Title

Spatial analysis based on variance of moving window averages

Journal

Journal of Phytopathology, 154(6)

ISSN

0931-1785

Authors

Wu, B M

Subbarao, K V

Ferrandino, F J

et al.

Publication Date

2006-06-01

Peer reviewed

Spatial Analysis Based on Variance of Moving Window Averages

B. M. Wu, K. V. Subbarao, F. J. Ferrandino, and J. J. Hao

First and second authors: Department of Plant Pathology, University of California, Davis, c/o U. S. Agricultural Research Station, 1636 East Alisal Street, Salinas 93905; third author: Plant Pathology and Ecology Dept., The Connecticut Agricultural Experiment Station, New Haven, CT 06504; and fourth author: Department of Plant Pathology, University of California, Davis, CA 95616.

Corresponding author: K. V. Subbarao, fax: (831)755-2814, email: kvsubbarao@ucdavis.edu

ABSTRACT

A new method for analyzing spatial patterns was designed based on the variance of moving window averages (VMWA), which can be directly calculated in geographic information systems (GIS) or a spreadsheet program (e.g. MS Excel). Different types of artificial data were generated to test the method. Regardless of data types, the VMWA method correctly determined the mean cluster sizes. This method was also employed to assess spatial patterns in historical plant disease survey data encompassing both airborne and soilborne diseases. The results obtained using the VMWA method were generally different from those obtained with Lloyd's index of patchiness and beta-binomial distribution methods, were in partial agreement with the results from spatial analysis by distance indices, and were highly consistent with the results from semivariogram and spatial autocorrelation analysis methods. Results demonstrated that the VMWA method can be applied to many types of data, including binomial diseased or healthy plant counts, incidence, severity, and number of diseased plants or pathogen propagules although directional and edge effects may limit its application.

Additional keywords: aggregation index, spatial dependency.

Spatial patterns are considered as a manifestation of biological processes, reflecting the interactions among different determining factors. Spatial patterns provide information regarding the mechanisms that bring about these patterns, which can then be applied to improve sampling techniques, and to develop strategies for the management of our environments. Methods to detect and assess spatial pattern and to quantify spatial distribution and association among different elements in an ecological system, have received increasing attention in ecological research since the beginning of modern ecology. This in turn has laid a strong foundation for a

better understanding of the biological processes that generate the patterns. Methods for analyzing spatial patterns have been developed in a wide variety of disciplines, such as plant communities, statistics, forestry and geostatistics, over the past decades (Clark, 1979, Campbell and Noe 1985, Madden and Hughes 1995, Real and McElhany 1996, Liu 2001).

A commonly used approach to determine spatial patterns is to compare the observed frequency distribution with theoretical frequency distributions such as Poisson, binomial, negative binomial, Neyman type A, and beta-binomial distributions (BBD) (Hughes and Madden 1993, Madden and Hughes 1994). Based on the best fit of observed frequencies to these distributions, a spatial pattern is considered aggregated, random or uniform (Campbell and Noe 1985, Campbell and Madden 1990, and Madden and Hughes 1995). It is generally accepted that for count data, a good fit to the Poisson distribution suggests random distribution (Fisher 1925), while a good fit to the negative binomial distribution implies heterogeneity (Madden and Hughes 1995). Similarly, for a binary variable a good fit to the binomial distribution indicates homogeneity while a good fit to the beta-binomial distribution suggests heterogeneity (Madden and Hughes 1995). The frequency distribution approach has also been extended to calculate indices of aggregation or dispersion. The indices that have been commonly used include, Fisher's variance-to-mean ratio (Fisher, 1925), David and Moore's index of clumping (David and Moore, 1954), the slope of the log (variance) to log (mean) line in the Taylor's empirical power law (Taylor 1961), Morisita's index, I_δ (Morisita 1962), Lloyd's mean crowding and indices of patchiness (LIP) (Lloyd, 1967), the parameter k of the negative binomial distribution (Elliott 1977), and the parameter θ of beta- binomial distribution (Irwin, J. O. 1954, Griffiths, D. A. 1973, Hughes and Madden 1993, Madden and Hughes, 1994 & 1995). Because the methods based on frequency distributions ignore the information about the locations of sample sites and

their spatial relationship, they reflect only the relationship among individuals within sample units. If the presence of one individual enhances the presence of other individuals within the same sample unit, then the frequency distribution would be determined as aggregated; otherwise it follows a uniform or random distribution. The aggregation determined using frequency distribution based methods has often been referred to as heterogeneity (Madden and Hughes 1995).

Other methods that take into consideration the location of sample sites have also been developed for characterization of spatial patterns. These methods include: spatial analysis by distance indices (SADIE), methods based on quadrat variance, spatial autocorrelation, distance based joint-counts/network method and geostatistics.

SADIE (Perry and Hewitt 1991, Alston 1994, Perry 1995) determines an index of aggregation by comparing the observed spatial arrangement with other arrangements derived from it, such as those where individuals are arranged as crowded as possible, those as random as possible, and those as regular as possible. Aggregation detected by SADIE method may be nonrandom in two different ways, nonrandom frequency distribution of counts regardless of the locations of sample units, and the nonrandom spatial arrangement of the sample units. Perry and Hewitt (1991) defined ‘move to crowding’ as the minimal total number of moves required to move all individuals, one by one and step by step, so that all the individuals finish in the same sample unit. ‘Move to randomness’ was also defined similarly, and a test based on the ratio of ‘move to crowding’ and ‘move to randomness’ was proposed as an index for detecting aggregation in spatial patterns. Lately, ‘distance to regularity’ (D) was suggested as a replacement to ‘move to randomness’ (Alston 1994). The approach was further extended to map

two-dimensional data, and a new method was developed to estimate the initial focus of a cluster based on the movements (Perry 1995).

Blocked quadrat variances (BQV) (Greig-Smith 1952) method determines cluster size by identifying the peak of mean “local variance” (mean squared difference between adjacent blocks) as quadrat sizes increase by combining two contiguous quadrats to form a new quadrat. The two major drawbacks with the BQV method are that the block sizes must be some power of 2, and that the starting position of blocking can affect the results. The two-term local quadrat variance (TTLQV) method (Hill 1973) offers more flexibility to the block size, and reduces the effects of the starting position. However, quadrats are still blocked along the belt transect and edge effects are important. Therefore, the variance is affected not only by distance (spacing) but also by the block size. This disadvantage was overcome in the paired-quadrat variance (PQV) and random paired quadrat variance (RPQV) methods, in which the variance was calculated for all possible paired quadrats along the belt at a given spacing for the former or for randomly selected paired quadrats for the latter (Goodall 1974). The quadrat variance methods provide informative results of aggregation and the size of clumps and are independent of the assumption of distributions. However, the calculations become complicated for two-dimensional data sets.

Spatial autocorrelation, distance based joint-counts methods and geostatistics are all similar. Spatial autocorrelation can be defined as a spatial property that the presence of some quality at one sample site (quadrat, or location) makes its presence at proximal sample sites more or less likely (Cliff and Ord 1973). Proximity can be determined by the connection between the two samples sites, such as distance, lag distance, vector or other criterion. Unlike methods based on negative binomial and Poisson distributions, which can only be used for discrete count data, the spatial autocorrelation analysis can be conducted on any type of data, continuous or discrete.

Data used in spatial autocorrelation can be point based, area based, or even without information on the exact shape of areas. The two most common coefficients in spatial autocorrelation analysis are Moran's *I*, which is based on sum over the cross product of deviations from mean, and Geary's *c*, which is based on sum of the squared difference (Moran 1950, Geary 1954). Joint-counts method, which has also been referred as distance-based methods and neighbor networks, is an analysis for discrete data based on point-pairs (Campbell and Madden 1990, Real and McElhany 1996, Liu 2001). There are a number of ways for point-pairing, such as Gabriel connectedness (Gabriel and Sokal 1969), the nearest neighbor (Clark and Evans 1954, Pielou 1959), and one- or two-dimensional distance class (Gray et al., 1986, Nelson et al., 1992, Ferrandino 1996). After establishing connectedness, the number (or average distance) of each type of joint is counted and this observed value is then compared with the expected one under the assumption of randomness. Geostatistical techniques, originally developed for use in mining (Clark 1979), can be used for both continuous and discrete variables, and require less stringent assumptions of stationarity compared with spatial autocorrelation techniques. In geostatistics, spatial dependence can be analyzed directionally or omnidirectionally, and represented as semivariogram or correlogram. Semivariogram calculates semivariance against each lag distance (a vector h), in which semivariance is defined as half of mean squared difference between all point pairs (p, $p+h$) separated by h. In a correlogram, the linear (Pearson's) correlation coefficients between point pairs separated by a vector h are plotted against vector h, or lag distance at different directions.

Recently, demand for spatial analysis has escalated because of advances in techniques for acquiring, transporting and processing data, which resulted in explosive growth of data available for spatial analysis. Spatial analysis tools in geographic information systems (GIS) have also

made traditional spatial analysis techniques requiring extensive computing easier than ever to implement. However, few spatial techniques have been developed based on GIS data. An instinctive method of determining whether a spatial data set is aggregated by human beings is to visually scan through the spatial data and determine whether the data is aggregated based on whether the densities of individuals or mean value varies "significantly" among different sub-areas. In GIS software, such as GRID module of Arc/Info (Arc/Info version 7, Environmental Systems Research Institute (ESRI), Redlands, CA) and ArcGIS (ArcGIS version 9, ESRI), a common function called smoothing filter or focal mean, in which the value at the center of the moving window is computed as a simple arithmetic average of the other values in the window, is very similar to the scanning movement of human eyes and with a mental calculation of densities or mean values. A spatial analysis method based on moving window averaging will eliminate the importance of the arbitrary and happenstantial positioning of contiguous quadrats which inflate the error for all fixed quadrat methods. Such a method will be easy to interface with GIS programs, and establish a foundation for developing other spatial analysis techniques in the future.

Our objective, therefore, was to develop a method that is based on moving window averaging. Variance of the resulted averages is used to determine if the average varies significantly at different locations. The relationship between the change of variance as the moving window changes size and the average cluster size of spatial data is analyzed. The method, though not described in detail, has been used previously (Bhat et al. 2003) in analysis of spatial patterns, and briefly discussed in a review (Wu and Subbarao 2005).

MATERIALS AND METHODS

Spatial dependence and variance of moving window averages (VMWA). Generally, for *n* variables (X_i, i=1,…,n) each with a variance *v*, the variance of their mean $Y = \sum_{i=1}^{n} X_i/n$ can be calculated as:

$$var(Y) = [\sum_{i=1}^{n} var(X_i) + \sum_{i \neq j} cov(X_i, X_j)]/n^2 \quad (1)$$

Where *var* (X)=*v*, and cov(X_i,X_j) denotes covariance between X_i and X_j (i≠j). This suggests that the variance of averages of *n* samples is *v*/*n* if the samples are all randomly distributed and have a variance *v* (cov(X_i,X_j) =0 for i≠j), the variance of averages is greater than *v*/*n* if they are aggregately distributed or positively correlated with each other (cov(X_i,X_j) >0 for i≠j), and smaller than *v*/*n* if they are uniformly distributed (cov(X_i,X_j) <0). Averaging by moving an *n*×*n* window is similar to calculating averages of n^2 variables with the same variance and mean, the deviation of variance of the averages from v/n^2, can therefore be used as a measure of spatial dependency within the window. To use an index (or a set of indices) as a measure of the spatial dependency over the whole data set, first of all, a 'stationary' assumption is needed: i.e., there exists a general or universal degree of dependency over the whole data set. The spatial dependency among the points within a moving window either does not change as the position of the moving window changes through the data set, or they are additive and their mean represents the degree of spatial dependency over the whole data set. This is the basis of using a single index (or a set of indices) for the whole dataset. Secondly, because a large window contains smaller windows, the spatial dependency at smaller windows will be carried over to larger windows. To quantify the carryover effects, an 'isotopic' assumption is needed: the spatial dependency between any point pair in a large window is the same as in the smaller windows as long as the minimum size of window to cover the two points remains the same, regardless of the

direction of the vector linking the two points, and the overall spatial dependency within a window can therefore be decomposed as the sum of the spatial dependency between all the possible point pairs. Based on these two assumptions, the new method of spatial analysis was developed using the variance of moving window averages (VMWA), the detailed mathematical derivations are presented in the Appendix at the end of the manuscript.

Calculation of spatial dependency indices. The calculation of dependency indices of VMWA can be carried out easily in GIS, or in Microsoft Excel following a simple procedure.

Step 1: Arrange the original two-dimensional data set, one sample point per cell;

Step 2: Move a window (size 1×1, 2×2, 3×3, etc.) cell by cell two-dimensionally and calculate the average of samples within the moving window to generate a new data set by assigning the average to the center cell (or left upper cell closest to the center) of the moving window;

Step 3: Calculate mean and variance of original data and the new data sets derived from different sizes of moving windows;

Step 4: Assess the edge effects based on the change of overall mean, dramatic changes in mean suggest strong edge effects, or the values at edges are very different from the values at other places;

Step 5: Calculate spatial dependency index at each window size based on variance of the averages at different window sizes by solving equations (equations 6 & 7 in the Appendix);

Step 6: Interpret the results: if the index > 0, then aggregated; if the index = 0, then random; otherwise (index < 0) uniform distribution.

Application of the VMWA method to artificial 0-1 data. The VMWA method was first tested with artificially generated binary data with regular, random and aggregated distributions.

Six regularly distributed binary data sets were generated for a 20×20 field. Among them, examples 1 and 2 consisted of alternatively arranged lines of 0s and 1s, diagonally and horizontally, respectively. In examples 3-6, the 1s were only located at the four corners of 3×3 to 6×6 windows, and the remainder of the cells were all 0s, so that no two 1s were located in any 2×2 (to 5×5) window. Random 0-1 data sets were generated based on an assumption that each sample site has a value of 0 or 1 randomly with $P(x=1)=0.1$ to 0.9 to equal 1 (500 data sets were generated at each P value), and that different sample sites are independent of each other. Aggregated 0-1 data were generated using commonly used Neyman-Scott cluster process (Cressie 1991), with the first step being the random generation of 1, 2, 4, 10 disease foci (infected plants have a value=1) in a field with 20×20 plants. Then, whether a plant becomes infected from each of the foci was determined (simulating 200 times of infection) according to the distance (r) from the disease focus: if the distance was farther than the influence range (denoted here the maximum distance that pathogen propagules could be disseminated) parameter R or the plant has already been infected, then no action was taken, otherwise the plant had the potential of getting infected at a probability of $P = e^{-r/R}/2\pi rR$ (van den Bosch et al. 1988). For each focus number, four hundred data sets were generated for each influence range from R=1 to R=7.

Application of the VMWA method to artificial count data. Artificial count data with different distributions were generated and used to test the VMWA method. Random distribution data sets were generated assuming that each sample unit consisted of 20 individuals and each of them randomly (independently) became diseased at a probability p=0.1-0.9, or otherwise remained healthy (500 data sets were generated at each p value). Aggregated data sets were generated using a Neyman-Scott cluster process similar to the above. First, 10 disease foci

(sample units) were randomly generated in a 20×20 field. Then, disease (unlimited) was reproduced at each sample unit around each of these foci according to a given disease gradient function based on its distance r from the center of these foci, $D(r)= e^{-r/R}/(2\pi rR)$. At each focus number, four hundred data sets were generated for each influence range from R=1 to R=16. Similar simulation was also done in a 40×40 field with influence range from R=1 to R=36.

Application of VMWA to historical survey data. The VMWA method was also employed to assess the spatial patterns in historical data from disease surveys and compared with the results obtained from other methods.

First, the number of lettuce plants infected by *Bremia lactucae* from historical disease surveys in three fields, in which each sample unit consisted of 10 plants, and sample sites were 20 m apart within a row and across rows (Wu et al. 2001), was analyzed using VMWA, BBD (Madden and Hughes 1994), SADIE (Perry 1995 & 1998), and spatial autocorrelation software LCOR2 (Gottwald et al. 1992).

Second, the number of microsclerotia of *Verticillium dahliae* per gram soil from cauliflower fields, the number of plants infected by *V. dahliae* and the disease severity in the same field sites from a previous study (Xiao et al. 1997) were analyzed using the VMWA method. Each site was divided into 8×8 contiguous quadrats, 2×2 m each (2-m length on two rows of cauliflower plants). The results obtained with the VMWA method were compared with those obtained with LIP, BBD, semivariogram and SADIE methods, as applicable to each data set.

The VMWA method was also used to analyze historical data from disease surveys on pepper wilt caused by *V. dahliae* (Bhat et al. 2003). The original data of healthy or diseased status for every individual plant, was used in this analysis instead of the 20-plant quadrats reported in the study (Bhat et al. 2003). No comparison of the results with LIP, BBD, LCOR2, SADIE, or

semivariogram methods was made because none was applicable to these large binary data sets with missing data at some sample sites.

RESULTS

Application of the VMWA method to artificial 0-1 data. For the six examples of regularly distributed 0-1 data, the index calculated using the VMWA method correctly reflected the spatial pattern. In examples 1 and 2, the spatial dependency index was negative at moving window size 2×2, and positive at 3×3, consistent with the one-line-wide gap between any two lines of 1s (Table 1). In examples 3 to 6, the indexes remained negative or almost zero when moving windows were smaller than the minimum window sizes (3×3 to 6×6 in examples 3 to 6) required to cover any two 1s, and became positive when the moving windows were at the minimum sizes (Table 1).

Regardless of the incidence, the VMWA method correctly classified the spatial patterns for randomly distributed 0-1 data sets since the value of the index was close to zero and varied little in the five hundred simulations at each incidence (Table 2).

As for aggregated 0-1 data, the VMWA method correctly reflected the ranges used in epidemic simulations when the ranges were shorter than 5 in a 20×20 field regardless of whether the number of disease foci was 1, 2, 4, or 10 (Fig. 1A-D). The indices were positive for moving window 2×2, remained positive till the window size was equal to influence range +1, and decreased to zero at the window size equal to range +2 (Fig. 1A-D). However, as the range increased beyond 5, the index curve became slightly flatter. For the small moving windows, the index topped off at range 3-4, and began to decrease as the range increased further (Fig. 1A-D) perhaps because the center of the disease foci became a plateau as the range increased.

Application of the VMWA method to artificial count data. Simulations of count data (number of diseased plants) showed that the VMWA method correctly identified the spatial pattern as disease incidence ranged from 10% to 90%. The indices were close to zero and varied little (Table 3). As for the simulated aggregated data, the VMWA method correctly reflected the ranges used in simulation of the epidemics when the ranges were small. The index curves changed little after the range increased beyond 10 in a 20×20 field (Fig. 2A) and 20 in a 40×40 field (Fig. 2B). These curves were slightly different from the curves for 0-1 data in Figs. 1A-D in that for count data, there was no decrease in the value of the index at small moving windows as the range increased. This difference was perhaps because the center of a disease focus never became a plateau (unlimited disease level) as the range increased in simulation, unlike the simulation of 0-1 data.

Application of VMWA to historical survey data. For the number of lettuce plants infected by *B. lactucae,* the VMWA method detected positive (positive index) association among the sample sites within a window size of 4 to 5 in all three fields (Table 4). Therefore, it determined that the distributions were aggregated. The field with 60% incidence had the lowest degree of aggregation, and the smallest cluster size among the three fields. These results were highly consistent with the results from the previous semivariogram analysis (Wu et al. 2001) and results from LCOR2 (Table 5). Results from SADIE (Perry 1995 & 1998) and BBD analysis (Madden and Hughes 1994) also defined that diseased plants in the three fields were all aggregated (Table 4), suggesting a positive association not only within sample units but also across sample units.

For the number of microsclerotia of *V. dahliae* per gram soil from cauliflower fields, weak aggregation (with positive, but small indexes) was detected by the VMWA method in sites A1, A2, B1, B2, B3, C2 and C3, but not in the other sites (Table 6), which suggests weak association

between the numbers of microsclerotia across sample sites. The results were highly consistent with the previous results from semivariogram analysis (Xiao et al. 1997), but different from the results of LIP tests, which showed a LIP value higher than 1 in all the sites, indicating an aggregated distribution of microsclerotia within sample units (Xiao et al. 1997). The results also differed from those obtained with SADIE, which showed aggregated spatial patterns of microsclerotia only in sites A2 and C2 (Table 6). As for the number of cauliflower plants infected by *V. dahliae* (Xiao et al. 1997), the VMWA method detected aggregated distributions not only in field sites A1, A2, and B3 in which BBD detected aggregation previously (Xiao et al. 1997), but also in sites B1, B2, C1, D1, and D3 (Table 6) in which aggregation was not detected previously by BBD (Xiao et al. 1997). The results were partly consistent with the results obtained from SADIE method, which also defined aggregated distribution for diseased plants in sites B1, C1, and D1, but not in sites B2, B3, and D3. Furthermore, no aggregation was found by the VMWA method (Table 6) in field sites A3 and C3 where aggregation was defined previously by BBD method (Xiao et al. 1997), but not by SADIE method (Table 6). When disease severity was analyzed, aggregation was detected in four out of the six field sites examined (Table 6). The LIP, BBD and SADIE methods were not suitable for these data sets.

As for the binary data of healthy or diseased (infected by *V. dahliae)* pepper plants, the VMWA method revealed strong aggregation in field sites 1A, 1B and 1C, weak aggregation in field sites 2A, 2B, and 3B (Table 7), and random distribution in the other three sites 2C, 3A, and 3C, highly consistent with the visual impression of the maps (Fig. 3).

DISCUSSION

Based on the above analyses, the VMWA method possesses certain advantages over currently available techniques for spatial pattern analysis. First, choice of proper software for

spatial pattern analysis has long been the bane for average users in spatial pattern analysis because of the variety of programs available, the level of computer knowledge required to use these programs and their applicability being restricted to certain types of data. This has resulted in many cases of inappropriate application of spatial analysis methods. The VMWA method can be applied to many types of data including binary diseased or healthy plant counts, disease incidence, and number of diseased plants or pathogen propagules. Therefore, VMWA method offers significant advantages over other methods that are only applicable for certain types of data. Second, the calculation of the VMWA method does not require additional sophisticated software, and can easily be performed in GIS or Microsoft Excel, in which a user will only need to paste the data into a spreadsheet, or run a macro program for multiple data sets. Furthermore, if the purpose is only to determine whether or not there is aggregation over the whole study area, the method can be further simplified by ignoring the locations of data points in the moving windows. Therefore, all COV_k at different window sizes k=1, 2, …, n (equation 6 in the appendix) can be considered equal. Aggregation can be defined as long as v_n is significantly greater than v/n.

Due to the similarity in principle between the VMWA method and spatial autocorrelation (semivariogram), the results from these methods are highly consistent with each other for most of the example data sets (when edge effect is ignorable) used in this paper. The VMWA method theoretically differs from spatial autocorrelation in that only the average association over all directions is given in the outputs of VMWA while the latter can also give the association for each direction, and that the VMWA method uses a window size that is slightly different from the distance (vector) used in autocorrelation (or semivariogram). Because of these, compared with spatial autocorrelation and semivariogram methods, the VMWA method is more suitable when

data sets are small, and when the topological relationship between any two sample sites is more important than the physical distance and orientation between them. It should also be noted that the number of calculations for moving average is considerably fewer than the number of calculations for distance between any two pairing points, as the size of a dataset increase. This is also a significant benefit of the VMWA method as more and more raster data are becoming available, and the resolution of the data increasing over years. For example, for a dataset composed of 1.0×10^8 (10000×10000) points, spatial autocorrelation needs to calculate distance for about $5.0*10^{15}$ point pairs while the VMWA method, at maximum, needs to calculate 1.0×10^{12} averages. Furthermore, if we are only interested in the spatial dependence at a smaller scale, such as within a 500×500 or smaller window, the number of averaging calculations can be further reduced to 5.0×10^{10} (for a 3-GHz CPU, it takes about 20 seconds, compared with 278 hours to do spatial autocorrelation). This would significantly reduce the computing time, and make it feasible to do spatial analysis for datasets covering large areas and with high resolutions.

The VMWA method is, to some extent, also similar to the quadrat variance methods. The latter calculate the 'local variance' between adjacent blocks or paired-quadrats, and then calculate the overall average (Greig-Smith 1952, Hill 1974, and Goodall 1974, Ludwig and Goodall 1978). For example, at block size 1, the 'local variance' can be calculated differently in different quadrat variance methods, such as $[(x_1-x_2)^2+(x_3-x_4)^2+\ldots...]/2$, or $[(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_4)^2+\ldots...]/2$, and more. Taking the first formula as an example and assuming that the mean of the data sets is m, it can be further rewritten as: $\Sigma(x-m)^2/2-[(x_1-m)(x_2-m)+ (x_3-m)(x_4-m)+\ldots]$. It becomes clear that the second part of the formula is a portion of total cross product at distance 1 (one-dimensional data). Because calculation of the 'local variance' becomes complicated as the block size increases, the quadrat variance methods become complicated as well. Besides, the

results of the methods are usually affected by the starting place of quadrat and block size, and it is difficult to use these methods in two-dimensional data sets, which are more common in nature. The VMWA method, in contrast, is intrinsically more suitable for two-dimensional data than the traditional quadrat variance methods because the calculation of average is very easy within each moving window. Moreover, moving the windows two dimensionally makes the results of the variance reflecting the whole data set better, and is less affected by local variation or by the starting position of the moving windows. The partition of variance for a big moving window into variance at smaller windows solved the carryover problem satisfactorily. These together make the VMWA method more stable as well as more accurate.

As with most other methods, there are also certain limitations with the VMWA method. As stated above, it cannot detect directional differences. Instead, it gives an average degree of spatial association between two samples. It is inappropriate to use this method when directional effects are significant because the calculation of spatial dependency index relies upon the 'isotopic' assumption. In addition to this limitation of directional effect, edge effect is another factor that may limit the use of VMWA method because a moving window consists of fewer pairs far apart at the edges than at the center, and the number of pairs affected increases as the window size increases. Although another method was developed as a comparison, it was found that the edge effect, in general, tends to cause the aggregation index to fluctuate. In cases where edge effect was not significant, the two methods ended in very similar results (results not shown).

The VMWA method has also great potential for interfacing with GIS because moving window averaging is a common function in GIS, such as in ArcGIS, a widely used GIS software. This feature makes VMWA very easy to carry out although more studies are needed to adapt this

method into GIS. To further increase its flexibility, it is also possible to use different shapes of moving windows, such as rectangle windows to determine the shape of clusters in addition to the average cluster size (see equations 12-14 in the Appendix).

Appendix: Mathematical derivations

Assuming that the original data set $\{y_{ij}\}$ ($1 \le i \le L$, and $1 \le j \le W$) has a mean m, and averages are calculated within an $n \times n$ moving window w consists of y_{w11}, y_{w12}, …, y_{w1n}, …y_{wn1}, y_{wn2}, …, y_{wnn} (composed of f_w points). The variance (S_n) of the new data set (with an average m_n) derived by moving window ($n \times n$) averaging can be calculated as:

$$S_n = \frac{\sum_{w=1}^{LW} \left[E\left(y_{wij}\right) - m_n \right]^2}{(LW - 1)} \quad (2)$$

$$= \frac{\sum_{w=1}^{LW} \left[\left(\sum_{i,j=1}^{n} y_{wij} \right) / f_w - m_n \right]^2}{(LW - 1)} \quad (3)$$

It can be rewritten as:

$$(LW-1)S_n + LW(m - m_n)^2 = \sum_{w=1}^{LW} \left\{ \left[\sum_{i,j=1}^{n} \left(y_{wij} - m\right) \right]^2 / f_w^2 \right\} \quad (4)$$

Based on the 'isotopic' assumption, product $(y_{i1,j1} - m) \cdot (y_{i2,j2} - m)$ was only determined by k, the minimum window size covering ($i1$, $j1$) and ($i2$, $j2$), two points in the moving window then:

$$(LW - 1) S_n + LW (m - m_n)^2 = \sum_{w=1}^{LW} \sum_{k=1}^{n} \left\{ (y_{w,1} - m)(y_{w,k} - m) \lambda_{w,n,k} / f_w^2 \right\} \quad (5)$$

Assuming also $(y_{w,1}-m)(y_{w,k}-m)$ is not dependent on the position of the moving window w (the stationary assumption), and defining $COV_k = \sum_{w}^{LW} \{(y_{w,1} - m)(y_{w,k} - m)\}/LW$, then,

$$S_n \bullet \frac{LW-1}{LW} + (m-m_n)^2 = \sum_{k=1}^{n} \delta_{n,k} COV_k \quad (6)$$

where $\delta_{n,k} = \sum_{w}^{LW} \left(\lambda_{w,n,k} / f_w^2\right) / LW$ (see equations 8-11 for calculation of $\delta_{n,k}$)

Therefore, we can calculate COV_k based on $S_1,\ldots,S_k$ by solving equations 6, and define:

$$I_n = COV_n / COV_1 = COV_n / v \quad (7)$$

as equivalents to spatial autocorrelation coefficients to measure the spatial dependency between sample points.

Calculation of coefficients $\delta_{n,k}$. When an n×n window is moving through a L×W (L≤W) data set, the number of the data points that the window contains is n×n in the middle and less than n×n along the edges and corners. The overall frequency ($F_{a,b}$) that a moving window is full or partially full (n-a) × (n-b) [n<L and a, b≤int(n/2)] can be calculated as:

$$F_{a,b} = \begin{cases} (L-n+1)(W-n+1)/LW & \text{if } a=b=0 \\ [2-\text{int}(2a/n)](W-n+1)/LW & \text{if } a>0, b=0 \\ [2-\text{int}(2b/n)](L-n+1)/LW & \text{if } a=0, b>0 \\ [2-\text{int}(2a/n)][2-\text{int}(2b/n)]/LW & \text{if } a,b>0 \end{cases} \quad (8)$$

In a full or partially full window (n-a)×(n-b), the coefficient $\lambda_{a,b,n,k}$ for point pairs that can only be placed in a window not smaller than $k \times k$, define A=n-a and B=n-b, can be calculated as:

$$\lambda_{a,b,n,k} = \lambda_{A,B,k} = \begin{cases} A \times B & \text{if } k=1 \\ 2(k-1)\times[(A-k+1)(2B-k+1)+(B-k+1)(2A-k+1)] & \text{if } k \le A,B \\ 2A \times A \times (B-k+1) & \text{if } B \ge k > A \\ 2B \times B \times (A-k+1) & \text{if } A \ge k > B \\ 0 & \text{if } A,B < k \end{cases} \quad (9)$$

Because $\sum_k \lambda_{a,b,n,k} = [(n-a)(n-b)]^2$ for any a, b, and n, therefore $\sigma_{a,b,n,,k}$ can be calculated as:

$$\sigma_{a,b,n,,k} = \lambda_{a,b,n,k} / f_w^{\,2} = \lambda_{a,b,n,k} / (n\text{-}a)^2 / (n\text{-}b)^2 \qquad (10)$$

Therefore, the weighted average number of the point pairs that can only fit in a window not smaller than $k \times k$ can be calculated by integrating formulae 8-10:

$$\delta_{n,,k} = \sum_{w}^{LW} \left(\lambda_{w,n,k} \; / \; f_w^{\,2} \right) / \; LW \;\; = \sum_{a,b=0}^{\mathrm{int}(n/2)} \left(\mathrm{F}_{\mathrm{a,b}} \times \sigma_{a,b,n,k} \right) \qquad (11)$$

$\delta_{n,,k}$ are dependent on the size of dataset (L and W), moving window size n, and k (the minimum window to cover the point pair), and therefore can be calculated with a program for using in equation 6.

More generally, if one uses n rectangle moving windows MW_1, …, MW_n, (with width w and length l not greater than n) and define matrices $C = \begin{pmatrix} COV_1 \\ COV_2 \\ \cdot \\ COV_n \end{pmatrix}$, $V = \begin{pmatrix} V_1 \\ V_2 \\ \cdot \\ V_n \end{pmatrix}$, $D = \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \cdot & \delta_{1,n} \\ \delta_{2,1} & \delta_{2,2} & \cdot & \delta_{2,n} \\ \cdot & \cdot & \cdot & \cdot \\ \delta_{n,1} & \delta_{n,2} & \cdot & \delta_{n,n} \end{bmatrix}$,

where $V_n = S_n \bullet \dfrac{LW-1}{LW} + (m - m_n)^2$, then the equations 6 can be expressed as DC=V. However, the calculation of $\delta_{n,,k}$ is slightly different from the one for square moving windows: First, frequency $F_{\mathrm{a,b}}$ for a full or partially full moving window (l-a) × (w-b) {where $l<L$, $w<W$, $a \leq \mathrm{int}(l/2)$, and $b \leq \mathrm{int}(w/2)$} can be calculated as:

$$F_{a,b} = \begin{cases} (L-l+1)(W-w+1)/LW & \text{if } a = b = 0 \\ [2-\mathrm{int}(2a/l)](W-w+1)/LW & \text{if } a > 0, b = 0 \\ [2-\mathrm{int}(2b/w)](L-l+1)/LW & \text{if } a = 0, b > 0 \\ [2-\mathrm{int}(2a/l)][2-\mathrm{int}(2b/w)]/LW & \text{if } a, b > 0 \end{cases} \qquad (12)$$

Defining $A=l-a$ and $B=w-b$, calculation of $\lambda_{a,b,l,w,k}$ will be the same as for that of $\lambda_{A,B,,k}$ in formula 9, then $\delta_{n,k}$ can be calculated based on the l and w of the n-th moving windows as:

$$\delta_{n,k} = \sum_{a,b=0} \left(\mathrm{F}_{\mathrm{a,b}} \times \lambda_{A,B,k} / [(l_n - a)(w_n - b)] \right)^2 \qquad (13)$$

As long as D^{-1} exists, $COV_1 \sim COV_n$ can be calculated based on $V_1 \sim V_n$:

$$C=D^{-1}V \quad (14)$$

REFERENCES

1. Alston R. (1994) Statistical analysis of animal populations. PhD thesis, University of Kent.
2. Bhat RG, Smith RF, Koike ST, Wu BM, Subbarao KV. (2003) Characterization of *Verticillium dahliae* isolates and wilt epidemics of pepper. Plant Dis. 87: 789-797.
3. Campbell CL, Madden LV. (1990) Introduction to Plant Disease Epidemiology. Wiley Interscience, New York. 532 pp.
4. Campbell CL, Noe JP. (1985) The spatial analysis of soilborne pathogens and root diseases. Ann. Rev. Phytopathol. 23: 129-148.
5. Clark I. (1979) Practical Geostatistics. London: Applied Science Publishers.
6. Clark PJ, Evans FC. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35: 445-453.
7. Cliff AD, Ord JK. (1973) Spatial autocorrelation. Pion Ltd., London. 178 pp.
8. Cressie N. (1991) *Statistics for Spatial Data.* John Wiley & Sons, New York, 904 pp.
9. David FN, Moore PG. (1954) Notes on contagious distributions in plant population. Ann. Bot. (Lond.) 18: 47-53.
10. Elliott JM. (1977) Some methods for the statistical analysis of samples of benthic invertebrates. Westmoreland, England: Freshwater Biology Association.
11. Ferrandino FJ. (1996) Two-dimensional distance class analysis of disease-incidence data: Problems and possible solutions. Phytopathology. 86: 685-691.
12. Fisher RA. (1925) Statistical Methods for Research workers. Hafner, New York. 362 pp.

13. Gabriel KR, Sokal RR. (1969) A new statistical approach to geographic analysis. Syst. Zool. 18: 259-270.
14. Geary RC. (1954) The contiguity ratio and statistical mapping. Incorporated Statist. 5: 115-145.
15. Goodall DW. (1974) A new method for the analysis of spatial pattern by random pairing of quadrats. Vegetatio 29: 135-146.
16. Gottwald TR, Richie SM, Campbell CL. (1992) LCOR2—Spatial correlation analysis software for the personal computer. Plant Dis. 76: 213-215.
17. Gray SM, Moyer JW, Bloomfield P. (1986) Two-dimensional distance class model for quantitative description of virus-infected plant distribution lattices. Phytopathology 76: 243-248.
18. Greig-Smith P. (1952) The use of random and contiguous quadrats in the study of the structure of plant communities. Ann. Bot. 16: 293-316.
19. Griffiths DA. (1973) Maximum likelihood estimation for beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29: 637-648.
20. Hill MO. (1973) The intensity of spatial pattern in plant communities. J. Ecol. 61: 225-236.
21. Hughes G, Madden LV. (1993) Using the beta-binomial distribution to describe aggregated patterns of disease incidence. Phytopathology 83: 759-763.
22. Irwin JO. (1954) A distribution arising in the study of infectious diseases. Biometrika 41: 266-268.
23. Liu C. (2001) A comparison of five distance-based methods for spatial pattern analysis. J. Veg. Sci. 12: 411-416.

24. Lloyd M. (1967) Mean crowding. J. Anim. Ecol. 36: 1-30.
25. Ludwig JA, Goodall DW. (1978) A comparison of paired-with blocked-quadrat variance methods for the analysis of spatial pattern. Vegetatio 38: 49-59.
26. Madden LV, Hughes G. (1994) BBD—Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78: 536-540.
27. Madden LV, Hughes G. (1995) Plant disease incidence: distribution, heterogeneity, and temporal analysis. Ann. Rev. Phytopathol. 33:529-64.
28. Moran PAP. (1950) Notes on continuous stochastic phenomena. Biometrics 37: 17-23.
29. Morisita M. (1962) I_δ-index, a measure of dispersion of individuals. Res. Popul. Ecol. 4: 1-7.
30. Nelson SC, Marsh PL, Campbell CL. (1992) 2 DCLASS, a two-dimensional distance class analysis software for the personal computer. Plant Dis. 76: 427-432.
31. Perry JN. (1995) Spatial analysis by distance indices. J. Anim. Ecol. 64: 303-314.
32. Perry JN. (1998) Measures of spatial pattern for counts. Ecology 79: 1008-1017.
33. Perry JN, Hewitt M. (1991) A new index of aggregation for animal counts. Biometrics 47: 1505-1518.
34. Pielou EC. (1959) The use of point-to-point distances in the study of the pattern of plant populations. J. Ecol. 47: 607-613.
35. Real LA, McElhany P. (1996) Spatial Pattern and process in plant-pathogen interactions. Ecology 77: 1011-1025.
36. Taylor LR. (1961) Aggregation, variance and the mean. Nature 189: 732-735.
37. van den Bosch F, Zadoks JC, Metz JAJ. (1988) Focus expansion in plant disease. I: The constant rate of focus expansion. Phytopathology 78: 54-58.

38. Wu BM, van Bruggen AHC, Subbarao KV, Pennings GGH. (2001) Spatial analysis of lettuce downy mildew using geostatistics and geographic information systems. Phytopathology 91: 134-142.

39. Wu BM, Subbarao KV. (2005) Analysis of spatial patterns in plant pathology. Pages 167-187 in: Recent Research Developments in Plant Pathology, Vol.3 (2004). S. G. Pandalai, ed. Research Signpost, Kerala, India.

40. Xiao CL, Hao JJ, Subbarao KV. (1997) Spatial patterns of microsclerotia of *Verticillium dahliae* in soil and Verticillium wilt of cauliflower. Phytopathology 87: 325-331.

TABLE 1. Spatial dependency indices at different window sizes calculated using a method based on variance of moving window averages (VMWA) for artificial binary data with a regular distribution

Window size	Example1[a]	Example2	Example3	Example4	Example5	Example6
2×2	-0.38[b]	-0.31	-0.33	-0.09	-0.03	0.02
3×3	0.23	0.39	0.33	-0.17	-0.12	-0.13
4×4	-0.10	-0.45	-0.33	0.26	-0.03	0.02
5×5	0.03	0.50	0.34	-0.09	0.21	-0.04
6×6	-0.01	-0.50	-0.34	-0.14	-0.07	0.16
7×7	0.02	0.48	0.32	0.21	-0.09	-0.04
8×8	-0.02	-0.48	-0.32	-0.10	-0.04	-0.05

[a] All example datasets were composed of 20×20 data points each with a value 0 or 1. In examples 1 and 2, 0s and 1s were arranged in alternative lines, diagonal and horizontal, respectively. In examples 3 to 6, 1s were only located at the four corners of 3×3 to 6×6 windows in the order, and the rest were all 0s, so that no two 1s were located in any 2×2 to 5×5 (in the same order) windows.

[b] Positive values of index indicate positive association, negative values indicate negative association, and zero values indicate random distribution, or independence.

TABLE 2. Averages and standard deviations of spatial dependency indices at different window sizes calculated using a method based on the variance of moving window averages (VMWA) for simulated binary data with a random distribution and different incidence levels

Incidence	Spatial dependency indices at different window sizes								
	2×2	3×3	4×4	5×5	6×6	7×7	8×8	9×9	10×10
10%	-0.002[a]	0.000	-0.003	-0.003	-0.002	-0.004	-0.003	-0.004	0.000
	±0.028[b]	±0.029	±0.023	±0.021	±0.024	±0.022	±0.027	±0.019	±0.031
20%	-0.004	-0.003	-0.001	-0.004	-0.002	-0.004	0.000	-0.003	-0.001
	±0.028	±0.025	±0.021	±0.002	±0.022	±0.020	±0.026	±0.020	±0.031
30%	-0.004	-0.003	-0.002	-0.004	-0.002	-0.004	-0.001	-0.004	-0.001
	±0.028	±0.023	±0.020	±0.020	±0.022	±0.020	±0.026	±0.019	±0.029
40%	-0.004	-0.002	-0.001	-0.005	0.000	-0.003	-0.003	-0.005	-0.003
	±0.028	±0.024	±0.020	±0.020	±0.021	±0.020	±0.024	±0.019	±0.027
50%	-0.004	-0.004	0.000	-0.003	-0.001	-0.004	0.000	-0.003	-0.001
	±0.026	±0.023	±0.021	±0.021	±0.023	±0.019	±0.024	±0.020	±0.029
60%	-0.001	-0.002	-0.002	-0.004	-0.002	-0.002	0.000	-0.004	-0.001
	±0.028	±0.024	±0.019	±0.019	±0.023	±0.019	±0.024	±0.020	±0.029
70%	-0.002	-0.001	0.000	-0.004	-0.003	-0.005	-0.002	-0.005	0.000
	±0.029	±0.022	±0.020	±0.021	±0.022	±0.020	±0.024	±0.019	±0.029
80%	-0.002	-0.003	-0.003	-0.004	-0.002	-0.002	-0.003	-0.004	-0.002
	±0.028	±0.026	±0.023	±0.020	±0.021	±0.019	±0.024	±0.018	±0.029
90%	-0.005	-0.001	-0.002	-0.003	-0.004	-0.006	0.002	-0.003	0.000
	±0.030	±0.029	±0.023	±0.021	±0.025	±0.023	±0.023	±0.020	±0.030

[a] The average of the index in 500 simulations. Positive values indicate positive association, negative values indicate negative association, and zero values indicate random distribution, or independence. Each data set was generated based on the assumption that in a field with 20×20 healthy plants (value =0), each plant (sample unit) has a probability (p=0.10% to 90%) to become diseased (with value 1) independently.

[b] The standard deviation of the index immediately above.

TABLE 3. Averages and standard deviations of spatial dependency indices at different window sizes calculated using a method based on the variance of moving window averages (VMWA) for simulated number of diseased plants with random distribution and different incidence levels

Incidence	Spatial dependency indices at different window sizes								
	2×2	3×3	4×4	5×5	6×6	7×7	8×8	9×9	10×10
10%	-0.002[a] ±0.029[b]	-0.003 ±0.027	-0.002 ±0.020	-0.003 ±0.020	-0.001 ±0.021	-0.004 ±0.019	-0.001 ±0.026	-0.004 ±0.020	-0.002 ±0.027
20%	-0.006 ±0.027	-0.002 ±0.024	-0.003 ±0.020	-0.004 ±0.020	0.000 ±0.022	-0.005 ±0.019	-0.001 ±0.025	-0.004 ±0.019	-0.001 ±0.029
30%	-0.003 ±0.028	-0.003 ±0.024	-0.001 ±0.021	-0.003 ±0.020	-0.002 ±0.021	-0.003 ±0.02	0.000 ±0.025	-0.004 ±0.020	-0.002 ±0.029
40%	-0.001 ±0.026	-0.003 ±0.024	-0.003 ±0.021	-0.003 ±0.019	-0.002 ±0.022	-0.004 ±0.020	-0.001 ±0.024	-0.004 ±0.019	-0.001 ±0.028
50%	-0.005 ±0.028	-0.003 ±0.024	-0.004 ±0.020	-0.002 ±0.020	-0.003 ±0.021	-0.004 ±0.020	0.000 ±0.024	-0.004 ±0.018	-0.003 ±0.029
60%	-0.004 ±0.027	-0.001 ±0.026	-0.002 ±0.020	-0.002 ±0.020	-0.002 ±0.022	-0.004 ±0.021	0.000 ±0.027	-0.002 ±0.020	-0.002 ±0.028
70%	-0.002 ±0.026	-0.003 ±0.026	0.000 ±0.022	-0.004 ±0.019	-0.001 ±0.021	-0.006 ±0.020	0.001 ±0.025	-0.004 ±0.019	-0.001 ±0.031
80%	-0.001 ±0.029	-0.002 ±0.025	-0.002 ±0.020	-0.004 ±0.021	-0.001 ±0.022	-0.005 ±0.02	0.000 ±0.026	-0.003 ±0.018	-0.001 ±0.029
90%	-0.004 ±0.028	-0.001 ±0.025	-0.004 ±0.022	-0.002 ±0.020	0.000 ±0.022	-0.004 ±0.019	0.000 ±0.025	-0.003 ±0.018	-0.003 ±0.029

[a] The average of spatial dependency indices in 500 simulations. Positive values indicate positive association negative values indicate negative association, and zero values indicate random distribution, or independence. Each data set, consisted of 20×20 sample units, was generated assuming that each sample unit consisted of 20 individuals and each of them randomly (independently) became diseased (with p=0.1-0.9) or remained healthy.

[b] The standard deviation of the index immediately above.

TABLE 4. Characterization of spatial patterns of lettuce plants showing downy mildew symptoms in 1998 surveys (Field A-C) using the VMWA, SADIE, and BBD methods

Method	Indices/ Coefficients	Field A (31.5%)	Field B (60.0%)	Field C (71.0%)
VMWA	I_2[a]	0.58	0.30	0.49
	I_3	0.51	0.02	0.18
	I_4	0.24	0.15	0.40
	I_5	0.18	-0.07	0.05
SADIE	I_a	3.32*[b]	1.56*	3.20*
BBD	θ	0.41*	0.16*	0.36*
	D	3.66*	2.26*	3.40*

[a] I_2 to I_5 denote spatial dependency indices at window sizes 2×2, 3×3, 4×4, and 5×5, respectively. Positive values indicate positive association, negative values indicate negative association, and zero values indicate random distribution, independence.

[b] I_a, θ and D values followed by asterisk were significant (P<0.05).

TABLE 5. Analysis of spatial patterns of lettuce plant showing downy mildew in 1998 surveys (Field A-C) using LCOR2

Axes	0	1	2	3	4	5	6	7	8	9
Field A	Incidence = 31.5%									
0	1.00**	0.61**	0.53**	0.29**	0.39**	0.40**	0.59**	0.43**	0.34*	0.27
1	0.61**	0.53**	0.41**	0.25*	0.29**	0.37**	0.44**	0.39**	0.38*	0.16
2	0.57**	0.42**	0.38**	0.15	0.27*	0.20	0.33*	0.20	0.17	0.08
3	0.53**	0.47**	0.37**	0.16	0.26*	0.25*	0.31*	0.22	0.33*	0.24
4	0.46**	0.32**	0.34**	0.18	0.25	0.22	0.39**	0.22	0.28	0.18
5	0.34**	0.23*	0.17	0.11	0.12	0.07	0.14	-0.08	-0.11	-0.29
6	0.17	0.09	0.11	-0.08	0.05	-0.05	0.01	-0.14	0.11	-0.06
7	0.14	0.12	0.18	0.14	0.06	-0.04	0.04	-0.14	0.19	-0.04
8	0.22	0.16	0.25	0.10	0.15	0.05	0.06	-0.14	0.33	0.31
9	0.09	0.00	-0.12	-0.08	-0.25	-0.10	0.09	-0.14	-0.01	-0.32
Field B	Incidence = 60.0%									
0	1.00**	0.25**	0.27**	0.20*	0.15	0.02	0.04	0.09	0.23	0.17
1	0.38**	0.19*	0.09	0.15	0.02	0.05	0.01	0.08	0.16	0.01
2	0.12	0.07	0.03	0.06	-0.06	0.10	-0.08	0.01	-0.11	-0.03
3	0.04	0.00	0.02	0.00	0.01	0.03	0.02	-0.12	0.04	0.00
4	-0.06	-0.21	-0.13	-0.10	-0.15	-0.10	-0.01	-0.17	-0.18	0.04
5	-0.11	-0.07	-0.22	-0.14	-0.34*	-0.25	-0.16	-0.09	-0.11	-0.06
6	-0.03	-0.02	-0.11	-0.10	-0.24	-0.02	0.01	0.01	0.24	0.06
7	-0.05	-0.05	-0.01	-0.08	0.14	0.08	0.15	-0.04	0.35	0.04
8	-0.27	-0.17	-0.03	-0.01	0.15	0.17	0.30	0.13	0.08	-0.03
9	-0.27	-0.17	-0.08	0.06	0.23	0.23	0.33	-0.04	0.19	-0.03
Field C	Incidence = 71.0%									
0	1.00**	0.32**	0.31**	0.46**	0.20	0.19	0.22	0.25	-0.02	0.33*
1	0.70**	0.29**	0.28**	0.43**	0.07	0.17	0.25*	0.20	-0.07	0.29
2	0.64**	0.34**	0.26*	0.44**	0.06	0.16	0.26*	0.01	0.02	0.23
3	0.57**	0.24*	0.21	0.43**	0.02	0.16	0.18	0.12	0.10	0.17
4	0.52**	0.15	0.23*	0.34**	-0.12	0.10	0.17	0.06	-0.16	0.10
5	0.49**	0.17	0.23	0.37**	0.02	0.16	0.13	0.17	-0.15	0.26
6	0.46**	0.18	0.25	0.38**	0.10	0.16	0.14	0.08	-0.33	0.20
7	0.45**	0.21	0.36*	0.38**	0.07	0.19	0.08	0.17	-0.08	0.31
8	0.51**	0.33*	0.37*	0.47**	0.22	0.32	0.04	0.08	0.05	0.30
9	0.46**	0.30	0.39*	0.43*	0.38	0.33	0.20	0.05	0.03	0.56

* Values followed by asterisk were significant (*: $P<0.05$, and **: $P<0.01$).

TABLE 6. Characterization of spatial patterns in number of microsclerotia of *Verticillium dahliae*, number of cauliflower plants infected by the pathogen and disease severity (Xiao et al. 1997) using the VMWA, LIP or BBD (Madden and Hughes, 1994) methods

Plot	Number of microsclerotia			Number of diseased plants			Disease severity
	VMWA[a]	LIP	I_a[c]	VMWA	θ[b]	I_a[c]	VMWA
A1	0.10	2.17	1.234	0.17	0.055*	1.682*	
A2	0.13	1.15	1.533*	0.44	0.145*	2.481*	
A3	-0.01	1.08	1.273	0.04	0.047*	0.941	
B1	0.07	1.10	1.101	0.15	0.019	1.543*	
B2	0.12	1.06	1.265	0.16	0.017	1.147	
B3	0.11	1.09	0.929	0.10	0.056*	1.272	
C1	-0.01	1.07	0.909	0.13	0.007	1.618*	0.38
C2	0.10	1.17	1.618*	-0.06	0.010	0.802	0.13
C3	0.07	1.11	1.120	-0.02	0.040*	1.095	0.12
D1	-0.08	1.09	1.129	0.27	0.004	2.069*	0.13
D2	-0.01	1.06	1.138	-0.18	0.011	0.823	0.04
D3	-0.06	1.04	1.175	0.07	0.018	1.308	-0.13

[a] Spatial dependency index I at window size of 2 (two nearby sample sites) using VMWA method. Positive values indicate positive association, negative values indicate negative association, and zero values indicate random distribution, independence.

[b] a θ value from BBD followed by "*" indicated an aggregation of diseased plants (Xiao et al. 1997).

[c] I_a from SADIE method followed by "*" indicates an aggregation (P<0.05).

TABLE 7. Spatial dependency indices at different window sizes for spatial patterns of pepper plants infected by *Verticillium dahliae* in 1998 surveys at Field 1, Field 2 and Field 3 calculated using the VMWA method

Window size	Field 1A	Field 1B	Field 1C	Field 2A	Field 2B	Field 2C	Field 3A	Field 3B	Field 3C
2×2	0.42[a]	0.37	0.45	0.07	0.08	0.05	0.04	0.11	0.02
3×3	0.39	0.35	0.39	0.03	0.04	-0.04	0.01	0.05	0.01
4×4	0.31	0.27	0.35	0.02	0.03	0.01	0.04	0.04	0.04
5×5	0.41	0.24	0.29	0.01	0.05	0.02	-0.02	0.06	0.00
6×6	0.14	0.15	0.24	0.03	-0.04	-0.02	0.03	-0.03	0.01
7×7	0.49	0.21	0.27	0.02	0.08	0.01	-0.04	0.10	0.01
8×8	-0.10	0.03	0.08	0.01	-0.08	0.01	0.08	-0.04	-0.01
9×9	0.52	0.23	0.22	0.02	0.05	0.00	-0.03	0.06	-0.01

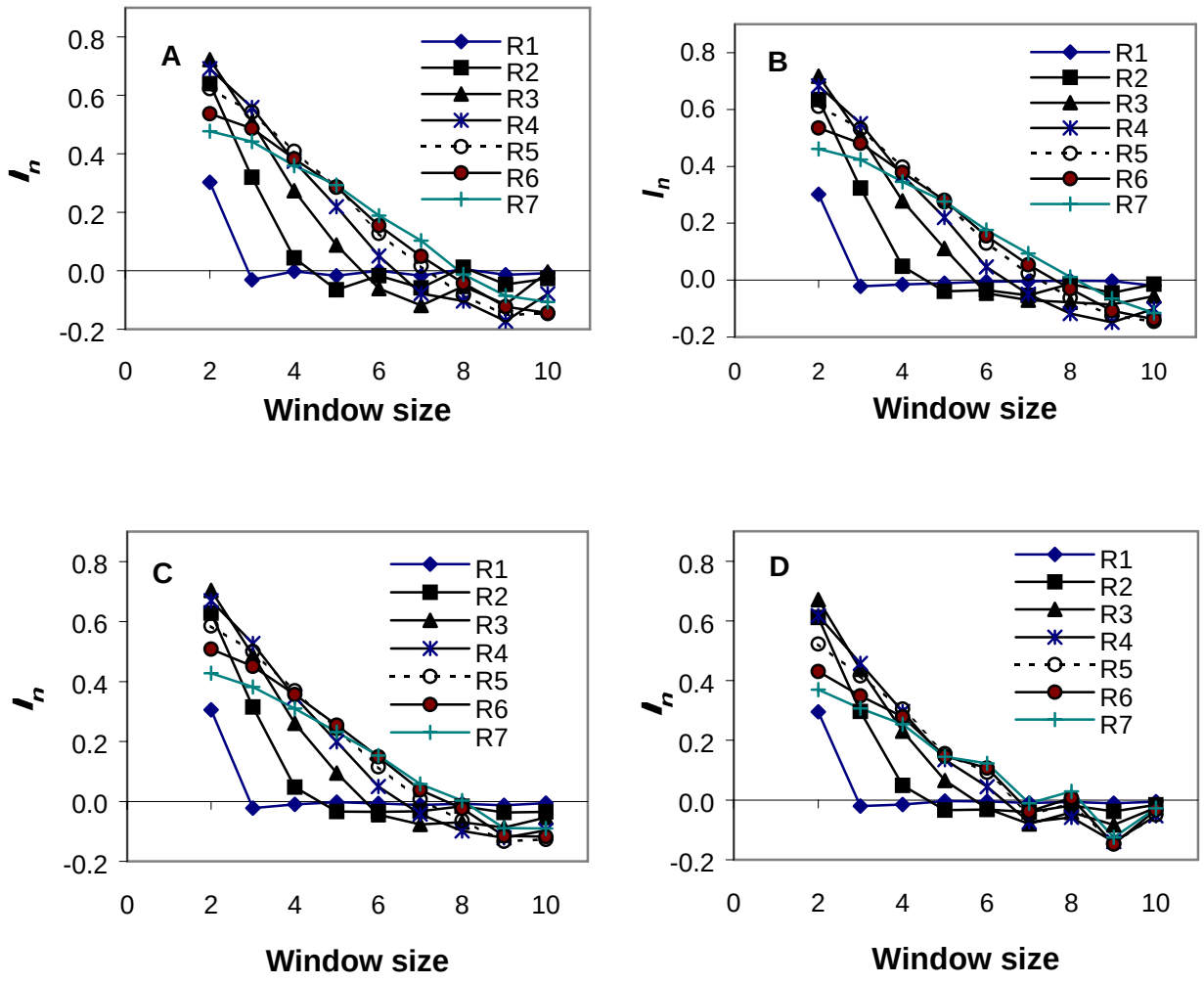
[a] Positive values indicate positive association, negative value indicate negative association, and zero values indicate random distribution, independence.

Figure Legends

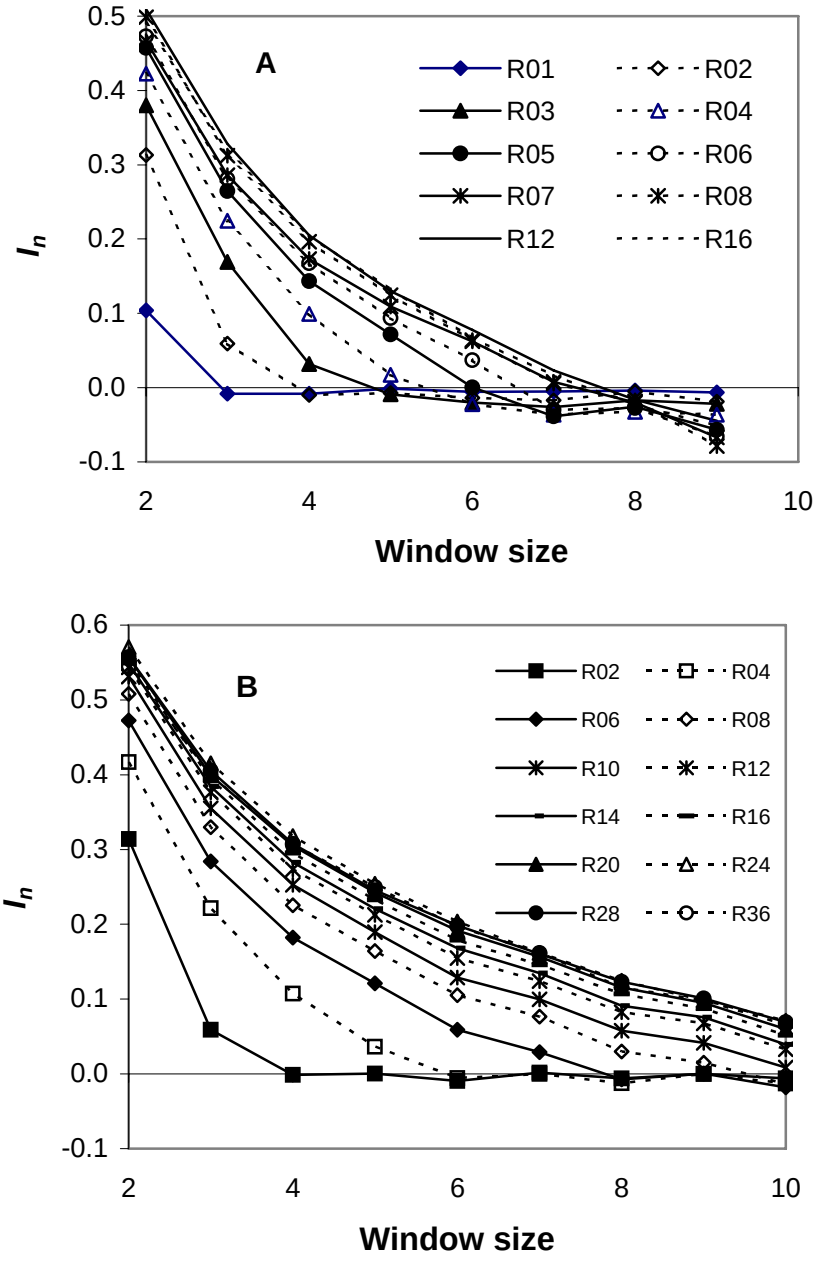
Fig. 1. Spatial analysis of the aggregated binary data using the VMWA method. Data were artificially generated using a Neyman-Scott cluster process with 1 (**A**), 2 (**B**), 4 (**C**), and 10 (**D**) disease foci. Each index (I_n) in the figure was the average in four hundred simulations. R1 to R7 in the figure denote the influence range R=1 to 7 used in the simulation. In each simulation, disease foci were first randomly initiated in a 20×20 field. Then, whether a plant became infected from each focus was determined (simulating 200 times of infection) according to the distance (r) from the focus: if the distance was farther than the influence range R or the plant has already been infected, then no action was taken, otherwise the plant had the potential of getting infected at a probability of $p = e^{-r/R}/2\pi rR$.

Fig. 2. Spatial analysis of the aggregated count data with VMWA method. Data were artificially generated in a field containing 20×20 (**A**) and 40×40 (**B**) sample units using a Neyman-Scott cluster process with 10 disease foci and different influence ranges 1 to 36 (denoted as R01 to R36). Each index (I_n) in the figure was the average in four hundred simulations. In each simulation, 10 disease foci were randomly fixed, and disease (unlimited) at each sample unit was then simulated around each disease focus.

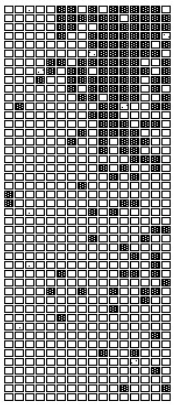
Fig. 3. Distribution of pepper plants infected by *Verticillium dahliae* in historical field surveys (Bhat et al. 2003) at three sites in Fields 1-3. H = healthy plant, D = diseased plant, and NA = missing.



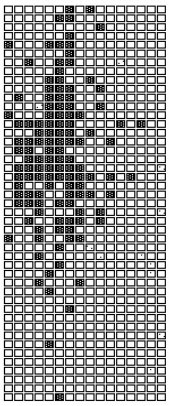
Wu et al., Fig. 1


A
R01
R02
R03
R04
R05
R06
R07
R08
R12
R16
I_n
Window size
B
R02
R04
R06
R08
R10
R12
R14
R16
R20
R24
R28
R36
I_n
Window size

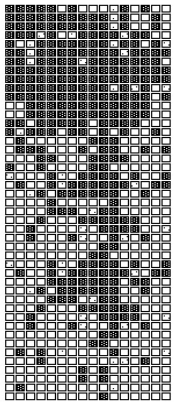
Wu et al., Fig. 2



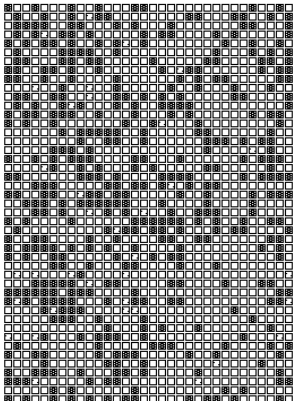
Field 1A



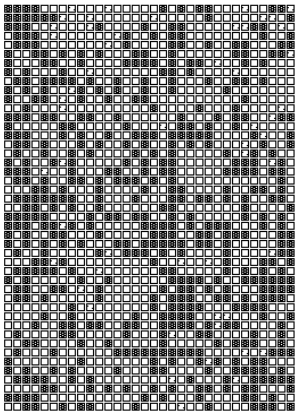
Field 1B



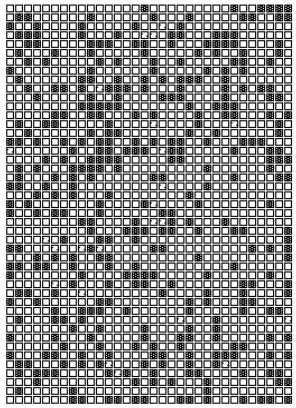
Field 1C



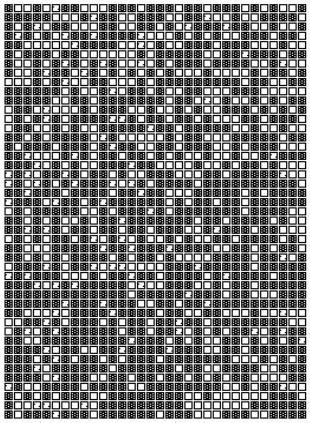
Field 2A



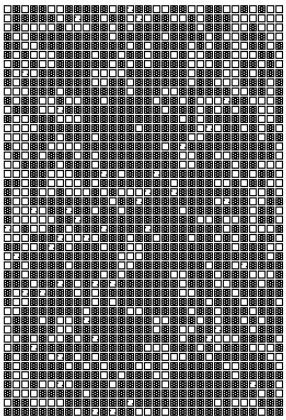
Field 2B



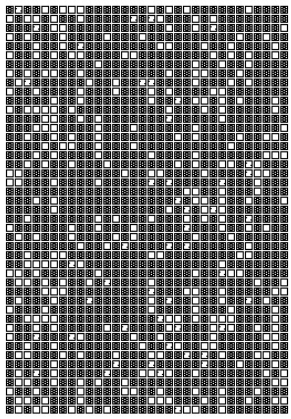
Field 2C



Field 3A



Field 3B



Field 3C

H D NA

Wu et al., Fig. 3