4.1.1 The Empirical Variogram

The empirical variogram provides a description of how the data are related (correlated) with distance. The semivariogram function, $\gamma(h)$, was originally defined by Matheron (1963) as half the average squared difference between points separated by a distance h. The semivariogram is calculated as

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z_i - z_j)^2$$

where $N(h)$ is the set of all pairwise Euclidean distances $i - j = h$, $|N(h)|$ is the number of distinct pairs in $N(h)$, and z_i and z_j are data values at spatial locations i and j, respectively. In this formulation, h represents a distance measure with magnitude only. Sometimes, it might be desirable to consider direction in addition to distance. In such cases, h will be represented as the vector $\mathbf{h}$, having both magnitude and direction.

Note: The terms semivariogram and variogram are often used interchangeably. By definition, $\gamma(h)$ is the semivariogram and the variogram is $2\gamma(h)$. For conciseness, however, this manual will refer to $\gamma(h)$ as the variogram.

The main goal of a variogram analysis is to construct a variogram that best estimates the autocorrelation structure of the underlying stochastic process. Most variograms are defined through several parameters; namely, the *nugget effect*, *sill*, and *range*. These parameters are depicted on the generic variogram shown in figure 4.1 and are defined as follows:

- *nugget effect*—represents micro-scale variation or measurement error. It is estimated from the empirical variogram as the value of $\gamma(h)$ for $h = 0$.
- *sill*—the $\lim_{h\to\infty} \gamma(h)$ representing the variance of the random field.
- *range*—the distance (if any) at which data are no longer autocorrelated.

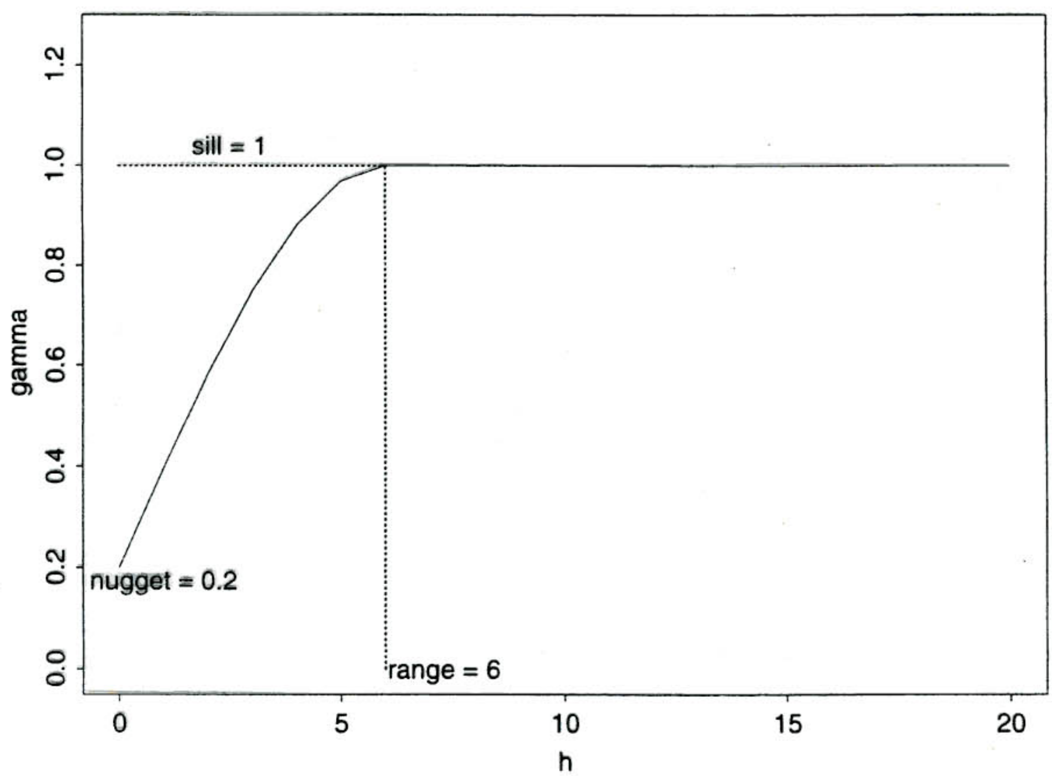

FIGURE 4.1. A generic variogram showing the *sill*, and *range* parameters along with a *nugget effect*.

Construction of a variogram requires consideration of the following:

- an appropriate *lag increment* for h;
- a *tolerance* for the lag increment; and
- the *number of lags* over which the variogram will be calculated.

The *lag increment* defines the distances at which the variogram is calculated. The *tolerance* establishes distance bins for the lag increments, to accommodate unevenly spaced observations. The *number of lags* in conjunction with the size of the lag increment will define the total distance over which a variogram is calculated.

There are two practical rules (Journel and Huijbregts, 1978) that should be considered when making your choices for lag increment and number of lags:

1. The experimental variogram should only be considered for distances h for which the number of pairs is greater than 30.
2. The *distance of reliability* for an experimental variogram is $h < D/2$, where D is the maximum distance over the field of data.

Uses of the Variogram.

Western, A. W., Blöschl, G., Grayson, R. B. (1998). Geostatistical characterization of soil moisture patterns in the Tarrawarra catchment. Journal of Hydrology, 205, 20-37.

Widespread use of the variogram in soil science: *interpolation* of spatial patterns, *estimation of the average* catchement soil moisture, *hydrological modeling* which sometimes accounts for spatial variability.

Analysis of the development of soil moisture in a mini-catchement.
Insertion of time domain reflectometry probes, 500 point measures on a 10 x 20 m grid.
(Plus two more detail grids of 1000 and 2000 points.)

In summer, when the soil is dry the moisture is relatively uniform, low sill.
In winter the sill is high due to topographic control on soil moisture.
Correlation lengths are lowest during the wet period (35-50m), lateral distribution leads to narrow wet strips
and highest during the dry periods (50-60m), moisture is uniform.

Suggested causes of the nugget effect:

random measurement errors,

data not collected at small enough spacing to reveal continuous behaviour (the apparent nugget effect).

Effects of sample size:

Use of offset replicated sampling of grids of 514 points gave almost identical variograms.

Use of different numbers of points. Suggests 300 points are needed -> but this depends upon the scale of the pattern relative to sampling frequency.

Note: conclusions based on fitted exponential variogram. Correlation lengths are ~ 1/3 of range.

In summer isotropic,
winter anisotropic with shorter range in n-s direction, short hillslope influence.

All variograms indicate soil moisture stationary. Sills close to the variance of the sample.

In summer when soil is dry moisture is uniform, sill is low: This is because soil moisture is being limited by water retention properties of the soil. Strong soil moisture control on evapotranspiration – lateral re-distribution unimportant.

In winter sill is high due to the topographic control on soil moisture – lateral redistribution of water under conditions of high soil moisture.

Variations in nugget do not follow seasonal patterns.

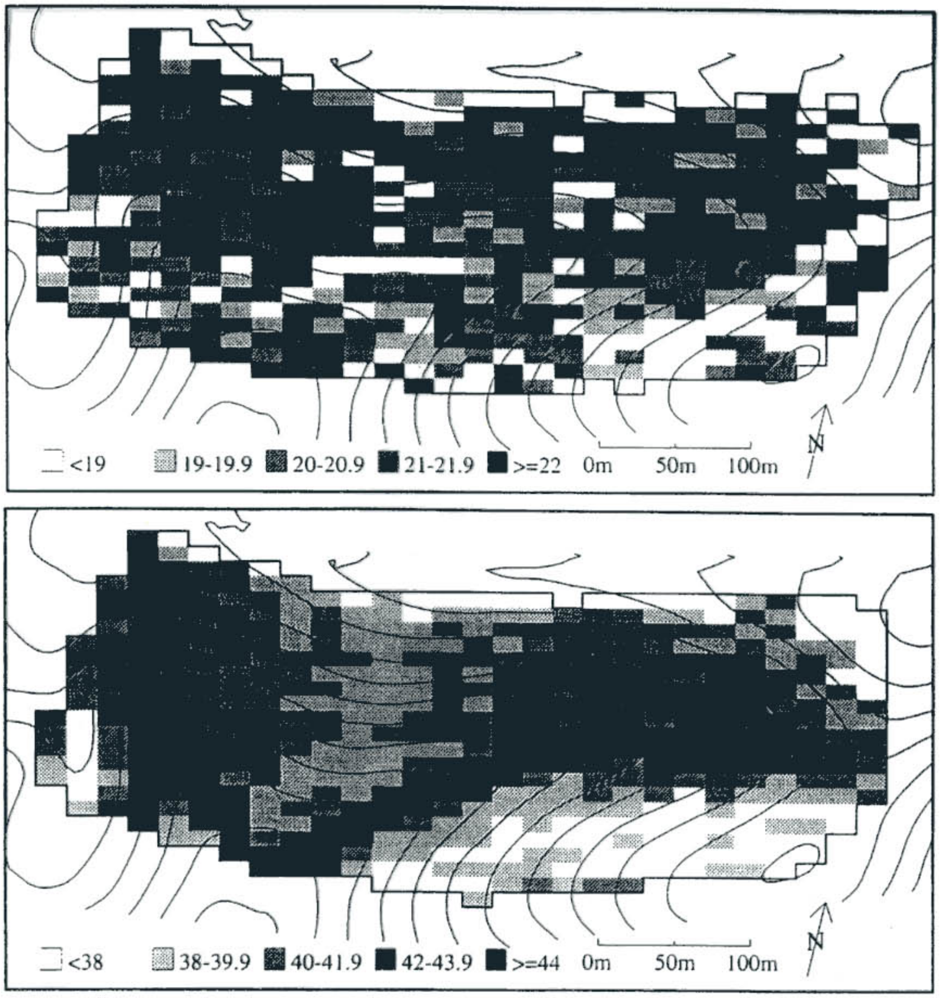

Fig. 2. Soil moisture patterns collected at Tarrawarra on (top) 23 February 1996 and (bottom) 2 May 1996. Note that each rectangle represents one point measurement of the average volumetric soild moisture in the top 30 cm of the soil profile.

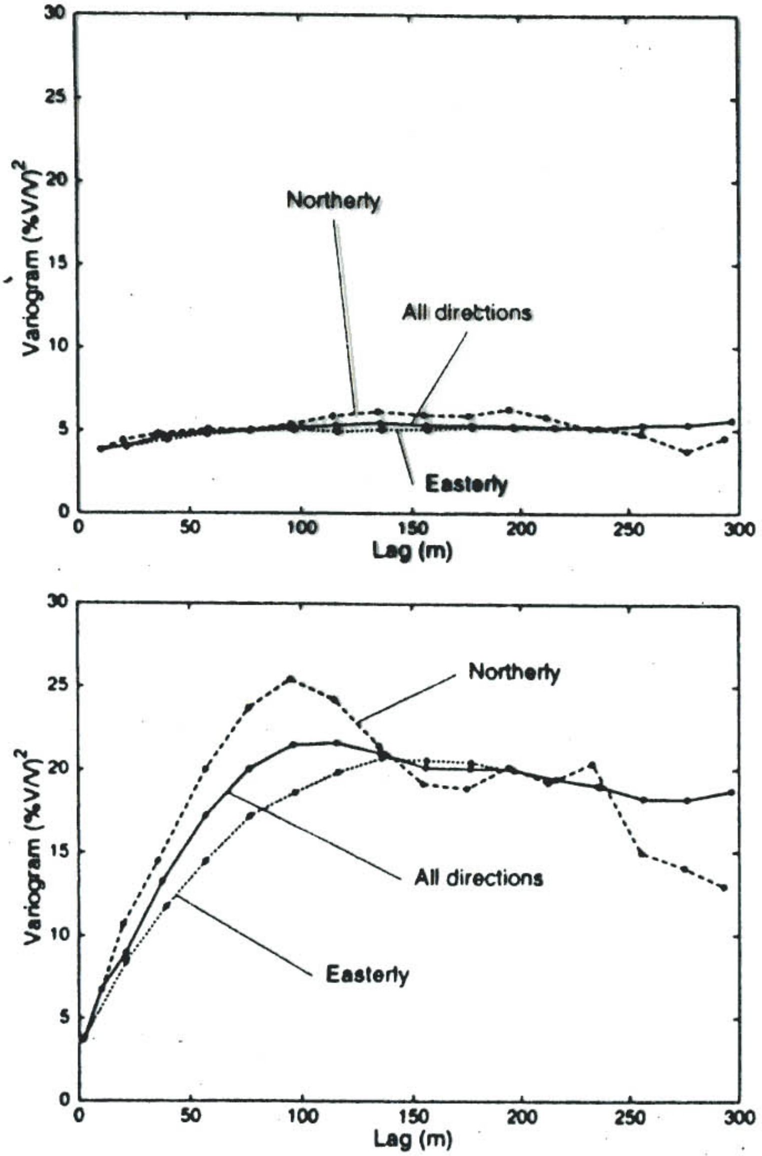

Fig. 4. North-south and east-west directional variograms and the omnidirectional variograms for Tarrawarra on (top) 23 February 1996 and (bottom) 2 May 1996.

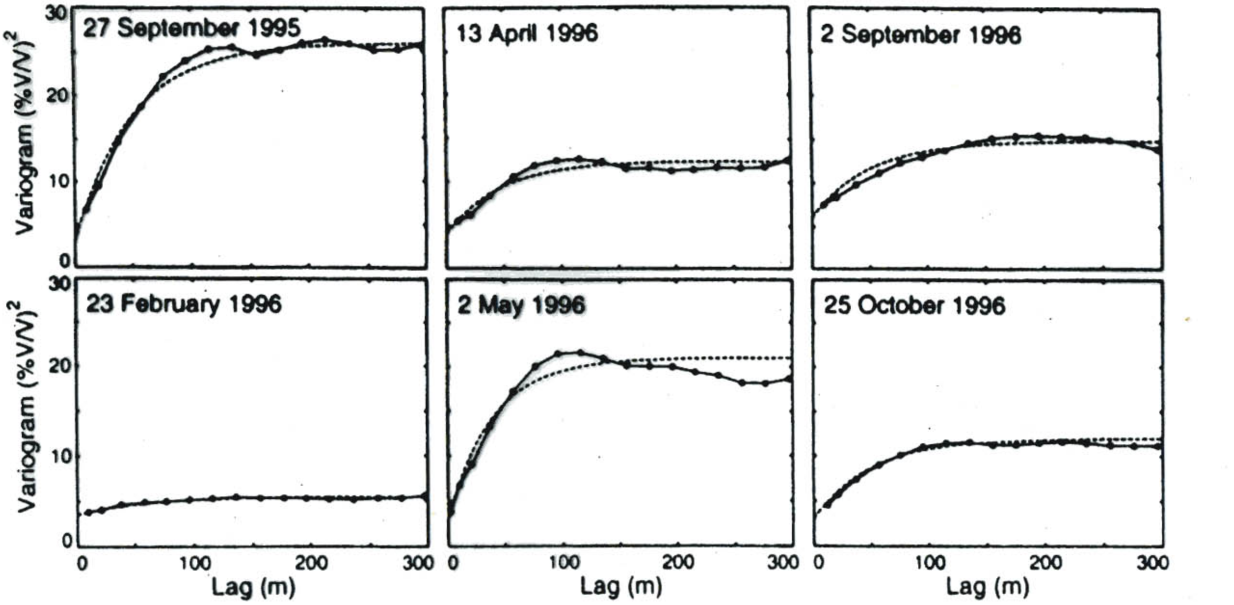

Fig. 5. Sample omnidirectional variograms (solid lines) for six representative soil moisture patterns from Tarrawarra. Exponential variograms (dashed lines) with a nugget have been fitted to the sample variograms.

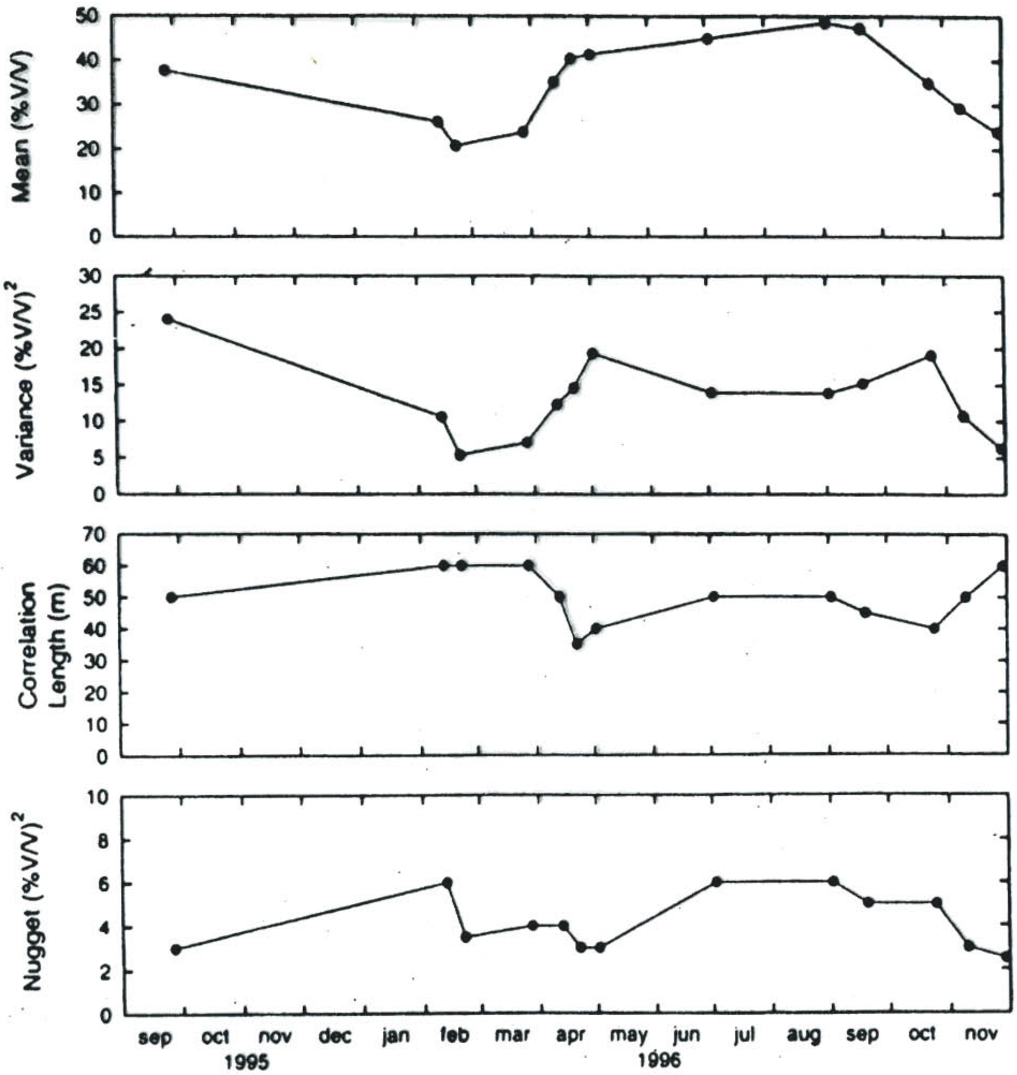

Fig. 6. The seasonal evolution of the geostatistical structure of the soil moisture pattern at Tarrawarra. The top graph shows the mean soil moisture, the second graph shows the spatial variance, the third graph shows the correlation length and the bottom graph shows the nugget.

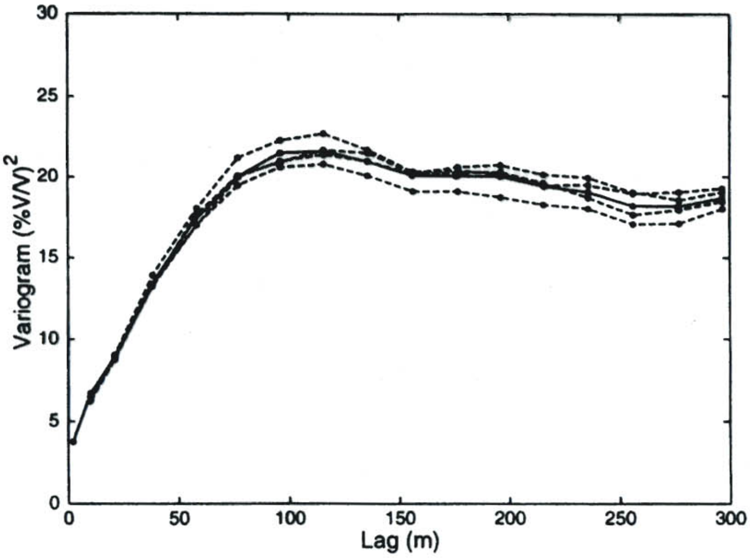

Fig. 8. Sample omnidirectional variograms for the four patterns of the 10 × 20 m grid from Run 7 (514 samples each) plotted with the sample variogram based on all 2056 moisture measurements.

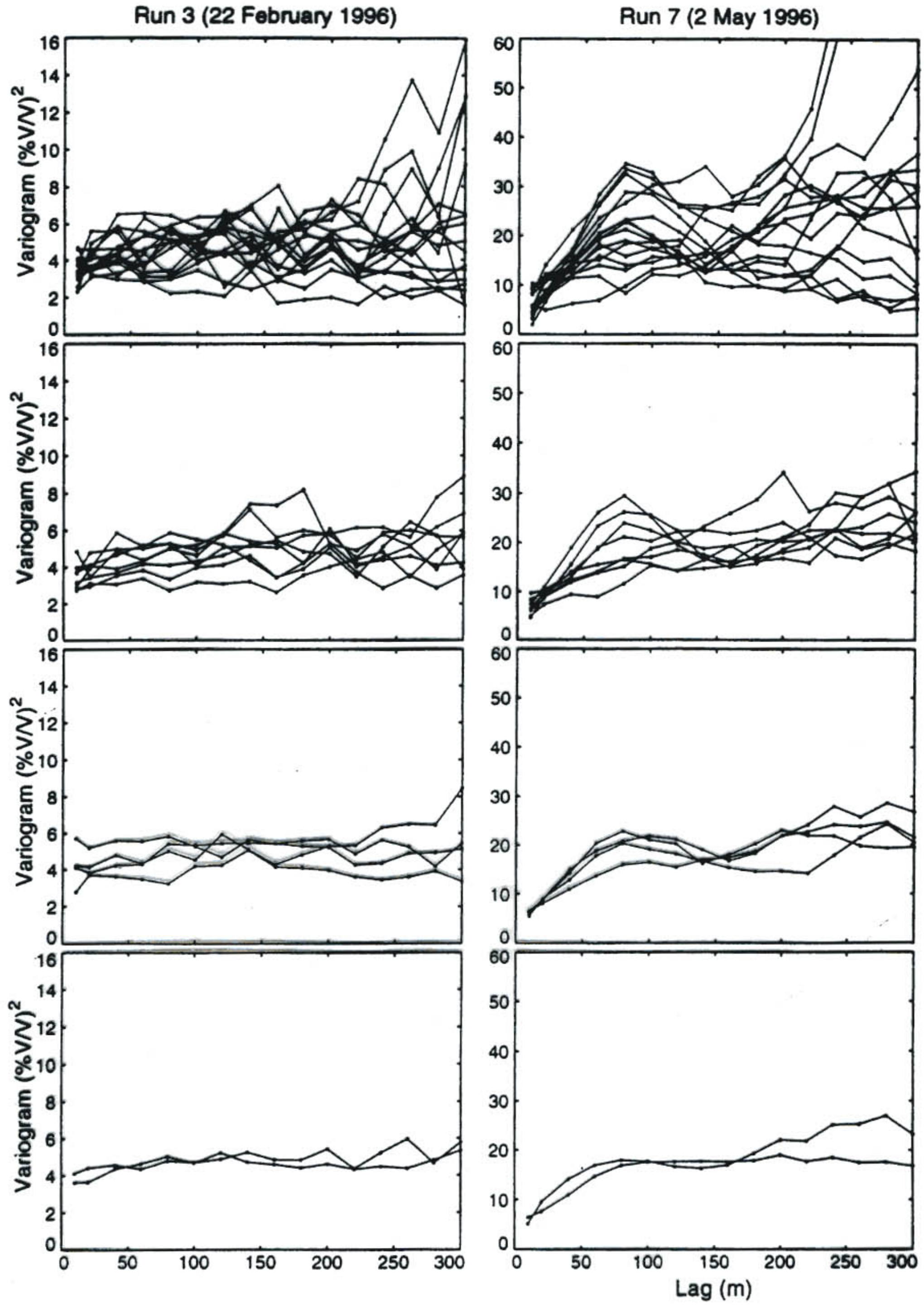

Fig. 9. The effect of sample size on the reliability of the variograms. Data from Tarrawarra on the 23 February 1996 (Run 3) and 2 May 1996 (Run 7) are used as an illustration. From the top to the bottom row, variograms are based on 44, 86, 164 and 296 points respectively.