

Variogram Analysis for Assessing Landscape Spatial Heterogeneity in NDVI: an Example Applied to Agriculture in the Jiansanjiang Reclamation area, Northeast China

Zhaofei Wen^{1,2}, Shengjun Wu¹, Feng Liu¹, Shuqing Zhang², Patricia Dale³

1. Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences
Chongqing 401120, P. R. China
wenzhaofei@cigit.ac.cn

2. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences
Changchun 130012, P. R. China

3. Environmental Futures Centre, Griffith School of Environment, Griffith University, Brisbane, Nathan Queensland 4111, Australia

Abstract—It is necessary to characterize and monitor the spatial heterogeneity of agricultural landscapes, to assist the spatial understanding of some related agricultural processes and to predict food production patterns. In the study, variogram models fitted to empirical variograms are proposed to retrieve spatial heterogeneity characteristics, and they are applied to NDVI data generated from Landsat TM imagery at a 30 m pixel scale. Four representative study sites with distinct landscape patterns, located in the Honghe farm of Jiansanjiang Reclamation Area, Northeast China, were selected for spatial heterogeneity testing and analyzing. The results provided quantitative agricultural landscape knowledge: (1) the agricultural landscape heterogeneity at different directions within the study area could be effectively quantified with the variogram model and then interpreted; (2) the spatial heterogeneity of the dry land matrix was normally larger than that of the paddy field matrix; and (3) the agricultural spatial heterogeneity was affected by two factors: the land use type and the distribution pattern of land use types. As remote sensing techniques can now provide various types of surface monitoring data, we argue that quantitative variogram analysis of the data for spatial heterogeneity can help identify explain related ecological phenomena. It could also be used to improve quantitative agricultural remote sensing monitoring in a spatial heterogeneity area as well.

Index Terms—semivariogram, geostatistical, spatial variability, spatial structure.

I. INTRODUCTION

Monitoring production and cultivation patterns in agricultural landscapes at a large scale is of paramount importance to assure food security, which is affected by increasingly uncertain climatic conditions [1], for a growing human population, especially for densely populated China. Sustainable agriculture is acknowledged by the Group on Earth Observation (GEO) as one of the critical societal benefit areas for international cooperation and collaboration and it requires an operational system for monitoring global agriculture [2].

However, agricultural monitoring requires high temporal frequency data, which are currently provided only by moderate spatial resolution sensors (e.g., TERRA/MODIS, SPOT/VEGETATION, etc.). At such moderate spatial resolutions, farmland that is heterogeneous within a pixel will be mixed and hence obscured. For instance, farmland might be a mosaic of crop fields (e.g., dry lands and paddy fields), which may differ not only in terms of size, shape and orientation but also in their spatial distribution, and thus may contain intra-pixel heterogeneity in medium scale imagery. Such heterogeneity would hence reduce the accuracy of crop biophysical parameters estimation (e.g., leaf area index (LAI)) [3-4]. To reduce these impacts, a suitable strategy is to incorporate the quantified corresponding intra-pixel spatial heterogeneity in the biophysical parameters estimation processes [5]. In addition, spatial heterogeneity caused by agricultural processes that occur in a spatial context, and this constrains the generation of landscape pattern [6].

Thus, to monitor agricultural processes and to predict productivity, we not only interested in spatial heterogeneity per se, but also in recognizing the leading to such spatial heterogeneity [7]. The goals of this study are, (1) to quantitatively measure agricultural spatial heterogeneity of the vegetation variable, NDVI, which is used as an indicator of relative biomass and greenness [8], at the landscape scale; (2) to compare these measures among disparate landscapes and explain the observed patterns through this explanatory variable (i.e., NDVI) and the influence of changing scales of observation; and (3) to provide these spatial heterogeneity measurements as necessary data inputs for agricultural monitoring.

II. STUDY AREA AND DATA DESCRIPTION

The Jiansanjiang Reclamation Area, located in the center of Sanjiang plain, northeast China, is an important basin of commercial grain production. Here, the Honghe State Farm, next to the Honghe National Nature Reserve (Honghe NNR,

see Fig. 1), is selected as our study area. The main land use types are dry land and paddy field which are divided by roads and/or ditches, at various sizes. As a state-run farm, agricultural activities are conducted under a uniform management regime, so that the same types of crop have similar cultivation practices. Within the farm we selected four typical study sites for experiments and analysis. These are Sites 1-4 shown in Fig. 1, and detailed in Table 1, with contrasting landscape patterns.

To capture spatial heterogeneity effectively using remotely sensed data, its resolution must be high enough to characterize the scale of the typical length of interest [9]. In this work, relatively high resolution data, Landsat 5 (i.e., TM imagery with 30 m spatial resolution) was used, as it is fine enough to describe the length scale of agricultural landscapes [10]. The image dataset was acquired at the best time in the growing season for crops (i.e., 31/8/2006). The atmospheric effects of the imagery were removed under the FLAASH module, a first-principles atmospheric correction tool that corrects wavelengths in the visible through near-infrared and shortwave

infrared regions, in ENVI software. The NDVI data was calculated directly from the reflectance values in the near-infrared (NIR) band and the visible red band (RED) from Landsat TM data for each pixel which was atmospherically corrected.

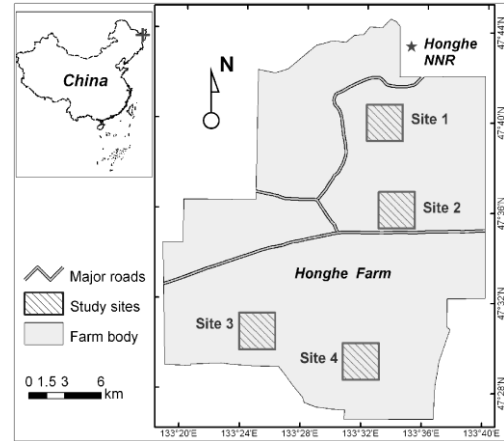


Fig. 1. Location of the study area and study sites

TABLE I. THE INFORMATION FOR THE FOUR STUDY SITES NUMBERED AS IN FIG. 1 (IMAGERY ACQUISITION DATE: 31/8/2006)

#	Central location	Size(m ²)	Landscape pattern	
			Land use type	Description
Site 1	47°39'52"N, 133°33'10"E	3000×3000	Dry land(dry land matrix)	Soya bean planted area
Site 2	47°36'02"N, 133°34'03"E	3000×3000	Dry land & paddy fields(dry land matrix)	Patches of dry lands and paddy fields distributed in concentrated and successive way respectively
Site 3	47°30'32"N, 133°24'59"E	3000×3000	Paddy fields(paddy field matrix)	Rice growing area
Site 4	47°29'21"N, 133°31'56"E	3000×3000	Dry land & paddy fields(paddy field matrix)	Patches of dry lands and paddy fields distributed alternatively or scattered

III. METHODS

A. Creating the empirical variogram

A pixel value of the NDVI image can be considered as a realization of regionalized variable $Z(X)$ [11], i.e., a function of location, describing the spatial distribution of an aspect of vegetation across the image domain I . The empirical variogram examines the relationship between distance and pairs of pixels ($z(x_i), z(x_j)$), and it can be calculated as [12]:

$$\gamma_e(h) = \frac{1}{2N(h)} \sum_h [z(x_i) - z(x_j)]^2 \quad (1)$$

where $N(h)$ is the number of pairs of pixels separated by distance h and the empirical variogram denoted as $\gamma_e(h)$ [3].

Variogram values may not be statistically reliable at large distances due to the decreasing number of pixel pairs N for computing the values and that would lose more information [12]. Hence, we computed the variogram up to a maximum h_{\max} distance equal to one-third the total distance at a specific direction [13]. In this study, we mainly analyze the spatial heterogeneity in both east-west and north-south directions ($h_{\max}=1000$ m) and characterize the omnidirectionality of the overall spatial heterogeneity ($h_{\max}=1412$ m).

B. Variogram modeling

The theoretical variogram model of $Z(x)$ in the follow expression:

$$\gamma(h) = \frac{1}{2} E[Z(x+h) - Z(x)]^2 \quad (2)$$

is estimated by fitting a valid mathematical function to each empirical variogram [14].

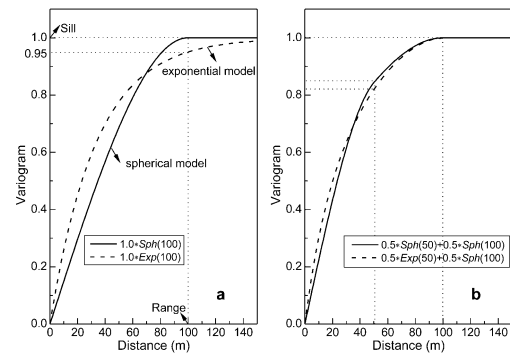


Fig.2 Theoretical models of the variogram. a: exponential and spherical elementary models with a range equal to 100 m. b: linear models of these two elementary models

Figure 2 shows typical theoretical variogram models (e.g., exponential, spherical model). Of particular interest are the estimates of two parameters, the *sill* and the *range*, which define the spatial heterogeneity of the landscape [14]. Practically, at an infinitesimally small separation distance, e.g., within a pixel distance, the difference between measurements often does not tend to zero [14]. This is called the *nugget effect* and the value is called the *nugget* denoted as c_0 , i.e., the intercept in regression, as shown on Fig. 2.

To account for the multi-scale spatial heterogeneity of the data, we used a linear model of regionalization defined as a linear combination of two or more functions as follows [3] (see Fig. 2):

$$\gamma(h) = c \sum_{i=1} w_i g_i(a_i, h) \quad (3)$$

where w_i is the fraction of variogram sill c , g_i denotes each elementary variogram function (e.g., the spherical model or exponential elementary model shown in Fig. 2b), and a_i is the variogram range related to each function g_i .

In this study, the GS+ version 7 was adopted to calculate the empirical variogram, and the parameters of the calculated variogram models were then fitted interactively under VARIOWIN. The IGF index (i.e., indicative goodness of fit), which was computed through an iterative procedure of VARIOWIN, is the standardized weighted average of the squared difference between the observed values and the fitted curves [15].

C. Quantifying the spatial heterogeneity

To summarize the spatial information of the variogram provided by the *range* and *sill*, a single parameter is used: the *integral range*, denoted A , which can be considered as the equivalent area of the influence of autocorrelation. In addition, the square root of A , denoted D_c , is thus the influence distance, which is related to the mean length scale of the image spatial structures [5]. The formula of the exponential and spherical variogram models and their corresponding integral range (A) are listed in Eq. 5 and Eq. 6 respectively [1, 3].

For the exponential model:

$$\gamma(h) = c[1 - \exp(-\frac{3h}{a})]; \quad A = \frac{2\pi a^2}{9} \quad (4)$$

For the spherical model:

$$\gamma(h) = \begin{cases} c[\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3], & h \leq a \\ c, & h > a \end{cases}; \quad A = \frac{\pi a^2}{5} \quad (5)$$

For a combinational linear variogram model as showed in Eq.4, the integral range A can then be computed as follows:

$$A = \sum_{i=1} w_i A_i \quad (6)$$

where A_i correspond to the integral range of sub model $g_i(a_i, h)$ in Eq. 3.

IV. RESULTS

A. Directional heterogeneity analysis

The directional empirical variograms with their respective fitted variogram models for the four study sites (i.e., four distinct landscape types) are shown in Fig. 3, and the corresponding detailed information for the models is listed in Table 2.

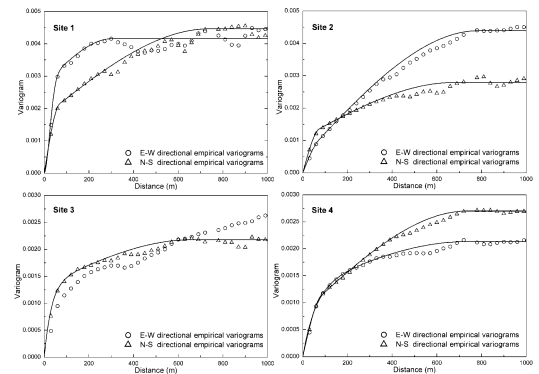


Fig.3 E-W and N-S directional empirical variograms with corresponding fitted variogram models of the four study sites

TABLE II. DETAILED INFORMATION OF THE FITTED VARIOGRAM MODELS

#	Direction	a_i		c	w_i		D_c (m)	IGF
		a_1 (m)	a_2 (m)		w_1 (%)	w_2 (%)		
Site 1	E-W	<i>Sph</i> (59)	<i>Sph</i> (306)	0.00415	66	34	/	0.00574
	N-S	<i>Sph</i> (70)	<i>Sph</i> (737)	0.00446	40	60	/	0.00369
	Omnidirection	<i>Sph</i> (85)	<i>Sph</i> (309)	0.00488	60	40	164	0.00065
Site 2	E-W	<i>Sph</i> (72)	<i>Sph</i> (786)	0.00439	11	89	/	0.00214
	N-S	<i>Sph</i> (75)	<i>Sph</i> (677)	0.00279	39	61	/	0.00145
	Omnidirection	<i>Exp</i> (272)	<i>Sph</i> (1142)	0.00437	46	54	683	0.00105
Site 3	E-W	/	/	/	/	/	/	/
	N-S	<i>Exp</i> (95)	<i>Sph</i> (658)	0.00218	61	39	/	0.00545
	Omnidirection	<i>Exp</i> (146)	<i>Sph</i> (703)	0.00240	68	32	331	0.00061
Site 4	E-W	<i>Exp</i> (169)	<i>Sph</i> (724)	0.00213	60	40	/	0.00560
	N-S	<i>Sph</i> (84)	<i>Sph</i> (737)	0.00270	30	70	/	0.00172
	Omnidirection	<i>Exp</i> (171)	<i>Sph</i> (780)	0.00253	52	48	441	0.00062

Sph and *Exp*: Spherical model and Exponential model respectively (see Eq. 5 and 6); "/": no data.

In site 1 (i.e., a mosaic of dry land), the first ranges (a_1) in the E-W and N-S direction are similar: 59 m and 70 m respectively, which exhibit similar small scale spatial structures in both directions. This can mainly be explained by the

influence of the roads or ditches. Although the width of roads or ditches ranges between 10 and 20 m [16], and this is smaller than the detected ranges (a_1) but, due to the influence of the point spread function (i.e., PSF) resulting from the high

contrast between roads or ditches and fields in the imaging process, these line features can affect two or more pixels, and thus be detected (see Fig. 4). 66% and 40% of the image spatial variance in the E-W and N-S direction is influenced by these spatial structures. The remaining 60% spatial variance (i.e., w_2) in the N-S direction is mainly caused by the spatial arrangement of agricultural land patches, as their average length accords with the second range ($a_2=737$ m) (see Table 2 and Fig. 4). But the second range in the E-W direction is only 306 m, which is smaller than that of the average field length in the same direction. This is because of the spatial heterogeneity both among agricultural land patches and within each patch, such as the effect of N-S plowing direction [17].

In contrast to Site 1, Site 2 consists of both dry land and paddy fields with a dry land matrix (see Table 1, Fig. 4). The empirical variogram graph for Site 2 is smoother than that of Site 1 (see Fig.3). Although the directional spatial heterogeneities caused by linear features (roads and ditches) as in Site 1 have been detected by the variogram models ($a_1=72$ m, 75 for E-W and N-S respectively; see Table 2), they contribute little to the overall spatial heterogeneity because their sill weights are generally smaller than those of Site1, except for the E-W direction where $w_1=11$. This could be attributed to the low contrast between these line features and paddy fields in the NDVI imagery. The second range (a_2) is 786 m and 677 m in E-W and N-S direction respectively, and the corresponding sill weight (w_2) is 89% and 61% (both larger than 50%). Thus a_2 explains the average length of big agricultural land patches in this landscape (see Fig. 4), whereas w_2 shows that the total spatial heterogeneity in Site 2 is related to the spatial arrangement of agricultural land patches, with different land use types (dry lands and paddy fields).

For Site 3, two scales ($a_1=95$ m and $a_2=658$ m) were detected, corresponding to ditch and/or road structure and paddy field arrangement respectively (see Table 2). The sill weight of the first spatial structure is 61%, which means that the main spatial heterogeneity in the N-S direction is caused by the patterns of ditch and road. In the E-W direction, however, spatial heterogeneity is not observed at this landscape scale (i.e., 3000 m), since the variogram model does not level off before h_{max} (see Fig. 3).

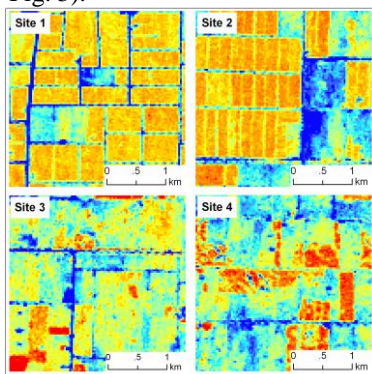


Fig.4 NDVI color images of the four study Sites

In Site 4, the first and the second range in the N-S direction are 84 m and 737 m respectively, which is similar to those in Site 2 (see Table 2). In the E-W direction, the second range

($a_2=724$ m) is similar that of Site 2, but the first range ($a_1=169$ m) is much larger ($a_1=72$ m in Site 2). Although both Sites 4 and 2 are composed of paddy field and dry land, their landscape patterns vary. The combination of paddy fields and dry lands with linear features (i.e., roads and/or ditches) is visually distinct in Site 2, but it is not in Site 4. Thus line features are not detected in this direction due to their low contrast with background features (e.g., paddy fields).

B. Omnidirection heterogeneity analysis

By integrating the heterogeneities in all directions, omnidirectional heterogeneities of the four study sites were analyzed. Values of the sill parameter c of the four study sites are ordered with Site 1 (0.00488) > Site 2 (0.00437) >> Site 4 (0.00253) > Site 3 (0.00240). Compared with Sites 4 and 3, more dry lands appear in Sites 1 and 2. Thus, it appears that dry lands may contain more spatial heterogeneity than paddy fields. But for parameter D_c (i.e., the integral length scale of spatial structure), the largest and smallest values occur in Sites 2 and 1: 683 m and 164 m respectively. This means that the spatial autocorrelation of Site 2 is highly positive and, of site 1 it is less positive. Although Sites 2 and 4 are composed of dry lands and paddy fields, their spatial autocorrelations differ a lot because of pattern description of agricultural land patches differences (see Table 1): the value of D_c of Site 4 is 441 m, which is much smaller than that of Site 2 (i.e., 683 m). Therefore, two factors, land use types and their patterns of distribution could all influence agricultural landscape spatial heterogeneity.

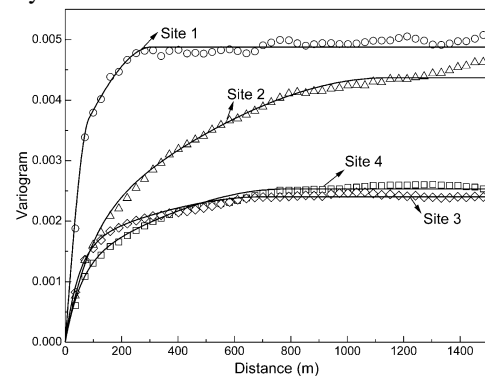


Fig.5 Omnidirectional empirical variograms with corresponding fitted models of the four study sites

V. DISCUSSION AND CONCLUSIONS

Four sites in Honghe farm were selected and tested in our study, with different agricultural landscapes including land use types and land use patterns, based on a Landsat NDVI image. First, by analyzing directional spatial heterogeneities in the N-S and E-W directions, it can be concluded that variogram models can quantify agricultural landscape patterns from two perspectives: spatial variance and length scale of the spatial structure. Causes of the landscape patterns can generally be recognized by comparison with basic characteristics observed on the ground. For example, in Site 1 (a soya bean growing area), the spatial heterogeneity in the E-W direction, is different from that in the N-S direction. It mainly results from the differences agricultural process (e.g., growth status) among

land patches due to different growing conditions such as soil moisture, soil fertility, solar energy distribution, etc, and from those within each field patch caused by plowing directions (i.e., N-S direction in this region). Second, considering the omnidirectional analysis, it was found that the spatial heterogeneity of the dry land matrix is normally larger than that of the paddy field matrix in this sample. This may be related to ridges (line features) contributing to the spatial variance in these dry land landscapes. Third, the multi-scale characteristics of agricultural landscape heterogeneity can generally be detected by the variogram analysis also, for instance, the main small and large spatial structures are respectively caused by the crisscross of roads/ditches and the different land use types between land patches.

Moreover, the quantified spatial heterogeneity also can be used to reduce the impacts of spatial heterogeneity that decreases the accuracy of quantitative agricultural remote sensing monitoring [3-4]. For instance, by integrating the spatial variability and the autocorrelation extent, D_c represents the mean extent of the image spatial structures at the landscape scale. Defining an appropriate pixel size at which the spatial structure and the spatial variability could be captured, is therefore possible, based on D_c , if the sampling space (i.e., spatial resolution in this study) is less than $D_c/2$, according to the Nyquist-Shannon sampling theorem [10]. In this study, we have the smallest D_c , i.e., 164 m, occurring in Site 1, thus the largest pixel size of monitoring remotely-sensed data should be less than 82 m [9](i.e., half of the smallest D_c) at the data acquisition time (i.e., 31/8/2006).

In landscape ecology, there is no single natural scale at which ecological phenomena should be studied; systems generally show characteristic variability on a range of spatial and temporal scales [6]. Fortunately, remote sensing now can provide various spatial resolution data which are temporarily related, ranging from micro scale (e.g., Quick Bird, IKONOS) to landscape scale (e.g., Landsat TM/ETM+, CBERS) and regional scale (e.g., MODIS, AVHRR), thus presenting a good opportunity for ecological process analysis. However, the agricultural landscape heterogeneities of the study area were analyzed based on only one temporal Landsat NDVI image. One NDVI image only represents the instantaneous heterogeneous characteristics of the crop, whereas the crop spatial heterogeneity at the landscape scale should be monitored during the whole of the growing season. What's more, the accuracy of the fitted variogram models is not fully investigated and needs to be further explored. Nevertheless, the proposed methodology, i.e., quantifying and analyzing spatial heterogeneity using variogram model on remote sensing variable data, is still giving insight into interesting spatial characteristics of the crops and could be extend to multi-temporal analyses.

ACKNOWLEDGMENT

This work is mainly supported by the Chongqing Municipal Science and technology research projects (No.

cstc2012ggB20001) and the National Natural Science Foundation of China (No. 40901290).

REFERENCES

- [1] Duveiller G, Defourny P (2010) A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sens Environ* 114:2637-2650
- [2] Justice CO, Becker-Reshef L (2007) Report from the Workshop on Developing a Strategy for Global Agricultural Monitoring in the framework of Group on Earth Observations (GEO). In: UN FA
- [3] Garrigues S, Allard D, Baret F, Weiss M (2006a) Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sens Environ* 105:286-298
- [4] Hu Z, Islam S (1997) A Framework for Analyzing and Designing Scale Invariant Remote Sensing Algorithms. *IEEE Trans Geosci Remote Sensing* 35:747-755
- [5] Garrigues S, Allard D, Weiss M, Baret F, Marni S and Jeanjean H (2002) Influence of spatial heterogeneity and scaling on leaf area index estimates from remote sensing data. In, *The first International Symposium on "Recent Advances Quantitative Remote Sensing"*. Torrent(Valencia),Spain
- [6] Levin SA (1992) The Problem of Pattern and Scale in Ecology. *Ecology* 73:1943-1967
- [7] Jenkins RB, Coops NC (2011) Landscape Controls on Structural Variation in Eucalypt Vegetation Communities: Woronora Plateau, Australia. *Aust Geogr* 42:1 – 17
- [8] Boone RB, Galvin KA, Smith NM, Lynn SJ (2000) Generalizing El Nino effects upon Maasai livestock using hierarchical clusters of vegetation patterns. *Photogramm Eng Remote Sens* 66:737-744
- [9] Garrigues S, Allard D, Baret F, Weiss M (2006b) Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens Environ* 103:81-96
- [10] De Cola L (1994) Simulating and mapping spatial complexity using multiscale techniques. *Int J Geogr Inf Sci* 8:411-421
- [11] Burrows SN, Gower ST, Clayton MK, Mackay DS, Ahl DE, Norman JM and Diak G (2002) Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems* 5:667-679
- [12] Cohen WB, Spies TA, Bradshaw GA (1990) Semivariograms of digital imagery for analysis of conifer canopy structure. *Remote Sens Environ* 34:167-178
- [13] Webster R (1985) Quantitative spatial analysis of soil in the field. *Adv Soil Sci* 3:1-70
- [14] Chilès J, Delfiner P (1999) *Geostatistics: Modeling Spatial Uncertainty*. John Wiley and Sons, New York
- [15] Pannatier Y (1996) *VARIOWIN: Software for Spatial Data Analysis in 2D*. Springer-Verlag, New York
- [16] Lu T, Ma KM, Fu BJ, Wei NH (2009) Effects of ditch network structure on landscape pattern in the San jiang Plain, Northeast China. *Acta Ecol Sinica* 28:2746-2752
- [17] Fu Q (2011) Effect of Different Ridge Directions on Yield of Soybean. *Heilongjiang Agri Sci* 4:29-31