European Journal of Plant Pathology **101**: 133–137, 1995.

Analysis of the spatial distribution of rust-infected leek plants with the Black-White join-count statistic

P. D. de Jong[1] and J. de Bree[2]
[1] *Research Institute for Plant Protection, P.O. Box 9060, NL-6700 GW Wageningen, The Netherlands*
[2] *Agricultural Mathematics Group, P.O. Box 100, NL-6700 AC, The Netherlands*

Accepted 23 September 1994

Key words: leek, *Allium porrum*, leek rust, *Puccinia allii*, spatial distribution, Black-White join-count statistic

Abstract

Leek rust, caused by *Puccinia allii* Rudolphi, is an important disease of leek (*Allium porrum* L.) in the Netherlands. For the development of a practical sampling method for early detection of leek rust in commercial fields, information on the spatial distribution of the disease is necessary. In this study, the spatial distribution of diseased plants during three naturally occurring epidemics of leek rust was observed. The observations were analysed with the Black-White join-count statistic. The spatial distribution of rust-infected leek plants was different for each of the three epidemics, ranging from random to highly clustered.These results show, that in the development of a practical sampling method for detection of leek rust, it is necessary to take into consideration a possibly clustered distribution of diseased plants.

Introduction

Rust of leek, caused by *Puccinia allii* Rudolphi, is an important disease of leek (*Allium porrum* L.) in The Netherlands. Decline in leek quality due to light infestation may cause severe economic loss to the grower. Therefore, chemical control of the disease must be started in an early stage of the infection. However, applying fungicides to a crop in the absence of the disease, is undesirable both from an environmental and an economic point of view. The most direct method of assessing the presence of the disease in the crop is sampling. For the development of a practical sampling method for early detection of leek rust in commercial fields, general information on the possible spatial distribution of the disease is necessary.

A straightforward and much used method for analysis of the spatial distribution of diseased plants is to assess presence or absence of the disease in all plants in a row or in a transect across rows and to use the ordinary runs statistic to test randomness of infected plants. [Campbell et al., 1984; Madden et al., 1982, 1987a, 1987b; Jeger et al., 1987]. The ordinary runs statistic is a particular join-count statistic. Join-count statistics are widely applied measures of spatial autocorrelation for nominal data [Cliff and Ord, 1981]. Another join-count statistic, the Black-White statistic (BW) has a flexibility not present in the ordinary runs statistic that permits application to a two-dimensional array of plants (e.g. a set of rows) and to detect spatial autocorrelation not only between nearest neighbours, but also between second and higher order neighbours.

Gray et al. [1986] proposed a method for spatial analysis of nominal data very similar to the use of join-count statistics, that they called two-dimensional distance class analysis, without referring to the join-count statistics. The two-dimensional distance class analysis as presented by Gray et al. [1986] requires computer simulation to obtain the probability distribution of the test statistic, whereas the probability distributions for the join-count statistics can be approximated with standard distribution functions [Cliff and Ord, 1981].

In this study, the BW join-count statistic is used to detect non-randomness in the spatial distribution of rust-infected leek plants in naturally occurring epidemics.

Materials and methods

Fields observations

The spatial distribution of diseased plants was assessed during three naturally occurring leek rust epidemics in fields nearby Wageningen: on 18 September 1991 in a 60 m × 45 m field, on 3 July 1992 in a 30 m × 35 m field and between 8 July and 6 August 1993 in a 30 m × 50 m field. The fields received treatment with pyrethroids to control insect damage but were not treated with fungicides. All fields were row planted with the highly susceptible cultivar Albana. Details on the fields are given in Table 1. Presence or absence of rust was assessed on consecutive plants within selected rows. Distances between selected rows were 5 rows in 1991 and 10 rows in 1992 and 1993. In 1992 and 1993 plants in transects across rows were also assessed; the distance between the transects being 5 m and 10 m respectively. In 1993, observations on the different dates were made on the same rows and approximately the same transects. The number of observed rows, transects and plants per row or transect are given in Table 1.

Table 1. Dimensions of the fields, between and within row planting distances and the number of observed rows, transects and plants per row or transect for the three observed epidemics

	1991	1992	1993
Dimensions (mxm)	60 × 45	30 × 35	30 × 50
Row length (m)	45	35	50
Between row planting distance (m)	0.50	0.50	0.50
Within row planting distance (m)	0.16	0.16	0.16
Number of observed rows	22	3	5
Number of plants per row	270	210	240
Number of observed transects	–	4	5
Number of plants per transect	–	56	54

Statistical analysis

Randomness in the one-dimensional spatial distribution of diseased plants was tested with the BW join-count statistic [Cliff and Ord, 1981]. The mathematical definition of the BW statistic is:

$$BW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}(x_i - x_j)^2 \qquad (1)$$

in which $x_i = 1$ if plant i is infected and $x_i = 0$ if plant i is healthy ($i = 1 \ldots n$), and $\{\delta_{ij};\ i, j = 1 \ldots n\}$ form a binary connection matrix, in the sense that $\delta_{ij} = 1$ if the plants i and j are defined to be neighbours, and $\delta_{ij} = 0$ otherwise (including $\delta_{ii} = 0$). In formula (1), the sum of squares is taken to avoid negative values and the factor $^1/_2$ is added, because every possible pair of plants occurs twice in the summation. Thus, the BW statistic is the number of times, that a healthy plant is neighboured by an infected plant. It must be emphasized, that with the BW statistic, neighbourship is merely a matter of definition. For calculation of the BW statistic for spatial autocorrelation between k^{th} order neighbouring plants, plants are regarded as neighbours if there are $k - 1$ plants in between.

The data presented in this study, consist of sets of rows (or transects) with no spatial connection between the rows (or transects). When calculating the BW statistic for first order neighbours, $\delta_{ij} = 1$ if plant i and plant j are in the same row (or transect) and there are no other plants in between and $\delta_{ij} = 0$ otherwise. For second order neighbours, $\delta_{ij} = 1$ if plant i and plant j are in the same row (or transect) and there is only one plant in between and $\delta_{ij} = 0$ otherwise. For third order neighbours, $\delta_{ij} = 1$ if plant i and plant j are in the same row (or transect) and there are two other plants in between and $\delta_{ij} = 0$ otherwise, etc.

The BW statistic is asymptotically normally distributed under the null hypothesis of no spatial autocorrelation. An approximate test of significance is therefore provided by evaluating the BW statistic as a standard normal deviate, as suggested by Cliff and Ord [1981]. The first two moments of the BW statistic are then used to specify mean (μ) and variance (σ^2) of the normal distribution. Computational formulae for the first two moments are given in the appendix. The appropriateness of the use of the normal distribution as an approximation of the probability distribution of the BW statistic has been confirmed by the results of Monte Carlo simulation, using 5 rows with each 50 plants and combinations of disease frequency of 5% or higher with order of neighbouring ranging from one to 25.

For every set of rows (or transects), BW statistics were calculated for any order of neighbouring less or equal to half the length of the row (or transect). With even higher order of neighbouring, the normal approximation of the probability distribution of the BW statistic appears to be less reliable. A left-sided test was applied to detect low values of the BW statistic, that indicate positive spatial autocorrelation between diseased plants (i.e. clustering) and a significance level of 0.05 was used. Characterization of the spatial distribution of diseased plants in a set of rows (or transects) as random or clustered was based on the outcome of the BW statistic for first order neighbouring. The BW statistics for higher order neighbouring were used to obtain an indication for the extent of the clustering.

Results

For all three years, observations were made when the number of pustules per infected plant was low and infection could only be perceived by close observation. Despite the low numbers of pustules per plant, there was a considerable percentage of the plants infected and diseased plants could be found throughout the fields.

The results of the analysis of the spatial distribution of diseased plants are given in Table 2. The frequency of infected plants was 30.4% for the observation on the '91 epidemic and 25.5% for the '92 epidemic, as estimated from the row observations. For the '93 epidemic, the frequency of infected plants increased from 6.2 % to 81.3 %, between 8 July and 6 August, as estimated from the row observations.

In the '91 epidemic, positive spatial autocorrelation within rows was significant for every order neighbouring tested, indicating a strongly clustered distribution of diseased plants. In the '92 epidemic, there was no significant positive autocorrelation within transects and significant positive spatial autocorrelation within rows was

Table 2. Frequency of diseased plants in the sample (f), the number of tests done (t), the number of cases of significant positive spatial autocorrelation ($\alpha = 0.05$) in the set of tests (s) and the order of neighbouring (k) over which significant positive spatial autocorrelation occurred

			f(%)	t	s	k
1991	18 Sep	rows	30.4	135	135	1–135
1992	3 Jul	rows	25.5	105	6	13, 21, 35, 36, 43, 96
		trans	24.1	28	–	–
1993	8 Jul	rows	6.2	120	14	1–4, 6, 13, 14, 21, 22, 67, 73, 81, 115, 117
		trans	7.0	27	2	1, 3
	12 Jul	rows	6.8	120	21	1–5, 7, 9, 11–14, 16, 18–22, 24, 25, 27, 29
		trans	12.2	27	3	1, 2, 16
	16 Jul	rows	7.8	120	40	1–11, 16–19, 21, 25, 27, 28, 38, 71, 74, 92, 94–100, 102–104, 106, 108, 113–115, 119, 120
	22 Jul	rows	16.0	120	19	1–10, 12, 14, 35, 40, 82, 90, 94, 95
		trans	14.4	27	5	1–5
	31 Jul	rows	39.3	120	28	1–24, 27, 96, 106, 117
		trans	46.3	27	11	1, 2, 4–9, 12, 13, 20
	6 Aug	rows	81.3	120	65	1–23, 25–30, 32–37, 40, 41, 43–50, 52, 55, 61, 63–65, 67, 73, 76, 81, 83, 84, 86, 87, 90, 92, 93, 103, 119, 120
		trans	86.3	27	4	1, 6, 7, 12

found for 13th, 21st, 35th, 36th, 43rd and 96th order neighbouring, but not for first order neighbouring. Therefore the spatial distribution was considered to be random both for rows and for transects. In the '93 epidemic, the number of cases of significant positive spatial autocorrelation within the sets of rows increased from 14 on 8 July to 65 on 22 July, always including first order neighbouring, and the spatial distribution of diseased plants along rows was therefore clearly clustered. The spatial distribution of diseased plants along the transects was also clustered for the whole period of observation.

Discussion

The fields analysed in this study consisted of at least 10.000 plants. Therefore the fields could not be assessed completely, but had to be sampled. The numbers of rows and transects included in the observations in this study were the largest possible, given the limitations in man power. The rows and transects were chosen to be evenly distributed over the fields. For efficient use of the BW statistic for characterization of spatial patterns, more information is needed on the minimal sample size that still gives representative results.

Although the BW join-count statistic permits analysis of two-dimensional spatial data, the use of the statistic was restricted in this study to one-dimensional analysis, due to the character of the fields. Gray et al. [1986] present examples of two-dimensional spatial analysis, but their examples are limited to small latices of at most 300 elements. An alternative way of sampling, i.e. assessment of all plants in a number of square patches in the field, will allow for a two-dimensional analysis, but will restrict the analysis to a low order of neighbouring. An additional problem is, that the leek plants were row-planted and not standing in an exact lattice.

Simultaneous statistical inference using BW statistics for several order of neighbouring poses intractable problems because of mutual dependence of the individual BW statistics. Therefore, no test for randomness using several BW statistics for different order neighbouring simultaneously can be constructed and characterization of the spatial distribution as random or clustered has to be based on the BW statistic for one particular order of neighbouring only, the first order neighbouring being a logical choice. The BW statistics for higher order neighbouring provide an indication for the extent of the clustering. The presence of significant positive spatial auto-correlation for first till k^{th} order neighbouring can be interpreted as an indication of cluster size. However, the significance level of such complex conclusions remains undefined, due to the already mentioned mutual dependence of the individual BW statistics.

The strong clustering found during the 1991 epidemic can be explained by the nature of the source of inoculum. In 1991, the sampled field was infected from a nearby source, consisting of a 5 m × 7 m leek plot at a distance of 5 m from the border of the field. It was in the vicinity of this point source, that there was a very large cluster of infected plants, reflected in the analysis by the presence of significant spatial autocorrelation for every order of neighbouring tested. Such a point source may occur in a commercial setting, when part of a previously planted neighbouring crop is left unharvested.

In 1992 and 1993, the sampled fields were infected from a source field measuring 15 m × 40 m at a distance of approximately 25 m and planted one month earlier than the observed fields. As the sources of inoculum were so identical for 1992 and 1993, other factors must have caused the difference in spatial distribution found in the 1992 and the 1993 epidemic. As leek rust is wind-borne, wind-direction and wind-speed are other important factors determining the spatial distribution of diseased plants. However, the data presented in this study are not detailed enough to draw any conclusion on the effect of wind-direction and wind-speed on the spatial distribution in the observed epidemics.

From the results of this study, it is clear, that the spatial distribution of rust-infected leek plants is highly variable, depending among other factors on the nature of the inoculum source.

In a forthcoming paper, the effect of the spatial distribution of the diseased plants on the efficacy of different sampling patterns for detection of leek rust will be discussed, and a practical sampling pattern is proposed with a high efficacy even under strongly clustered distribution of diseased plants.

Acknowledgements

Thanks are due to Dr. Ir. R. Daamen, Prof. Dr. Ir. R. Rabbinge, two anonymous reviewers and many others, who critically reviewed the manuscript.

Appendix

Mean and variance of the BW statistic

Under assumption of nonfree sampling, i.e. situation that the probability of an individual plant to be infected is not known from an independent sample:

$$\mu = \frac{S_0 n_d n_h}{n^{(2)}} \tag{2}$$

$$\sigma^2 = \frac{1}{4}\left[\frac{2S_1 n_d n_h}{n^{(2)}} + \frac{(S_2 - 2S_1)n_d n_h(n_d + n_h - 2)}{n^{(3)}} + \frac{4(S_0^2 + S_1 - S_2)n_d^{(2)} n_h^{(2)}}{n^{(4)}}\right] - \mu^2 \tag{3}$$

where n_d is the number of diseased plants, n_h is the number of healthy plants, n is the total number of plants and

$$n^{(b)} = n(n-1)\ldots(n-b+1) \tag{4}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_{ij} \tag{5}$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\delta_{ij} + \delta_{ji})^2 \tag{6}$$

$$S_2 = \sum_{i=1}^{n} (\delta_{i.} + \delta_{.i})^2 \tag{7}$$

where

$$\delta_{i.} = \sum_{j=1}^{n} \delta_{ij} \tag{8}$$

$$\delta_{.i} = \sum_{j=1}^{n} \delta_{ji} \tag{9}$$

For a set of rows with no spatial connection between the rows, it follows from formulae (1) to (9), that S_0, S_1 and S_2 can be calculated as:

$$S_0 = 2N(M - K) \tag{10}$$

$$S_1 = 4N(M - K) \tag{11}$$

$$S_2 = N(16(M - 2K) + 8K) \tag{12}$$

where N is the number of rows, M is the number of plants per row and K is the order of neighbouring ($K \leq {}^1/_2 M$).

Source of formulae (1) to (9) is Cliff and Ord (1981).

References

Campbell CL, Jacobi WR, Powell NT and Main CE (1984) Analysis of disease progression and the randomness of occurrence of infected plants during tobacco black shank epidemics. Phytopathology 74: 230–235

Cliff AD and Ord JK (1981) Spatial processes. Models & applications. Pion, London, 266 p

Gray SM, Moyer JW and Bloomfield P (1986) Two-dimensional distance class model for quantitative description of virus-infected plant distribution lattices. Phytopathology 76: 243–248

Jeger MJ, Kenerley CM, Gerik TJ and Koch DO (1987) Spatial dynamics of *Phytomatotrichum* root rot in row crops in the Blackland region of north central Texas. Phytopathology 77: 1647–1656

Madden LV, Louie R, Abt JJ and Knoke JK (1982) Evaluation of tests for randomness of infected plants. Phytopathology 72: 195–198

Madden LV, Louie R and Knoke JK (1987a) Temporal and spatial analysis of maize dwarf mosaic epidemics. Phytopathology 77: 148–156

Madden LV, Pirone TP and Raccah B (1987b) analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77: 1409–1417