Chin. Geogra. Sci. 2017 Vol. 27 No. 4 pp. 552–568
doi: 10.1007/s11769-017-0887-5

Springer Science Press
www.springerlink.com/content/1002-0063

Equality Testing for Soil Grid Unit Resolutions to Polygon Unit Scales with DNDC Modeling of Regional SOC Pools

YU Dongsheng[1, 2], PAN Yue[1, 2], ZHANG Haidong[1], WANG Xiyang[1, 2], NI Yunlong[1], ZHANG Liming[3], SHI Xuezheng[1]

(1. *State Key Laboratory of Soil and Sustainable Agriculture*, *Institute of Soil Science*, *Chinese Academy of Sciences*, *Nanjing* 210008, *China*; 2. *University of Chinese Academy of Sciences*, *Beijing* 100393, *China*; 3. *College of Resource and Environment*, *Fujian Agriculture and Forestry University*, *Fuzhou* 350002, *China*)

Abstract: Matching soil grid unit resolutions with polygon unit map scales is important to minimize the uncertainty of regional soil organic carbon (SOC) pool simulation due to their strong influences on the modeling. A series of soil grid units at varying cell sizes was derived from soil polygon units at six map scales, namely, 1∶50 000 (C5), 1∶200 000 (D2), 1∶500 000 (P5), 1∶1 000 000 (N1), 1∶4 000 000 (N4) and 1∶14 000 000 (N14), in the Taihu Region of China. Both soil unit formats were used for regional SOC pool simulation with a DeNitrification-DeComposition (DNDC) process-based model, which spans the time period from 1982 to 2000 at the six map scales. Four indices, namely, soil type number (*STN*), area (*AREA*), average SOC density (*ASOCD*) and total SOC stocks (*SOCS*) of surface paddy soils that were simulated by the DNDC, were distinguished from all these soil polygon and grid units. Subjecting to the four index values (*IV*) from the parent polygon units, the variations in an index value (*VIV*, %) from the grid units were used to assess its dataset accuracy and redundancy, which reflects the uncertainty in the simulation of SOC pools. Optimal soil grid unit resolutions were generated and suggested for the DNDC simulation of regional SOC pools, matching their respective soil polygon unit map scales. With these optimal raster resolutions, the soil grid units datasets can have the same accuracy as their parent polygon units datasets without any redundancy, when *VIV* < 1% was assumed to be a criterion for all four indices. A quadratic curve regression model, namely, $y = -0.80 \times 10^{-6}x^2 + 0.0228x + 0.0211$ ($R^2 = 0.9994$, $P < 0.05$), and a power function model $\hat{R} = 10.394\hat{S}^{0.2153}$ ($R^2 = 0.9759$, $P < 0.05$) were revealed, which describe the relationship between the optimal soil grid unit resolution (y, km) and soil polygon unit map scale (1∶10 000x), the ratio ($\hat{R}$, %) of the optimal soil grid size to average polygon patch size ($\hat{S}$, km^2) and the $\hat{S}$, with the highest R^2 among different mathematical regressions, respectively. This knowledge may facilitate the grid partitioning of regions during the investigation and simulation of SOC pool dynamics at a certain map scale, and be referenced to other landscape polygon patches' mesh partition.
Keywords: soil organic carbon (SOC); soil grid unit resolutions; soil polygon unit map scales; DeNitrification-DeComposition (DNDC) model; SOC pools

Citation: Yu Dongsheng, Pan Yue, Zhang Haidong, Wang Xiyang, Ni Yunlong, Zhang Liming, Shi Xuezheng, 2017. Equality testing for soil grid unit resolutions to polygon unit scales with DNDC modeling of regional SOC pools. *Chinese Geographical Science*, 27(4): 552–568. doi: 10.1007/s11769-017-0887-5

1 Introduction

Soil organic carbon (SOC) is the largest terrestrial carbon pool (Schlesinger, 1997), with stocks approximately four times greater than the biotic pool and approximately three times greater than the atmospheric pool

Received date: 2016-11-23; accepted date: 2017-03-11
Foundation item: Under the auspices of Special Project of National Key Research and Development Program (No. 2016YFD0200301), National Natural Science Foundation of China (No. 41571206), Special Project of National Science and Technology Basic Work (No. 2015FY110700-S2)
Corresponding author: YU Dongsheng. E-mail: dshyu@issas.ac.cn

(Lal, 2004). Relatively modest changes in SOC storage can significantly alter the atmospheric CO_2 concentration (Davidson and Janssens, 2006). Therefore, an accurate SOC pool estimation has become an important requirement to assess the global carbon balance and climate change.

Agricultural soils are a highly sensitive part of the global carbon cycle (Shi *et al*., 2010; Wang *et al*., 2011), carbon sequestration by agricultural soils presents an immediately viable option to increase the soil carbon pool, reducing atmospheric CO_2 and mitigating global warming (Sun *et al*., 2010). Because the complexities of human activities and tillage practices affect agricultural soil (Wu *et al*., 2015), SOC dynamic changes are increasingly simulated over broad space and time scales by process-based models (Giltrap *et al*., 2010; Xu *et al*., 2012), such as DeNitrification-DeComposition (DNDC) (Li *et al*., 2003).

The DNDC model which was developed by Li *et al*. (1992a; 1992b) can simulate C and N biogeochemical cycles that occur in agricultural systems and are driven by both the environmental factors (i.e., soil organic matter, texture, pH, bulk density, hydraulic properties, daily temperatures, precipitation) and management practices (i.e., crops, tillage, fertilization, manure application, grazing). This method has been validated internationally through long-term applications at the plot scale, including many sites in North America, Europe, Asia, *etc*. (Pathak *et al*., 2005; Li *et al*., 2006; Tonitto *et al*., 2007) and is one of the most widely accepted biogeochemical models in the world (Tang *et al*., 2006; Li, 2007; Li *et al*., 2010).

The DNDC model has also been utilized to upscale estimates of SOC pools from the plot to region scale. At the region scale, the initial DNDC modeling that has been conducted has used counties as basic simulation units, where the minimum and maximum soil parameter values for each county were derived from soil maps to simulate an upper and lower estimate of several C and N pools (Cai *et al*., 2003; Li *et al*., 2004). However, county scale model simulations are subject to great uncertainties as soil properties are averaged for each county, largely ignoring the nonlinear impacts of soil heterogeneity therein (Rüth and Lennartz, 2008; Zhang *et al*., 2014a).

For recent DNDC up-scaled utilization, a region is partitioned into many simulation units, within which all soil properties are assumed to be as homogeneous as they are at the plot scale (Li *et al*., 2005, Zhang *et al*., 2012). This homogeneity assumption is a possible major source of error when extending DNDC modeling from the plot to the region scale (Li *et al*., 2002; 2004). As the area of the basic simulation unit increases, so does the soil property variability or heterogeneity (Qin *et al*., 2015), calling into question the accuracy and uncertainty of its capture (Smith and Dobbie, 2001; Bouwman *et al*., 2002).

Soil polygons that are derived from soil vector maps are used as basic simulation units, which is one way to reduce the effects of soil heterogeneity on DNDC modeling as much as possible (Zhang *et al*., 2012; Xu *et al*., 2013; Yu *et al*., 2013). Even so, the soil heterogeneity within a soil polygon unit still exists and depends on the soil vector map scale, with a smaller map scale resulting in higher heterogeneity (Yu *et al*., 2013). For different broad regions, the multi-scales of the polygon units for DNDC modeling range widely from 1∶50 000 to 1∶14 000 000 and greatly affect the accuracy and uncertainty of the modeling (Xu *et al*., 2011; 2013; Yu *et al*., 2013; Zhang *et al*., 2014a).

Another way to reduce the effects of the soil heterogeneity on the DNDC modeling is to use soil grid cells as the basic simulation units (Huang *et al*., 2004; Yu *et al*., 2007c; Shi *et al*., 2010; Yu *et al*., 2011). The cell size or resolution of the soil grid units is one ruler with which to scale the soil heterogeneity, with lower resolution or larger cell size resulting in higher intracellular heterogeneity. The cell size or resolution also greatly affects the accuracy and uncertainty of the soil grid unit simulation with DNDC (Yu *et al*., 2011; Pogson and Smith, 2015).

Soil grid units are more often applied in the simulation of SOC pools (Qiu *et al*., 2004; Tang *et al*., 2006; Liu *et al*., 2011; Yu *et al*., 2011) as they are more easily manipulated for spatial model simulation, geo-statistics and spatial analysis than soil polygon units (Huang *et al*., 2004; Li *et al*., 2005). Soil grid units are often derived from soil polygon units by data conversion, but the grid resolution choice varies by researcher even if the soil polygon units have the same map scale and fall within same region (Yu *et al*., 2007c; Shi *et al*., 2010; Yu *et al*., 2012). For example, the soil polygon units that were compiled in the Soil Database of China (Yu *et al*., 2007a) at the map scale of 1∶1 000 000 have been con-

converted into grid units with resolutions of 1 km × 1 km (Yu *et al*., 2007b) and 10 km × 10 km (Yu *et al*., 2007c; Yu *et al*., 2012) to simulate and estimate agricultural SOC pools in China, respectively. Soil grid units with resolutions of 2 km × 2 km (Shen *et al*., 2003) and 50 km × 50 km (Wan *et al*., 2011), which were converted from soil polygon units at an original map scale of 1∶4 000 000, were used for the grid simulation of SOC dynamics in different regions. Soil polygon units at an original map scale of 1∶50 000 were converted into grid units with resolutions of 100 m × 100 m (Shi *et al*., 2010) and 30 m × 30 m (Su *et al*., 2012) for the grid simulation of SOC dynamics in agro-ecosystems.

Our concern is whether these soil grid units at different cell sizes have equivalent accuracy or granularity to their parent soil polygon units for DNDC modeling; in other words, whether these soil grid unit datasets regulate coarser data or contain redundant data, in contrast to their parent soil polygon unit dataset at a certain map scale. A coarser or redundant dataset affects the inner homogeneity of the simulation unit's soil properties and the common outcome, because the modeling error will be lower if all the features within the simulation unit are more homogeneous (Cai *et al*., 2003; Yu *et al*., 2011; Yu *et al*., 2013).

In fact, the accuracy and redundancy are two important issues during the dataset conversion of soil simulation units from polygon to grid format, which are often neglected during modeling at the regional scale. The accuracy of the grid unit dataset determines the reliability and uncertainty of SOC grid simulation (Batjes, 2000; Ni, 2001), and the redundancy of the dataset muddles the data accuracy, redundant workload and cost of the simulation (Yu *et al*., 2011; Yu *et al*., 2013). Some researchers have focused on data accuracy but neglected the data redundancy (Yu *et al*., 2007b; Shi *et al*., 2010), while others neglected the data accuracy (Batjes, 2000; Yu *et al*., 2007c) when conducting the dataset conversion, always searching for an individual solution in every case.

Given the variety of datasets and number of simulations in combination with data accuracy, redundancy and computational costs (Schmidt *et al*., 2008), two important questions are raised. How sensitive is DNDC modeling to different simulation units at varying vector map scales or raster grid resolutions? Which raster resolution is optimal for DNDC grid simulation at a fixed soil map scale for error and cost controls? Matching the grid resolutions with polygon map scales in equality of soil simulation units is an essential issue to DNDC grid modeling.

In the present study, paddy soil polygon simulation units at six vector map scales from 1∶50 000 to 1∶14 000 000 were converted to grid simulation units at varied raster resolutions in the Taihu Region of China. Soil organic carbon pools were simulated by polygon and grid simulations with a DNDC model at varying vector map scales and raster resolutions, respectively. The objectives of the study were to: 1) reveal the impact of the vector map scale and raster resolution of soil simulation units on DNDC modeling; 2) determine an optimal raster resolution for grid simulation units at fixed soil vector map scales based on an assessment of the simulation units' data accuracy and redundancy metrics, and 3) construct a relationship between the soil polygon units map scale or average patch size and the optimal raster resolution of grid units for DNDC modeling at the regional scale. The results will serve as a reference for soil simulation unit conversion from the polygon to grid format to support soil carbon cycle modeling at the regional scale.

2 Materials and Methods

2.1 Study area

The Taihu Region (29°56′N–32°16′N, 118°50′E–121°54′E) is located in the middle and lower reaches of the Yangtze River in China and covers a watershed area of 36 500 km^2, including parts of Jiangsu and Zhejiang provinces and the entire Shanghai Municipality administrative area (Fig. 1).

The terrain of the region is generally flat plains that are broken by a high amount of rivers. A northern subtropical monsoon climate prevails in the region, with a mean annual temperature and precipitation of 16℃ and 1100–1400 mm, respectively (Xu *et al*., 1980). The soil types in the region are mainly paddy, fluvo-aquic and red soils, which cover 90% of the total area. Paddy soils, which are the largest single proportion of any soil type in the Taihu Region, occupy 23 200 km^2, approximately 66% of the total area (Yu *et al*., 2014). Derived from loess, alluvium and lacustrine deposits, the paddy soils in the Taihu Region are recognized as the most typical of their type in China (Li, 1992), with a long history of

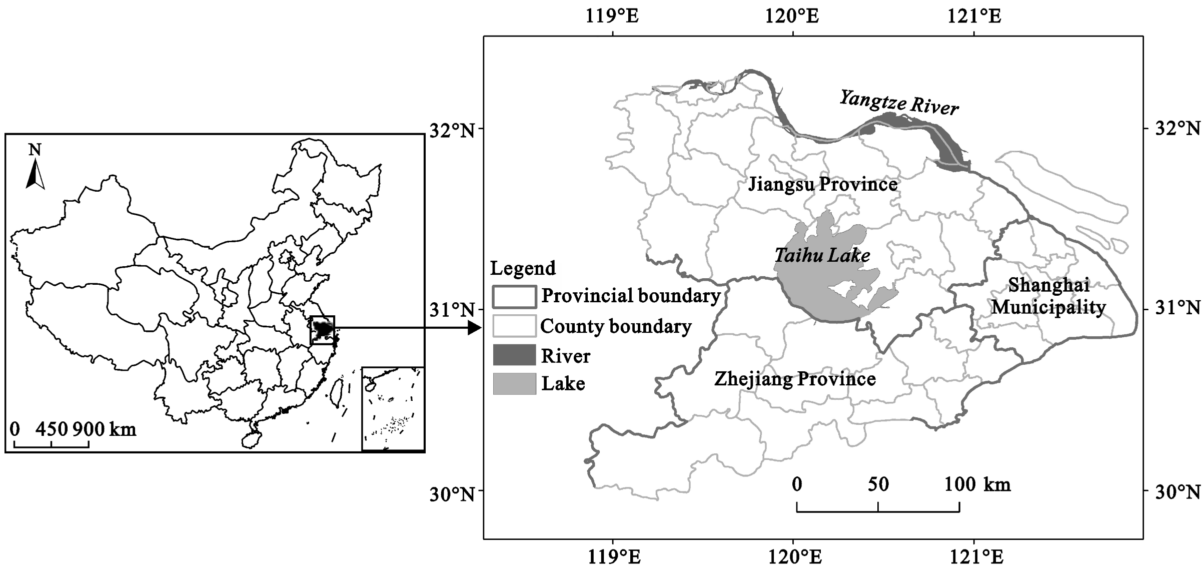

Fig. 1 Location of the Taihu Region in China

rice cultivation spanning several centuries (Xu *et al.*, 1980). Six soil subgroups, namely, bleached, gleyed, percogenic, degleyed, submergenic and hydromorphic, are included in the paddy soils. These subgroups are cross-referenced in US Soil Taxonomy (ST) as typic epiaquepts (bleached, percogenic, hydromorphic) and typic endoaquepts (gleyed, degleyed, submergenic) (Soil Survey Staff, 1994; Shi *et al.*, 2006). A double crop rotation has been managed with summer rice (planted in June and harvested in October) and winter wheat (planted in November and harvested in May); the cultivated soil layer often reaches to a depth of 20 cm in the region (Zhang *et al.*, 2012; Zhang *et al.*, 2014a).

2.2 Development of polygon and grid simulation unit datasets at different map scales

First, paddy polygon unit datasets for DNDC simulation at six soil vector map scales, e.g., 1∶50 000 (C5), 1∶200 000 (D2), 1∶500 000 (P5), 1∶1 000 000 (N1), 1∶4 000 000 (N4) and 1∶14 000 000 (N14), were developed in the Taihu Region. These datasets were generated by vector overlays from paddy polygons in the datasets and polygons that depict county boundaries at a scale of 1∶50 000, respectively. All the simulation units of the paddy polygons at one certain map scale within one county have the same feature input value for DNDC modeling such as crops, agricultural management and climate, except for soil features such as soil types, soil organic matter content, clay content, bulk density, rock fragments content, soil layer thickness, pH, and hydraulic properties (Yu *et al.*, 2013).

The paddy polygon unit datasets at the six map scales were developed by a Geographic Information System (GIS) Linkage technique based on soil type (Yu *et al.*, 2005; Yu *et al.*, 2007a; Yu *et al.*, 2007b), namely, the Pedological Knowledge Based (PKB) method (Zhao *et al.*, 2006), from soil vector maps at their corresponding map scales. The soil vector maps were compiled by using a standard soil mapping system that was formulated from the Second National Soil Survey of China, which was conducted in the 1980s (Office for the Second National Soil Survey of China, 1994). The three paddy polygon unit datasets C5, D2, and P5 are representatives of digital maps at regional scales, describing soil features at the county, district and province levels, respectively. The other three paddy polygon unit datasets N1, N4 and N14 all describe soil features at national scales (Yu *et al.*, 2013).

For the soil mapping at the six different map scales, the soil species is the basic mapping unit for C5 and D2, that for P5 and N1 is the soil family, and that for N4 and N14 is the soil subgroup (Yu *et al.*, 2014). The soil properties attributed to all the paddy polygons were derived from representative soil profiles, which were surveyed, compiled and authorized in the Second Soil Survey of China in the 1980s (Shi *et al.*, 2006). The number of soil profiles attributed to paddy polygons at the C5, D2 and P5 scales totaled 1107, 136 and 127, which were taken from the three books: 'Soils of County', 'Soils of District' and 'Soils of Province', respectively. The soil properties attributed to the paddy polygons at the national map scale (N1, N4 and N14) originated from the 49 soil profiles described in the book 'Soils of China' (Shi *et al*., 2006; Yu *et al*., 2014).

Second, paddy grid unit datasets were developed from the above paddy polygon unit datasets at the six map scales for DNDC simulation. Each polygon unit dataset was converted into a series of grid unit datasets with different grid cell sizes. The grid cell ranged from a default to a maximum size, with the size increment set to approximately 10% of the default. The polygon unit dataset at the N14 scale was also converted to grid unit datasets at different cell sizes ranging from the default to a minimum size, with an approximate decrement of 10% of the default. The default was determined by the soil vector map scale and the lowest mapping unit size (2 mm × 2 mm) described and exhibited in a hard copy of the map (Yu *et al*., 2014). For the conversions of the six polygon unit datasets (C5, D2, P5, N1, N4 and N14), the default grid cell sizes are 100 m, 400 m, 1 km, 2 km, 8 km and 28 km, respectively. The maximum grid cell sizes were that at which the difference in the paddy SOC pool that was simulated by DNDC with the grid unit dataset exceeds the simulation from their parent polygon unit dataset by 30%, and the minimum grid cell size was that at which the difference is less than 0.8%. All the data conversions were conducted using the Polygons to Raster Conversion Tools (PRCT), a component of the ArcGIS 9.0 software, with the grid cell value assignment type option of Maximum-Area (Yu *et al*., 2014).

Finally, the attributes of all the simulation units that were rendered as vector (polygon unit) and raster (grid unit) datasets, inluding the soil properties, daily weather, cropping systems, and agricultural management practices, are required to initialize the DNDC model at regional scale (Yu *et al*., 2011; Yu *et al*., 2013). Each simulation unit has its own data records, which were used as the input and output for the DNDC modeling of SOC dynamics (Zhang *et al*., 2009a; Zhang *et al*., 2009b; Zhang *et al*., 2014a).

2.3 DNDC modeling and validation

The DNDC (DeNitrification-DeComposition) model is a process-based model of carbon (C) and nitrogen (N) biogeochemistry in agro-ecosystems (Li *et al*., 1992a; 1992b), and can simulate soil C and N biogeochemical cycles in paddy rice ecosystems, depending on the series of anaerobic processes that are supplemented in the model (Li *et al*., 2002; Li *et al*., 2004; Li, 2007). Soil organic matter, clay content, pH and soil bulk density are sensitive parameters that serve as input for SOC simulation with DNDC (Li *et al*., 2002; Levy *et al*., 2007; Zhang *et al*., 2012; Zhang *et al*., 2014a).

Farming management scenarios were compiled for DNDC modeling of SOC dynamics based on the five assumptions from Zhang *et al*. (2009a; 2009b; 2012; 2014a), which did not vary with the soil simulation unit within a county. The five assumptions were as follows: 1) nitrogen fertilizer consisted of 40% urea, 40% NH_4HCO_3, and 20% $NH_4H_2PO_3$; 2) 15% of the above-ground crop residue was returned to the soil; 3) 20% livestock waste and 10% human waste were added as manure to the soil; 4) one midseason drainage event and shallow flooding event were each applied to summer rice; and 5) for the rice-wheat rotation, tillage was conducted twice before 1990 at 20 cm for rice and 10 cm for wheat on the planting dates; no-tilling was applied for wheat after 1990.

The DNDC modeling spans the time period from 1982 to 2000, a duration of 19 years. A total of 65 340 paddy polygon unit simulations and a great deal of paddy grid unit simulations were executed. DNDC version 9.1 was run in the present study.

To validate and assess the performance of DNDC modeling, the observed values of SOC content, which were acquired in 2000 from 1033 soil sampling sites within paddy polygon units at the C5 map scale, were used against the modeling values from this year (Zhang *et al*., 2014a). The observed SOC content of the top layer (0–15 cm) varied from 1.9 to 36 g/kg, and the simulated SOC ranged from 5.1 to 34 g/kg in 2000, where 99.6% of the simulated polygon units in C5 were within the ranges produced by the observed values. Four statistical criteria, namely, the correlation coefficient (r), the relative error (E), the mean absolute error (MAE) and the root mean square error ($RMSE$), were employed to evaluate the model's performance. The r was 0.5 at a significance level $P < 0.01$, the E was 6.4%, the MAE was 4.0 g/kg and the $RMSE$ was 5.0 g/kg, all indicating that the modeled results were encouragingly consistent with the observations and the DNDC model was acceptable for the SOC modeling of paddy soils in the

Taihu Region (Zhang *et al*., 2014a). A more complete discussion of DNDC model validation and error assessment for the region can be observed in Zhang *et al*. (2012; 2014a).

2.4 Data calculation and analysis

The simulated SOC density (*SOCD*, kg C/m^2) of a paddy polygon or grid unit (j) is calculated according to the following equation (Yu *et al*., 2014):

$$SOCD_j = \sum_{i=1}^{n}(1 - \delta_i\%) \times \rho_i \times C_i \times T_i \ / \ 100 \qquad (1)$$

where n is the number of soil pedogenic layers, δ_i% represents the volumetric percentage of the > 2 mm fraction (rock fragments), ρ_i is the bulk density (g/cm^3), C_i is the simulated soil organic C content (g/kg) in 2000, and T_i represents the thickness (cm) of the layer i.

The simulated *SOCD* of the surface paddy soil is calculated to a depth of 20 cm, which is the bottom of the cultivated layer in the region (n = 1).

Four indices of the surface paddy soil, namely, the paddy soil area (*AREA*, Mha), the paddy soil type number (*STN*), the simulated SOC stocks (*SOCS*, Tg) and the average *SOCD* (*ASOCD*, kg C/m^2) in 2000, in which the DNDC modeling was validated with 1033 observed SOC content values, were selected to assess the data accuracy between a paddy grid unit dataset and its parent polygon unit dataset. The four index values (*IVs*) that were determined from each polygon unit dataset are recognized as a benchmark for a comparison with the values from their affiliated grid unit datasets. Except for the index value (*IV*) of *STN*, which was obtained by accounting, the *IVs* of *AREA*, *SOCS* and *ASOCD* were calculated as follows:

$$IV\ (AREA) = \sum_{j=1}^{m}(AREA_j) \qquad (2)$$

$$IV\ (SOCS) = \sum_{j=1}^{m}(SOCD_j \times AREA_j) \qquad (3)$$

$$IV\ (ASOCD) = IV\ (SOCS)\ /\ IV\ (AREA) \qquad (4)$$

where $AREA_j$ is the area of the paddy polygon or grid unit; $SOCD_j$ is the simulated SOC density of a paddy polygon or grid unit, and j is the number of paddy polygons or grid units, m is the total number of paddy polygons or grid units (Yu *et al*., 2014).

Variation in an index value (*VIV*, %), which were obtained from a grid unit dataset (IV_{grid}) and its parent polygon unit dataset ($IV_{polygon}$), are recognized as a ruler to scale the magnitude of the consistency between the two datasets. The accuracy of the two datasets formats may be consistent or identical, but only if the absolute values of all these indices *VIV*s are less than 1% (Yu *et al.,* 2014). The VIV is calculated as follows:

$$VIV\ (\%) = \text{abs}(100 \times (IV_{grid} - IV_{polygon})\ /\ IV_{polygon}) \qquad (5)$$

where $IV_{polygon}$ is an index value obtained from a polygon unit dataset, and IV_{grid} is the index value obtained from an affiliated grid unit dataset.

The optimal soil grid unit size for a polygon unit dataset conversion to grid unit dataset is the maximum grid cell size to which the two datasets are scaled identically.

3 Results

3.1 Variations in input soil parameters among simulation unit datasets

The spatial distribution characteristics of the input soil parameters, such as SOC and clay content, pH and soil bulk density, which are depicted by various simulation unit datasets differ from each other (Table 1, Table 2). For instance, the mean values of the initial SOC content at different map scales are different. The smallest mean value is 15.69 g/kg at N4 scale and the largest mean value is 33.38 g/kg at N14 scale. The CV of the initial SOC content is the highest at N1 scale and the CV is the lowest at N14 scale (Table 2). The mean SOC values of the grid units at different resolutions are all smaller than the values of their parent polygon units at the map scales of C5, P5 (Table 1) and N1 (Table 2) but not for the map scales of D2 (Table 1), N4 and N14 (Table 2). In general, the mean SOC value and CV of the grid units varies with the grid cell size, CVs of pH and bulk density are less than that of clay and SOC depicted by various simulation unit datasets.

3.2 Index values determined from simulation polygon units at different map scales

The basic mapping unit type, paddy soil type number, polygon unit number and soil area, which were determined from the six paddy polygon unit datasets at different map scales, describe the physical characteristics of these soil datasets, differ from each other (Table 3).

Table 1 Soil parameter inputs from different resolution units at regional map scales in the Taihu Region of China

Map scale	Simulation unit (km)	Clay (%)		pH		SOC (g/kg)		Bulk density (g/cm^3)	
		Mean	CV	Mean	CV	Mean	CV	Mean	CV
C5 (1∶50 000)	Polygon	29.00	37.07	6.65	9.77	16.81	33.26	1.18	10.17
	0.1	29.01	37.06	6.67	9.75	16.79	33.34	1.18	10.17
	0.2	28.85	37.54	6.67	9.90	16.56	33.54	1.18	10.17
	0.3	28.35	38.77	6.67	10.04	16.30	33.45	1.19	10.08
	0.4	27.71	40.24	6.67	10.19	16.00	33.38	1.20	10.00
	0.5	27.20	41.62	6.67	10.34	15.74	33.43	1.20	10.00
	1	25.61	44.01	6.71	10.88	14.94	33.66	1.22	9.84
	2	24.55	44.77	6.74	11.13	14.46	34.14	1.22	9.84
	3	24.45	44.09	6.77	10.78	14.33	34.52	1.22	9.84
D2 (1∶200 000)	Polygon	26.77	40.79	6.96	10.06	29.42	33.92	1.16	8.62
	0.4	26.64	40.99	6.96	10.20	29.50	33.86	1.16	8.62
	0.5	26.72	41.65	6.93	10.25	29.37	33.95	1.16	8.62
	0.6	26.75	42.21	6.90	10.43	29.14	33.87	1.16	9.48
	0.7	26.71	43.13	6.88	10.47	29.03	34.14	1.16	9.48
	0.8	26.67	43.94	6.86	10.64	28.93	34.39	1.16	9.48
	1	27.20	39.74	6.99	9.87	28.89	33.16	1.16	8.62
	2	27.13	38.85	7.00	10.00	28.46	31.59	1.16	8.62
	4	26.80	41.75	6.95	10.22	28.45	29.77	1.16	9.48
P5 (1∶500 000)	Polygon	25.26	44.02	6.97	10.47	17.50	32.85	1.16	8.62
	1	25.61	47.36	6.79	11.63	17.06	34.12	1.17	10.26
	2	25.63	49.63	6.67	11.99	16.83	34.46	1.18	11.02
	3	25.84	50.23	6.67	12.29	16.70	35.02	1.19	10.92
	4	26.31	50.29	6.63	12.37	16.67	36.77	1.19	10.08
	5	26.56	49.10	6.63	12.22	17.06	34.53	1.20	10.83
	6	26.86	50.22	6.59	12.14	17.06	34.87	1.20	10.83

Notes: Mean: mean value; CV: coefficient of variation

Table 2 Soil parameter inputs from different resolution units at national map scales in the Taihu Region of China

Map scale	Simulation units (km)	Clay (%)		pH		SOC (g/kg)		Bulk density (g/cm^3)	
		Mean	CV	Mean	CV	Mean	CV	Mean	CV
N1 (1∶1 000 000)	Polygon	27.01	35.99	6.56	10.98	29.01	37.61	1.15	7.83
	2	26.36	37.78	6.53	11.03	28.71	38.84	1.15	8.70
	3	25.95	37.88	6.56	10.82	28.28	39.53	1.15	8.70
	4	27.54	38.31	6.61	12.41	24.45	41.76	1.12	8.93
	5	26.78	37.49	6.61	10.29	28.50	37.61	1.15	7.83
	6	26.69	36.98	6.63	10.26	28.78	36.55	1.16	7.76
	7	27.32	36.20	6.66	9.91	28.29	35.49	1.16	7.76
N4 (1∶4 000 000)	Polygon	25.95	26.51	6.43	6.84	15.69	33.69	1.12	6.25
	8	26.31	25.54	6.41	7.18	15.50	33.86	1.13	6.19
	9	25.99	26.63	6.39	6.73	15.42	35.95	1.13	6.19
	10	27.00	23.78	6.41	6.71	15.85	33.96	1.13	6.19
	12	27.23	23.32	6.47	5.10	16.01	33.83	1.12	5.36
	14	27.52	23.33	6.50	4.77	15.75	32.81	1.12	4.46
	16	27.89	20.04	6.51	4.45	15.58	28.84	1.12	4.46
	18	28.97	16.98	6.53	4.29	16.14	26.55	1.13	4.42

Continued table

Map scale	Simulation units (km)	Clay (%)		pH		SOC (g/kg)		Bulk density (g/cm^3)	
		Mean	CV	Mean	CV	Mean	CV	Mean	CV
N14 (1∶14 000 000)	Polygon	33.46	18.29	6.51	7.37	33.38	31.34	1.14	2.63
	17	32.77	18.31	6.61	6.51	32.34	31.08	1.14	2.63
	18	32.91	18.14	6.60	6.67	32.11	30.36	1.14	2.63
	19	33.39	18.75	6.58	6.53	33.39	32.32	1.14	2.63
	20	32.12	16.69	6.58	6.99	31.23	29.84	1.14	2.63
	21	33.10	18.31	6.59	6.83	32.27	31.14	1.14	2.63
	22	33.35	17.96	6.54	7.03	33.09	29.07	1.14	2.63
	23	33.09	18.68	6.59	6.68	32.33	32.11	1.14	2.63
	24	33.07	17.96	6.54	6.88	32.79	30.16	1.14	2.63
	25	33.12	17.30	6.52	6.75	33.06	29.19	1.14	2.63
	26	33.17	18.66	6.57	6.85	32.87	31.15	1.14	2.63
	27	32.60	18.10	6.61	6.96	31.29	31.80	1.14	2.63
	28	33.13	17.27	6.54	6.88	32.13	30.10	1.14	2.63

Notes: Mean: mean value; CV: coefficient of variation

Table 3 Index values determined from the DNDC simulations with paddy polygon units at different map scales in the Taihu Region of China

Map scale	SPN	Index values from vector simulation units ($IV_{polygon}$)						
		SOCS (Tg)	*AREA* (Mha)	*ASOCD* (kg C/m^2)	*STN*			
					S1	S2	S3	S4
C5 (1∶50 000)	52 304	144.78	2.32	6.24	622	137	6	1
D2 (1∶200 000)	7 263	168.78	2.60	6.48	127	78	6	1
P5 (1∶500 000)	4 766	172.04	2.53	6.71		68	6	1
N1 (1∶1 000 000)	967	161.21	2.59	6.24		48	6	1
N4 (1∶4 000 000)	32	167.55	2.74	6.12			6	1
N14 (1∶14 000 000)	8	207.73	2.80	7.42			2	1

Notes: *SOCS*: SOC stocks of surface paddy soil; *AREA*: paddy soil area; *ASOCD*: average SOC density of surface paddy soil; *STN*: paddy soil type number; SPN: paddy soil unit number; S1: soil species; S2: soil family; S3: soil subgroup; S4: soil great group (paddy soil)

For instance, four of the six paddy soil subgroups, namely, bleached, percogenic, degleyed and submergenic paddy soil, are not described in the N14 polygon unit dataset, but every paddy soil subgroup is described in other five datasets.

Additionally, the *IVs* of *SOCS* and *ASOCD* for the surface paddy soils that were simulated by DNDC with the six polygon unit datasets display pronounced differences from each other (Table 3). The *IV* of *SOCS* increased as the map scale of polygon unit dataset decreased. The highest *IV* of *SOCS* was simulated with the N14 polygon unit dataset beacuse the hydromorphic paddy soils have the largest area and the highest simulated *SOCD* (> 8 kg C/m^2) mapped in the dataset. The area of hydromorphic paddy soils, which was mapped in the N14 dataset with the darkest polygons of the *SOCD* that was simulated by DNDC (Fig. 2f), is approximately 7 times that in the C5 dataset (Fig. 2a). The spatial distribution maps of *SOCD* that were simulated with these polygon unit datasets display differences from each other (Fig. 2). The *SOCD* maps of the surface paddy soil, which were synthesized cursorily and simulated by DNDC with the N4 and N14 polygon unit datasets, differ distinctly from the others.

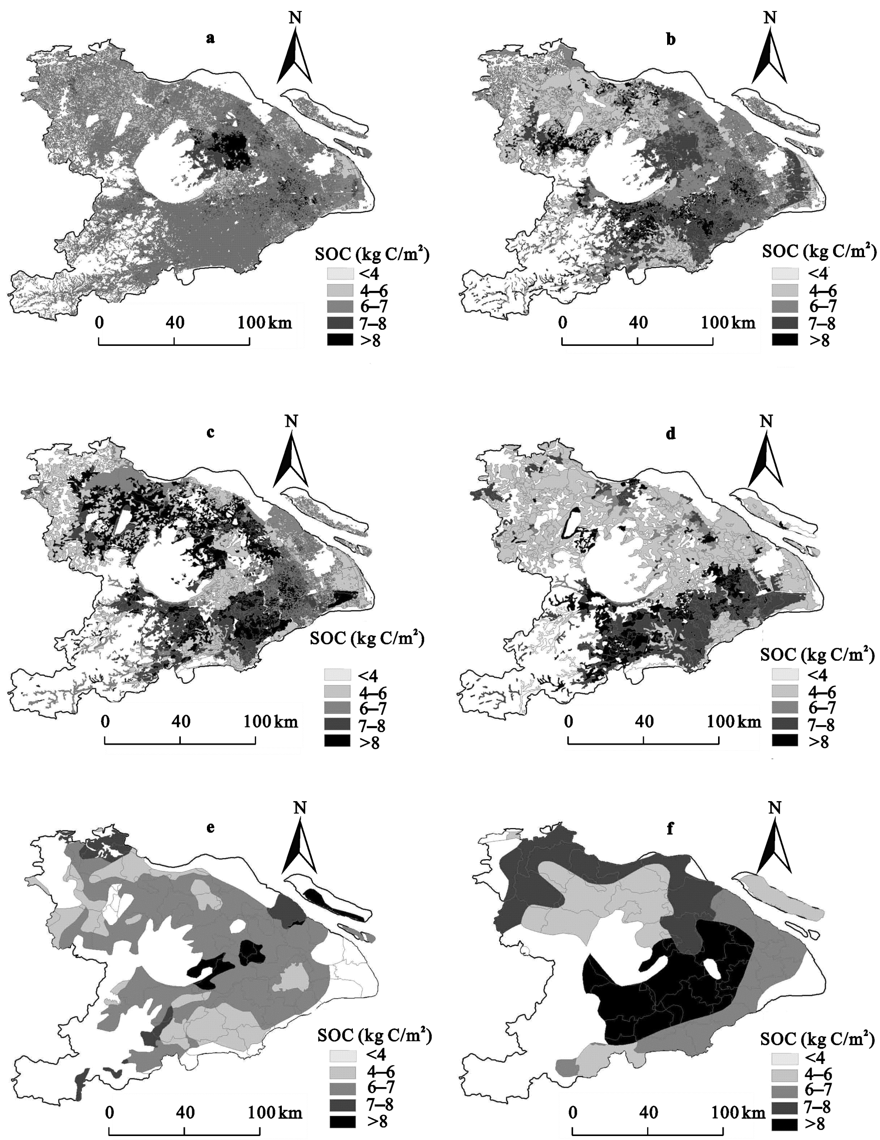

Fig. 2 Soil organic carbon density (SOCD) map simulated by DeNitrification-DeComposition (DNDC) model from vector paddy soil units at different map scales in the Taihu Region of China (a, C5 (1 ∶ 50 000); b, D2 (1 ∶ 200 000); c, P5 (1 ∶ 500 000); d, N1 (1 ∶ 1 000 000); e, N4 (1 ∶ 4 000 000); f, N14 (1 ∶ 14 000 000))

3.3 Optimal soil grid unit resolutions for SOC modeling at regional map scales

The *VIVs* of the four assessment indices (*STN*, *AREA*, *SOCS* and *ASCOD*), which were determined from the grid unit datasets and their parent polygon unit datasets at regional map scales, generally increase with increasing grid cell size (Fig. 3a, 3b and 3c). The *VIV* magnitude and trend vary with grid cell size, dataset and index. For the C5 polygon unit dataset and affiliated grid unit datasets, the *VIVs* of the four indices are all > 1% when the grid cell size is set as > 0.5 km. Only the *VIV* of *ASOCD* is < 1% when the grid unit resolution ranges from 0.3 to 0.5 km. When the grid cell size decrease to 0.3 km, three of the four *VIVs* are all < 1%, except for the SOCS index. The *VIVs* of the four indices are all < 1% only when the grid cell size is ≤ 0.2 km (Fig. 3a) and *STN* index is depicted with soil species (Table 3). The 0.2 km × 0.2 km resolution is optimal for the conversion of the C5 dataset from polygon to grid units, as the grid and parent polygon unit datasets are roughly equivalent in their information content and the data redundancy is minimal (Fig. 3a) when simulating regional SOC pools with DNDC at this cell dimension.

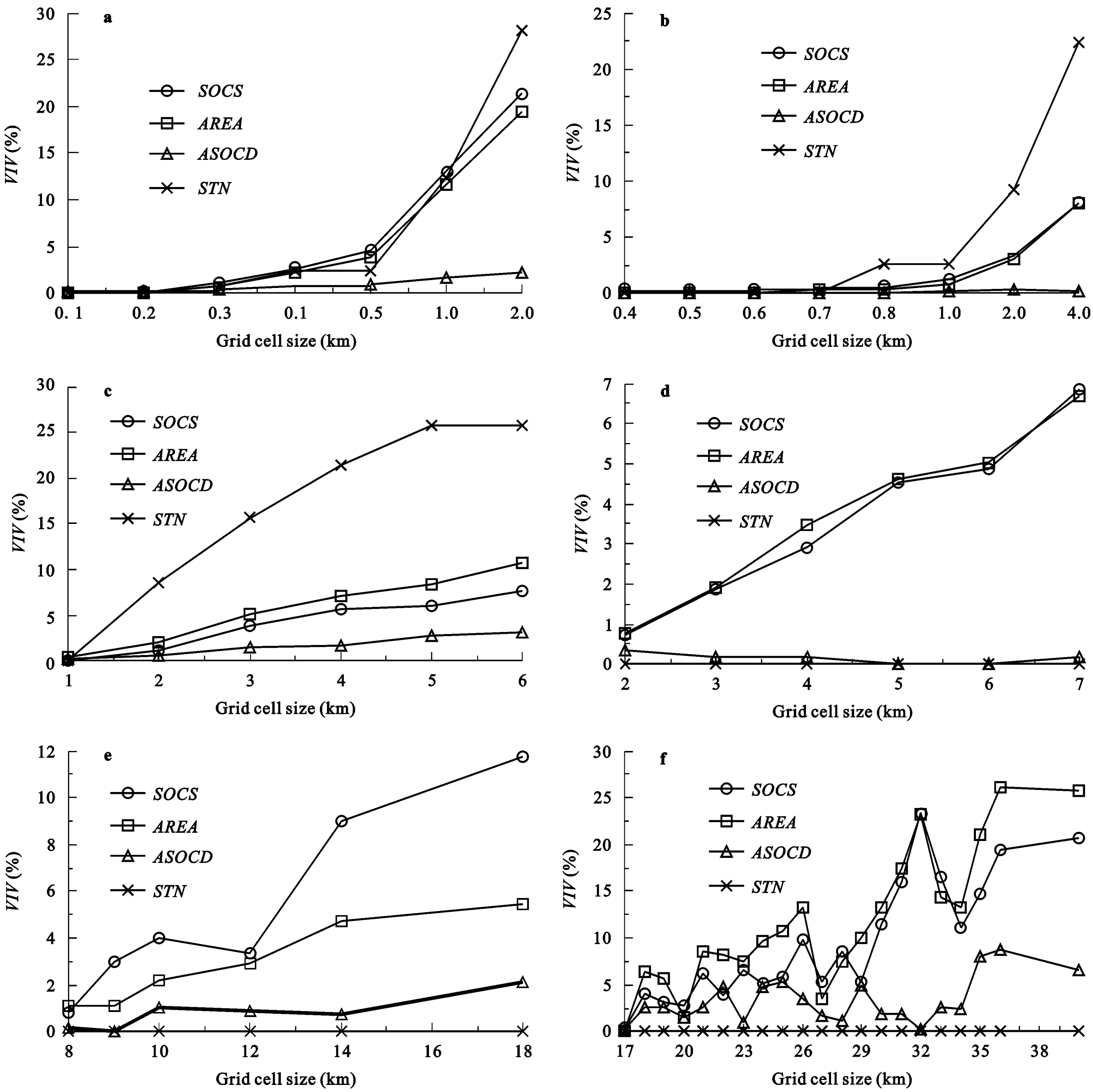

Fig. 3 *VIVs*, which vary with grid unit resolutions at different soil unit map scales in the Taihu Region of China (*VIV*, variation of an index value; *SOCS*, soil organic carbon stocks simulated by DNDC; *AREA*, soil area; *ASOCD*, average soil organic carbon density simulated by DNDC; *STN*, soil type number; a, C5 (1∶50 000); b, D2 (1∶200 000); c, P5 (1∶500 000); d, N1 (1∶1 000 000); e, N4 (1∶4 000 00); f, N14 (1∶14 000 000))

Similarly, for the conversion of the D2 and P5 dataset, only the *VIV* of *ASOCD* is < 1% when the grid unit resolution is 2 km, respectively. However, when the grid cell size for D2's conversion decreases to 0.8 km, all the *VIVs* are < 1% except for the *STN* index of soil species (Fig. 3b); all the *VIVs* > 1% for P5's conversion when the grid cells size increases over 2 km (Fig. 3c) and the *STN* index is depicted with the soil family (Table 3). The *VIVs* of the four indices that were derived from the conversion of the D2 and P5 datasets are all < 1% only when their grid cell sizes are ≤ 0.7 km and ≤ 1 km, respectively. At their maximum cell dimensions, the grid and parent polygon unit datasets are nearly identical, which minimizes the numbers of grid cells (NGS) and the time and cost of simulation (Fig. 3b and 3c). The optimal grid unit resolution for D2 and P5's conversion when simulating regional SOC pools with DNDC is 0.7 km × 0.7 km and 1 km × 1 km, respectively.

3.4 Optimal soil grid unit resolutions for SOC modeling at national map scales

Generally, almost all the *VIVs* of the four assessment indices (*STN*, *AREA*, *SOCS* and *ASCOD*) from these grid unit datasets and their parent polygon unit datasets at national map scales increase with increasing grid cell size except for N14 (Figs. 3d, 3e and 3f). The *VIVs* of three indices (*SOCS*, *AREA* and *ASOCD*) from the conversion of the N14 dataset vary with grid cell size in the random scatter diagram, except for the *STN* index of the soil subgroup when the grid cell size ranges from 18 to 36 km, which is around the center of its default grid cell size (28 km). The random scatter diagram complicates the selection of an optimal grid unit resolution as the *VIV* for the four indices are not consistent with varying grid cell size. To simulate regional SOC pools with DNDC, the optimal grid resolution for the conversion of the N14 dataset was determined to be 17 km × 17 km, as all the *VIVs* are < 1% when the maximum grid cell size is ≤ 17 km (Fig. 3f).

The results for the conversion of the N1 and N4 datasets demonstrate that the *VIVs* of *ASOCD* and *STN* are < 1% and the *VIVs* of *SOCS* and *AREA* are > 1% when the grid cell size is > 2 km (Fig. 3d) and > 8 km (Fig. 3e) and the *STN* index is depicted with the soil family and subgroup (Table 3), respectively. The *VIVs* of the four indices that were obtained from their grid unit datasets are only < 1% when the maximum grid cell size is ≤ 2 km and ≤ 8 km, respectively. Accordingly, the grid resolutions of 2 km × 2 km and 8 km × 8 km are optimal for the converson of the N1 and N4 datasets from paddy polygon to grid units for the DNDC grid modeling of regional SOC pools (Figs. 3d and 3e).

3.5 Relationship between polygon unit map scale and corresponding optimal grid unit resolution for simulation of regional SOC pools

Correlation analysis indicated a statistically significant relationship between the paddy polygon unit map scale (1∶10 000*x*) and the corresponding optimal grid unit resolution (*y*, km). Using several common regression model such as linear, logistic, polynomial, power and exponential regression, the relationship can be described with $R^2 > 0.90$ at a significance level of $P < 0.05$ as follows:

$$y = -0.80 \times 10^{-6}x^2 + 0.0228x + 0.0211 \ (R^2 = 0.9994, P < 0.05) \quad (6)$$

$$y = 0.0546x^{0.7999} \ (R^2 = 0.9909, P < 0.05) \quad (7)$$

$$y = 0.012x + 0.8759 \ (R^2 = 0.9690, P < 0.05) \quad (8)$$

All the mathematical regression curves deviate from a default linear regression, but the quadratic curve regression Equation (6) describes the relationship with the highest R^2 (Fig. 4) among them, with a significance level of $P < 0.05$. Although non system with underlying and fixed mathematics to reveal the relationship, the quadratic curve model was put forward from Yu *et al.* (2014) study to reveal such relationship for the calculation of regional SOC pools, too.

4 Discussion

4.1 Effect of map scale and grid resolution of simulation units on DNDC modeling of SOC pools

The *IVs* of the *STN* and *AREA* vary with the map scales of the polygon unit datasets (Table 3), and the *VIVs* that were determined from their grid unit datasets and parent polygon unit datasets vary with grid cell size (Fig. 3). The *VIV* of *STN* from the C5 and D2 datasets vary with grid cell size and can be described best by an exponential curve (Yu *et al*., 2011), and the *VIV* from P5 varies as a logarithmic curve (Yu *et al*., 2014). Both the map scale and grid resolution affect the soil type number and soil area, much lower the numbers of polygon units or grid cells, which affects the time and cost of simulation (Yu *et al*., 2011; Yu *et al*., 2013).

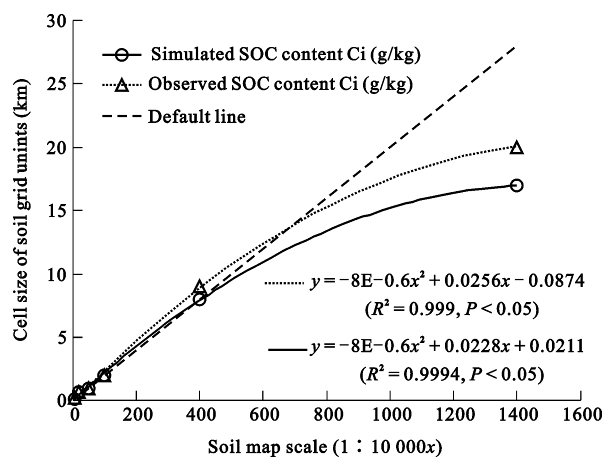

Fig. 4 Relationship between paddy polygon unit map scale and matched optimal grid unit resolution for the SOC simulation with DNDC in the Taihu Region of China (*y*: value of vertical ordinate; *x*: value of horizontal ordinate)

Additionally, the decreasing map scale or raster resolution changed the estimated contents of soil features such as SOC, clay content, pH and soil bulk density (Table 1, Table 2), which are all sensitive input parameters in DNDC modeling. The difference in the input parameter values directly affect results of the modeling (Valade *et al.,* 2014; Zhang *et al*., 2014b; Zhu and Zhuang, 2014). Thus, both the map scale (Xu *et al*., 2011; Xu *et al*., 2013; Yu *et al*., 2013; Zhang *et al*., 2014a) and grid resolution (Yu *et al*., 2011; Pogson and Smith, 2015) of the soil units significantly affected the simulation results of the regional SOC pools, which is why the *IVs* of the simulated *SOCS* and *ASOCD* and their *VIVs* varied with map scale and grid cell size (Table 3, Fig. 3).

Understandably, the C5 paddy polygon unit dataset in this study, which contains the greatest numbers of soil polygon units, soil families and species (Table 3), is the most detailed and accurate database in the Taihu Region (Zhang *et al*., 2012; Yu *et al*., 2013). Thus, the *IVs* of *STN*, *AREA*, *SOCS* and *ASOCD* that were obtained in the C5 dataset are considered to be the most reliable and are recognized as benchmarks for comparisions with other datasets at different map scales or raster resolutions in the region (Yu *et al*., 2011; Yu *et al*., 2013).

The variability in the weather data (i.e., precipitation, maximum and minimum air temperature) and farming management scenarios (i.e., sowing method, nitrogen fertilizer application rates, livestock, planting and harvest dates) among these simulation unit datasets were almost not found and could be neglected for the purposes of this analysis. Because these datasets were from the same weather and farming management database at the county scale and were overlain with these soil polygon datasets, all polygon and grid simulation units at any map scale or resolution within one county have the same feature input value for DNDC modeling except for soil features. Soil attributes, area, and changes in the soil type are the main source of the SOC variability in this study that was simulated by DNDC and are associated with the simulation units' scale and resolution (Yu *et al*., 2013; Pogson and Smith, 2015).

4.2 Comparison of simulation grid unit resolutions at different map scales among referenced re-searches

The optimal grid cell sizes for C5 and D2 conversions were larger than their default sizes, which maybe indicated that the actual accuracy of soil mapping at county and district level did not meet C5 and D2's requirement, respectively. The optimal grid cell sizes for N14 conversions was far smaller than its default size, which did not imply that soil mapping of the N14 was conducted in the acurracy of much higher than the 1∶14 000 000 map scale, but indicated that oversized grid cell could not reflect soil features spatial situation anymore at any map scale. Default sizes are not proper for the three (C2, D2, N14) polygon datasets conversions except other three datasets (P5, N1, N4), resulting all mathematical regression curves (Equations (6)–(8)) deviate from a default linear regression.

At the C5 map scale (1∶50 000), the original soil polygon units were converted to grid cells with sizes of 100 m × 100 m (Yang *et al*., 2009; Shi *et al*., 2010) and 30 m × 30 m (Su *et al*., 2012) as basic assessment units to simulate the SOC dynamics of agro-ecosystems. Compared to the optimal resolution (200 m × 200 m) suggested here, both soil grid unit datasets are redundant. Similarly, the soil grid unit datasets at a cell size of 1 km × 1 km, which was converted from the 1∶1 000 000 (N1) scale polygon unit dataset (Yu *et al*., 2007b), and 2 km × 2 km, which was converted from the 1∶4 000 000 (N4) scale (Shen *et al*., 2003), contain significant redundancy compared to the optimal resolution 2 km × 2 km and 8 km × 8 km, respectively, which were achieved in this study. Although the grid unit datasets that were used by these researchers retained the same data content as their parent polygon unit dataset, the grid cell size is not a real raster resolution that matches with their map scales due to the data redundancy. The workload and cost of the regional SOC investigation and simulation would triple due to the higher number of grid cells if the grid cells were designed as soil sampling and simulation units.

By contrast, the grid units with a cell size of 50 km × 50 km, which was converted from the 1∶4 000 000 (N4) scale (Wan *et al*., 2011) and 10 km × 10 km, which was converted from the 1∶1 000 000 (N1) scale (Yu *et al*., 2007c) soil polygon unit dataset, were used as assessment units to model the SOC dynamics in different regions. The grid simulations can be anticipated to have higher uncertainty than their parent polygon units' simulations because the grid unit datasets is coarser than their parent polygon unit datasets. If the grid cells are designed as soil sampling and simulation units, the regional SOC investigation and simulation will not have equivalent accuracies to their map scales.

The harmonized world soil database (HWSD), which was completed by FAO/IIASA/ISRIC/ISSCAS/JRC in 2009, was produced at a cell size of approximately 1 km × 1 km from an original polygon unit dataset, which contains over 16 000 different soil mapping units and was derived from the Soil Map of the World (1∶5 000 000), the regional Soils and Terrain Digital Database (SOTER) (1∶1 000 000 to 1∶5 000 000), the European Soil Map and the Soil Map of China (1∶1 000 000, digitized and compiled by authors YU Dongsheng *et al.*) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009). Based on the relationship developed here and assuming a scale of 1∶5 000 000, the effective resolution of the grid unit dataset would be approximately 10 km × 10 km, rather than 1 km × 1 km. Utilizing the HWSD database at a grid cell size of 1 km × 1 km for global SOC pool research would be subject to elevated data redundancy at the map scale and data uncertainty with the raster resolution, although it is currently the best global soil database in the world (Yu *et al.*, 2013).

4.3 Comparison of the optimal soil raster unit resolutions between the calculation and simulation of regional SOC pools

Yu *et al.* (2014) conducted a similar study in the same region as this study. A different method that was adopted in Yu *et al.* (2014) compared to this study was that the SOC content (C_i, g/kg) in Equation (1) was observed data in 1982, which is one of the input parameters for DNDC modeling in this study, while the C_i (g/kg) in this study was simulated data in 2000 by the DNDC modelling, which creates slight differences in the results. For example, the optimal grid cell sizes that correspond to the 1∶4 000 000 (N4) and 1∶14 000 000 (N14) map scales were 9 km and 20 km when C_i (g/kg) was observed (Yu *et al.,* 2014) and 8 km and 17 km when C_i (g/kg) was simulated in this study. The main reason for the difference was that more soil feature data, such as soil clay content and pH, were used to simulate the C_i (g/kg) in this study, but these data were not included in Yu *et al.* (2014). The additional soil features that were included in this study imply more rigorous criteria to assess data consistency between grid unit datasets and their parent polygon unit datasets, and further increase the optimal raster resolution, even if the same indices and criteria as in Yu *et al.* (2014) were applied. Maybe the different year (1982 and 2000, respectively) of the SOC pools estimation was another reason for the difference, but it should be incredible that both the optimal grid cell sizes were smaller than that from the previous study. Fortunately, this slight difference only occurred during the polygon unit dataset conversions for the small map scales of N4 and N14, indicating that these additional soil features that were included in this study affected the assessment of dataset consistency through the C_i (g/kg) simulated by DNDC at these two map scales, even if the model was well validated.

However, the optimal grid sizes corresponding to the other four map scales (C5, D2, P5, P1) did not show any difference whether C_i (g/kg) was observed or simulated, which indicates that these additional soil features did not affect the assessment of dataset consistency at these four map scales through the simulated C_i (g/kg) and proves that the model was well validated in the region. The relationships between the optimal grid cell size (y, km) and map scale (1∶10 000x) that were revealed in the two studies are all described best in a quadratic curve model Equation (6) when C_i (g/kg) was simulated and Equation (9) ($R^2 = 0.9982$, $P < 0.05$) when C_i (g/kg) was observed (Yu *et al.*, 2014) with slightly different regression coefficients.

$$y = -0.803\times10^{-6} x^2 + 0.025 x - 0.0874 \quad (9)$$

The quadratic model Equation (6) implies that the optimal grid cell size may be larger than the default when the map scale is larger than 1∶4 000 000 (N4) (Fig. 4). The default-size soil grid units that were converted from polygon units at these map scales will result in data redundancy. When the map scale is less than 1∶4 000 000 (N4), the optimal grid cell size is less than the default, and the deviation increases with decreasing map scale (Fig. 4). The conversion of soil polygon units at these map scales to grid units at the default cell size will decrease the data accuracy and increase the simulation's uncertainty. Thus, the quadratic model is more important to soil polygon unit dataset conversion at map scales that are less than N4 compared to the other map scales.

The quadratic model Equation (6) also can substitute Equation (9) when C_i (g/kg) is observed during the calculation of regional SOC pools, because the optimal grid unit resolution that is determined from Equation (6) may be higher than that from Equation (9) at a certain map scale (Fig. 4). The accuracy of the soil assessment unit dataset and the certainty of the results are more critical than the dataset's redundancy.

4.4 Application of quadratic curve regression model for DNDC modeling at different map scales

Almost all the map scales of the soil polygon unit datasets that are commonly used for China, which were generated from the Second National Soil Survey of China, are included in this study. The six soil map scales were designed for soil mapping at different administrative levels, including county, district, province and the whole country (Shi *et al*., 2006).

The Taihu Region is a typical area in China where paddy soil prevails. Although the region is located in the Yangtze Delta plain in East China, where rice fields are integrated with a high density of rivers or ponds, gardens and urban land, the spatial distribution pattern of rice fields is similar to hilly or mountain regions where rice fields coexist with crops, grass, shrubs and forests and urban land (Yu *et al*., 2011). We may assert with some degree of confidence that the knowledge obtained in this present study can be applied elsewhere in East and South China, where 95% of China's rice fileds are distributed (Li, 1992c).

In the North and West China, the mean size of the soil vector mapping unit is larger than that of East and South China at various map scales, because of simpler natural conditions and reduced spatial variability. We may conclude that the optimal grid cell size that is determined from the quadratic model Equation (6) can be smaller than the real optimal size in the region (Yu *et al*., 2013). The applied optimal grid cell size will create slight redundancy in the grid unit dataset but not lower its accuracy contrasting to their corresponding soil polygon units' map scales. Although the quadratic model was obtained from a specific case study and the regression coefficients would vary with the research region slightly, this knowledge can be referenced as a guideline for soil unit conversion from polygon to grid and to optimize field sampling strategies for the regional simulation of SOC pool dynamics in China.

Within China, a few administrative regions are different from those used here due to their historical anthropogeography and physical geography, resulting in additional soil datasets with non-traditional map scales, such as the 1∶75 000, 1∶100 000 or 1∶150 000 soil mapping scales for the county level, 1∶250 000 or 1∶350 000 for the district level, and 1∶750 000 or 1∶1 500 000 for the province level (Shi *et al*., 2006). The optimal grid resolutions for the conversion of soil polygon units for DNDC modeling at these map scales, can also be inferred from these guidelines.

4.5 Implication for conversion of other landscape polygons to grid units

The paddy soil polygon units in the study can be thought as paddy soil landscape patches, the landscape patches and indices suffer both natural and artificial controls, which is similar with other landscape, i.e., vegetation and landuse landscape (Fu *et al*., 2011). In fact the optimal grid resolution for polygon landscape patch conversion depends on the landscape indices, such as patch density, patch size, patch area and patch shape. Relative size and position of mesh cell to polygon patch affect assigned cell attibuties when partition the polygon patch, especially for those cells acrossing patches border.

According to the series paddy soil polygon patches data, the average patch area ($\hat{S}$, km^2) was calculated, namely 0.446 km^2 (C5), 3.585 km^2 (D2), 6.683 km^2 (P5), 21.222 km^2 (N1), 172.152 km^2 (N4) and 731.491 km^2 (N14), respectively. Ratio ($\hat{R}$, %) of the optimal mesh cell size obtained in this study to $\hat{S}$ was 9.0%, 13.7%, 15.0%, 18.8%, 37.2% and 39.5%, respectively. Both $\hat{S}$ and $\hat{R}$ increased with polygon map scale decrease. A significant relationship between the $\hat{R}$ and the mean polygon landscape $\hat{S}$ was found and shown as Equations (10) and (11), which explained why the quadratic or power function model relationship exist between optimal grid resolution and soil polygon unit map scale.

$$\hat{R} = 10.394\hat{S}^{0.2153} \quad (R^2 = 0.9759,\ P < 0.05) \tag{10}$$

$$\hat{R} = -0.0002\hat{S}^2 + 0.1788\hat{S} + 12.601 \quad (R^2 = 0.9737,\ P < 0.05) \tag{11}$$

The Equation (10) with the highest R^2 was constructed from landscape patch index ($\hat{S}$), so it can be referenced to other landcape polygon patches' mesh partition.

5 Conclusions

The DNDC model has been utilized to upscale estimates of SOC pools from the plot to region scale. For up-scaled DNDC utilization, a region is partitioned into many simulation units, e.g., soil vector polygon units or raster grid units, within which all properties are assumed to be as homogeneous as at plot scale. The homogeneity assumption is a possible major source of error when extending DNDC modeling from the plot to region scale. The homogeneity within the simulation units is linked to their map scale or grid resolution, which has a strong influence on the results of SOC pool simulation. This study reveal the impact of the vector map scale and raster resolution of soil simulation units on DNDC modeling.

Soil grid units are more often applied to SOC pool simulation, as they are more easily manipulated for spatial model simulation, geo-statistics and spatial analysis than soil polygon units. Most are derived by data conversion from soil polygon units, but the choice of grid unit resolution varies by researcher even if these units are derived from a certain vector polygon unit dataset. An optimal raster resolution corresponding to a certain map scale was proposed in this study for soil polygon unit conversion to grid units. The optimal raster resolution is the maximum grid cell size for which the soil grid unit dataset and the vector polygon unit dataset can be identically scaled. The quadratic curve regression model $y = -0.80 \times 10^{-6} x^2 + 0.0228\ x + 0.0211$ (R^2 = 0.999 4, $P < 0.05$), and the power function model $\hat{R} = 10.394\ \hat{S}^{\,0.2153}$ (R^2 = 0.975 9, $P < 0.05$) were distinguished in this study. The quadratic curve model best describes the relationship between the optimal soil grid unit resolution (y, km) and soil polygon unit map scale (1∶10 000x), and the power function model best describes the relationship between the ratio ($\hat{R}$, %) of optimal mesh cell size to average polygon patch size ($\hat{S}$, km^2) and the $\hat{S}$, with the highest R^2 at the significance level of $P < 0.05$ respectively.

For the investigation and simulation of regional SOC pools, the quadratic curve model is more important to the conversion of soil polygon units at map scales that are less than 1∶4 000 000 (N4) compared to the other map scales. Although this quadratic curve model was determined from a specific case study and the regression coefficients would vary with the investigated region slightly, this knowledge can be referenced as a guideline for the conversion of soil assessment unit datasets from polygon to grid cells, which is efficient for optimizing field sampling strategies, minimizing the uncertainty of the investigation and simulating regional SOC pools at different map scales. The power function model can be referenced to other landcape polygon patches' mesh partition. Matching grid resolutions with polygon map scales in equality of soil simulation units rationally should be emphasized and enhanced in modeling of regional biogeochemical cycles.

References

Batjes N H, 2000. Effects of mapped variation in soil conditions on estimates of soil carbon and nitrogen stocks for South American. *Geoderma*, 97: 135–144.

Bouwman A F, Boumans L J M, Batjes N H, 2002. Modeling global annual N_2O and NO emissions from fertilized fields. *Global Biogeochemical Cycle*, 16(4): GB1080. doi: 10.1029/2001GB001812

Cai Z C, Sawamoto T, Li C S *et al*., 2003. Field validation of the DNDC model for greenhouse gas emissions in East Asian cropping systems. *Global Biogeochemical Cycle*, 17(4): GB1107. doi: 10.1029/2003GB002046

Davidson E A, Janssens I A, 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, 440: 165–73.

FAO/IIASA/ISRIC/ISSCAS/JRC, 2009. *Harmonized World Soil Database* (*version 1.1*). FAO, Rome, Italy and IIASA, Laxenburg, Austria, 1–38.

Fu B J, Chen L D, Ma K M *et al*., 2011. *Landscape ecology: Principles and Applications* (*Second Edition*). Beijing: Science Press, 1–396. (in Chinese)

Giltrap D L, Li C S, Saggar S, 2010. DNDC: a process-based model of greenhouse gas fluxes from agricultural soils. *Agriculture, Ecosystems & Environment*, 136: 292–300.

Huang Y, Zhang W, Zheng X, 2004. Modeling methane emission from rice paddies with various agricultural practices. *Journal of Geophysics Research*, 109: D08113, doi: 10.1029/2003JD004401.

Lal R, 2004. Soil carbon sequestration tomitigate climate change. *Geoderma*, 123 (2): 1–22.

Levy P E, Mobbs D C, Jones S K *et al*., 2007. Simulation of fluxes of greenhouse gases from European grasslands using the DNDC model. *Agriculture, Ecosystems and Environment*, 121: 186–192.

Li C S, 2007. Quantifying greenhouse gas emissions from soils: Scientific basis and modeling approach. *Soil Scicence and Plant Nutrient*, 53: 344–352.

Li C S, Frolking S, Frolking T A, 1992a. A model of nitrous oxide evolution from soil driven by rainfall events: I. Model structure and sensitivity. *Journal of Geophysics Research*, 97 (D9): 9759–9776.

Li C S, Frolking S, Frolking T A, 1992b. A model of nitrous oxide evolution from soil driven by rainfall events: II. Model applications, *Journal of Geophysics Reserch*, 97(D9): 9777–9783.

Li C S, Frolking S, Xiao X M, 2005. Modeling impacts of farming management alternatives on CO_2, CH_4, and N_2O emissions: a case study for water management of rice agriculture of China. *Global Biogeochemical Cycle*, 19: GB3010. doi: 10.1029/2004GB002341

Li C S, Mosier A, Wassmann R *et al*., 2004. Modeling greenhouse gas emissions from rice-based production systems: Sensitivity and upscaling. *Global Biogeochemical Cycle*, 18: GB1043. doi: 10.1029/2003GB002045

Li C S, Qiu J J, Frolking S *et al*., 2002. Reduced methane emissions from large-scale changes in water management in China's rice paddies during 1980–2000. *Geophysical Research Letter*, 29(20): 1972. doi: 10.1029/2002 GL015370

Li C S, Salas W, DeAngelo B *et al*., 2006. Assessing alternatives for mitigating net greenhouse gas emissions and increasing yields from rice production in China over the next twenty years. *Journal of Environment Quality*, 35: 1554–1565.

Li C S, Zhuang Y H, Frolking S *et al*., 2003. Modeling soil organic carbon change in croplands of China. *Ecological Applications*, 13: 327–336.

Li H, Qiu J J, Wang L G *et al*., 2010. Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat-maize rotation system in China. *Agriculture, Ecosystems and Environment*, 135: 24–33.

Li Qinkue, 1992. *Paddy Soil of China*. Beijing: Science Press, 1–680. (in Chinese)

Liu Z P, Shao M A, Wang Y Q, 2011. Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture, Ecosystems and Environment*, 142: 184–194.

Ni J, 2001. Carbon storage in terrestrial ecosystems of China: estimates at different spatial resolutions and their responses to climate change. *Climate change*, 49: 339–358.

Pathak H, Li C S, Wassmann R, 2005. Greenhouse gas emissions from Indian rice fields: calibration and upscaling using the DNDC model. *Biogeosciences*, 2: 113–123.

Pogson M, Smith P, 2015. Effect of spatial data resolution on uncertainty. *Environmental Modelling & Software*, 63: 87–96.

Qin Falyu, Shi Xuezheng, Xu Shengxiang *et al*., 2015. Zonal differences in correlation patterns between soil organic carbon and climate factors at multi-extent. *Chinese Geographical Science*, 25(5): 537–548. doi: 10.1007/s11769-015-0736-3

Qiu Jianjun, Wang Ligang, Tang Junhua *et al*., 2004. Studies on the situation of soil organic carbon storage in croplands in northeast of China. *Scientia Agricultura Sinica*, 37 (8): 1166–1171. (in Chinese)

Rüth B, Lennartz B, 2008. Spatial variability of soil properties and rice yield along two catenas in southeast China. *Pedosphere*, 18(4): 409–420.

Schlesinger W H, 1997. *Biogeochemistry*: *An Analysis of Global Change*. San Diego: Academic Press, 1–102.

Schmidt K, Behrens T, Scholten T, 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. *Geoderma*, 146: 138–146.

Shen Yu, Huang Yao, Zong Lianggang *et al*., 2003. Simulation and prediction of soil organic carbon dynamics in Jiangsu province based on model and GIS techniques. *Scientia Agricultura Sinica*, 36 (11): 1312–1317. (in Chinese)

Shi X Z, Yang R W, Weindorf D C *et al*., 2010. Simulation of organic carbon dynamics at regional scale for paddy soils in China. *Climatic Change*, 102: 579–593.

Shi X Z, Yu D S, Warner E D *et al*., 2006. Cross-reference system for translating between genetic soil classification of China and Soil Taxonomy. *Soil Scicence Society of America Journal*, 70: 78–83.

Smith K A, Dobbie K E, 2001. The impact of sampling frequency and sampling times on chamber-based measurements of N_2O emissions from fertilized soils. *Global Change Biology*, 7(8): 933–945. doi: 10.1046/j.1354-1013.2001.00450.x

Soil Survey Staff, 1994. *Keys to Soil Taxonomy* (*6th edition*). Washington, DC: United States Goverment Printing Office, 1–437.

Su S L, Zhang Z G, Xiao R *et al*., 2012. Geospatial assessment of agroecosystem health: development of an integrated index based on catastrophe theory. *Stoch Environment Research and Risk Assessement*, 26: 321–334.

Sun W J, Huang Y, Zhang W *et al*., 2010. Carbon sequestration and its potential in agricultural soils of China. *Global Biogeochemical Cycle*, 24: GB3001. doi: 10.1029/2009GB 003484

Tang H, Qiu J, Eric V R *et al*., 2006. Estimations of soil organic carbon storage in cropland of China based on DNDC model. *Geoderma*, 134: 200–206.

Tonitto C, David M B, Li C S *et al*., 2007. Application of the DNDC model to tile-drained Illinois agroecosystems: model comparison of conventional and diversified rotations. *Nutrient Cycling in Agroecosystems*, 78(1): 65–81. doi: 10.1007/s 10705-006-9074-2

Valade A, Ciais P, Vuichard N *et al*., 2014. Modeling sugarcane yield with a process-based model from site to continental scale: uncertainties arising from model structure and parameter values. *Geoscientific Model Development*, 7: 1225–1245.

Wan Y F, Lin E D, Xiong W *et al*., 2011. Modeling the impact of climate change on soil organic carbon stocks in upland soils in the 21st century in China. *Agriculture*, *Ecosystems and Environment*, 141: 23–31.

Wang S H, Shi X Z, Zhao Y C *et al*., 2011. Regional simulation of soil organic carbon dynamics for dry farmland in east China by Coupling a 1:500 000 soil database with the Century model. *Pedosphere*, 21(3): 277–287.

Wu Y, Liu S, Tan Z, 2015. Quantitative attribution of major driving forces on soil organic carbon dynamics. *Journal of Advances in Modeling Earth Systtems*, 7: 21–34. doi: 10.1002/

2014MS000361

Xu Qi, Lu Yanchun, Liu Yuanchang *et al*., 1980. *Paddy Soil of Taihu Region in China*. Shanghai: Science Press, 1–68. (in Chinese)

Xu S X, Shi X Z, Zhao Y C *et al*., 2011. Modeling carbon dynamics in Paddy soil in Jiangsu Province of China with soil databases differing in spatial resolution. *Pedosphere*, 21(6): 696–705.

Xu S X, Shi X Z, Zhao Y C *et al*., 2012. Spatially explicit simulation of soil organic carbon dynamics in China's paddy soils. *Catena*, 92: 113–121.

Xu S X, Zhao Y C, Shi X Z *et al*., 2013. Map scale effects of soil databases on modeling organic carbon dynamics for paddy soils of China. *Catena*, 104: 67–76.

Yang Ruwei, Xu Qi, Shi Xuezheng *et al*., 2009. The research of paddy soil organic carbon simulation at regional scale. *Anhui Agricultural Science Bulletin*, 15(13): 126–128. (in Chinese)

Yu D S, Ni Y L, Shi X Z *et al*., 2014. Optimal soil raster unit resolutions in estimation of soil organic carbon pool at different map scales. *Soil Science Society of America Journal*, 78(3): 1079–1086. doi: 10.2136/sssaj 2013.07. 0262

Yu D S, Shi X Z, Wang H J *et al*., 2007a. National scale analysis of soil organic carbon stocks in China based on Chinese soil taxonomy. *Pedosphere*, 17(1): 11–18.

Yu D S, Shi X Z, Wang H J *et al*., 2007b. Regional patterns of soil organic carbon stocks in China. *Journal of Environment Management*, 85: 680–689.

Yu D S, Yang H, Shi X Z *et al*., 2011. Effects of soil spatial resolution on quantifying CH_4 and N_2O emissions from rice fields in the Taihu Region of China by DNDC model. *Global Biogeochemical Cycle*, 25: GBC2004. doi: 10.1029/2010 GB003825

Yu D S, Zhang L M, Shi X Z *et al*., 2013. Soil assessment unit scale affects quantifying CH4 emissions from rice fields. *Soil Science Society of America Journal*, 77(2): 664–672.

Yu Dongsheng, Shi Xuezheng, Sun Weixia *et al*., 2005. Estimation of China soil organic carbon storage and density based on 1:1 000 000 soil database. *Chinese Journal of Applied Ecology*, 16: 2279–2283. (in Chinese)

Yu Y Q, Huang Y, Zhang W, 2012. Modeling soil organic carbon change in croplands of China, 1980–2009. *Global and Planetary Change*, 82–83: 115–128.

Yu Yongqiang, Huang Yao, Zhang Wen *et al*., 2007c. Modeling farmland soil organic carbon dynamics in eastern China: spatio-temporal pattern. *Geography and Geo-Information Science*, 23(1): 97–100. (in Chinese)

Zhang L M, Yu D S, Shi X Z *et al*., 2009a. Quantifying methane emissions from rice fields in the Taihu Lake region, China by coupling a detailed soil database with biogeochemical model. *Biogeosciences*, 6: 739–749. doi: 10.5194/bg-6-739-2009

Zhang L M, Yu D S, Shi X Z *et al*., 2009b. Simulation of global warming potential (GWP) from rice fields in the Tai-Lake region, China by coupling 1:50 000 soil database with DNDC model. *Atmospheric Environment*, 43: 2737–2746.

Zhang L M, Yu D S, Shi X Z *et al*., 2012. Simulation soil organic carbon change in China's Tai-Lake paddy soils. *Soil & Tillage Research*, 121: 1–9.

Zhang L M, Yu D S, Shi X Z *et al*., 2014a. Effects of soil data and dimulation unit resolution on quantifying changes of soil organic carbon at regional scale with a biogeochemical process model. *PLoS ONE*, 9(2): e88622. doi: 10.1371/ journal.pone. 0088622

Zhang W, Zhang Q, Huang Y *et al*., 2014b. Uncertainties in estimating regional methane emissions from rice paddies due to data scarcity in the modeling approach. *Geoscicentific Model Development*, 7: 1211–1224. doi: 10.5194/gmd-7-1211-2014

Zhao Y C, Shi X Z, Weindorf D C *et al*., 2006. Map scale effects on soil organic carbon stock estimation in north China. *Soil Science Society of America Journal*, 70: 1377–1386.

Zhu Q, Zhuang Q, 2014. Parameterization and sensitivity analysis of a process-based terrestrial ecosystem model using adjoint method. *Journal of Advances in Modeling Earth Systtems*, 6: 315–331. doi: 10.1002/2013MS000241