CHINESE GEOGRAPHICAL SCIENCE
Volume 12, Number 2, pp. 182 – 185, 2002
Science Press, Beijing, China

SPATIAL CORRELATION ANALYSIS OF CROP YIELD IN THE MIDDLE AND WEST OF JILIN PROVINCE

LI Lin-yi[1], LIU Zhao-li[1], LI Chun-lin[2], WAN En-pu[1]
(*1. Changchun Institute of Geography, the Chinese Academy of Sciences, Changchun 130021, P. R. China; 2. Jilin Agricultural University, Changchun 130118, P. R. China*)

ABSTRACT: In this paper, spatial correlation of crop yield in the middle and west of Jilin Province is analyzed by using the method of geostatistics semivariogram, taking the NDVI of NOAA/AVHRR spectrum data as the regionalized variable, aiming to provide theory and practical basis for field sampling of crop yield estimation using remote sensing. The ratio of nugget variance and sill of semivariograms are 21.1% and 9.7% in the west and middle regions in Jilin Province respectively. This shows that the crop yields are spatially correlated. The degree and range of correlation are far different in the different situations. In the west test region, the range is 49.9km and the sill is 0.00019. In the middle test region, the range is 16.5km and the sill is 0.00453. The dissimilarity in the western test region is larger than that in the middle one. The range in which the correlation existed of the former is far larger than the later. Different characteristics of spatial correlation of crop yield are decided by the environmental factors. Samples for crop yield estimation should be extracted according to the characteristic of spatial distribution of crop yield to promote the efficiency of sampling.
KEY WORDS: spatial correlation; crop yield; semivariogram; geostatistics

CLC number: S127 Document code: A Article ID: 1002-0063(2002)02-0182-04

1 INTRODUCTION

Field sampling is a key link of crop yield estimation using remote sensing, and it is essential for obtaining precise results. In the past years, the determination of the number of observation and positioning the location of the sample points are all on the basis of probability sampling theory, which is random sampling based on an assumption of independent observations.

In fact, crop growth and yield are geo-referenced data, they are the values with spatial location change. Moreover, observations are correlated strictly due to their location positioning. Geo-referenced data should be sampled according the characteristics of their spatial distribution.

The observation is not independent. The redundancy increases as the degree of locational dependence increase(SANDRA and DANIEL, 1996). In order to obtain an average value of crop growth or yield with the same accuracy, a larger sample size is needed for the crop field with weak correlation than the one with strong correlation.

Disadvantages of analyzing geo-referenced data using probability theory and methods have been realized in some studies(SANDRA, 1996). Probability theory applied to geo-referenced data failing to capture locational information, but there were no effective solution owning to the limitation of many factors. Recently the big advances of spatial analysis theories and rapid development of technology such as remote sensing, geographic information system, global positioning system provide means to obtaining and processing geo-referenced data and make it possible to analyze the spatial correlation or dependence of these data. In these paper, spatial correlation of crop growth and yield in the middle and west of Jilin Province are analyzed by using the method of semivariogram of geostatistics, taking the NDVI computing from NOAA/AVHRR spectrum data as the regionalized variable.

Received date: 2002-03-12
Foundation item: Under the auspices of Major Program of the Chinese Academy of Sciences (KZ95T-03-01).
Biography: LI Lin-yi (1968 –), female, a native of Jilin Province, doctor candidate. Her research interest includes land use change monitoring and crop yield estimation using '3S' techniques.

2 STUDIED AREA

Jilin Province is situated in the northeast part of China between Heilongjiang and Liaoning provinces. In Jilin, from southeast to northwest, the topographic feature varies from mountain to hilly land, tableland and plain. Jilin Province belongs to temperate zone continental monsoon climate. Seasonal distribution of precipitation is uneven. Rainfall is concentrative in summer, which account for about 60% of annual precipitation.

Jilin Province is an important commodity grain base of China. Arable land is mainly distributed over the middle and west of the province, and 85% arable land is dryland. The main land cover in the east of Jilin is forest land. In the middle part of Jilin, soil is full of humus and well structured, and the climate is proper for crop growth with 500 – 600mm of rainfall annually and 2800 – 3000℃ of ≥ 10℃ accumulated temperature. The nature condition in the west part of Jilin is inferior to that in the middle part. Soil is infertile and salinization in some areas. Annual precipitation is 400 – 500mm, and even less than 400mm in some places. ≥ 10℃ accumulated temperature is more than 3000℃. The arid climate is the limitation to the agriculture in this part.

3 METHODS

Semivariogram of geostatistics are used to analyze the correlation of crop yield in this paper. The principle and method of semivariogram were elaborated in many papers and books. For a regionalized variable $Z(x)$, semivariogram $\gamma(h)$ is defined as half the expected square difference between values at places x and $x + h$,

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

where $\gamma(h)$ is the semivariogram, h is the interval between observations, $\gamma(h)$ can be computed by

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i}^{N(h)} [Z(x) - Z(x+h)]^2$$

where $\hat{\gamma}(h)$ is an estimate of $\gamma(h)$, $N(h)$ is pairs of observations separated by the same interval h.

$\hat{\gamma}(h)$ is a useful measure of dissimilarity between spatial separate observations. $\hat{\gamma}(h)$ increases with the distance between observations h theoretically. The larger γ, the less similar at the observations. Namely, semivariogram can be used to analyze the correlation between observations, including the degree of spatial correlation and the range in which spatial correlation exist (WANG, 1999; PAUL and PETER, 1998).

4 SPATIAL CORRELATION ANALYSIS

4. 1 Selection of Regionalized Variable

Vegetation index is the reflection of the capacity of photosynthesis and dry biomass production of plants. It is reported that the normalized difference vegetation index (*NDVI*) is well correlated to the leaf area index (*LAI*) and crop yield (MURPHY, 1996). *NDVI* increase as crop yield increase (HAMAR, 1996; MICHIO and TSUYOSHI, 1991). So the characteristics of spatial correlation of crop yield can be revealed by analyzing the correlation of *NDVI*. In this study, *NDVI* was taken as the regionalized variable. *NDVI* was computed using the channel 1 and channel 2 spectrum data of NOAA/AVHRR with the formula as follows:

$$NDVI = \frac{(CH2 - CH1)}{(CH2 + CH1)}$$

where $CH1$ is the red radiance and $CH2$ is the near infrared radiance.

4. 2 Data of Sample

NOAA/AVHRR spectrum data in August 25, 1995 covered Jilin Province was taken as the source data after being rectified, strengthened, edge filtered in this study. *NDVI* was computed in PC ENVI environment. Two 48 × 48 pixels regions of dry land with fewer combined pixels were chosen on the *NDVI* image in the middle and west of Jilin Province respectively, whose corresponding area on the ground are two 48km × 48km regions, which are representative of environmental conditions such as climate, soil and geomorphic type of the middle and western Jilin.

Pixels have been extracted in regular interval as shown in Fig. 1. The total number of samples is 576 in every test region. The descriptive statistics of the *NDVI* of these samples is listed in Table 1.

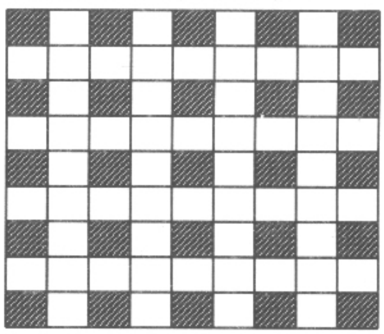

Fig. 1 Style of sampling for NOAA/AVHRR *NDVI*

Table 1 Summary statistics of samples in two test regions

Test region	Sample size	Mean	Sample variance	Standard deviation	Minimum value	Maximum value
Middle	576	0.6607	0.00010	0.0116	0.6200	0.6900
West	576	0.6602	0.00045	0.0212	0.5765	0.7059

* $F_{0.01}(1, 8) = 11.26$

4.3 Analysis of Spatial Correlation of Crop Yield

Experimental semivariograms are computed on geostatistics software GS + using the samples data of two test regions mentioned above. Theoretical model such as spherical model, exponential model, power model, linear model and Gaussion model are all fitted on the basis of experimental semivariogram. By comparison, spherical semivariograms are the optimal model for both of the test regions which are shown in Fig. 2 and Fig. 3. Their formula are:

Semivariogram of the middle test region

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 0.00004 + 0.00015\left(\frac{3}{2}\frac{h}{49.9} - \frac{1}{2}\left(\frac{h}{49.9}\right)^3\right) & 0 < h \leqslant 49.9 \\ 0.00004 + 0.00015 & h > 49.9 \end{cases}$$

Semivariogram of the middle test region

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 0.000044 + 0.000409\left(\frac{3}{2}\frac{h}{16.5} - \frac{1}{2}\left(\frac{h}{16.5}\right)^3\right) & 0 < h \leqslant 16.5 \\ 0.000044 + 0.000409 & h > 16.5 \end{cases}$$

And the relevant parameters are listed in Table 2.

Results shows that the crop yield in the two test regions are spatially correlated. Sill, maximum value of $\gamma(h)$, is consist of two parts, structured variance C and nugget variance C_0. The former represents a component of the variation that is spatially correlated, the later represents spatially uncorrelated component. Generally, the ratio of nugget variance and sill is the measurement of spatial correlation. If the ratio is less than 25%, spatial correlation of this variable is strong. If the ratio is between 25% to 75%, the correlation is middle. If the ratio is more than 75%, the correlation is weak (WANG *et al.*, 2000). In this study, the ratio of nugget variance and sill are both less than 25% which are 21.1% and 9.7% respectively. So the crop yields in the two test regions are strong spatial correlated. The sill indicates the degree of spatial variability, the higher the value of sill, the larger the spatial dissimilarity of the crop yield. The sill is 0.00019 in the middle test region and 0.00453 in the western test region. The dissimilarity in the western test region is larger than the middle one.

The range can be used to represent the range of the spatial correlation existed. Places closer than the range are correlated statistically, and places further apart are

Table 2 The best-fitted semivariogram of maize yield and corresponding parameters

Test region	Optimal semivariogram	Nugget C_0	Sill $C_0 + C$	Proportion $C/(C_0 + C)$	Range a	*RSS*	R^2	F^*
Middle	Spherical model	0.00004	0.00019	0.789	49.9	2.179E-10	0.990	792
West	Spherical model	0.000044	0.000453	0.903	16.5	4.388E-08	0.638	14.4

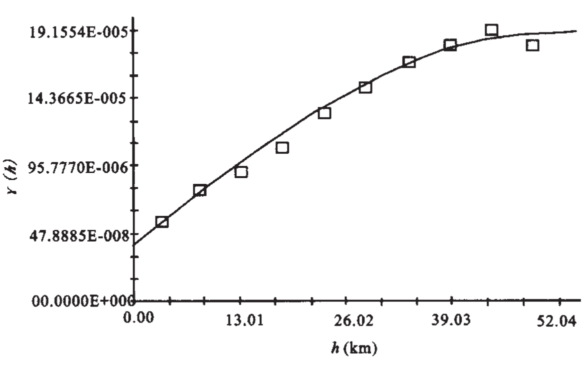

Fig. 2 Semivariogram of the test region in the middle of Jilin Province

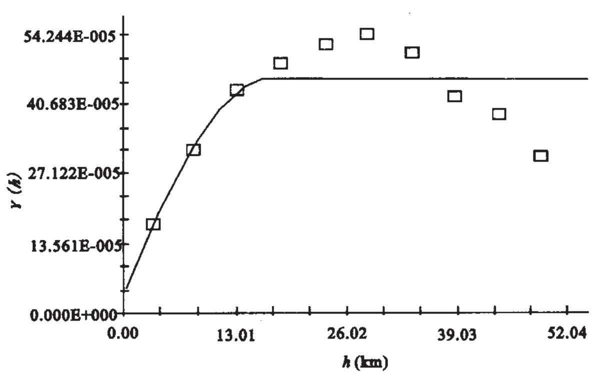

Fig. 3 Semivariogram of the test region in the west of Jilin Province

not. In the west test region, the range is 49.9km, and in the middle test region, the range is 16.5km, the former is far larger than the later.

Moreover, the yield maps are interpolated using kriging methods in these two areas. Comparing the yield map with soil nutrient map and geomorphologic map, it is found that the crop yield distribution are closely related to the characteristic of these environmental elements. Namely, the variability of crop yield are mainly caused by the different environmental conditions such as soil nutrient, micrometeorological and soil moisture conditions.

5 CONCLUSIONS

(1) Sampling in regular interval with support of 1km, crop yield spatial correlation are analyzed using semivariogram method of geostatistics on two 48km × 48km areas in the middle and west of Jilin Province respectively. Crop yield which reflected by *NDVI* are tested to be spatial correlated on this scale.

(2) The characteristics of spatial correlation of crop yield are quite different between regions tested. The differences lies in the degree of correlation and the range correlation exist. The similarity between two places in the middle test region is larger than the similarity between two places with the same distance within ranges. The range in which spatial correlation exist is larger in the middle test region than that in the west one. Different characteristics of spatial correlation of crop yield are decided by the environmental conditions under which crop grows including geomorphologic type, soil nutrient and micrometeorological condition

(3) Crop yield are spatially correlated, sampling method based on probability theory will lead to redundancy. Sampling for crop yield estimation should be designed and carried out according to the distribution of crop yield at different area so as to obtain precise result with less observations.

REFERENCES

HAMAR D, 1996. Yield estimation for corn and wheat in the Hungarian Great Plain using Landsat MSS data[J]. *International Journal of Remote Sensing*, (17): 1689 – 1699.

MICHIO Shibayama, TSUYOSHI Akiyama, 1991. Estimating grain yield of maturing rice canopies using high spectral resolution reflectance measurements[J]. *Remote Sensing of Environment*, (36): 45 – 53.

MURPHY C, 1996. Improved ground sampling and crop yield estimation using satellite data[J]. *International Journal of Remote Sensing*, (17): 954 – 956.

PAUL J Curra, PETER M Atkinson, 1998. Geostatistics and remote sensing[J]. *Progress in Physical Geography*, (22): 61 – 78.

SANDRA Lach Arlinghaus, DANIEL A Griffith, 1996. *Spatial-Statistics*[M]. New York, London, Tokyo: CRC Press, Inc. Boca Raton, 31 – 41.

WANG Jun, FU Bo-jie, QIU Yang, 2000. Spatial variability of soil moisture in small catchment on Loess Plateau—Semivariograms[J]. *Acta Geographica Sinica*, 55: 428 – 438. (in Chinese)

WANG Zheng-quan, 1999. *Geostatistics and Its Application in Eecology*[M]. Beijing: Science Press, 35 – 52. (in Chinese)