UNIVERSAL WISER
PUBLISHER

Research Article

Spatial Characteristics of the Fungus Powdery Mildew (*Erysiphe neolycopersici*) on Tomatoes and its Spread in Industrial Greenhouses

Anastasia Sokolidi[1*] , **Richard Webster**[1] , **Alice E. Milne**[1], **Martin Bielik**[2], **Phil Morley**[2], **John P. Clarkson**[3], **Jon S. West**[1]

[1]Rothamsted Research, Harpenden AL5 2JQ, United Kingdom
[2]APS Produce, Arreton, Isle of Wight PO30 3AR, United Kingdom
[3]Warwick University, Coventry CV4 7AL, United Kingdom
 E-mail: anastasia.sokolidi@rothamsted.ac.uk

Abstract: In regions with cool temperate climates, tomatoes are grown on an industrial scale in large greenhouses. There the crops are susceptible to infection by powdery mildew, the fungus *Erysiphe neolycopersici*, which is introduced largely as fungal spores from outside the greenhouses and spread by wind within them. We have monitored the spread of the disease and mapped its distribution in four commercial greenhouses throughout the growing season to understand its aetiology. We modelled the patterns of infection geostatistically, each comprising a deterministic long-range trend plus a short-range spatially correlated random residual. We identified three main kinds of pattern; one consisted of a constant plus a spatially correlated residual, second comprised a linear trend throughout the greenhouse plus a correlated random residual, and third, the trend had the form of a bell akin to a Gaussian surface plus, again, a correlated random residual. Here, we show three examples of these distributions and the detail of their geostatistical analysis using both the traditional method of moments (MoM) estimation of variograms and residual maximum likelihood (REML) to separate the deterministic and random components. The analytical modelling is followed by ordinary punctual kriging in the first case, by universal kriging in the second, and by regression kriging in the third case to display the infection as isarithmic ('contour') maps. We interpret the first form of distribution as arising from numerous foci as spores landed on the leaves from various sources spread by air currents and the movement of workers along the paths through the greenhouse. In the second case, the disease seemed to have spread from an infection introduced through the main door in one corner of the greenhouse and spread from there by the workers and air currents. The third infection arose near the centre of the greenhouse by the main path and spread outwards from there. In all three examples, the main pathways seemed important routes along which the fungus spread.

Keywords: tomatoes, greenhouses, powdery mildew, *Erysiphe neolycopersici*, geostatistics, kriging

1. Introduction

Tomatoes are an important crop in many countries and are grown commercially on an industrial scale. In regions with cool temperate climates, such as the United Kingdom (UK), outdoor production is limited to a short summer season. To extend the season, the tomatoes are instead grown in polytunnels and greenhouses. The greenhouses are huge, typically 1 hectare (ha) in extent, and in many instances are built into larger blocks separated by plastic or glass

barriers to make effectively 1 ha individual houses. In the UK, the season begins when the tomatoes are placed in the greenhouses as seedlings from a nursery.

As the plants grow, they become susceptible to infection by the fungus, *Erysiphe neolycopersici*, due to the increase in leaf and stem area. The initial symptoms of the disease appear as small white spots on the leaves. These spots later develop into larger patches covered with the fungus's spores, which give them a white powdery appearance. Figure 1 is a typical example of the fungus on tomato leaves. If the plants are left untreated, the leaves eventually turn yellow and die, and the fruit is of poor quality with a smaller yield [1]. The disease tends to be most prevalent in summer when the plants are at their peak of growth. The disease can be halted by treatment with fungicides. Growers consider that prevention is better than cure, however, and with this aim they spray their crops with fungicides as prophylactics at regular intervals.

Figure 1. Powdery mildew on tomato

Infection by *E. neolycopersici* begins when spores land on the plants. Fungal hyphae grow from each spore into the plants. The fungus then colonizes the surface of the leaf whilst producing its fruiting bodies, conidiophores, bearing more conidia, which are readily detached by wind or mechanical disturbance when they are mature; and the cycle begins again from many more foci when the spores land on the plants. It takes only about 10 days from the initial infection to the first visible signs of the disease [1]. Infection within any one greenhouse seems to be introduced from elsewhere by wind through vents and doors. Workers can introduce the disease as they move along pathways, most frequently during the peak season to tend and harvest the crop.

There have been many investigations of the distributions of weeds, plant parasites and crop diseases in the field with attempts to model them statistically and map them with a view to identifying the processes that have brought them about. Recent examples in which the most up-to-date methods of spatial analysis have been applied include bacterial blight in rice [2], virus disease in tomatoes [3], rust in coffee [4], crown atrophy in coconut [5] and weed infestation in cereal crops [6, 7]. The most relevant recent example in the context of our investigation is that by Liu et al. [8] on microclimatic conditions combined with theoretical disease spread in greenhouses. Earlier, Boulard et al. [9, 10] investigated the role of airflow in greenhouses and the exchange of air from outside of them on the spread of a fungal disease of roses, and they combined it with the fluid dynamics of the airflow and the movements of spores within the air.

Combining these dynamics of infection and the complexity of the spatial distribution of the evident symptoms has

proved problematic. We are investigating powdery mildew, *E. neolycopersici*, in large commercial greenhouses. Our aim is to assess its severity, map its distributions within the greenhouses and understand the origins of infection and its spread. As far as we know, this has not been done before. Only the investigators mentioned in references [2, 8, 9] seem to have approached the problem, though with other diseases. Here, we describe quantitatively the spatial distribution of the disease at particular times, taking into account its evident spatial correlation, and map it in individual greenhouses. We describe the geostatistical techniques we are using to model the spatial correlation and then to interpolate by kriging to produce maps.

2. Methods

2.1 *Monitoring*

We monitored the fungus, *E. neolycopersici*, in four commercial greenhouses, each of ≈ 1 ha, on the Isle of Wight from June to the end of the crop in November. To show the nature of the spatial variation in the disease, we selected the observations from two of the greenhouses, namely H11 and H13, on three occasions only in 2021, which were 22 July (occasion 2, OB2), 19 August (occasion 4, OB4) and 2 September (occasion 5, OB5). The severity of the disease was scored from 0 to 9 in accordance with International Plant Genetic Resources Institute (IPGRI) [11], the principal points on which are as follows:

0: Very low (no visible signs of infection);
3: Low (small patches < 2 cm across, little sporulation and mycelium);
5: Medium (approximately 50% of the leaves have visible symptoms of disease);
7: High (large patches affecting ≈ 70% of the leaves and abundant mycelium).

Severities between these points were scored with intermediate values. The disease was scored along rows every 4.5 m. The distance between rows was 1.5 m and every 8th row was assessed. The greenhouses have paths through their middles, approximately 4 m wide, for the movement of heavy machinery and produce. The pathway is denoted in Figure 2 with a blue dashed line going through the greenhouse at 42 m on the eastings axis. The tomato varieties differed in the two greenhouses. In greenhouse H13, the sole variety was Piccolo, which is highly susceptible to the fungus. Greenhouse H11 grew five varieties, only one of which, Graziano, has any resistance to the disease. Tomato plants were planted at a density of 1 m apart.
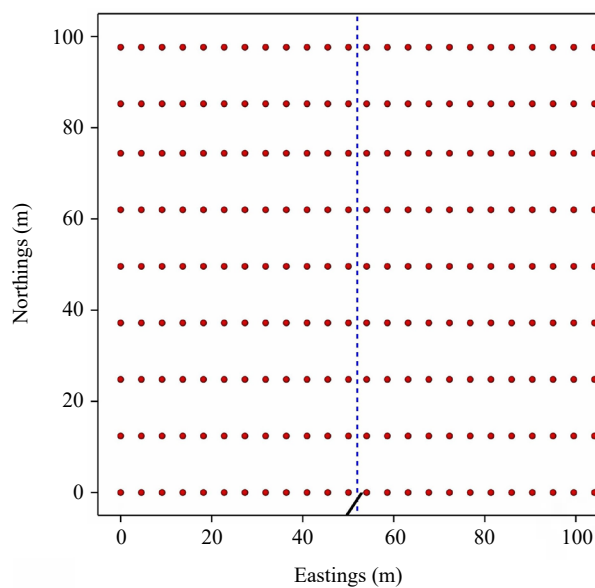

Figure 2. The sampling grid of disease in the greenhouses H13 and H11 (the short diagonal line at the bottom of the greenhouse indicates the door)

3. Implementations and results

3.1 *Data summary*

Figures 3a, 4a and 5a show the scores, the data, as 'bubble plots'; they are respectively for greenhouse H13 on OB2, greenhouse H13 on OB5 and greenhouse H11 on OB4. The diameters of the 'bubbles' are proportional to the scores. We mention here that the two upper rows of bubbles in Figure 5a are the scores on the somewhat resistant variety Graziano. The dashed blue lines running from top to bottom of the bubble plots mark the 4 m wide paths.

The small bubbles outnumber the large ones by far; there are large proportions of zeros, i.e. no infection, and progressively fewer sampling points as scores increased from 1 to maxima in the range 5 to 8; the distributions of the scores are strongly positively skewed - see Table 1. To stabilise the variances for statistical analysis we transformed the scores to common logarithms as $\log_{10}(\text{score} + 1)$. These values thus became the data for all subsequent analyses. Table 1 summarises them.

Table 1. Data summaries

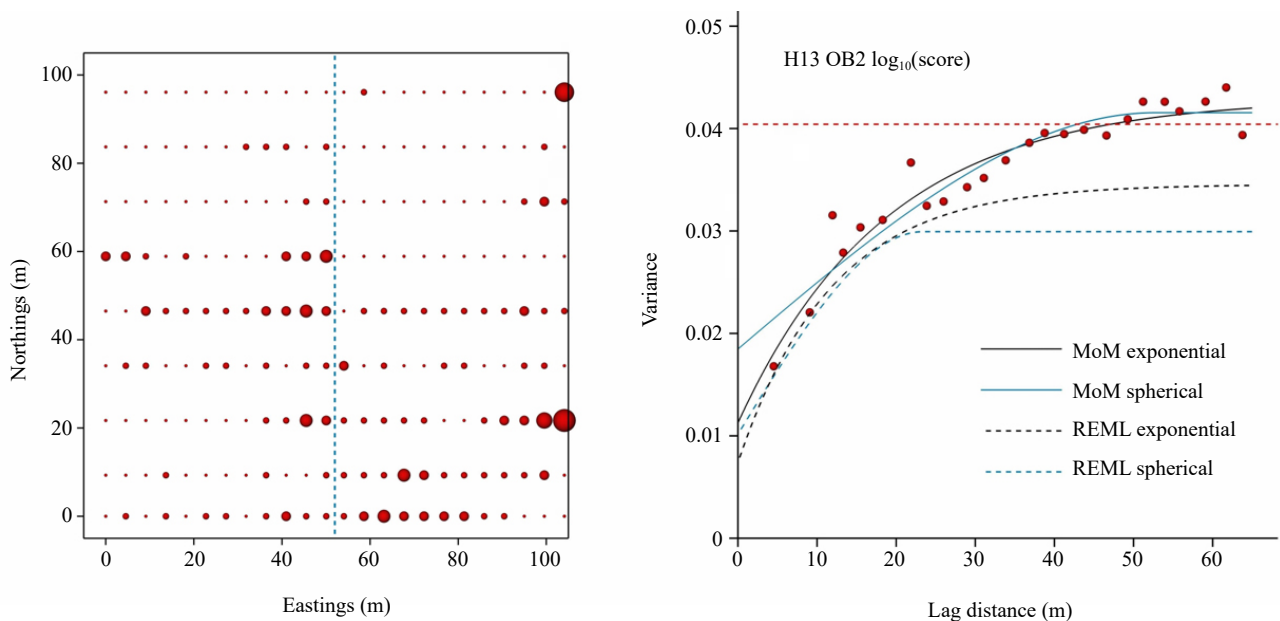| Greenhouse	Scores			Log_{10}transforms		
	Mean	Maximum	Skew	Mean	Maximum	Skew
H13 OB2	0.657	6	2.10	0.168	0.04042	0.77
H13 OB5	1.68	8	0.95	0.357	0.06481	−0.17
H13 OB5 residuals				0	0.02844	0.16
H11 OB4	0.556	5	1.97	0.132	0.04387	1.26
H11 OB4 residuals				0	0.01300	0.96

Figure 3. (a) Bubble plot of infection in greenhouse H13 on OB2; (b) Experimental variogram and fitted functions
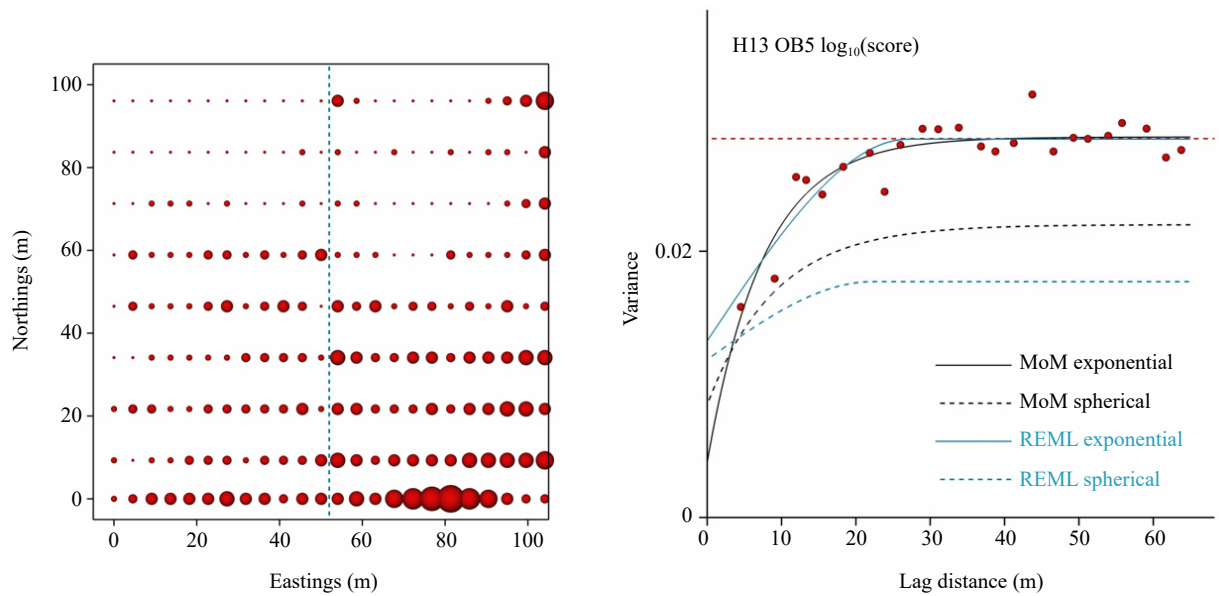*Note: MoM = method of moments; REML = residual maximum likelihood

Figure 4. (a) Bubble plot of infection in greenhouse H13 on OB5; (b) Experimental variogram of MoM residuals and fitted function
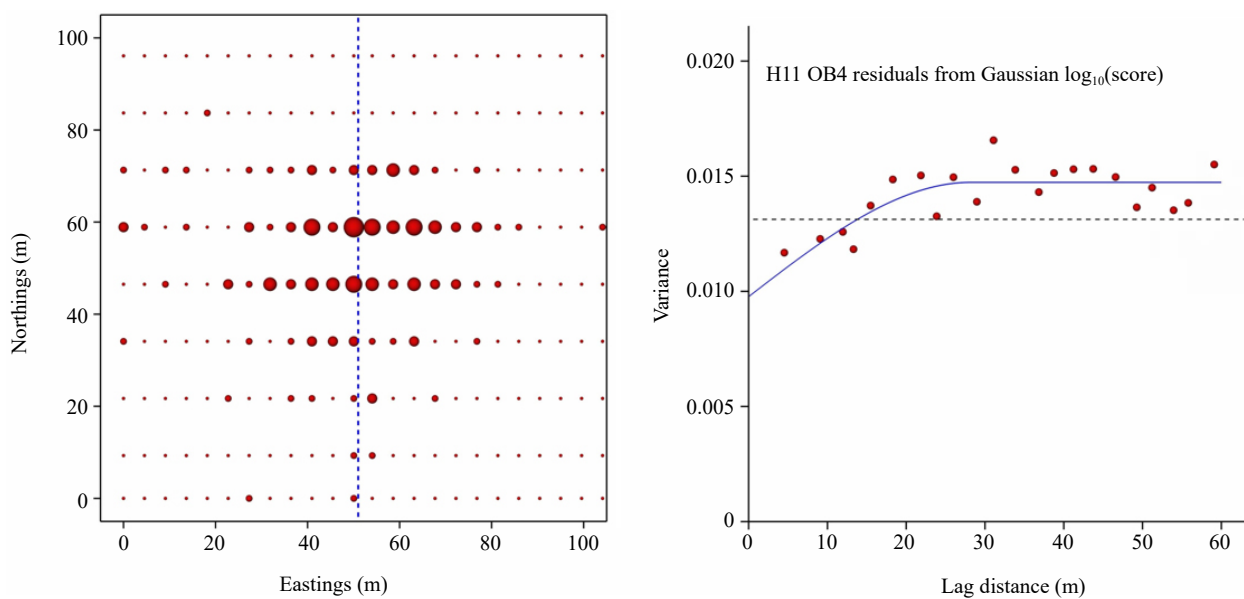*Note: MoM = method of moments; REML = residual maximum likelihood

Figure 5. (a) Bubble plot of infection in greenhouse H11 on OB4; (b) Experimental variogram of MoM residuals and fitted function and REML estimate of the variogram
*Note: MoM = method of moments; REML = residual maximum likelihood

3.2 Geostatistical modelling

Figures 3a, 4a and 5a show general patterns of infection in the greenhouses, including much point-to-point fluctuation. Following Cressie [12] we can express this combination as shown in Equation (1):

$$Z(\mathbf{x}) = \text{long-range variation} + \text{short-range variation}, \tag{1}$$

in which $\mathbf{x} \equiv \{x_1, x_2\}$ denotes the spatial coordinates of any position in a greenhouse in two dimensions and $Z(\mathbf{x})$, a

random variable, is the score there. To understand the aetiology of the infection we need to consider both terms on the right-hand side of Equation (1). Figure 3a shows a fairly uniform spread of the disease about which the scores fluctuate over short distances. In that equation, the long-range variation would be represented by a constant. Figure 4a shows a trend extending from one corner of the greenhouse, bottom right in the figure, into the rest of the greenhouse. Figure 5a has a maximum near the centre of the greenhouse from which the infection appears to have spread and which diminishes with increasing distance from the maximum. In both of these, there is a long-range component of the variation that is clearly not constant. The scores displayed in Figure 3a are evidently correlated spatially. So too is the short-range variation in Figures 4a and 5a once the long-range variation has been filtered out.

To display the infections simply, we wanted isarithmic ('contour') maps showing the main patterns, taking into account the short-range correlation in the data. For this, we interpolated logarithms of the scores on fine grids by punctual kriging and threaded isarithms through the grids. We, therefore, needed formal models of Equation (1) from which to formulate and estimate the variograms. We treated the long-range component of variation as deterministic, a fixed effect, and the short-range component as an autocorrelated random residual from the trend. By modelling the variation in this way, we should be able both to map the variation and to understand the way infection spreads.

The example of H13 on OB2 illustrated in Figure 3a, is the simplest to model. As above, we treat the trend as constant and the residual as a spatially correlated intrinsically stationary random process as shown in Equation (2):

$$Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}). \tag{2}$$

Here μ is the mean of the process, and ε is a spatially correlated random variable with mean zero and variance σ^2. The variogram is then a sufficient expression of the correlation between all places $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ separated by the vector $\mathbf{h}$, the lag, in distance and direction. It is defined as shown in Equation (3):

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\right\}^2\right] \text{ for all } \mathbf{h}, \tag{3}$$

in which E denotes the expected value (of the squared difference).

Where there is an evident trend in fungal infection within the crops the means, μ, can no longer be treated as constant; the trend depends on $\mathbf{x}$, so that the underlying model of Equation (2) must be elaborated to Equation (4):

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + \varepsilon(\mathbf{x}). \tag{4}$$

The combination of linear trend with correlated residuals in Figure 4a for H13 on OB5 can be expressed as shown in Equation (5):

$$Z(\mathbf{x}) = \beta_0 + \beta_{1x1} + \beta_{2x2}\,\varepsilon(\mathbf{x}). \tag{5}$$

It is a standard model of regression in which β_1 and β_2 are coefficients of the spatial coordinates x_1 and x_2, β_0 is a constant, and $\varepsilon(\mathbf{x})$ is the residual. It is a mixed-effects model of the variation comprising the fixed effects of the β_i, $i = 0, 1, 2$, and the random ε with variogram

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left\{\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{h})\right\}^2\right] \text{ for all } \mathbf{h}, \tag{6}$$

The trend in Figure 5a with its peak near the centre of the greenhouse has a bell shape akin to a two-dimensional Gaussian surface. We modelled it as shown in Equation (7):

$$Z(\mathbf{x}) = \beta_0 + \beta_1 \frac{1}{2\pi\alpha_1\alpha_2}\exp\left[-\left\{\frac{(x_1 - u_1)^2}{\alpha_1^2} + \frac{(x_2 - u_2)^2}{\alpha_2^2}\right\}/2\right] + \varepsilon(\mathbf{x}), \tag{7}$$

in which u_1 and u_2 represent the position of the peak of the surface in the two dimensions and α_1 and α_2 are the distances between the peak and the points of inflexion in those dimensions, β_0 is a constant and β_1 is a coefficient.

3.2.1 *Estimating the variogram*

Traditional practice has been to estimate the variogram from observed values, $z(\mathbf{x}_i)$, $i = 1, 2, ...,$ by the method of moments (MoM). The formula is as shown in Equation (8):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{j=1}^{m(\mathbf{h})} \left\{ z(\mathbf{x}_j) - z(\mathbf{x}_j + \mathbf{h}) \right\}^2, \tag{8}$$

where m is the number of paired comparisons at lag $\mathbf{h}$. By incrementing $\mathbf{h}$ in steps, one obtains an ordered set of semivariances which constitute the experimental or sample variogram. To this one fits a plausible valid function, usually nowadays by non-linear least-squares approximation – non-linear because the most suitable functions such as the spherical and exponential are non-linear in their parameters.

An alternative means of estimation that has gained some popularity in recent years is residual maximum likelihood (REML). It takes into account all possible paired comparisons, whereas the MoM procedure tends to disregard comparisons at the largest lag distances because they are unreliable. Neither method is necessarily better than the other.

Where there is a trend the observed values in Equation (8) must be replaced by the residuals, $\varepsilon(\mathbf{x})$. Early practitioners obtained them by trend-surface analysis, i.e. ordinary least-squares regression on the spatial coordinates, and disregarded the bias in the variograms, which increased with increasing lag distances [12]. The estimated trend surface itself was no longer a minimum-variance estimate because of the failure to take into account the spatial correlation in the residuals. The introduction of REML has made good these shortcomings; it enables one to estimate both the coefficients of the trend and the parameters of the variogram of the residuals simultaneously and without bias [13, 14]. It is now best practice.

Unfortunately, REML can take into account only fixed effects that are linear combinations of the spatial coordinates; it cannot cope with non-linear ones such as the bell-shaped surface of Equation (7). We have therefore had to fall back on the earlier technique of separating the trend from the residuals and estimating their coefficients and parameters independently thereafter. We spell out the procedure below.

3.2.2 *Kriging*

Where data $z(\mathbf{x}_i)$, $i = 1, 2, ...,$ appear to be drawn from a stationary random process as represented by Equation (2) an ordinary kriged prediction $\hat{Z}(\mathbf{x}_0)$ at any new point $\mathbf{x}_0$ is a weighted average, as shown in Equation (9):

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=0}^{n} \lambda_i z(\mathbf{x}_i). \tag{9}$$

The weights, λ_i, $i = 1, 2, ..., n$ sum to 1 to avoid bias and are chosen to minimize the kriging error variance by solution of equations that incorporate the semi-variances from the variogram. The mathematics are well documented in the example by Webster & Oliver [14]. The number of points, n, in the summation may embrace all the data, but in practice, only a few data closest to the target carry sufficient weight to influence the result. The solution of the kriging system also provides the prediction error variance.

Where there is a trend, as represented by Equation (5), for example, kriging is somewhat more complex. Matheron [15] augmented the kriging system with coefficients of the trend in what he called 'universal kriging'. The semi-variances in the system are still drawn from the variogram of the random process, but that variogram is now that of the residuals from the trend, i.e. the $\varepsilon(\mathbf{x}_i)$, $i = 1, 2, ...,$ not that of the original data. What Matheron did not do was provide the means of estimating that variogram. Thanks to REML, we can now do that and incorporate semi-variances from it in the universal kriging systems.

For our third example with the Gaussian trend surface of Equation (7) we proceeded in stages as follows:

a. Fit a trend surface to the data by ordinary least-squares regression on the spatial coordinates as predictors;

b. Compute an experimental variogram of the residuals from the trend, and fit a plausible function to that variogram;

c. Interpolate values of the residuals on a fine grid by ordinary punctual kriging with semi-variances drawn from the variogram function;

d. Add to those kriged residuals predicted values from the trend-surface regression equation.

The whole process became known as regression kriging. The kriged predictions are unbiased, but the calculated prediction error variances underestimate the true error variances, often seriously, as Lark & Webster [16] discovered when re-analysing the data of Moffat et al. [17], who used regression kriging to map the depths of geological strata. The technique has come in for a lot of criticism on this account. Part of the reason is that the variogram itself is biased. Perhaps equally serious for our investigation is that the trend function might not be the best fit to the data because of the spatial correlation in the residuals. The situation is not necessarily as bad in practice as it might seem, however, because as Cressie [12] points out, the biases approach zero with increasing numbers of data. Further, by suitably weighting the $\hat{\gamma}(\mathbf{h})$ of Equation (8) when modelling the experimental variogram one can diminish the bias in the fitted function. Also, differences between the variograms computed from the residuals as described above and those from REML at short lag distances are small, and the semi-variances at these short distances are typically the only ones that enter the kriging equations.

With these considerations in mind and with 216 scores on each occasion, we have adopted the above procedure where the trend appeared bell-shaped.

3.3 *Direct application of the geostatistical models*

We shall report the results of our investigation in full elsewhere. Here we present the selection mentioned above to show the main forms of spatial variation in the fungal infection, how we modelled them geostatistically and the inferences we can draw from the modelling. Table 1 summarizes the data.

3.3.1 *Constant trend*

The bubble plot of the scores in greenhouse H13 on OB2 (Figure 3a) shows little evidence of a trend, and we have treated data as deriving from a stationary process represented by Equation (2). Table 1 summarises the data. The experimental variogram computed by the MoM, Equation (8), is shown by the red discs in Figure 3b. We fitted both exponential and spherical functions to the experimental variogram using the directive FITNONLINEAR in GenStat [12]; both fit well, accounting for 89% of the variance. Their equations are shown in Equation (10) and (11):

$$\text{Exponential} \qquad \gamma(h) = c_0 + c_1 \left\{ 1 - \exp\left(-\frac{h}{a} \right) \right\} \qquad \text{for } 0 < h$$
$$= 0 \text{ for } h = 0, \tag{10}$$

$$\text{Spherical} \qquad \gamma(h) = c_0 + c_1 \left\{ \frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r} \right)^3 \right\} \qquad \text{for } 0 < h < r$$
$$= c_0 + c_1 \text{ for } h \geq r$$
$$= 0 \text{ for } h = 0. \tag{11}$$

The parameters are c_0 the nugget variance, c_1 the sill variance of the correlated variance, and r and a are the distance parameters of the functions. Their values are listed in Table 2.

We show in addition the functions fitted by REML for comparison, and Table 3 lists the conventional leave-one-out cross-validation statistics of the differences between the true values and the kriged predictions when the points where the true values are omitted from the kriging systems:

$$ME = \frac{1}{N}\sum_{i=1}^{N} z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i),$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} \left\{ z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i) \right\}^2,$$

$$MSDR = \frac{1}{N}\sum_{i=1}^{N} \frac{\left\{ z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i) \right\}^2}{\hat{\sigma}_{OK}^2(\mathbf{x}_i)}.$$

In these equations N is the total number of observations, ME, the mean error, is the mean the difference between the observed values and the predicted ones, mean squared error (MSE) is the mean of the squared differences, and MSDR is the mean squared deviation ratio in which the squared differences are divided by the ordinary kriging error variances, $\hat{\sigma}_{OK}^2$.

The two functions for the MoM procedure have remarkably similar statistics; both have mean square deviation ratios, MSDRs, close to the ideal of 1. The models fitted by REML are not quite so good in that respect but would be acceptable in the absence of other information.

Figure 6a maps were made by ordinary punctual kriging of the data with the MoM variogram model. The kriging interval was 2.5 m, and the results were passed to MATLAB for the final display.
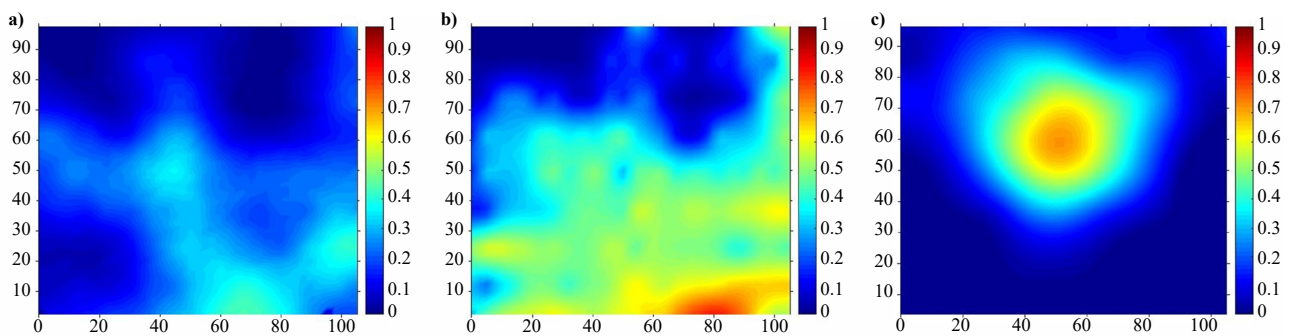

Figure 6. (a) Kriged map of infection in greenhouse H13 on OB2; (b) Kriged map of infection in greenhouse H13 on OB5; (c) Kriged map of infection in greenhouse H11 on OB4

Table 2. Variogram parameters

Greenhouse	Model	Parameters		
		Nugget	Sill	Distance (m)
H13 OB2	MoM exponential	0.01125	0.03174	18.77
H13 OB2	MoM spherical	0.01847	0.02308	52.73
H13 OB2	REML exponential	0.00739	0.02717	11.87
H13 OB2	REML spherical	0.01015	0.01979	23.44
H13 OB5 MoM residuals	Exponential	0.00415	0.02443	7.63
H13 OB5 REML residuals	Exponential	0.01354	0.008457	9.13
H11 OB4 MoM residuals	Spherical	0.00977	0.004964	28.18

Table 3. Cross-validation statistics

Greenhouse	Model	Mean	Mean deviation	Mean squared deviation	MSDR
H13 OB2	MoM exponential	0.168	−0.000108	0.023988	1.066
H13 OB2	MoM spherical	0.168	−0.000319	0.02577	0.984
H13 OB2	REML exponential	0.168	−0.000048	0.02371	1.176
H13 OB2	REML spherical	0.168	0.000351	0.02358	1.198
H13 OB5 MoM residuals	Exponential	0.357	0.001204	0.02138	1.149
H13 OB5 REML residuals	Exponential	0.357	0.001385	0.021254	1.169
H11 OB4 MoM residuals	Spherical	0.132	0.000644	0.01295	1.028

3.3.2 Linear trend

The scores in greenhouse H13 on OB5 showed a strong trend from north to south (Figure 4a). As above, we have two options for analysing the data geostatistically: the earlier technique of separating the trend from the residuals and analysing them separately and the current best procedure by REML. We have done both for comparison; first an ordinary least-squares (OLS) regression, Equation (5), and second REML. The coefficients were as follows:

$$\text{OLS} \quad \beta_0 = 0.510 \quad \beta_1 = 0.00215 \quad \beta_2 = -0.00565,$$
$$\text{REML} \quad \beta_0 = 0.368 \quad \beta_1 = 0.00301 \quad \beta_2 = -0.00471.$$

Figure 4b shows the experimental variogram computed by the MoM as the red discs to which we fitted an isotropic exponential function by non-linear least-squares approximation using the directive FITNONLINEAR in GenStat [18]. The function is the dashed line obtained by REML. Table 2 lists the parameter values and Table 3 the cross-validation statistics.

Figure 6b is the map made by universal punctual kriging of the data with the REML variogram of the residuals and the spatial coordinates. The kriging interval was again 2.5 m, and the results were transferred to MATLAB for the final display.

3.3.3 Gaussian trend

The scores in greenhouse H11 on OB4 (Figure 5a) exemplify the Gaussian trend with the form defined by Equation (7). We fitted the surface with non-linear least-squares approximation of the transformed scores again using the directive FITNONLINEAR in GenStat. The trend surface accounted for 70.4% of the variance with the following values:

$$\text{Peak positions} \quad u_1 = 48.9 \text{ m} \quad u_2 = 54.0 \text{ m}$$
$$\text{Distances to inflexions} \quad \alpha_1 = 19.3 \text{ m} \quad \alpha_2 = 16.2 \text{ m}$$
$$\text{Coefficients} \quad \beta_0 = 0.00922 \quad \beta_1 = 1450$$

We subtracted the trend from the data and analysed the residuals. We computed the experimental variogram of the residuals by the MoM, Equation (8), to which we fitted an isotropic spherical function. Figure 5b shows the resulting experimental variogram as red discs and the fitted function. Table 2 lists the estimates of the parameters, and Table 3 lists the cross-validation statistics. As expected, the ME is close to zero because kriging is unbiased. The MSE seems modest, and the MSDR is very close to 1.

Figure 6c shows the map made by regression kriging following the steps in Section 3.3. The Gaussian surface was first subtracted from the data. The experimental variogram of the residuals was computed by the MoM and modelled with a spherical function to give the parameter values listed in Table 2. The residuals were kriged at intervals of 2.5 m, the Gaussian surface was added to the kriged predictions, and the results were then passed to MATLAB for the final

display.

4. Discussion and inference

The three examples of spatial variation in *E. neolycopersici* in the two greenhouses provide insight into the behaviour of the disease: its origins, its establishment and its spread.

The trend evident in Figure 6b is perhaps most readily explained. The infection is most serious in one corner of the greenhouse, bottom right in the figure, and declines in an apparently linear fashion from there. The doorway to the greenhouse is at that corner, and it seems most likely that currents of air, bearing spores, entered there to infect the plants, and that the disease then spread from those infected plants. The pattern of disease displayed in Figure 6b is dominated by the trend. But we must bear in mind that kriging smooths; it loses fine detail. The bubble plot, Figure 3b, shows individual monitoring positions where the infection exceeds the general trend. They are isolated exceptionally large scores by the side of the main pathway and along the rows, and it seems likely that the disease was spread to these sites by workers as they travelled along these routes.

Figure 6c has near the centre of the greenhouse a single dominant peak, away from which the infection declines in a bell-shaped fashion. Almost certainly the disease initially infected one or more plants close to that peak and then spread from there in all directions. We note, however, that the decline is most marked towards the top of the figure, and we believe the reason is that the top two rows that were scored were of, Graziano, a cultivar that is more resistant to *E. neolycopersici* than the other varieties. As with the previous example, there are several isolated scores evident in the bubble plot, Figure 3c, that stand out from the trend, notably alongside the main path from south to north. The combination of the Gaussian trend surface, which accounts for 70.4% of the variance, and the kriged smoothing obscure these exceptional scores.

Figures 3a and 6a show more varied patterns of disease with several foci. There are several points of infection along the main pathway leading from the door to the middle of the greenhouse. These suggest that the disease was spread mainly along the pathway by air currents from the door and perhaps by the workers. One of the main foci is immediately to right of the doorway which would have admitted air currents bearing spores and then spread them along the rows. Other foci at the edge of the greenhouse could have resulted from convection currents rising from the centre of the greenhouse and falling at the walls.

It remains for us to interpret the correlation among the residuals, the $\varepsilon(\mathbf{x})$ of Equation (4). The residuals comprise the short-range variation, and the correlation among them, which extends for ≈ 20 to 60 m, almost certainly arises as the infection spreads between neighbouring plants.

5. Conclusions

The patterns of the disease differ in the two greenhouses and from time to time in one greenhouse. All, however, seem to comprise two components, namely a deterministic trend or constant and a spatially correlated residual that can be treated as random. We modelled the distributions of the observed scores of the diseases severity geostatistically. In particular, we characterised quantitatively and located the trends, and we could relate them to plausible sources of infection. In the case of the linear trend (OB5), the infection seems to have spread from the spores entering the greenhouse from the corner, bottom right in Figure 4. The Gaussian trend (OB2) seems to have arisen from the spread of spores from infected plants near the centre of the greenhouse close to the central gangway (Figure 5). The most likely explanation is that it was introduced along that gangway by the workers as they moved their equipment to attend to the crop. The pattern displayed in Figure 3 is more complex. We could not separate a trend analytically. What is apparent, however, is the greater severity of the disease close to the central gangway from which the disease has spread along the rows. It seems that this gangway, the principal pathway through the greenhouse, plays an important role in the spread of infection. Greenhouse managers and crop workers need to be aware of this and take precautions as best they can to prevent the spread of disease by that route.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Jones H, Whipps JM, Gurr SJ. The tomato powdery mildew fungus *Oidium neolycopersici*. *Molecular Plant Pathology*. 2002; 2(6): 303-309. https://doi.org/10.1046/j.1464-6722.2001.00084.x

[2] Niones JT, Sharp RT, Donayre DKM, Oreiro EGM, Milne AE, Oliva R. Dynamics of bacterial blight disease in resistant and susceptible rice varieties. *European Journal of Plant Pathology*. 2022; 163: 1-17. https://doi.org/10.1007/s10658-021-02452-z

[3] Quartezani WZ, Hell LR, Cunha Junior JDO, Moraes WB, Belan LL, Moraes SDPCB, et al. Geostatistical analysis of tospovirus in tomato crop. *Summa Phytopathologica*. 2018; 44(1): 51-55. https://doi.org/10.1590/0100-5405/180383

[4] Silva MG, Pozza EA, Vasco GB, Freitas AS, Chaves E, Paula PVAA, et al. Geostatistical analysis of coffee leaf rust in irrigated crops and its relation to plant nutrition and soil fertility. *Phytoparasitica*. 2019; 47: 117-134. https://doi.org/10.1007/s12600-019-00714-5

[5] Gama ADS, Farias PRS. Geostatistical modeling of 'lethal coconut palm crown atrophy', a new disease in the State of Pará. *Semina-Ciencias Agrarias*. 2021; 42(6SUPL2): 3603-3616. https://doi.org/10.5433/1679-0359.2021v42n6SUPL2p3603

[6] Izquierdo J, Milne AE, Recasens J, Royo-Esnal A, Torra J, Webster R, et al. Spatial and temporal stability of weed patches in cereal fields under direct drilling and harrow tillage. *Agronomy*. 2020; 10(4): 452. https://doi.org/10.3390/agronomy10040452

[7] Martin CS, Milne AE, Webster R, Storkey J, Andujar D, Fernández-Quintilla C, et al. Spatial analysis of digital imagery of weeds in a maize crop. *ISPRS International Journal of Geo-Information*. 2018; 7(2): 61. https://doi.org/10.3390/ijgi7020061

[8] Liu R, Wang H, Guzmán JL, Li M. A model-based methodology for the early warning detection of cucumber downy mildew in greenhouses: An experimental evaluation. *Computers and Electronics in Agriculture*. 2022; 194: 106751. https://doi.org/10.1016/j.compag.2022.106751

[9] Boulard T, Chave M, Fatnassi H, Poncet C, Roy JC. *Botrytis cinerea* spore balance of a greenhouse rose crop. *Agricultural and Forest Meteorology*. 2008; 148(3): 504-511. https://doi.org/10.1016/j.agrformet.2007.11.014

[10] Boulard T, Roy JC, Fatnassi H, Kichah A, Lee I-B. Computer fluid dynamics prediction of climate and fungal spore transfer in a rose greenhouse. *Computers and Electronics in Agriculture*. 2010; 74(2): 280-292. https://doi.org/10.1016/j.compag.2010.09.003

[11] International Plant Genetic Resources Institute. *Descriptors for tomato (Lycopersicon spp.)*. Rome, Italy: International Plant Genetic Resources Institute; 1996. https://cgspace.cgiar.org/handle/10568/73041

[12] Cressie NAC. *Statistics for spatial data, revised edition*. New York, United States: John Wiley & Sons; 1993.

[13] Lark RM, Cullis BR, Welham SJ. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*. 2006; 57(6): 787-799. https://doi.org/10.1111/j.1365-2389.2005.00768.x

[14] Webster R, Oliver MA. *Geostatistics for environmental scientists*. 2nd ed. Chichester, England: John Wiley & Sons; 2007.

[15] Matheron G. *Universal kriging. Notebooks of the center of mathematical morphology of Fontainebleau, volume 1 [Le krigeage universel. Cahiers du centre de morphologie mathématique de Fontainebleau, fascicule 1]*. Paris, France: Ecole des Mines de Paris; 1969.

[16] Lark RM, Webster R. Geostatistical mapping of geomorphic variables in the presence of trend. *Earth Surface Processes and Landforms*. 2006; 31(7): 862-874. https://doi.org/10.1002/esp.1296

[17] Moffat AJ, Catt JA, Webster R, Brown EH. A re-examination of the evidence for a Plio-Pleistocene marine transgression on the Chiltern hills. I. structures and surfaces. *Earth Surface Processes and Landforms*. 1986; 11(1): 95-106. https://doi.org/10.1002/esp.3290110110

[18] Payne RW (ed.). *The guide to GenStat release 20.1 - part 2: Statistics*. Hemel Hempstead, United Kingdom: VSN International; 2021.