Spatial characteristics of the fungus powdery mildew (*Erysiphe neolycopersici*) on tomatoes and its spread in industrial greenhouses

Anastasia Sokolidi[a]*, Richard Webster[a], Alice E. Milne[a], Martin Bielik[b], Phil Morley[b], John P. Clarkson[c], & Jon S. West[a]

[a]*Rothamsted Research, Harpenden AL5 2JQ, UK*

[b]*APS Produce, Arreton, Isle of Wight PO30 3AR, UK*

[c]*Warwick University, Coventry CV4 7AL, UK*

*Corresponding author. 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, a second comprised a linear trend throughout the greenhouse plus a correlated random residual, and in a 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 traditional method-of-moments 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 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 infection introduced through the main door in one corner of the greenhouse and spread from there by the workers and air currents. In 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 UK, outdoor production is limited to a short summer season. To extend the season the tomatoes are instead grown in poly-tunnels and greenhouses. The greenhouses are huge, typically 1 hectare 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 smaller yield [1]. The disease tends to be most prevalent in summer

2

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. Photograph of 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 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 air flow in greenhouses and exchange of air from outside them on the spread of a fungal disease of roses, and they combined it with the fluid dynamics of the air flow and the movement 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 to 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

87 distribution of the disease at particular times, taking into account its evident spatial

88 correlation, and to map it in individual greenhouses. we describe the geostatistical

89 techniques we are using to model the spatial correlation and then to interpolate by

90 kriging to produce maps.

2. Methods

2.1. Monitoring

93 We monitored the fungus, *E. neolycopersici*, in four commercial greenhouses,

94 each of ≈ 1 ha, on the Isle of Wight from June to the end of the crop in November.

95 To show the nature of the spatial variation in the disease, we selected the

96 observations from two of the greenhouses, namely H11 and H13, on three occasions

97 only in 2021, which were 22 July (OB2), 19 August (OB4) and 2 September (OB5).

98 The severity of the disease was scored from 0 to 9 in accordance with IPGRI [11], the

99 principal points on which are as follows.

100 0: Very low (no visible signs of infection).

101 3: Low (small patches < 2 cm across, little sporulation and mycelium).

102 5: Medium (approximately $50\,\%$ the leaves have visible symptoms of disease).

103 7: High (large patches affecting $\approx 70\,\%$ of the leaves and abundant mycelium).

104 Severities between these points were scored with intermediate values. The disease was

105 scored along rows every 4.5 m. The distance between rows was 1.5 m, and every 8th

106 row was assessed. The greenhouses have paths through their middles, approximately

107 4 m wide, for the movement of heavy machinery and produce. The pathway is

108 denoted in Figure 2 with a blue dashed line going through the greenhouse at 42 m on

109 the eastings axis. The tomato varieties differed in the two greenhouses. In greenhouse

110 H13 the sole variety was Piccolo, which is highly susceptible to the fungus.

111 Greenhouse H11 grew five varieties, only one of which, Graziano, has any resistance

112 to the disease. Tomato plants were planted at a density of 1 m apart.
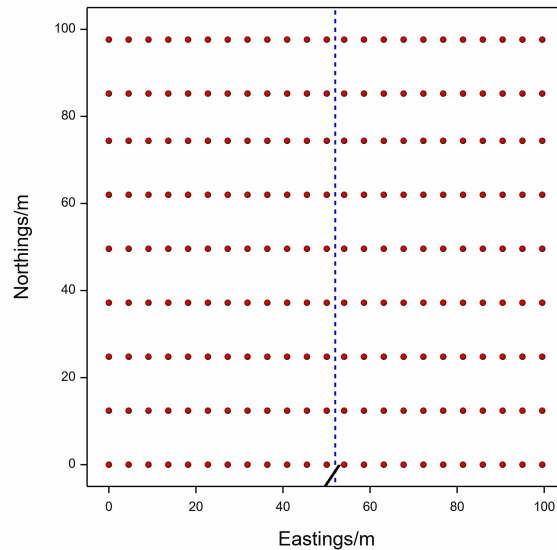
Figure 2. Shows the sampling grid of disease in the greenhouses H13 and H11.The short diagonal line at the bottom of the greenhouse indicates the door.

3. Implementations and Results

3.1 Data summary

Figures 3(a), 4(a) and 5(a) 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 the scores. We mention here that the upper two rows of bubbles in Fig. 5(a) 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 stabilize 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 summarizes them.

6
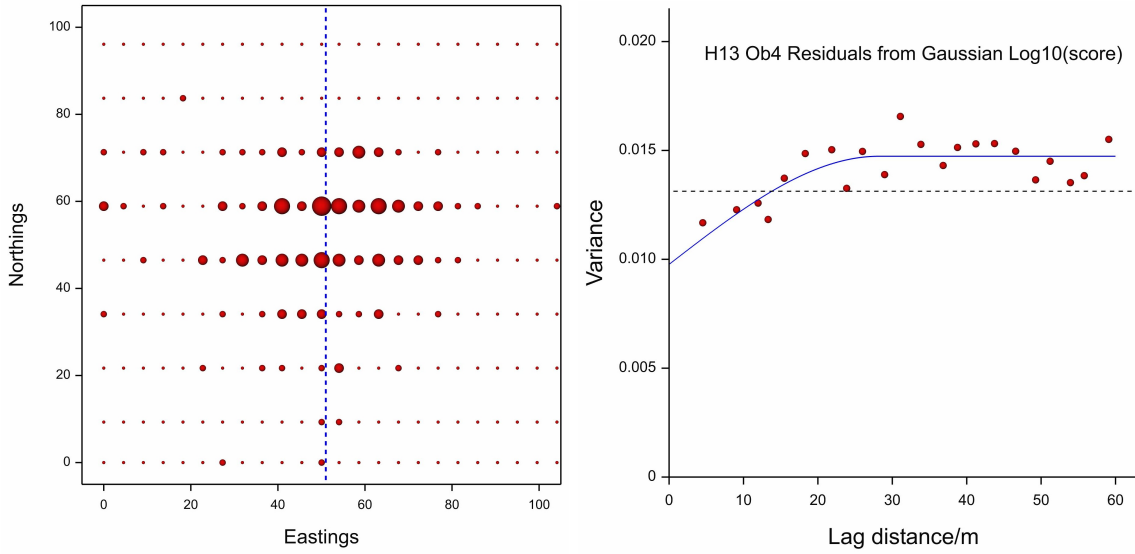
130
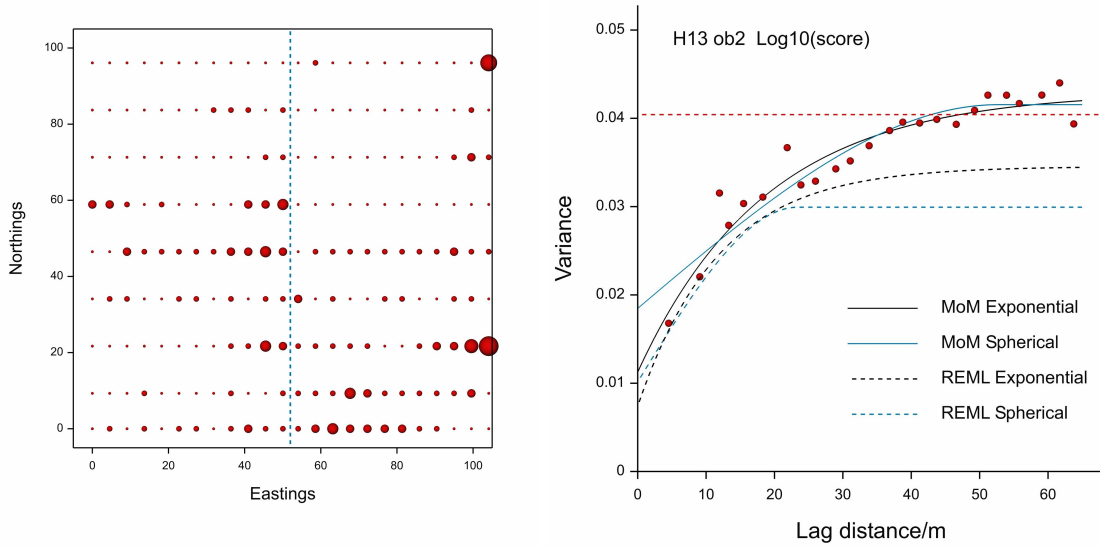
Figure 3. (a) Bubble plot of infection in Greenhouse H13 on occasion 2; (b) experimental variogram and fitted functions.

133
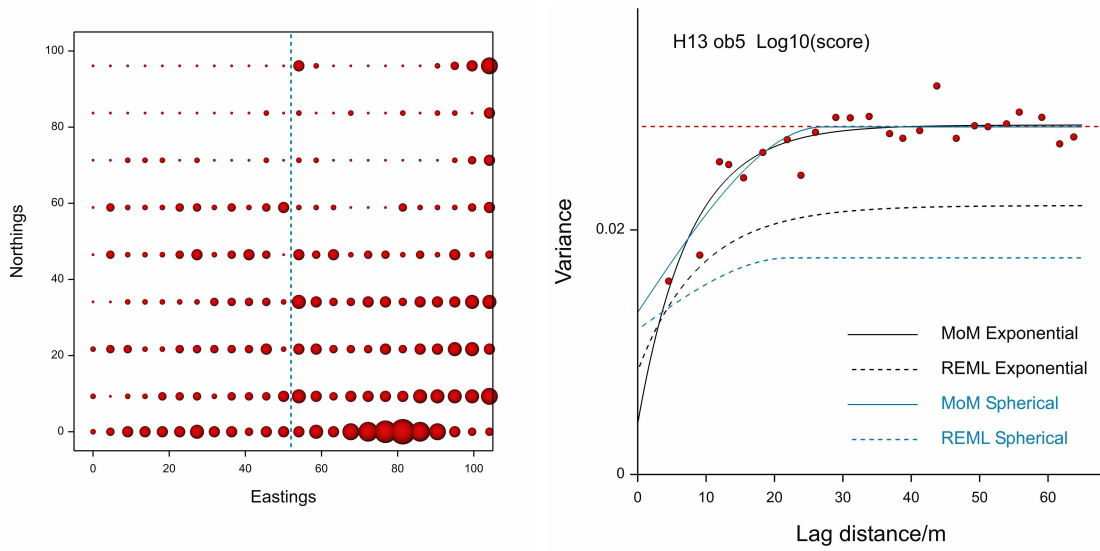
Figure 4. (a) Bubble plot of infection in Greenhouse H11 on occasion 4; (b) experimental variogram of M-o-M residuals and fitted function.

Figure 5. (a) Bubble plot of infection in Greenhouse H13 on occasion 5; (b) experimental variogram of M-o-M residuals and fitted function and REML estimate of the variogram.

Table 1 Data summaries

	Scores			Log$_{10}$transforms		
Glasshouse	Mean	Max.	Skew	Mean	Variance	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

3.2 Geostatistical modelling

Figures 3(a), 4(a) and 5(a) show general patterns of infection in the greenhouses, including much point-to-point fluctuation. Following Cressie [6] we can express this combination as

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

8

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 3(a) 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 4(a) shows a trend extending from one corner of the greenhouse, bottom right in the figure, into the rest of the greenhouse. Figure 5(a) 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 these there is a long-range component of the variation that is clearly not constant. The scores displayed in Fig. 3(a) are evidently correlated spatially. So too is the short-range variation in Figs 4(a) and 5(a) 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 (OB2), illustrated in Fig. 3(a), is the simplest to model. As above, we treat the trend as constant and the residual as a spatially correlated intrinsically stationary random process:

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

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

9

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

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

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

Where there is 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

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

The combination of linear trend with correlated residuals in Fig. 4(a) for H13 (OB5) can be expressed as

$$Z(\mathbf{x}) \;=\; \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon(\mathbf{x}) \;. \tag{5}$$

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

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

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

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

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, $z(\mathbf{x}_i)$, $i = 1, 2, \ldots$, by the method of moments. The formula is

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

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 by residual maximum likelihood (REML). It takes into account all possible paired comparisons, whereas the method-of-moments 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 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

219 Where data $z(\mathbf{x}_i)$, $i = 1, 2, \ldots$, appear to be drawn from a stationary random

220 process as represented by Equation (2) an ordinary kriged prediction $\widehat{Z}(\mathbf{x}_0)$ at any

221 new point $\mathbf{x}_0$ is a weighted average:

$$\widehat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i) . \tag{9}$$

222 The weights, λ_i, $i = 1, 2, \ldots, n$ sum to 1 to avoid bias and are chosen to minimize the

223 kriging error variance by solution of equations that incorporate the semivariances

224 from the variogram. The mathematics are well documented—in for example Webster

225 & Oliver [14]. The number of points, n, in the summation may embrace all the data,

226 but in practice only the few data closest to the target carry sufficient weight to

227 influence the result. Solution of the kriging system also provides the prediction error

228 variance.

229 Where there is trend, as represented by Equation (5), for example, kriging is

230 somewhat more complex. Matheron [15] augmented the kriging system with

231 coefficients of the trend in what he called 'universal kriging'. The semivariances in

232 the system are still drawn from the variogram of the random process, but that

233 variogram is now that of the residuals from the trend, i.e. the $\varepsilon(\mathbf{x}_i)$, $i = 1, 2, \ldots$, not

234 that of the original data. What Matheron did not do was to provide the means of

235 estimating that variogram. Thanks to REML we can now do that and incorporate

236 semivariances from it in the universal kriging systems.

237 For our third example with the Gaussian trend surface of Equation (7) we

238 proceeded in stages as follows.

239 1. Fit a trend surface to the data by ordinary least-squares regression on the spatial

240 coordinates as predictors.

241 2. Compute an experimental variogram of the residuals from the trend, and fit a

242 plausible function to that variogram.

3. Interpolate values of the residuals on a fine grid by ordinary punctual kriging with semivariances drawn from the variogram function.

4. 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 $\widehat{\gamma}(\mathbf{h})$ 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 semivariances at these short distances are typically the only ones that enter in 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 Glasshouse 13 (OB2), Figure 3(a), shows little evidence of trend, and we have treated data as deriving from a stationary process represented by Equation (2). Table 1 summarizes the data. The experimental variogram computed by the methods of moments, Equation (8), is shown by the red discs in Figure 3(b). 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

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

and

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

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:

$$
\begin{aligned}
\text{ME} \ &= \ \frac{1}{N} \sum_{i=1}^{N} z(\mathbf{x}_i) - \widehat{Z}(\mathbf{x}_i) \ , \\
\text{MSE} \ &= \ \frac{1}{N} \sum_{i=1}^{N} \left\{ z(\mathbf{x}_i) - \widehat{Z}(\mathbf{x}_i) \right\}^2 \ , \\
\text{MSDR} \ &= \ \frac{1}{N} \sum_{i=1}^{N} \frac{\left\{ z(\mathbf{x}_i) - \widehat{Z}(\mathbf{x}_i) \right\}^2}{\widehat{\sigma}_{\text{OK}}^2(\mathbf{x}_i)} \ .
\end{aligned}
$$

14

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

The two functions for the M-o-M 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 6(a) maps were made by ordinary punctual kriging of the data with the M-o-M variogram model. The kriging interval was 2.5 m, and the results were passed to MATLAB for the final display.

Table 2 Variogram parameters

Glasshouse	Model	Parameters		
		Nugget	Sill	Distance/m
H13 OB2	M-o-M Exponential	0.01125	0.03174	18.77
H13 OB2	M-o-M 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 M-o-M residuals	Exponential	0.00415	0.02443	7.63
H13 OB5 REML residuals	Exponential	0.01354	0.008457	9.13
H11 OB4 M-o-M residuals	Spherical	0.00977	0.004964	28.18

Table 3 Cross-validation statistics

Glasshouse	Model	Mean	Mean deviation	Mean squared deviation	MSDR
H13 OB2	M-o-M Exponential	0.168	−0.000108	0.023988	1.066
H13 OB2	M-o-M 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 M-o-M 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 M-o-M residuals	Spherical	0.132	0.000644	0.01295	1.028

3.3.2 Linear trend

The scores in Glasshouse 13 OB5 showed a strong trend from north to south, Figure 4(a). 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 regression (OLS), 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 4(b) shows the experimental variogram computed by the method of moments 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 is that obtained by REML. Table 2 lists the parameter values, Table 3 the cross-validation statistics.

Figure 6(b) 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

318 was again 2.5 m, and the results were transferred to MatLab for the final display.

3.3.3 Gaussian trend

320 The scores in Glasshouse 11 OB4, Figure 5(a), exemplify the Gaussian trend

321 with the form defined by Equation (7). We fitted the surface by non-linear

322 least-squares approximation of the transformed scores again using the directive

323 `FITNONLINEAR` in GenStat. The trend surface accounted for 70.4 % of variance with

324 the following values.

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

326 We subtracted the trend from the data, and analysed the residuals. We

327 computed the experimental variogram of the residuals by the method of moments,

328 Equation (8), to which we fitted an isotropic spherical function. Figure 5(b) shows

329 the resulting experimental variogram as red discs and the fitted function. Table 2

330 lists the estimates of the parameters, and Table 3 lists the cross-validation statistics.

331 As expected, the mean error is close to zero because kriging is unbiased. The mean

332 squared error seems modest, and the mean squared deviation ratio is very close to 1.0

333 Figure 6(c) shows the map made by regression kriging following the steps in

334 section 3.3. The Gaussian surface was first subtracted from the data. The

335 experimental variogram of the residuals was computed by the method of moments

336 and modelled with a spherical function to give the parameter values listed in Table 2.

337 The residuals were kriged at intervals of 2.5 m, the Gaussian surface added to the

338 kriged predictions, and the results then passed to MATLAB for the final display.

339
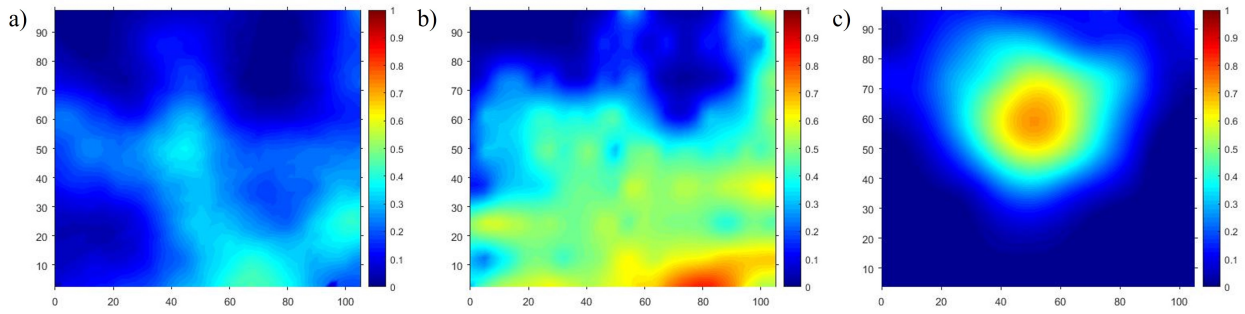
Figure 6. (a) Kriged map of infection in Greenhouse H13 on occasion 2; (b) Kriged map of infection in Greenhouse H13 on occasion 5; (c) Kriged map of infection in Greenhouse H11 on occasion 4.

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 6(b) 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 apparent 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 Fig. 6(b) is dominated by the trend. But we must bear in mind that kriging smooths; it loses fine detail. The bubble plot, Fig 3(b), shows individual monitoring positions where the infection exceeds the general trend. The are isolated exceptional 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 6(c) 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* that the other varieties. As with the previous example, there are several isolated scores evident in the bubble plot, Fig. 3(c), 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 3(a) and 6(a) 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 $\approx$ 20 to 60 m, almost certainly arises as infection spreads between neighbouring plants.

5. Conclusions

The patterns of the disease differ in the two greenhouses and from time to time in the 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 disease's severity geostatistically. In particular, we characterized 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

19

the greenhouse from the corner, bottom right in Figure 4. The Gaussian trend (OB2) seems to have arisen by 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 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 and infection 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

References

[1] Jones, H., Whipps, J.M., Gurr, S.J. The tomato powdery mildew fungus *Oidium neolycopersici*. *Molecular Plant Pathology*. 2021; 2:303–309.

[2] Niones, J.T., Sharp, R.T. Donayre, D.K.M., Oreiro, E.G.M., Milne, A.E., 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, W.Z., Hell, L.R., Cunha Junior, J. de O., Moraes, W.B., Belan, L.L., Moraes, S. de P.C.B., Jesus Junior, W.C., Furtado, E.L. Análise geostatística do

416 'vira-cabeç na cultura do tomateiro. *Summa Phytopathologica.* 2018; 44:51–55.

[4] Silva, M.G., Pozza, E.A., Vasco, G.B., Freitas, A.S., Chaves, E., Paula, P.V.A.A,
Dornelas, G.A., Alves, M.C., Silva, M.L.O., Pozza, A.A.A. Geostatistical
analysis of coffee leaf rust in irrigated crops and its relation to plant nutrition
and soil fertility. *Phytoparasitica.* 2019; 47:117–134.

[5] Gama, A.D., Farias, P.R.S. Geostatistical modeling of 'lethal coconut palm crown
atrophy', a new disease in the State of Para. *Semina-Ciencias Agrarias.* 2021;
42:3603–3616.

[6] Izquierdo, J., Milne, A.E., Recasens, J., Royo-Esnal, A., Torra, J., Webster, R.,
Baraibar, B. Spatial and temporal stability of weed patches in ceral fields under
direct drilling and harrow tillage. *Agronomy-Basel.* 2020; 10(4):452.

[7] San Martín, C., Milne, A.E., Webster, R., Storkey, J., Andújar, D.,
Fernández-Quintilla, C., Dorado, J. Spatial analysis of digital imagery of weeds
in a maize crop. *ISPRS International Journal of Geo-Information.* 2018;
7(61): 20. doi: 10.3390/ijgi7020061.

[8] Liu, R., Wang, H., Guzman, J.L., 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(0168-1699). https://doi.org/10.1016/j.compag.2022.106751

[9] Boulard, T., Chave, M., Fatnassi, H., Poncet, C., Roy, J.C. *Botrytis cinerea*
spore balance of a greenhouse rose crop. *Agricultural and Forest Meteorology.*
2008; 148:504–511.

[10] Boulard, T., Roy, J.C., 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: 280–292.

[11] IPGRI (International Plant Genetic Resources Institute). *Descriptors for tomato (Lycopersicon spp.).* International Plant Genetic Resources Institute. 1996. https://cgspace.cgiar.org/handle/10568/73041 [accessed 03.05.2022].

[12] Cressie, N.A.C., *Statistics for Spatial Data*, revised edition. New York, John Wiley & Sons; 1993.

[13] Lark, R.M., Cullis, B.R., Welham, S.J. On optimal prediction of soil properties in the presence of spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science.* 2006; 57: 787–799.

[14] Webster, R., Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd edition. Chichester: John Wiley & Sons. 2007.

[15] Matheron, G. *Le krigeage universel.* Cahiers du Centre de Morphologie Mathématique de Fontainebleau, No 1. Fontainebleau: Ecole des Mines de Paris. 1969.

[16] Lark, R.M., Webster, R. Geostatistical mapping of geomorphic variables in the presence of trend. *Earth Surface Processes and Landforms*. 2006; 31:862–874.

[17] Moffat, A.J., Catt, J.A., Webster, R., Brown, E.H. 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; 11:95–106.

[18] Payne, R.W. (editor). *The Guide to GenStat Release 20.1 – Part 2: Statistics.* Hemel Hemstead, UK: VSN International. 2021.