soil organic matter, pH value, and soil particle size were all smaller than those of the Kriging interpolation, indicating that the overall accuracy of the three soil properties predicted by the soil-landscape model was higher than that of the Kriging interpolation.

Table 6. Error comparison of the results predicted by the soil-landscape model and Kriging interpolation

Soil property	Soil-landscape model				Kriging interpolation			
	Measured value	Predicted value	MAE	MRE	Measured value	Predicted value	MAE	MRE
Organic matter (g/kg)	17.0	18.0	0.21	0.14	17.5	18.1	0.27	0.19
pH	6.6	6.6	0.23	0.03	6.6	6.6	0.28	0.04
Soil particle size (%)	30.7	33.3	8.53	0.48	30.7	34.3	9.86	0.63

The spatial distribution of the errors of the two prediction results (the MRE of the soil-landscape model - the MRE of the Kriging interpolation) were subtracted and further compared the spatial distribution of the accuracy of the two prediction results. The regions in which the MRE difference was less than 0 represent regions where the soil-landscape model was more accurate, and regions in which the MRE difference was higher than 0 represent regions where the Kriging interpolation was more accurate. The prediction accuracy of the three soil properties were partitioned based on these results,

which is shown in *Figure 6*. In the figure, GIP represents geostatistical interpolation optimum and SLP represents soil-landscape optimum.
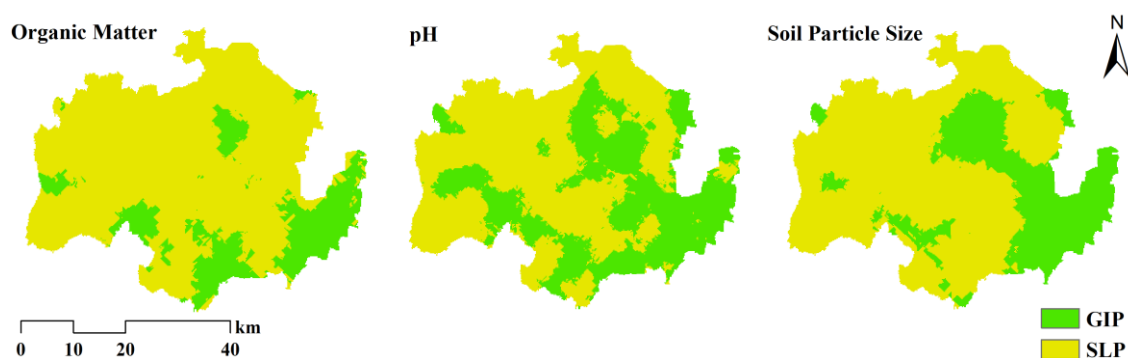

Figure 6. *Partitioned prediction accuracy of the soil-landscape model and geostatistical interpolation (the yellow colour represents areas where the prediction accuracy of the soil-landscape model is better, and the green colour represents areas where the prediction accuracy of the geostatistical interpolation is better)*

Based on the partitioned accuracy map, the prediction accuracy for organic matter, pH value, and soil particle size by the soil-landscape model was higher than that of the geostatistical interpolation method in most regions. For organic matter, the regions in which the soil-landscape prediction model was more accurate were mainly in the centre and west of the study region and constituted 80.59% of the total area. The regions in which the geostatistical interpolation was more accurate were mainly in the southeast region. For the pH value, the regions in which the soil-landscape prediction model was more accurate were mainly in the centre and west of the study region and constituted 60.82% of the total area. The regions in which the geostatistical interpolation was more accurate were distributed in the east and southwest regions. For the soil particle size, the regions in which the soil-landscape prediction model was more accurate were also mainly in the centre and west regions and constituted 70.58% of the total area. The regions in which the geostatistical interpolation was more accurate were mainly in the east region. Therefore, the soil-landscape model was more accurate for the three studied properties in the centre and west of the study region, and the relative error of the prediction results was smaller than that of the geostatistical interpolation by more than 20%. In the east study region, however, the prediction accuracy of the geostatistical interpolation was slightly higher than that of the soil-landscape model, and the relative error of the prediction results was smaller than that of the soil-landscape model by approximately 10%.

Discussion

Comparison with previous study on soil properties prediction

As mentioned before, Donghai County belongs to hilly region in Jiangsu Province, which usually makes it difficult to predict soil properties. To solve this problem, Former researchers took data from terrain attributes to raise the accuracy of predictions. Compared with their work such as results form Florinsky et al. (2002), higher R^2 can be achieved (0.561 vs 0.37) is this research. It is mainly because our soil-landscape model

use terrain data in addition with landscape data such as TWI and NDVI. In addition. Human activity data is taken consideration in this research. This finding shows that landscape data and human activity data can improve the performance of soil prediction model in hilly regions.

Accuracy comparision between soil properties

In this research, it can be found that no matter what method is used, the prediction error of soil particle size is greater than that of organic matter and pH. It can be seen from *Table 1* that the variation coefficient of soil particle size is large yet variabilitys of soil organic matter and ph are small. At the same time, soil particle size is affected by precipitation, agricultural activities as well as organisms (Tyler and Wheatcraft, 1992), far r exceeding the coverage of the factors constructed in this study. While the physical and chemical processes of soil organic matter and ph are clearer and the changes are more stable. The above reasons explain why the MAE and MRE of soil particle size is always slightly worse than soil organic matter and pH.

Error analysis of two prediction methods

When it comes to the results from soil-landscape model, results indicated that the MAE and MRE of the three soil properties were relatively small. The prediction accuracy of the soil properties by the prediction model was good, which verified and further extended the stepwise linear relationship between the soil organic carbon content and pH value and the gradient angle observed by previous studies (Mishra et al., 2009). The mean absolute errors of the soil pH and particle size in the dry land region were all less than those in the paddy fields, which is mainly because paddy fields and dry land present different types of human activities. The differences among the three soil properties were significant. The dry land region is in the west hilly regions of the study region, and values for the three soil properties were significantly lower in this region compared with that in the paddy fields. Therefore, under the same error conditions, the absolute errors in the paddy fields will be naturally higher than those in the dry land. The relative error of the organic matter content in the paddy field was smaller than that in the dry land, which indicated that the prediction results for the organic matter in the paddy field by the prediction models were more accurate than those in the dry land. External factors (e.g., land management measures, such as fertilizing) reduce the effect of internal factors (e.g., soil-forming factors, such as soil-forming parent material) (Yasrebi et al., 2009). Therefore, the present study introduced land-use types to represent the effect of external factors to improve the model prediction accuracy.

When it comes to the results from kriging, the MAE and MRE of the interpolation results for the organic matter content and pH value in the paddy field were slightly smaller than those in the dry land, which indicated that the Kriging interpolation results for the organic matter and pH value in the paddy field were more accurate than those in the dry land. The MAE and MRE of the interpolation results for the soil particle size in dry land were smaller than those in the paddy field, indicating that the Kriging interpolation results for this soil property were more accurate for the dry land region. The main reason for this result is that the organic matter and pH value in the paddy fields were high and the spatial variance was weak; therefore, the prediction results by the Kriging interpolation were better. The dry land region presented the opposite results, with the organic matter and pH values showing significant spatial variations; thus, the

accuracy of the interpolation prediction was low. The overall soil particle size in the dry land was low and presented small variations and weak spatial variance; thus, the accuracy of the interpolation prediction was good. These results are consistent with the findings of Yasrebi et al. (2009), who indicated that the prediction accuracy of Kriging interpolations was generally affected by the variation coefficient.

Comparison between two prediction methods

As proposed by previous studies, soil property prediction methods that consider topographic properties predict more smooth values and show a reduced effect of outliers on the prediction performance (Junxiao et al., 2016). The soil-landscape model in this study considered topographic factors. Thus, when comparing the accuracy of the soil property predictions based on the soil-landscape model and geostatistical interpolation, the accuracy of the prediction results by the soil-landscape model was overall higher than that of the geostatistical interpolation results.

Specifically, for the organic matter, pH, and soil particle size, the regions in which the soil-landscape prediction model was more accurate were all concentrated in the centre and west of the study region, whereas the regions in which the geostatistical interpolation was more accurate were mainly in the east region, which is because the centre and west of the study region are hilly regions with large topographic variations. The landscape factors in the soil-landscape model reflected this topographic difference, whereas the geostatistical method only predicted the soil properties based on data similarity. Therefore, the prediction results by the soil-landscape model in the hilly region were more accurate compared with the prediction results by geostatistical interpolation. Inversely, the east study region is a flat plain, and the soil properties were relatively homogenous. Therefore, the prediction accuracy of the geostatistical interpolation was better than that of the soil-landscape model.

Conclusions

The following specific conclusions can be reached.
- The study object was farmland soil in which human activities have strongly affected soil properties. Therefore, when the soil-landscape model was constructed, land-use types were added to represent the effect of human activities on the soil properties. The study results indicated that in the hilly region with large topographic variations, introducing the human activity variable improved the soil property predictions.
- The study region belongs to the Huanghuaihai hilly region. A comparison of the accuracy of the soil property predictions based on the soil-landscape model and geostatistical interpolation in the study region showed that the soil-landscape model was more accurate than the geostatistical interpolation. Specifically, for the organic matter content, pH, and soil particle size, the regions in which the soil-landscape prediction model was more accurate were all concentrated in the central and west hilly regions in the study region, whereas the regions in which the geostatistical interpolation was more accurate were mainly in the east plain region. Therefore, the prediction results by the soil-landscape model in the hilly region were more accurate compared to the prediction results by geostatistical interpolation.

However, this study attempted to introduce a human activity variable into soil-landscape models to improve prediction accuracy; therefore, the variables in this model must be selected more rigorously, comprehensively, and appropriately. For example, taking into account the geo-environmental factors of the soil itself, the influence of human activities and the similar area with the same topography in the study area, we need to select variables that are suitable to local conditions. These issues require further study.

REFERENCES

[1]   Barling, R. D., Moore, I. D., Grayson, R. B. (1994): A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. – Water Resources Research 30: 1029-1044.

[2]   Brus, D. J., Heuvelink, G. B. M. (2007): Optimization of sample patterns for universal kriging of environmental variables. – Geoderma 138: 86-95.

[3]   Carter, M. R. (1993): Soil Sampling and Methods of Analysis. – CRC Press, Boca Raton, FL.

[4]   Cressie, N. (1988): Spatial prediction and ordinary kriging. – Mathematical Geology 20: 405-421.

[5]   Dent, D., Young, A. (1981): Soil Survey and Land Evaluation. – Allen & Unwin, London, UK.

[6]   Ferreyra, R. A., Apezteguía, H. P., Sereno, R., Jones, J. W. (2002): Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. – Geoderma 110: 265-289.

[7]   Florinsky, I. V., Eilers, R. G., Manning, G. R., Fuller, L. G. (2002): Prediction of soil properties by digital terrain modelling. – Environmental Modelling & Software 17: 295-311.

[8]   Gessler, P. E., Chadwick, O. A., Chamran, F., Althouse, L., Holmes, K. (2000): Modeling soil–landscape and ecosystem properties using terrain attributes. – Soil Science Society of America Journal 64: 2046-2056.

[9]   Grunwald, S. (2016): Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. – CRC Press, Boca Raton, FL.

[10]   Gunstensen, A. K., Rothman, D. H., Zaleski, S., Zanetti, G. (1991): Lattice Boltzmann model of immiscible fluids. – Physical Review A 43: 4320.

[11]   Hack, J. T. (1973): Stream-profile analysis and stream-gradient index. – Journal of Research of the US Geological Survey 1: 421-429.

[12]   Hengl, T., Rossiter, D. G. (2003): Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. – Soil Science Society of America Journal 67: 1810-1822.

[13]   Jenny, H. (1941): Factors of Soil Formation: A System of Quantitative Pedology. – McGraw-Hill, New York, NY.

[14]   Jones, M. J. (1973): The organic matter content of the savanna soils of West Africa. – Journal of Soil Science 24: 42-53.

[15]   Kerry, R., Oliver, M. A. (2004): Average variograms to guide soil sampling. – International Journal of Applied Earth Observation and Geoinformation 5: 307-325.

[16] Krige, D. G. (1951): A statistical approach to some basic mine valuation problems on the Witwatersrand. – Journal of the Southern African Institute of Mining and Metallurgy 52: 119-139.

[17] Minasny, B., McBratney, A. B. (2006): A conditioned Latin hypercube method for sampling in the presence of ancillary information. – Computers Geosciences 32: 1378-1388.

[18] Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D., Van Meirvenne, M. (2009): Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. – Soil Science Society of America Journal 73: 614.

[19] Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J. (1995): Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. – Geoderma 67: 215-226.

[20] Roddier, F. (1988): Curvature sensing and compensation: a new concept in adaptive optics. – Applied Optics 27: 1223-1225.

[21] Simbahan, G. C., Dobermann, A. (2006): Sampling optimization based on secondary information and its utilization in soil carbon mapping. – Geoderma 133: 345-362.

[22] Townshend, J. R., Justice, C. O. (1986): Analysis of the dynamics of African vegetation using the normalized difference vegetation index. – International Journal of Remote Sensing 7: 1435-1445.

[23] Tyler, S. W., Wheatcraft, S. W. (1992): Fractal scaling of soil particle-size distributions: analysis and limitations. – Soil Science Society of America Journal 56: 362-369.

[24] van Groenigen, J. W., Pieters, G., Stein, A. (2000): Optimizing spatial sampling for multivariate contamination in urban areas. – Environmetrics 11: 227-244.

[25] Wang, J., Wang, X., Zhou, S., Wu, S., Zhu, Y., Lu, C. (2016): Optimization of sample points for monitoring arable land quality by simulated annealing while considering spatial variations. – International Journal of Environmental Research and Public Health 13: 980.

[26] Xia, J., Li, T., Deng, L. (2000): The application of the principal component analysis method in quality evaluation of cultivated land. – Southwest China Journal of Agricultural Sciences 13: 51-55.

[27] Yasrebi, J., Saffari, M., Fathi, H., Karimian, N., Moazallahi, M., Gazni, R. (2009): Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some soil chemical parameters. – Research Journal of Biological Sciences 4: 93-102.

[28] Yuan, X., Zhao, G., Zhu, X. (2008): Linkage of evaluation index system for cultivated land fertility evaluation in plain and hill regions. – Transactions of Chinese Social Agriculture Engineering 7: 65-71.

[29] Zevenbergen, L. W., Thorne, C. R. (1987): Quantitative analysis of land surface topography. – Earth Surface Processes and Landforms 12: 47-56.

[30] Zhou, B., Wang, F. (2004): Spatial prediction of soil properties based on decision tree modeling. – Chinese Journal of Soil Science 35: 385-390.

[31] Zhu, A. X. (2008): Fine Digital Soil Survey Model and Method. – Science Press, Beijing.