Journal of the Meteorological Society of Japan, Vol.79, No.1B, pp.555–574, 2001

Spatial Variation of Soil Moisture in China: Geostatistical Characterization

By Suxia Liu

Department of Hydrology, Institute of Geographical Sciences and Natural Resources Research,
Chinese Academy of Sciences, Beijing, P.R. China

Xingguo Mo

Station for Agro-ecology and Environmental Technology, Institute of Geographical Sciences
and Natural Resources Research, Chinese Academy of Sciences, Beijing, P.R. China

Haibin Li

Department of Hydrology, Institute of Geographical Sciences and Natural Resources Research,
Chinese Academy of Sciences, Beijing, P.R. China

Gongbing Peng

Department of Climatology, Institute of Geographical Sciences and Natural Resources Research,
Chinese Academy of Sciences, Beijing, P.R. China

and

Alan Robock

Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey, U.S.A.

(Manuscript received 14 February 2000, in revised form 5 October 2000)

Abstract

We analyze the large-scale spatial variation of soil moisture in eastern China using geostatistical techniques with observations at 99 stations for the top 0.1 m and top 1 m from 1987 to 1989. Sample variograms are found to have a clear sill and a nugget in many cases. A spherical variogram model, including a nugget in some cases, fits the sample variograms closely. Using a quantitative method to select the separation interval for variogram analysis, we find that the average range is 200–400 km for the top 0.1 m and 400–700 km for the top 1 m. The averaged coefficient of variation of soil moisture in the top 0.1 m is larger than for the top 1 m, showing that the range for the top 0.1 m is less than for the top 1 m. The range in summer is less than in winter. By calculating the ratio of the nugget effect to the spatial variance, we find that the ratios for the top 0.1 m data are smaller than that for the top 1 m data, showing that in most cases the spatial variation of the top 0.1 m is more strongly autocorrelated than of the top 1 m, and that the measurement errors are much larger in the top 1 m data than in the top 0.1 m. For more than half of the measured dates, the ratio of the nugget effect to the variance is less than 20 percent, indicating that the spatially correlated variation on those dates can explain more than 80 percent of the total variance.

1. Introduction

Knowing the spatial scale of soil moisture variations is important for understanding patterns of

---

Corresponding author: Suxia Liu, Current address: School of Mathematics, The University of New South Wales, Sydney NSW 2052, Australia.
E-mail: suxia@maths.unsw.edu.au

climate change, for developing and evaluating land surface models, for designing surface soil moisture observation networks, and for determining the appropriate resolution for satellite-based remote sensing of soil moisture. Vinnikov et al. (1996) developed a statistical model of spatial variations of soil moisture that partitions the variations into red noise and white noise components. They used Russian soil moisture observations to show that while there is a certain amount of small-scale variability related to soils, topography, vegetation, and root structure, the red noise component of spatial variability represents most of the soil moisture variance and reflects the statistical properties of the monthly averaged precipitation field. This demonstrates atmospheric control of soil moisture variability, and the scale of spatial correlation of this component is about 500 km. The estimates of scales of spatial correlation did not differ significantly for water content in the top 0.2 m and 1 m layers of soil. Entin et al. (2000) extended this analysis using soil moisture observations from Illinois (U.S.A.), Mongolia, Russia, and China. They showed that the spatial scale for all these regions is about 500 km for both the top 0.1 m and top 1 m. In particular, for China the mean scale was 525 km for both layers. Here we re-examine these results using a different statistical technique.

As an example of the detailed interactions on a smaller scale, Entakhabi and Rodriguez-Iturbe (1994) developed an analytical model of soil-water balance that is distributed in space and time. Surface hydrology is forced with a stochastic model of rainfall that captures the arrivals of intense precipitation cells within clusters that are embedded in squall lines and rainbands. An analysis of spatial and temporal fluctuation in both the rainbands and soil moisture fields made in the frequency domain showed that averaging over specific periods and spatial scales resulted in significantly altered statistical structures in the rain intensity and soil moisture fields. Averaging over 10–100 km by formulating a lumped model resulted in a complete alteration of the dynamics of the hydrological processes. This has implications for zero-dimensional regional water balance and hydrological parameterizations in mesoscale meteorological models and GCMs. Averaging in blocks larger than 100 km did not deteriorate the signal any further, since the structured portion of the rain intensity and soil moisture spectra was at smaller scales. It is this larger scale that we examine with in situ soil moisture observations in China.

Kriging, developed at Ecole des Mines de Paris (Matheron 1971), is of particular interest in research on the spatial scale of hydrological variables. It permits the evaluation of a variable at any ungauged site of a domain and yields the variance of the estimate using gauged stations. It is also meaningful for the interpolation of spatial patterns using point data, estimation of average catchment soil moisture, and for distributed hydrological modeling in general (Abbott and Refsgaard 1996). The spatial structure of soil moisture is greatly influenced by vertical and horizontal processes of water movement and energy transfer in the soil-vegetation-atmosphere system. These processes may be modulated by many factors such as topography, soils, geology and vegetation. The resultant soil moisture pattern is likely to reflect these influences through variations at scales linked to the processes controlling the soil moisture pattern. To discuss its geostatistical correlation structure in space is one way of characterizing this variability.

Deutsch and Journel (1992) theoretically found that indicator variograms could be used for a similar range of geostatistic estimation techniques as standard variograms in ordinary kriging. Indicator variograms are expected to be more flexible, as they allow different ranges for small and large values of a variable and can represent connectivity of high values in spatial fields. However, Western et al. (1998a) showed that the ability of the indicator approach to capture connectivity could not be shown conclusively, even when using their large number of soil moisture measurements from the 10.5 ha Tarrawarra catchment in southeast Australia. Based on many successful examples of the application of the ordinary kriging method on hydrological variables (Villeneuve et al. 1979; Wang and Guo 1988; Liu and Liu 1994; Ashraf et al. 1997), the ordinary kriging method works quite well for the spatial scale analysis of soil moisture.

Yang and Lei (1993) arranged 100 sampling sites along a line at a one meter interval at the Experimental Station of Guan County in Shandong Province in China and found that the spatial scale is very small, about 16 meters, and concluded that the spatial relationship among the field soil-water content is of little significance. Vinnikov et al. (1996) and Entin et al. (2000) explain this as the portion of the variance controlled by small-scale variations in the land surface, but demonstrates for every region examined that there is also large-scale variation controlled by atmospheric variations. Shi et al. (1997) estimated the regional distribution of soil moisture in a large area of 65,100 km^2 in the West Liao River Plains by using the kriging method. Their calculation shows that the spatial correlation scale of soil moisture in the top 1 m is about 200 km based on two groups of data collected in a dry year and a humid year. Western et al. (1998b) measured the spatial soil moisture patterns in the 10.5 ha Tarrawara catchment and found that the geostatistical structure evolved seasonally. Exponential variogram models, including a nugget, fit the sample variograms closely. High sills (15-25%^{2}), and low correlation lengths (35-50 m), were observed during the wet winter periods. During the dry summer period sills are smaller (5-15%^{2}) and correlation lengths are longer (50-60 m). Both a nugget effect due to measurement error and variability at small scales contribute to the variability at the 10 m scale, which is the smallest scale in most of the data sets. Western et al. (1998b) gives a good summary of previous studies of spatial soil moisture variability with geostatistical methods, all of which focus small areas, the largest being less than 10 km^2.

In this paper we explore the spatial scale of soil moisture in China based on two groups of data specially chosen from our large observed database over a large area of China for 11 years (Robock et al. 2000) with geostatistical analysis. We first describe the data we use, and then identify the basic characteristics of spatial variation by calculating the coefficient of variation. After explaining the theory of variogram analysis, we propose a quantitative method to determine the separation interval and the method to determine the longest separation distance, which are two of the key aspects of variogram analysis. The type, the scale, and the degree of spatial variation resulting from the variogram analysis are then discussed in detail. Finally we give a short conclusion.

2. Study area and data set

We have established a data set of observed soil moisture over 102 stations from 1981 to 1991 in China. The data set is described by Robock et

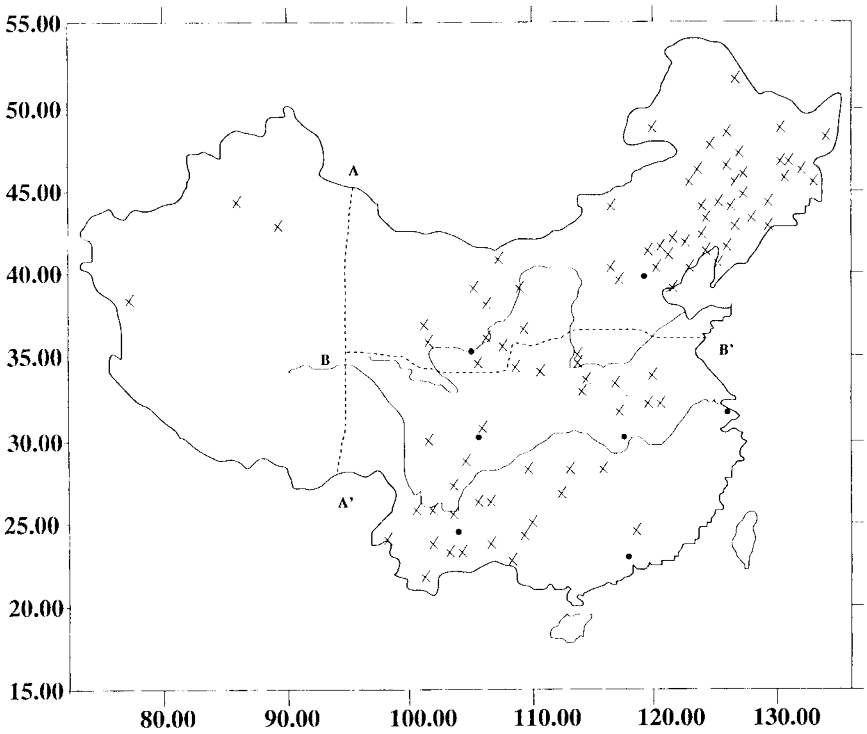

Fig. 1. The distribution of soil moisture stations. Zone A is the region east of the line A–A' and Zone B is the portion of Zone A south of the line B–B'. The Yellow River and Yangtze River are shown in the north and south respectively. Crosses show the location of soil moisture stations. The big solid dots show main cities such as Shanghai, Beijing, Wuhan, Guangzhou, Lanzhou, Chengdu and Kunming from east to west and from north to south.

al. (2000), and was used by Entin et al. (1999, 2000) and Srinivasan et al. (2000). In this paper we use the data from 1987 to 1989 for geostatistical analysis because the sample size within these three years reaches the maximum in the 11-year period of observations. Total soil moisture observations for the top 0.1 m and top 1 m soil layers (in cm) from east China for 99 stations (Fig. 1, east of dashed line A–A') are used in this analysis. In other words, 3 stations in the western half of China were discarded because the sparse network gives a poor representation for geostatistical analysis. Because in winter there are one to two months having no records of soil moisture measurement in the North because of frozen soil, we choose two periods for analysis. Period A is from May 28 to Oct. 28, and period B is from Dec. 18 to Feb. 18 of the next year. The sample sizes within period A and period B do not vary in time, as shown in Fig. 2. There are 48 measurement days in period A and 23 measurement days in period B over the three years. Within period A, the sampling sites are distributed all over eastern China, as shown east of the dashed line A–A' in Fig. 1. Within period B, however, the sampling sites are limited to the southern part of the region. For convenience, the area corresponding to period A is called Zone A (area east of the dashed line A–A' in Fig. 1) and the area corresponding to period B is called Zone B (area south of the dashed line B–B' in Fig. 1). Obviously Zone B is included in Zone A. The basic facts of the data set are summarized in Table 1.

The mean and coefficient of variation (C_v) of soil moisture over Zone A and Zone B and antecedent precipitation for each measurement day are shown in Table 2. The definition of C_v for a series x_i, $i = 1, ..., n$, is:

$$C_v = \frac{1}{\overline{x}} \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{1/2}, \tag{1}$$

where

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{2}$$

is the mean (Mn) of the series x_i. Series can be time series or spatial series. In this paper we use Cvt and Mnt as the coefficient of variation and mean of a time series, while Cvs and Mns are the coefficient of variation and mean of a spatial series. For convenience, we normalized x_i as follows:

$$x_i' = (x_i - \overline{x})/\sigma, \tag{3}$$

where

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2, \tag{4}$$

and σ is the standard deviation. The results of the variation of Mns and Cvs of soil moisture with measurement dates are shown in Fig. 3. It appears from the first two figures in Fig. 3 that in both Zones A and B the variation of soil moisture in the top 0.1 m with time is larger than the variation of

Table 1. Basic Facts, C_v and Mean of datasets.

Data set		Group A	Group B
Averaged No. of Sampling sites (top 0.1 m)		84	34
Averaged No. of Sampling sites (top 1 m)		66	25
Sampling Periods		May 28-Oct. 28 (period A)	Dec. 18-Feb. 18 (period B)
Coverage		Zone A (South+North China)	Zone B (South China)
C_v	Antecedent Rainfall	1.19	1.83
	Soil moisture in the top 0.1 m	0.38	0.33
	Soil Moisture in the top 1 m	0.31	0.19
Mean	Antecedent Rainfall (mm)	33.04	6.79
	Soil moisture in the top 0.1 m (cm)	2.22	2.32
	Soil Moisture in the top 1 m (cm)	27.82	32.05

Table 2. (a) The dates, number of sampling points, spatial mean, coefficient of variation (C_v) of antecedent rainfall, soil moisture in the top 0.1 m and top 1 m soil layer for Zone A.

Date	Antecedent Rainfall (10 days)			Soil Moisture in the Top 0.1 m			Soil Moisture in the Top 1 m		
	Mean (mm)	Sample Number	C_v	Mean (cm)	Sample Number	C_v	Mean (cm)	Sample Number	C_v
05/28/87	33.10	93	1.08	2.25	90	0.36	27.17	63	0.31
06/08/87	45.56	93	0.90	2.42	89	0.35	29.06	69	0.29
06/18/87	21.88	93	0.94	2.13	83	0.36	28.57	64	0.28
06/28/87	27.97	93	1.12	2.11	85	0.36	26.99	68	0.29
07/08/87	52.45	93	0.94	2.33	86	0.41	28.48	69	0.33
07/18/87	44.18	93	0.88	2.29	82	0.36	28.07	67	0.30
07/28/87	48.57	93	0.88	2.26	84	0.38	27.70	66	0.30
08/08/87	55.65	93	0.94	2.38	83	0.37	28.11	64	0.31
08/18/87	47.99	93	0.82	2.42	84	0.36	28.47	67	0.32
08/28/87	67.53	93	0.92	2.53	80	0.35	29.31	64	0.31
09/08/87	42.91	93	0.71	2.46	85	0.36	29.13	67	0.32
09/18/87	13.10	93	1.61	2.15	85	0.36	28.24	71	0.31
09/28/87	32.01	93	1.11	2.36	87	0.38	29.53	72	0.30
10/08/87	13.26	93	1.77	2.11	88	0.42	28.42	73	0.32
10/18/87	28.32	93	1.13	2.49	80	0.37	30.14	64	0.31
10/28/87	7.25	91	1.58	2.32	87	0.37	30.14	70	0.31
05/28/88	32.35	91	1.02	2.18	87	0.37	27.36	62	0.32
06/08/88	26.45	91	0.82	2.11	86	0.37	27.19	66	0.30
06/18/88	32.28	91	1.16	2.17	86	0.45	27.13	70	0.33
06/28/88	27.41	91	1.91	1.85	83	0.44	25.78	66	0.35
07/08/88	50.70	91	1.02	2.31	79	0.36	28.42	61	0.27
07/18/88	41.20	91	1.03	2.11	86	0.39	27.64	67	0.32
07/28/88	52.08	91	0.91	2.05	82	0.42	27.23	67	0.32
08/08/88	42.56	91	1.14	2.13	83	0.41	28.09	65	0.31
08/18/88	45.50	91	0.82	2.38	81	0.31	28.16	64	0.29
08/28/88	52.03	91	0.85	2.42	75	0.35	28.83	61	0.28
09/08/88	51.26	91	0.83	2.52	76	0.31	29.54	60	0.25
09/18/88	30.27	91	1.01	2.40	80	0.36	29.93	65	0.26
09/28/88	9.66	91	2.70	2.11	85	0.36	29.25	67	0.25
10/08/88	15.55	91	1.12	2.27	88	0.36	29.22	69	0.25
10/18/88	12.28	89	1.20	2.27	86	0.36	29.12	69	0.24
10/28/88	9.46	88	2.18	2.18	85	0.34	29.97	68	0.25
05/28/89	16.06	91	1.50	1.85	87	0.45	26.27	67	0.35
06/08/89	44.34	91	0.82	2.26	82	0.39	27.41	62	0.34
06/18/89	37.11	91	0.63	2.33	83	0.37	28.20	63	0.32
06/28/89	24.86	91	1.60	2.16	87	0.42	27.32	67	0.33
07/08/89	39.32	91	1.05	2.14	85	0.41	26.56	66	0.35
07/18/89	49.20	91	1.03	2.31	79	0.38	27.74	63	0.33
07/28/89	55.88	91	0.88	2.35	83	0.35	27.87	65	0.33
08/08/89	30.40	91	1.79	1.87	85	0.46	25.71	66	0.35
08/18/89	26.31	91	0.98	1.95	83	0.43	26.78	63	0.34
08/28/89	32.62	91	0.99	2.07	84	0.40	25.43	66	0.36
09/08/89	34.28	91	1.00	2.16	89	0.40	25.99	69	0.36
09/18/89	18.71	93	1.49	2.02	91	0.41	26.17	72	0.36
09/28/89	27.87	93	1.19	2.30	88	0.34	26.49	70	0.34
10/08/89	10.29	93	1.54	2.10	88	0.35	25.99	71	0.34
10/18/89	16.26	93	1.04	2.23	89	0.32	25.83	70	0.32
10/28/89	9.45	87	2.59	2.12	89	0.38	25.38	68	0.33

Table 2. (b) The dates, number of sampling points, spatial mean, coefficient of variation (C_v) of antecedent rainfall, soil moisture in the top 0.1 m and top 1 m soil layer for Zone B.

Date	Antecedent Rainfall (10 days)			Soil Moisture in the Top 0.1 m			Soil Moisture in the Top 1 m		
	Mean (mm)	Sample Number	C_v	Mean (cm)	Sample Number	C_v	Mean (cm)	Sample Number	C_v
01/08/87	16.20	37	0.87	2.54	34	0.33	33.11	25	0.18
01/18/87	4.55	37	1.22	2.32	33	0.36	31.76	23	0.20
01/28/87	3.54	38	1.39	2.25	37	0.35	31.56	28	0.21
02/08/87	5.70	39	1.67	2.25	37	0.29	31.19	28	0.20
02/18/87	15.25	40	1.30	2.23	33	0.44	30.85	24	0.25
12/08/87	4.41	38	3.17	2.38	35	0.30	32.82	26	0.18
12/18/87	3.29	37	1.76	2.32	35	0.30	32.48	26	0.18
12/28/87	0.16	37	3.91	2.13	34	0.30	31.69	26	0.18
01/08/88	3.53	36	1.44	2.19	34	0.32	32.56	25	0.18
01/18/88	5.70	36	1.14	2.28	32	0.32	33.12	23	0.15
01/28/88	1.51	36	2.19	2.22	34	0.33	32.62	25	0.17
02/08/88	5.38	36	2.18	2.17	34	0.35	31.70	25	0.23
02/18/88	6.72	37	1.12	2.20	34	0.36	31.50	25	0.21
12/08/88	1.47	37	1.94	1.93	34	0.32	29.85	24	0.16
12/18/88	0.75	37	2.71	1.88	35	0.33	29.41	25	0.17
12/28/88	2.76	37	1.42	2.01	34	0.33	30.11	25	0.17
01/08/89	20.29	36	0.99	2.64	32	0.30	33.19	23	0.18
01/18/89	19.46	36	1.04	2.92	30	0.31	35.04	21	0.16
01/28/89	5.14	36	1.33	2.69	34	0.27	34.02	25	0.19
02/08/89	10.21	36	1.83	2.68	33	0.29	34.03	23	0.18
02/18/89	14.26	37	1.46	2.67	35	0.31	32.77	25	0.19
12/08/89	0.86	40	4.76	2.12	38	0.33	30.38	28	0.19
12/18/89	5.12	39	1.29	2.35	36	0.37	31.41	27	0.19

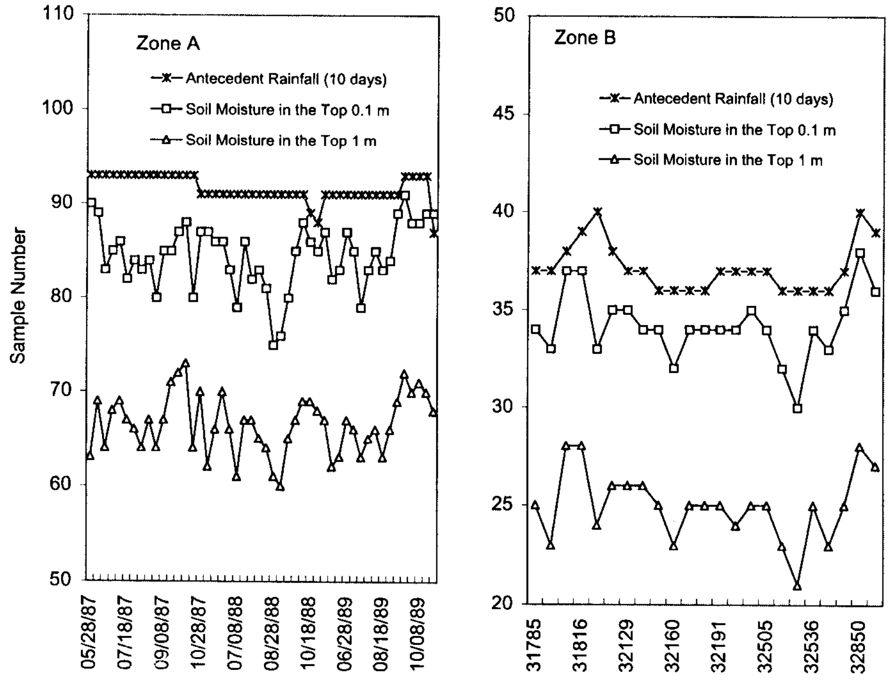

Fig. 2. The temporal change of sampling size in Zone A and Zone B.

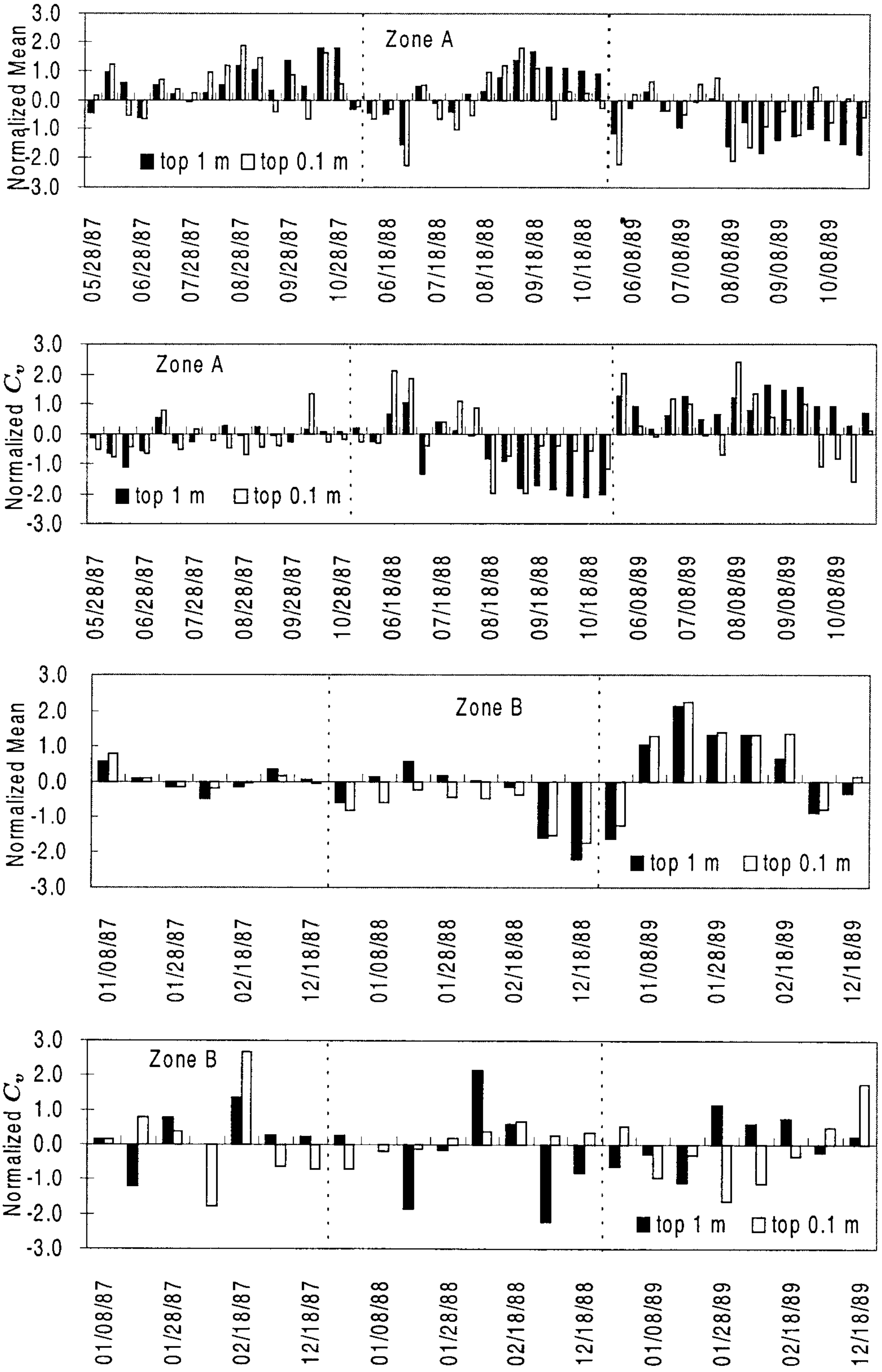

Fig. 3. The normalized mean and the coefficient of variation of soil moisture in the top 0.1 m and top 1 m over Zone A and Zone B for each measurement date. The dash lines separate years.

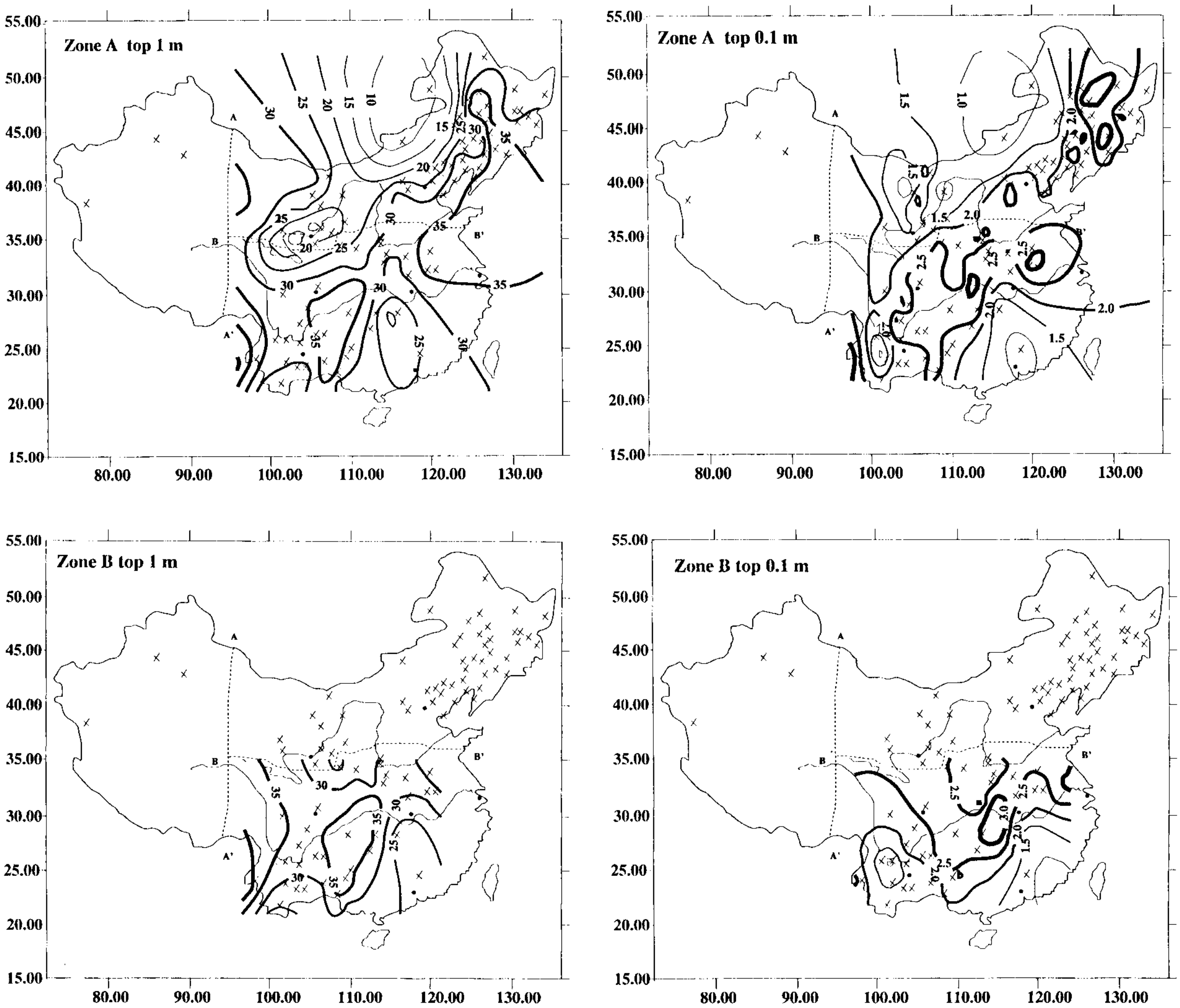

Fig. 4. Map of mean soil moisture for Zone-A-top-1- m, Zone-A-top-0.1- m, Zone-B-top-1- m and Zone-B-top-0.1- m. Units are cm of total soil moisture.

soil moisture in the top 1 m. It can also be seen that both in the top 0.1 m and top 1 m the variation of soil moisture in Zone A is larger than that in Zone B. The last two figures in Fig. 3 shows similar results, especially for spatial variations. The average C_v in Zones A and B both for the top 1 m and top 0.1 m shown in Table 1 supports these conclusions, as do the soil moisture maps shown in Fig. 4.

The impression from Fig. 3 is that there is no variation pattern over period A and period B, both from the point of view of the amount of soil moisture and spatial variation of soil moisture. Any measurement date from period A or B could represent the averaged case for the same period. This means that the data sets from both period A and period B are good for spatial analysis. Figure 4 shows the variation of Mnt and Cvt with observing stations.

The averages of C_v in Zones A and B both for the top 0.1 m and top 1 m are shown in Table 1. In both Zone A and Zone B, the average C_v in the top 0.1 m is larger than C_v for the top 1 m. For both the top 0.1 m and the top 1 m, the averaged C_v in Zone A is larger than that in Zone B.

Figure 5 shows the variation of antecedent rainfall, soil moisture in the top 0.1 m, and soil moisture in the top 1 m for periods A and B. A higher antecedent rainfall corresponds to wetter soil and vice versa, as expected, but the response of the av-

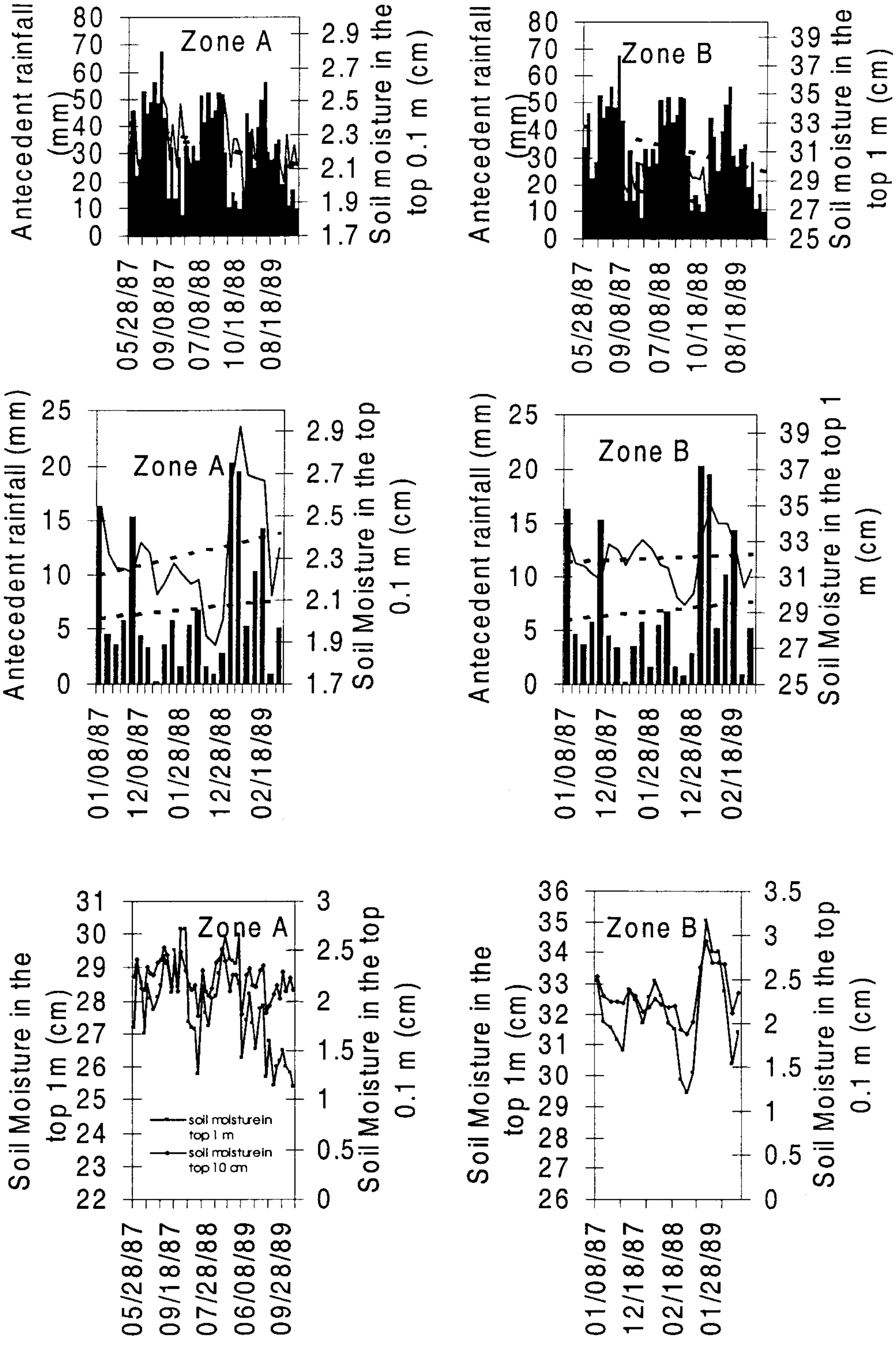

Fig. 5. The variation of antecedent rainfall, soil moisture in the top 0.1 m and top 1 m for each measurement date. (dashed line: tendency; curved line: soil moisture, column: antecedent precipitation)

erage amount of soil moisture to the average amount of precipitation is not monotonic. The difference of the averaged antecedent rainfall between Zones A and B seen in Table 1 did not cause a corresponding variation of soil moisture. This shows why it is important to analyze actual soil moisture observations. Figure 5 also shows a strong correlation between the variation of soil moisture in the top 0.1 m and in the top 1 m. The correlation coefficient between soil moisture in the top 1 m, and soil moisture in the top 0.1 m is larger than 0.5 and reaches 0.9 in Zone B.

3. Method of analysis

3.1 Variogram analysis

Variogram analysis was performed using standard geostatistical techniques (e.g., Journel and Huijbregts 1978; Isaaks and Srivastava 1989). The method used to explore the spatial variability (scale) of soil moisture in China in this paper is to first draw a variogram and then identify the structural parameters of geostatistical variograms, such as the sill, the correlation length (or range), and the nugget. Generally, omni-directional variograms were used in this analysis and sample variograms were calculated using all pairs separated by lags up to the longest separation distance, 2079 km. This is about half of the maximum separation distance as adopted elsewhere (e.g., Western et al 1998b). Later in this paper we discuss the longest separation distance in more detail. The sample variogram, $\gamma_s(h)$, at a given lag, h, is:

$$\gamma_s(h) = \frac{1}{2N(h)} \sum_{i,j} (\theta_i - \theta_j)^2, \tag{5}$$

where N is the number of pairs, θ_i and θ_j are the soil moisture in the top 0.1 m or 1 m soil layer at points i and j respectively, and the summation is conducted over all i, j pairs in that lag bin. The values of $\gamma_s(h)$ increase with increasing distance until they level off at a certain distance. This implies that the data at a small separation distance have similar values and that data at a large separation distance are likely to have quite different values. Thus, the variogram numerically describes the spatial continuity of the variation of variables. Pairs were grouped into lag "bins" and the equation (5) was used to calculate the variogram for that bin. The mean lag of all the pairs in a particular bin was used as the representative lag for that bin. For example, if the separation interval is 50 km, then those pairs for which the separation distance is between 25 km and 75 km are grouped into the calculation. The average of the results is considered as the value for that bin, the mean lag being 50 km. A detailed discussion of the choice of separation interval and longest separation distance is presented in sections 3.2 and 3.3 below.

Theoretically, a variogram of a random function should be close to zero when the separation distance h approaches zero. However, in reality a discontinuity is often present. This discontinuity is called the nugget effect (c_0) and is thought to result from measurement errors or insufficient sampling interval for assessing the underlying structure. Clark (1980) suggests that the nugget effect to some extent reflect the built-in random nature of a random function. This random variation cannot be predicted by any methods. If this random variation, as reflected by the nugget effect, accounts for a large proportion of the variogram, a geostatistical estimation may be of limited use.

Vinnikov et al. (1996) and Entin et al. (2000) discuss the same effect, but identify it as the small scale noise produced by hydrological variability as discussed in section 1. Vinnikov et al. (1996) use the symbol a_0 for the nugget, but did not label it as such.

The distance at which the sample variogram levels off is called the range (a). The variability beyond this range does not depend on the separation distance and the variables are no longer related spatially. The range can be viewed as a zone of influence of a variable or a transition from a state of spatial correlation to a state of absence of correlation. In this study, the range is regarded as the scale of spatial correlation.

The value of the sample variogram where it levels off is called the sill. The sill should be equivalent to the variance of the random function, because when the variogram reaches its sill, the covariance of variables approaches zero. Although the sill itself is of limited significance, the difference between the sill and nugget is of considerable interest because it represents the amount of variance due to the spatial correlation. A sample variogram figure with the important elements is shown in Fig. 6.

It is important here to recognize the difference between sample properties and the properties of the true underlying distribution (i.e., the population). Strictly speaking, the sample variogram computed from the data gives the properties of the sample only. It is then statistically inferred, by fitting a smooth curve (a theoretical variogram)

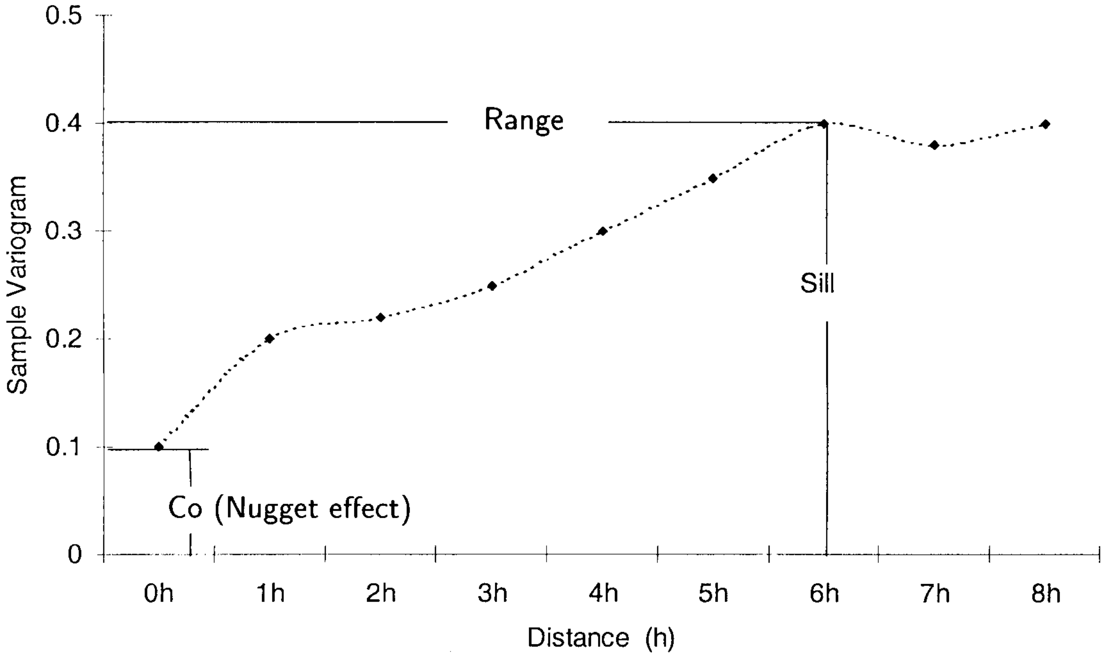

Fig. 6. An example of a sample variogram with sill, range and nugget effect. The nugget effect is the value of the sample variogram when the separation distance approaches zero; the range is the distance at which the sample variogram levels off and the sill is the value of the sample variogram where it levels off.

to the sample variogram that gives an estimate of the properties of the population. Whether the sill, range, and the nugget inferred from the sample are representative of the true values depends on the nature of the variability and the sampling regime. We fitted variogram models to the sample variograms by visual inspection. Theoretical spherical models with a nugget were used for all fitted variograms:

$$\gamma_t(h) = \begin{cases} c_0 & h = 0 \\ c_0 + c_1(3/2 \cdot h/a - 1/2 \cdot (h/a)^3) & 0 < h < a \\ \delta^2 & h \geq a \end{cases} \quad (6)$$

where $\gamma_t(h)$ is the fitted variogram, c_0 is the nugget, δ^2 is the sill, a is the range, and $c_1 = \delta^2 - c_0$. We use the direct method (Hou and Huang 1982) to get the parameters a, c_0 and δ^2 as follows:

(1) Draw a tangent line through the first two or three points of the sample variogram. The intercept of this line on the y-axis is an estimate of the nugget effect.
(2) Draw the sill. Since the sill should be close to the variance of the data, the variance is calculated first. The variance is used as an estimate of the sill.
(3) For a spherical model, the intersection of the tangent line and the sill is an estimate of two thirds of the range.

Usually, the quality of the variogram fit is characterized using the root mean square error (RMSE) calculation:

$$RMSE = \left[\frac{\sum(\gamma_t(h) - \gamma_s(h))^2}{n_h} \right]^{1/2}, \quad (7)$$

where n_h is the number of lag bins in the sample variogram (Western et al. 1998b).

3.2 Separation interval

With a given set of samples, it is always possible to compute a sample variogram. However, the procedure does not always provide an accurate estimation, especially when samples are located in a scattered pattern, as in our case. This is because the variogram depends on separation interval. A poor choice of separation interval could produce a sample variogram characterized by very erratic behavior (Armstrong 1984), resulting in an inappropriate fit to the underlying variogram. The separation interval must be chosen carefully before one calculates the sample variogram.

Generally speaking, using a small separation interval to calculate a sample variogram could help to accurately identify the nugget effect (Kim and Knudsen 1977). However, if sample variograms are calculated using smaller separation intervals, they become less smooth due to the decrease in the number of sample pairs used to calculate each point on the variogram.

On the other hand, experience tells us that there exists a maximum separation interval, over which an erratic variogram results. Usually $\gamma_s(h = 0)$ should be less than $\gamma_s(h = \Delta)$, where Δ is the separation interval. Experience also tells us that there exists a $\Delta_{\max}$, where $\gamma_s(h = 0)$ turns to be larger than $\gamma_s(h = \Delta_{\max})$, which produces an erratic variogram. We can find this $\Delta_{\max}$ by experiments beginning with the minimum separation interval, 50 km in this paper (the minimum distance of 102 station points is 41 km in China). By increasing the separation interval with a search step 25 km in this paper each time, we can reach $\Delta_{\max}$. By interactive plotting we can find $\Delta_{\max}$. According to the principle that the smaller the separation interval the less smooth the sample variograms, we take $\Delta_{\max}$–25 km as our optimal separation interval because it can give us the smoothest correct sample variogram. From trial and error, we found that the sample variograms, with a separation interval of 75 km, always have null nugget effect. However, a zero nugget does not necessarily always correspond to this separation interval. Because the soil moisture data in the top 1 m in Zone A have more sampling errors with non-nugget effects, the separation interval corresponding to the null nugget effect is 125 km. The relationship between the separation interval and null nugget effect is shown in Fig. 7.

3.3 Longest separation distance

Another problem associated with estimating an accurate sample variogram is that the sample variogram is highly variable for large values of separation distance, that is, if the sample variograms calculated using different subsets of the data are compared, there could be a significant difference between them (Armstrong 1984). This occurs because the number of data pairs used in calculating the values of a sample variogram at a long distance is relatively small. If data are sampled from an irregular distribution in space, this problem is even more severe because the number of data pairs for large separation distances could vary tremendously. Inspecting the distribution of the number of data pairs along different separation distances, and determining the longest separation distance for variogram calculations are necessary for estimating accurate sample variograms.

Figure 8 shows the variation of number of data pairs used for estimating the values of sample variograms on four typical dates which respectively represent Zone A 0.1 m, Zone A 1 m, Zone B 0.1 m, and Zone B 1 m. As shown in Shao (1992), the values of a mean sample variogram calculated using less than 50 pairs of data for large values of separation distance are given little weight in variogram estimation. She used this criterion to decide the largest separation distance based on the variation. However, it is hard to make a decision on the largest separation distance only from Fig. 8 because the largest separation distance varies month to month and year by year. Together with Fig. 8, we will use all of the sample variograms to make a decision.

Four typical variograms are shown in Fig. 9. The larger the separation distance is the less the number of data pairs and the more the sample variogram fluctuates. For soil moisture data in the top 0.1 m in Zone B, on an average, the largest separation distance is about 1500 km. When the separation distance is larger than 1500 km, the sample variogram varies quite a bit. The largest separation distance is 3000 km for soil moisture data in the top 0.1 m in Zone A. For the top 1 m, the separation distance is also 1500 km in Zone B, but is 2600 km in Zone A. Obviously, these largest separation distances correspond to different data pairs for each measurement date. Therefore we did not take the criterion of the number of data pairs as adopted in Shao (1992) to select the largest separation distance.

4. Spatial variability

Figure 9 shows four typical semi-variograms. Each sample from the study period can be fitted with a spherical model and nugget effect. This indicates that spatially correlated variations exist for both Zone A and Zone B. However, the nugget effect and the parameters of the fitted model (i.e., range and sill) varied between the dates, as shown in Table 3. This implies that the scale and degree vary with time. However, the average scale and degree give us interesting results, as shown below.

4.1 Types of spatial variation

The spherical model we used implies that from

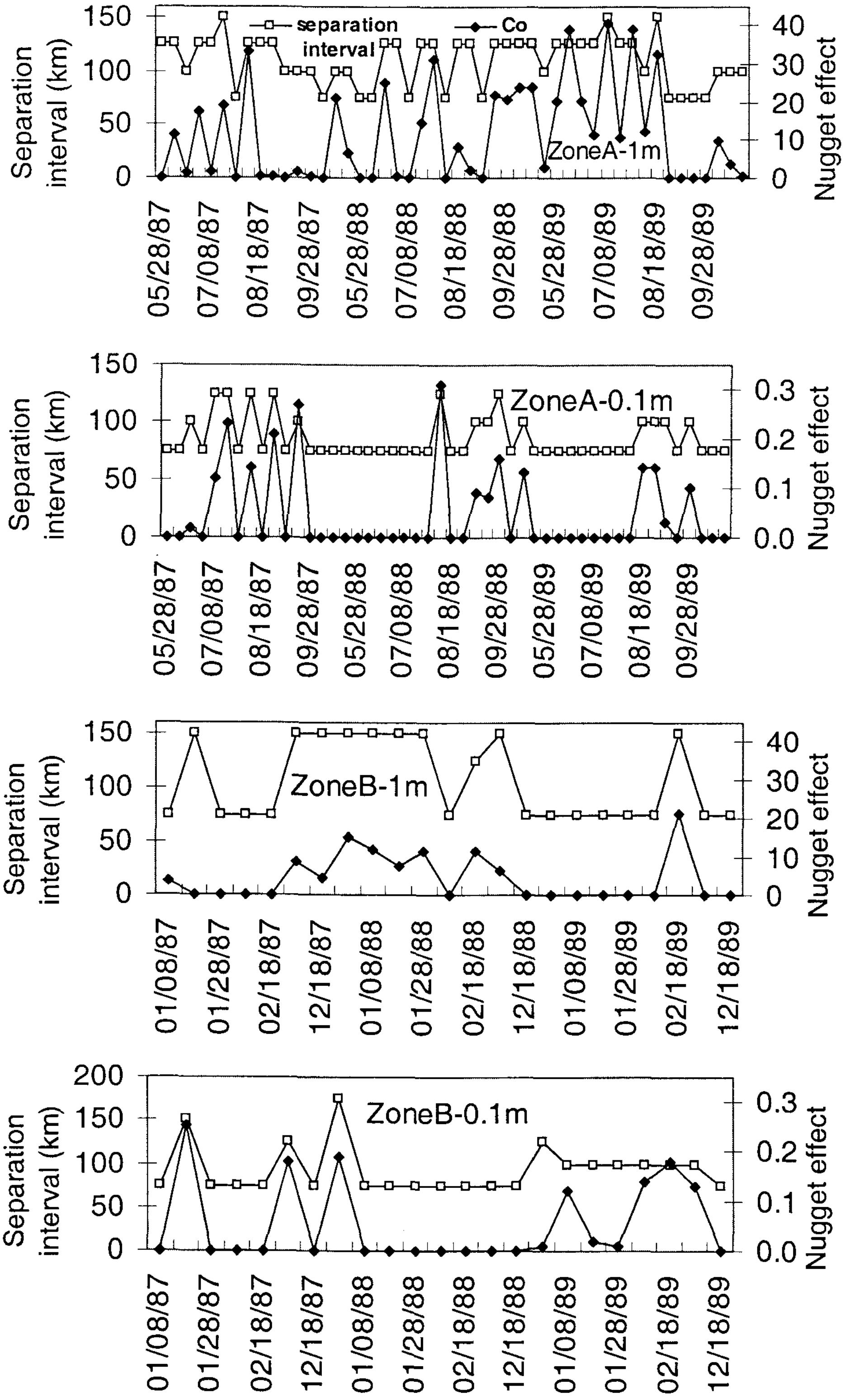

Fig. 7. Nugget values (C_o) and separation intervals for each date.

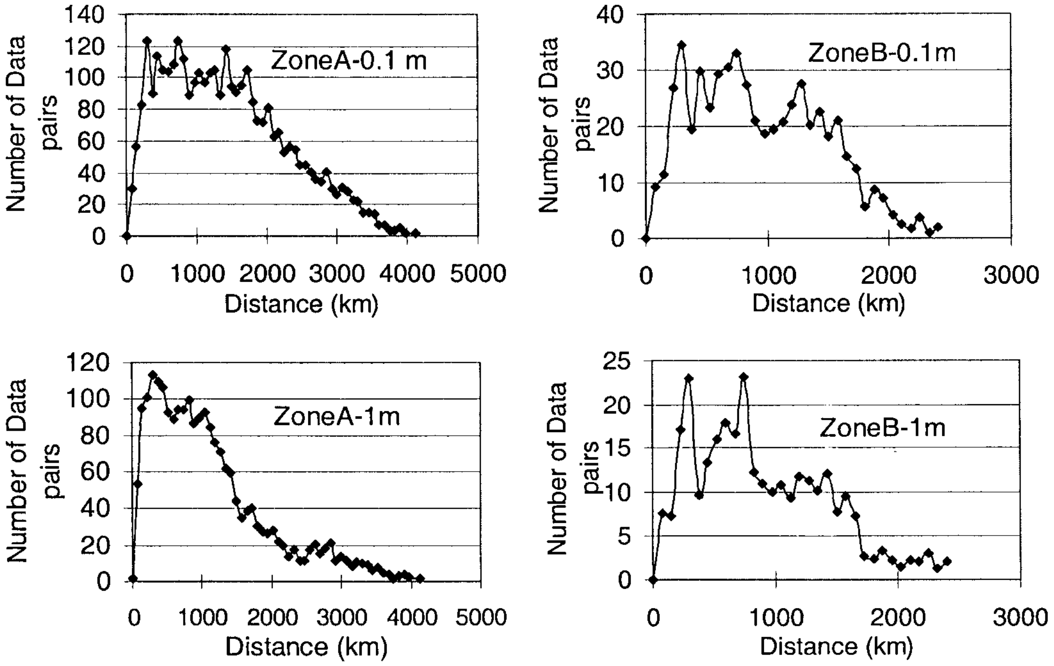

Fig. 8. The variation of number of data pairs used for estimating the values of sample variograms.

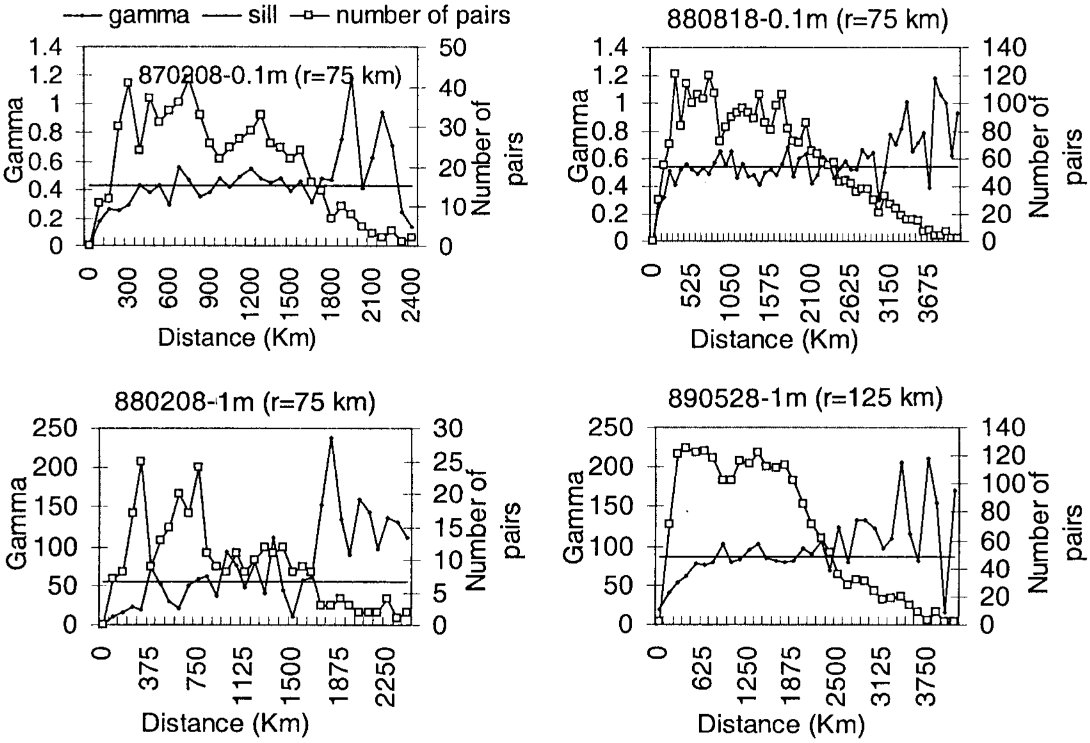

Fig. 9. Four typical sample variograms (gamma: sample variogram; r: separation interval).

Table 3. (a) Summary of Geostatistical Structure for Soil Moisture Data in the top 1 m and top 0.1 m Soil Layer for Zone A.

	SI(km)*		Co (%^{2})		Sill (%^{2})		Range (km)	
Date	Top1m	Top0.1m	Top1m	Top0.1m	Top1m	Top0.1m	Top1m	Top0.1m
05/28/87	125	75	0.06	0.00	69.77	0.65	450	300
06/08/87	125	75	11.18	0.00	71.64	0.72	600	360
06/18/87	100	100	1.02	0.02	62.10	0.60	450	375
06/28/87	125	75	17.35	0.00	62.94	0.56	630	150
07/08/87	125	125	1.74	0.12	87.94	0.90	675	630
07/18/87	150	125	19.05	0.23	71.70	0.68	870	675
07/28/87	75	75	0.00	0.00	70.56	0.75	435	300
08/08/87	125	125	33.17	0.14	76.39	0.77	2100	570
08/18/87	125	75	0.59	0.00	83.67	0.76	375	375
08/28/87	125	125	0.40	0.21	82.98	0.80	563	975
09/08/87	100	75	0.03	0.00	86.62	0.79	450	285
09/18/87	100	100	1.51	0.27	76.60	0.61	450	525
09/28/87	100	75	0.55	0.00	80.11	0.79	375	375
10/08/87	75	75	0.00	0.00	80.97	0.80	450	360
10/18/87	100	75	21.16	0.00	89.97	0.84	900	450
10/28/87	100	75	6.60	0.00	89.87	0.74	450	375
05/28/88	75	75	0.00	0.00	75.78	0.64	420	150
06/08/88	75	75	0.00	0.00	68.16	0.60	420	225
06/18/88	125	75	24.99	0.00	81.69	0.96	825	450
06/28/88	125	75	0.51	0.00	79.58	0.67	720	300
07/08/88	75	75	0.00	0.00	58.49	0.70	420	225
07/18/88	125	75	14.55	0.00	80.58	0.68	563	225
07/28/88	125	75	31.13	0.00	73.96	0.73	1125	300
08/08/88	75	125	0.00	0.31	76.08	0.76	563	525
08/18/88	125	75	8.07	0.00	64.76	0.54	600	225
08/28/88	125	75	1.89	0.00	66.18	0.72	375	300
09/08/88	75	100	0.00	0.09	56.29	0.60	338	600
09/18/88	125	100	21.76	0.08	59.15	0.76	900	300
09/28/88	125	125	20.63	0.16	54.46	0.59	975	563
10/08/88	125	75	23.85	0.00	50.29	0.66	375	225
10/18/88	125	100	23.85	0.13	50.29	0.66	1200	525
10/28/88	100	75	2.79	0.00	50.84	0.54	450	300
05/28/89	125	75	20.39	0.00	86.08	0.69	630	225
06/08/89	125	75	38.91	0.00	87.94	0.77	1200	375
06/18/89	125	75	20.42	0.00	79.68	0.76	900	150
06/28/89	125	75	11.50	0.00	82.33	0.82	585	300
07/08/89	150	75	40.42	0.00	88.00	0.78	1875	330
07/18/89	125	75	10.63	0.00	82.92	0.76	675	270
07/28/89	125	75	38.87	0.00	86.47	0.69	1050	330
08/08/89	100	75	12.32	0.00	81.62	0.75	480	300
08/18/89	150	100	32.57	0.14	81.84	0.69	1650	900
08/28/89	75	100	0.00	0.14	86.04	0.68	480	375
09/08/89	75	100	0.00	0.03	87.66	0.73	330	300
09/18/89	75	75	0.00	0.00	90.20	0.70	450	300
09/28/89	75	100	0.00	0.10	81.69	0.61	300	450
10/08/89	100	75	9.61	0.00	78.90	0.53	450	225
10/18/89	100	75	3.84	0.00	68.54	0.52	450	225
10/28/89	100	75	0.30	0.00	72.08	0.66	330	150

* SI: Separation Interval

Table 3. (b) Summary of Geostatistical Structure for Soil Moisture Data in the top 1 m and top 0.1 m Soil Layer for Zone B.

	SI(km)*		Co (%^{2})		Sill(%^{2})		Range (km)	
Date	Top1m	Top0.1m	Top1m	Top0.1m	Top1m	Top0.1m	Top1m	Top0.1m
01/08/87	75	75	3.42	0	34.44	0.69	300	240
01/18/87	150	150	0	0.25	38.53	0.70	630	1890
01/28/87	75	75	0	0	45.41	0.63	360	270
02/08/87	75	75	0	0	40.09	0.43	720	285
02/18/87	75	75	0	0	57.21	0.97	600	795
12/08/87	150	125	8.73	0.18	34.44	0.52	675	810
12/18/87	150	75	4.47	0	33.42	0.49	570	225
12/28/87	150	175	15.22	0.19	32.06	0.41	570	870
01/08/88	150	75	11.94	0	32.53	0.48	735	225
01/18/88	150	75	7.55	0	24.98	0.53	405	180
01/28/88	150	75	11.7	0	31.93	0.53	375	180
02/08/88	75	75	0	0	54.77	0.59	825	165
02/18/88	125	75	11.63	0	45.57	0.62	780	195
12/08/88	150	75	6.24	0	21.81	0.39	375	150
12/18/88	75	75	0	0	24.41	0.38	225	150
12/28/88	75	125	0	0.01	26.41	0.45	150	270
01/08/89	75	100	0	0.12	37.21	0.61	300	285
01/18/89	75	100	0	0.02	31.85	0.83	225	300
01/28/89	75	100	0	0.01	41.76	0.53	375	240
02/08/89	75	100	0	0.14	38.72	0.59	375	330
02/18/89	150	100	21.32	0.18	39.91	0.70	1260	285
12/08/89	75	100	0	0.13	33.6	0.48	675	300
12/18/89	75	75	0	0	34.19	0.76	525	150

* SI: Separation Interval

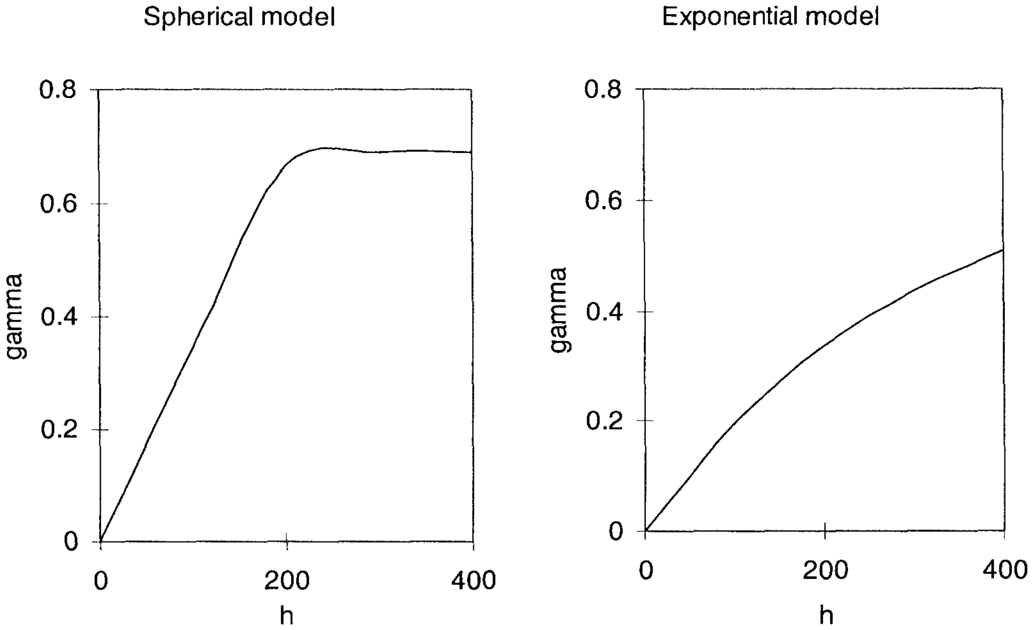

Fig. 10. The difference between a spherical model and exponential model.

1987 to 1989 the statistical relationship expressed by the variogram among nearby soil moisture sites is linear. However, as the separation distance between the data points got longer (up to one half the range), the linear relationship was replaced by a curved relationship. This is the most obvious difference between the spherical model and exponential model, as shown in Fig. 10. All of the variograms indicate that the soil moisture field is stationary since they exhibit clear sills.

4.2 Scale of structural spatial variation

The ranges for the variograms are well defined. The scale of the structured spatial variation varied from date to date, which can be seen by the variation of the range. The range represents the average maximum distance over which two samples are correlated. It is a direct measurement of the scale of spatially correlated variation. The larger the range, the larger the scale of the correlated spatial variation.

As shown in Table 3, for soil moisture data in the top 1 m soil layer, the range varied from 150 km to 825 km for Zone B and from 300 km to 1650 km for Zone A. For soil moisture data in the top 0.1 m soil layer, the range varied from 150 km to 870 km for Zone B and from 150 km to 900 km for Zone A. The arithmetic average ranges for all the measure dates in Zone A and Zone B are shown in the first line of Table 4. Considering the representativeness of the results, we discard the largest value when taking account of the maximum of the range. The averaged ranges calculated by discarding the largest value are shown in the second line of Table 4.

We compare the histogram of the range in Zones A and B for both 0.1 m and 1 m layers in Fig. 11. The probability density functions are not symmetric except for soil moisture data in the top 1 m for Zone B. Statistically, we calculate the average range by discarding those values of the range, which do not fall in the average region. The values of averaged range by this method are shown in the third line of Table 4.

It is interesting to see that the temporal change of range follows the change of nugget well, with the correlation coefficients between 0.55 and 0.70. The average range based on those data with zero nugget effect is in the fourth line in Table 4.

As shown in Table 4, for both Zones A and B the averaged ranges of soil moisture in the top 0.1 m, calculated by all four methods, are less than the ranges of soil moisture in the top 1 m. The conclusion agrees with both common sense and the relationship between the averaged C_v of soil moisture in the top 0.1 m and the averaged C_v of soil moisture in the top 1 m shown in Table 1. These results agree with those of Entin et al. (2000) who found a spatial scale of about 525 km for the top 1 m, but they also found the same scale for the top 0.1 m. Entin et al., however, used fewer stations but with longer records, so we would not expect the results to agree exactly.

The relationship between the averaged ranges of soil moisture in Zones A and B is similar no matter what kind of method used to get the average. This gives strong support to the average scale of about 500 km for soil moisture variation, which is in agreement with previous results by Vinnikov et al. (1996) and Entin et al. (2000).

4.3 Degree of the structured spatial variation

The difference in the degree of structured spatial variation is indicated by the variations of the nugget effect, and of the sill. A large sill accompanied by a small nugget effect represents a high degree of spatial correlation. The models fitted to the data had zero nugget effect on some dates. This does not mean that there were no measurement errors in the data, nor was there no random variation in the variable. The zero nugget effects probably occurred due to small random variations and measurement errors in those dates, and because the space interval for data collection was too

Table 4. The average range (km) using different methods.

Averaging Method	0.1 m Zone A	0.1 m Zone B	1 m Zone A	1 m Zone B
(1)	369	382	673	523
(2)	356	314	643	490
(3)	345	233	593	490
(4)	286	247	419	460

Notes: (1) Using all data; (2) Discarding the largest; (3) Statistical approximation;(4) Discarding values with non-zero nugget.

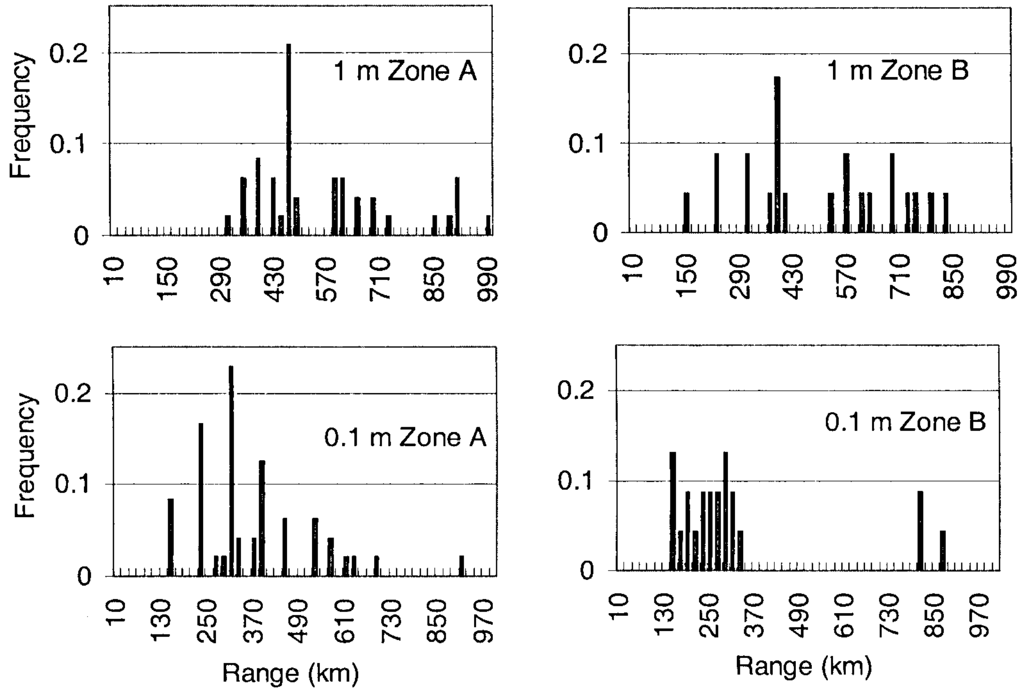

Fig. 11. Histogram in Zone A and Zone B for soil moisture data in the top 0.1 m and top 1 m soil layer.

large to trap these random variations and measurement errors. In this sense it is better to interpret the zero nugget effect as a small nugget effect when discussing the degree of spatial correlation. The largest nugget effect occurred in Zone A on August 8, 1988 for the top 0.1 m data and on July 8, 1987 for the top 1 m data.

Because the difference between the sill and the nugget effects represents the variance caused by structured spatial variability, this difference can be used to represent the degree of spatial correlation. Assuming that a spatial variation surface can be decomposed into spatially correlated variation and random variation, the ratio of the nugget effect to the spatial variance was computed and plotted to indicate the degree of the spatial variance. Since the nugget effect represents the variance due to pure random variations, the larger the ratio, the lower the degree of the spatial correlation. The differences of the ratio between the top 1 m data and top 0.1 m data are shown in Fig. 12.

The most striking feature in Fig. 12 is the magnitude of the ratio of difference between the top 1 m data and the top 0.1 m data. The ratios of the top 0.1 m data were smaller than those of the top 1 m data. Generally speaking, this feature suggests that spatial variation of the top 0.1 m was more strongly correlated than in the top 1 m and that the measurement errors were much larger in the top 1 m data than the top 0.1 m data. However, this does not apply on every day.

The degree of spatial correlation varied greatly among dates. Random variation was as high as 53.4% (maximum) and as low as almost zero on the next date, as shown in 1 m Zone B data. For most of the dates, the random variations accounted for less than 50% of total variance. For more than half of the measured dates, the ratio of the nugget effect to the variance was less than 20%, indicating that the spatially correlated variation in those dates can explain more than 80% of the total variance.

5. Conclusions

Two groups of data sets have been chosen from our 11-year Chinese soil moisture data set of 102 stations for geostatistical analysis in this paper. The selection principle, to keep the sample size for each measurement date as stable as possible, produces data set A representing the summer season and data set B representing the winter season. We found that soil moisture in the top 0.1 m is more variable than that in the top 1 m. Soil moisture in Zone A is more variable than in Zone B. As the coverage of Zone A and Zone B is different, it is better

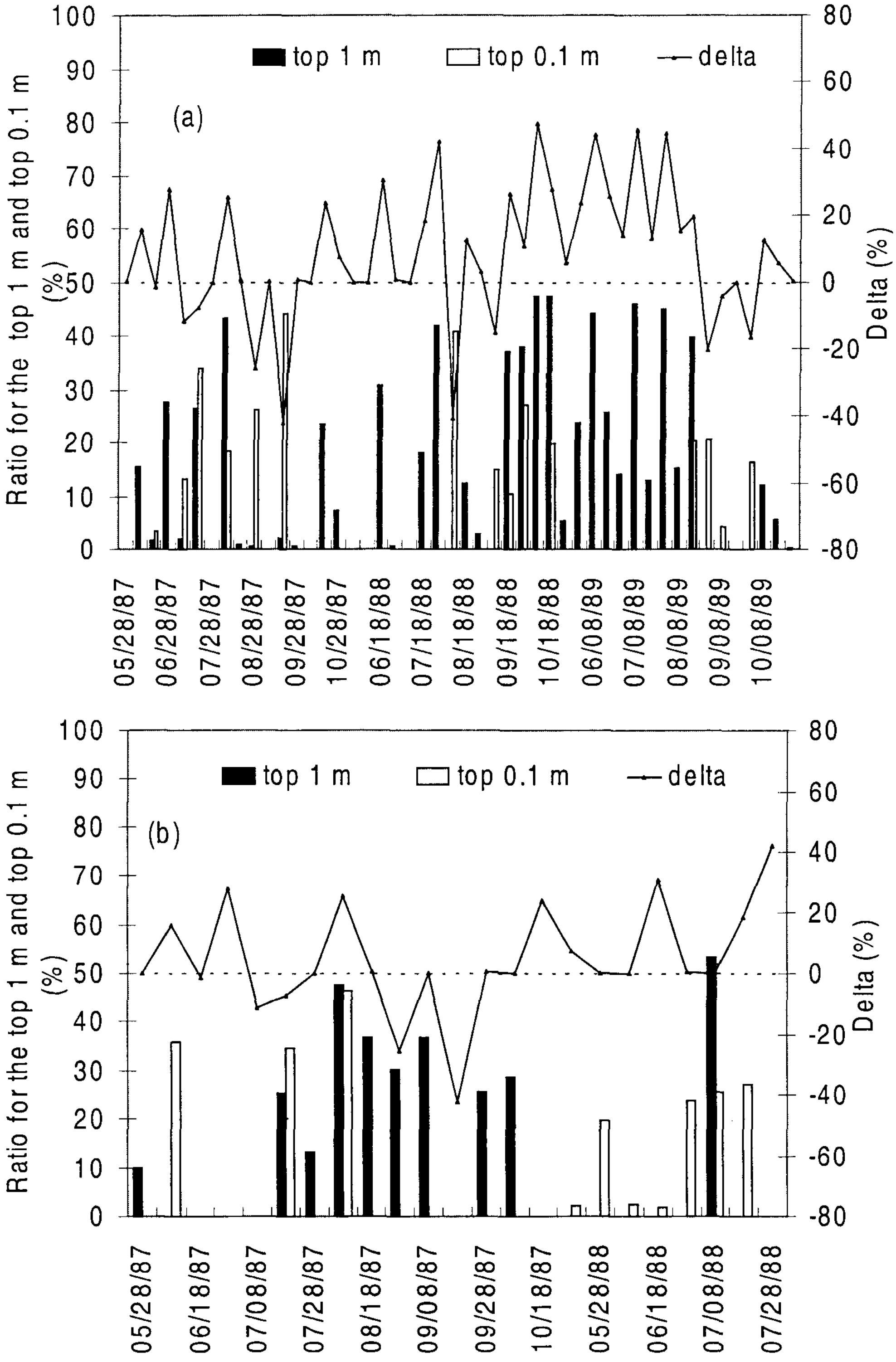

Fig. 12. The ratio of the nugget effect to the spatial variance between the top 1 m data and top 0.1 m data in Zone A (a) and Zone B (b). Delta is the difference between the ratio in the top 1 m and the ratio in the top 0.1 m. The dash line shows that the difference is zero, i.e., delta = 0.

to make comparisons from Table 1 and Fig. 3, than from the distribution maps in Fig. 4.

The variograms are found to have a clear sill and a nugget in many cases. Spherical models including a nugget fit the model closely. The spatial scale in the top 1 m in both Zones A and B is about 500 km, agreeing with the spatial scale reported by Entin et al. (2000) using the autocorrelation method.

Acknowledgments

The first author thanks Dr. Xuemei Shao for her great help in this work. This work is supported by Knowledge Innovation Project of the Chinese Academy of Sciences KZCX2-310, Chinese National Natural Science Foundation projects 49771019 and 49890330, Ministry of Science and Technology of the People's Republic of China project G1999043601, Knowledge Innovation Project of Institute of Geographic Sciences and Natural Resources Research CX10G-C00-05-01, NOAA grants NA56GP0212 and GC99-443b, and the New Jersey Agricultural Experiment Station. Two anonymous reviewers are thanked for their pertinent comments.

References

Abbott, M.B. and J.C. Refsgaard (eds.), 1996: *Distributed Hydrological Modelling*, Kluwer Academic Publishers, Netherlands.

Armstrong, M., 1984: Common problems seen in variograms, *Math. Geol.*, **16**, 305–313.

Ashraf, M., J.C. Loftis, and K.G. Hubbard, 1997: Application of geostatistics to evaluate partial weather station networks, *Agr. For. Meteor.*, **84**, 255–271.

Clark, I., 1980: *The Semivariogram in Geostatistics*, McGraw-Hill, New York, 17–28.

Deutsch, C.V., and A.G. Journel, 1992: *GSLIB Geostatistical Software Library and User's guide*. Oxford Univ. Press, New York, 340pp.

Entekhabi, D., and I. Rodriguez-Iturbe, 1994: Analytical framework for the characterization of the space-time variability of soil moisture. *Adv. Water Resour.*, **17**, 35–45.

Entin, J.K., A. Robock, K.Y. Vinnikov, V. Zabelin, S. Liu, A. Namkhai, and T. Adyasuren, 1999: Evaluation of Global Soil Wetness Project soil moisture simulations. *J. Meteor. Soc. Japan*, **77**, 183–198.

———, ———, ———, S.E. Hollinger, S. Liu, and A. Namkhai, 2000: Temporal and spatial scale of observed soil moisture variations in the extratropics. *J. Geophys. Res.*, **105**, 11865–11877.

Hou, J., and J. Huang, 1981: *Geostatistics and its Application in Ore Reserve Calculation*. Geological Publishing House, Beijing, 42–43. (in Chinese)

Isaaks, E.H and R. M. Srivastava, 1989: *An Introduction to applied geostatistics*. Oxford University press, New York, pp.561.

Journel, A.G. and C.J. Huijbregts, 1978: *Mining Geostatistics*, Academic Press, London.

Kim, Y.C and H.P. Knudsen, 1977: Geostatistical Ore Reserve estimation for a roll front type Uranium deposit (Practitioner's guide), U.S. Energy Research and Development Administration, open file No. GJBX-3 (77).

Liu, S. and C. Liu, 1994: The concept of representative elementary length and its application to flooding forecast of Yanghe reservoir basin. *Yangtze River*, **85(8)**, 19–22. (in Chinese)

Metheron, G. 1971: Theory of regionalized variables and its applications. *Cah. Centre Morrphol. Math.*, **5**, 211pp.

Robock, A., K.Y. Vinnikov, G. Srinivasan, J.K. Entin, S.E. Hollinger, N.A. Speranskaya, S. Liu, and A. Namkhai, 2000: The Global Soil Moisture Data Bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281–1299.

Shao, X., 1992: *Statistical relationships between tree growth and climate in western north America.* Ph. D. dissertation, Univ. Arizona, 86–87.

Shi, H., Y. Chen, K. Cai and X. Zhang, 1997: Study of the spatial variability of soil moisture and its large area estimation and prediction for the West Liao River Plains. *J. Arid Land Resources and Environment*, **11(4)**, 36–43. (in Chinese)

Srinivasan, G., A. Robock, J.K. Entin, L. Luo, K.Y. Vinnikov, P. Viterbo, and Participating AMIP Modeling Groups, 2000: Soil moisture simulations in revised AMIP models. *J. Geophys. Res.*, in press.

Villeneuve, J.P., G. Morin, B. Bobee, and D. Leblanc, 1979: Kriging in the design of streamflow sampling networks, *Water Resour. Res.*, **15(6)**, 1833–1840.

Vinnikov K.Y., A. Robock, N.A. Speranskaya, and C.A. Schlosser, 1996: Scales of temporal and spatial variability of midlatitude soil moisture. *J. Geophys. Res.*, **101**, 7163–7174.

Wang, G., and C. Guo, 1987: Application of kriging interpolation method in geography. *Acta Geographica Sinica*, **42(4)**, 366–374. (in Chinese)

Western, A.W., G. Bloschl, and R.B. Grayson, 1998a: How well do indicator variograms capture the spatial connectivity of soil moisture? *Hydrological Processes*, **12**, 1851–1868.

———, ——— and ———, 1998b: Geostatistical characterization of soil moisture patterns in the Tarrawarra catchment, *J. Hydrology*, **205**, 20–37.

Yang, S., and Z. Lei, 1993: The spatial structure of soil water content and sampling number determination in the field. *Acta Geographica Sinica*, **48(5)**, 447–455. (in Chinese)