Plant Pathology (2018) Doi: 10.1111/ppa.12860

Modelling the spatiotemporal dynamics of *Phytophthora infestans* at a regional scale

B. Firester[a], D. Shtienberg[b] and L. Blank[b]*

[a]*Hunter College Campus Schools, New York, NY, USA; and* [b]*Department of Plant Pathology and Weed Research, ARO, Volcani Center, HaMaccabim Road 68, Rishon LeZion, 7528809, Israel*

Potato late blight caused the Irish Potato Famine and still causes billions of dollars of annual crop damage. Tracking and predicting its spread remains problematic. Growers overspray potato fields with fungicide as a precaution, irrespective of the prevalence or actual risk of the disease. This study created a new weather-based mathematical model for the spread of late blight at a regional scale using empirical data, and validated it using a novel approach for presence-only data. The model was tested using a contrapositive 'proof' by comparing predicted to actual weather patterns to examine its accuracy. The model was then used to create risk maps showing the likelihood of future outbreaks in the region. Such risk maps can help growers optimize late blight suppression and fungicide use by alerting them to the most probable day at which they are at the most risk, using real-time weather data from the previous few days. These risk maps would be updated daily to account for conditions needed for sporulation. Overall, this work offers a methodology to understand and model a disease's spread in time and space.

Keywords: potato late blight, risk map, spatial modelling

Introduction

The dispersal of pathogens in agricultural systems over time and space is a complex and dynamic process. Abiotic (e.g. wind, temperature and relative humidity) and biotic (e.g. host sensitivity, transmitting vectors) factors may affect this process, and human actions (e.g. human dispersal of infected propagation material) are influential as well (Thébaud *et al.*, 2006; Blank *et al.*, 2016). Despite advances in research on the spread of diseases in recent years, quantitative knowledge on the factors affecting the heterogeneity of pathogen dispersal over time and space remains limited. The dynamic and spatial aspects of epidemiological processes pose unique challenges for research and disease management, because crop–pathogen systems include diverse interactions (of which not all are known) and involve many environmental variables whose effects are not necessarily linear (Metz *et al.*, 2012). Furthermore, the heterogeneous natures of biotic and abiotic variables that drive the dynamics of epidemiological processes are difficult to measure and operate on a variety of spatial scales (Levin, 1992; Thébaud *et al.*, 2006; Haverkort *et al.*, 2008; Metz *et al.*, 2012).

Information on spore dispersal has been applied to facilitate prediction of transport over various scales using different methodologies. Lagrangian stochastic simulation models have been the subject of a large body of research in plant pathology. This approach provides a framework for understanding transport and mixing of passive particles (e.g. Aylor, 2005; Aylor *et al.*, 2006, 2011; Meyer *et al.*, 2017). However, this approach suffers from two major weaknesses: (i) exhaustive computation time, and (ii) complex requirements in terms of input data compared to simpler dispersion models. Thus, implementing Lagrangian stochastic models in decision-making poses a challenge. In a different approach, pathogen movement through the landscape is done by evaluating field-by-field dispersal (Shaw & Pautasso, 2014). The recent development of network modelling approaches promises new insights into spatial and temporal components of disease spread. In the last decade, network models have been used to model plant epidemics (Jeger *et al.*, 2007; Brooks *et al.*, 2008; Jeger & Pautasso, 2008; Sutrave *et al.*, 2012). Jeger *et al.* (2007) suggested the use of epidemiological models based on networks to study individual hosts as a set of vertices. This helped elucidate management strategies that provide effective control through better understanding of how the disease spreads over the host landscape. However, this has not been done for *Phytophthora infestans*, the causal agent of late blight in potato and tomato.

Late blight devastates potato and tomato crops all over the world, causing billions of dollars in crop damage, as well as hundreds of millions in costs in attempts at prevention (Haverkort *et al.*, 2008; Small *et al.*, 2015). Dating back to the mid-19th century (Fry *et al.*, 2012), late blight has continued to decimate potato fields. Late blight is mainly managed by application of fungicides (Fry *et al.*, 2012; Small *et al.*, 2015; Fry, 2016) in the hope of killing the pathogen before it infects the plants. In most cases, fungicides are applied as a

*E-mail: liorb@volcani.agri.gov.il

Plant Pathology

precautionary measure, irrespective of the prevalence of late blight in the region and the actual risks. Strains of *P. infestans* have developed resistance to fungicides, specifically the expensive and otherwise highly effective mefenoxam/metalaxyl (Kirk *et al.*, 2001; Hu *et al.*, 2012).

Dispersal events on a regional scale are known to be important for the large-scale dynamics of aerially transmitted pests (Brown & Hovmøller, 2002). For *P. infestans*, sporulation occurs when infected plant tissues are exposed to wet conditions (RH > 90% or surface moisture) for a number of hours at moderate temperatures (10–20 °C). The effects of specific combinations of these factors have been well described and included in the LATEBLIGHT simulation model (Andrade-Piedra *et al.*, 2005). When humidity drops below high levels, the sporangia are easily dislodged and can travel and disperse in various ways to a range of distances (Fry *et al.*, 2012). If dislodged and dry, the sporangia can also travel further distances via the wind (Skelsey *et al.*, 2009a). They can survive in the wind as long as the temperature and solar radiation are not too high, which usually restricts the spread to a few hours in the day, usually in the morning, for them to disperse (Fry *et al.*, 2012). Long distance spread (between regions) is usually due to human interaction, e.g. by planting infected potato tubers or transplanting diseased tomato seedlings (Fry *et al.*, 2012), but long distance spread may also result from many smaller leaps by sporangia spread from field to field over time.

Sporangia travelling through wind is a difficult process to quantify or study. Some studies have tried to estimate the distance that pollen can travel in the wind (Hoyle & Cresswell, 2007), but many of the characteristics used in these studies are still unknown for *P. infestans*. To date, there are no quantitative data in the literature on probable dispersal distances and the maximum distance that *P. infestans* can be dispersed via wind. This is not surprising given the clear difficulties in conducting large-scale experiments with this disease. Zwankhuizen *et al.* (1998) published a study that attempted to quantify the subregional dispersal of *P. infestans*; they showed that within a period of 2 weeks, pathogen spores dispersed over an area of 25 km^2 (Zwankhuizen *et al.*, 1998). Aylor *et al.* (2001) reported that the pathogen could be dispersed over a distance of a few dozen kilometres, depending on the environmental conditions. To date, only a few studies (mostly based on modelling or on circumstantial evidence) have suggested the distance that the sporangia could be dispersed through the atmosphere, mostly using modelling or simulations (Skelsey *et al.*, 2008, 2009a). Aylor *et al.* (2011) collected sporangia up to 500 m downwind from a source of inoculum. Skelsey *et al.* (2009b) modelled the severity of future late blight epidemics based on the genotype, environment, landscape and management. However, detailed scouting to quantify the general dynamics of the disease progress was not conducted, and as far as is known, there have been no studies that quantified the dispersal of *P. infestans* sporangia over a regional scale.

The importance of predicting the occurrence is well documented as a powerful tool in late blight prevention and can help reduce the excessive spraying of fungicides (Fry, 2016). To optimize late blight suppression and fungicide use, various forecast systems and decision support systems (DSSs) have been developed (Krause *et al.*, 1975; Andrade-Piedra *et al.*, 2005; Small *et al.*, 2015; Fry, 2016). These tools help farmers gain more information as to when they are at risk, but are rarely used commercially. One reason for the limited adoption of forecast and DSS tools is growers' perception of risk. Optimizing fungicide usage is based on reducing the risk of exposing crops to an uncontrolled epidemic. In particular, growers attempt to minimize the occurrence of a 'false negative action', that is, skipping a fungicide application that a posteriori analysis would have shown to be essential (Shtienberg, 2013). As spatiotemporal aspects of disease dynamics are not considered in existing DSSs, modelling these aspects could improve the prediction of disease onset and reduce the risk of uncontrolled epidemics. Of particular value would be a tool that quantifies the infection risk levels based on current and past conditions of late blight prevalence in the area. Unlike currently used tools and programs (i.e. LATEBLIGHT or BlightPro; Andrade-Piedra *et al.*, 2005; Small *et al.*, 2015), such a tool could potentially improve the prediction of disease onset and optimize fungicide use. Additionally, current tools used to predict late blight development do not help in understanding issues such as probability of field-to-field, regional and long distance dispersal, spreading tendencies, or other characteristics of disease spread.

The goal of the current study was to develop a spatiotemporal dynamic model for *P. infestans* at a regional scale. The model was developed based on a data set recorded in the northwestern Negev region in Israel. The model was verified with the data used for its construction and then validated with multiple independent data sets recorded in the same region in different seasons.

Materials and methods

The predictive model is based on the combination of two probability functions based on proximity to infected fields and the direction between fields as compared to the wind, as these are the two main factors used to determine the risk of infection. Model training refers to the creation of these two probability functions. This is called 'training', as it can be redone to match a new region if the model were to be used elsewhere, i.e. it can be 'retrained' to match a new location. Model development refers to the parameterization of these two functions and the creation of risk maps – essentially finding which function is more important to the risk (i.e. what is the risk to a field that is close to an infected field, but not in the direction of the wind, versus to a field that is further away, but in the direction of the wind). Risk maps are the product of the model that will be used by farmers. They model the total probability of infection of every location in the region. The model was validated to check how it performed at predicting the presence and absence of late blight.

Data used in the work

The data used in this study to build, verify and validate the spatiotemporal dynamic model were recorded in commercial fields in the northwestern Negev region in Israel. The size of the studied area is about 20 × 35 km (*c.* 500 km^2); this area has abundant potato fields alongside other open-field crops such as carrots, peanuts, wheat and small radishes, and protected crops such as tomatoes (in its southern section) and peppers. The climate of the region is semi-arid, and the mean annual precipitation is 250 mm, but subject to substantial yearly variation. Potatoes are grown extensively, and all fields are regularly fertilized, irrigated and sprayed with fungicides, mainly against *P. infestans* and *Alternaria solani* (the causal agent of early blight). Each potato field was inspected twice each week by trained scouts who record the onset of early and late blight and the intensity of the disease, if observed. The exact date of each infection was recorded. For model training and development, the actual dates the fields were found infected were used.

The model training data, used to fit the model parameters, were collected between October and March 2005–2006. The data used for model development were collected between October and March 2004–2005, and the data used to validate the model were collected between October and March of 2007–2008, 2014–2015, and 2015–2016. No data were reused in another section. The model used the following input data: the date when late blight symptoms were first observed in each field, the location of the field, and the weather data (temperature, relative humidity and wind direction) recorded in 10 min intervals by the single closest weather station located inside the region and operated by the Israeli Meteorological Service.

Model training

Using the training data, two separate probability functions were created to estimate the probabilities of a diseased field causing subsequent infections in other fields in the region. One probability function is based on the distances between the fields, and the other is based on the angle between the vector from a diseased field and the wind vector. These probability estimates were used in two ways: (i) to create a spatial model inferring the routes by which an infection could have spread based on the empirical data; and (ii) to create risk maps predicting the chances that an existing set of diseased fields would spread the disease to nearby locations in the future.

The distance probability function

The 2005–2006 data was used to calculate the probability of pathogen spread as a function of distance from previously diseased fields. In the first week of December 2005, the disease was first identified in three nearby locations. The centroid of these three locations was used as the point of origin for this infection. In the following week, the disease had spread to 10 new fields in similar directions, scouted on the same date. Based on the wind direction, it was highly probable that these 10 infected locations originated from one of the three diseased locations in the previous week. Furthermore, this was supported by the fact that all the newly infected fields were located in the direction of the cumulative wind vector (see below), and because disease symptoms were apparent on the upper leaves (indicative of airborne infections) and not on the lower stem sections (indicative of tuber infection). Starting at the first detected location of the disease, concentric circles were laid out with radii increments of 1000 m. The probability of infection at a given distance from the point of origin was estimated to be proportional to the fraction of the area of the potato fields at that distance that were diseased in the following week. An average potato field's area in this region was used for this estimate. These probabilities varied inversely with distance, so a log-linear, best-fit regression was used to approximate the probability of infection as a function of distance from the source. This function had to be created from a single infection, where nearby fields (three in this case) spread to many others of varying distances and angles. It had to be sufficiently clear that the second wave of infected fields was not infected by an internal source (e.g. infected tubers).

The angle probability function

These same data were also analysed to model the wind angle from previously diseased fields. For that, a program was created to analyse the local weather data and identify the appropriate conditions that are necessary for sporangia to be released, be transported by the wind, and then cause a subsequent infection. First, based on published data (Andrade-Piedra *et al.*, 2005) it was assumed that sporulation occurs when the infected tissues are exposed to wet conditions (RH > 90% or surface moisture) for at least 10 h at temperatures above 10 °C. Next, the sporangia must be dry to be dispersed by the wind and they must survive the heat and strong solar radiation that are lethal to them. Lastly, relative humidity above 90% is needed for sporangia germination and for the subsequent infection (Krause *et al.*, 1975). These conditions generally occur 5–8 days prior to the observation of disease symptoms in a specific field. The model searched for days with suitable weather during 5–8 days before an infected field was observed, and used the weather data from only those viable hours.

In this model, the term 'vector' refers to an entity with both magnitude and direction, not a method of pathogen transport. An algorithm was built and run to evaluate the wind and relative humidity data (given the prevalence of appropriate temperature conditions). This algorithm searched through the weather data to find the times when sporulation occurs according to the criteria listed above. Once that was established, the program analysed the wind direction in the three hours after the relative humidity dropped below 90%. In this region, high relative humidity persists only during the nights, and the drop in humidity occurs during the early morning. In these morning hours, radiation increases and dries the sporangia. After the 3 h time frame, it was assumed that the radiation was too intense and damaged the sporangia. The winds during this time frame could have carried the sporangia without the risk of them being killed by solar radiation (Mizubuti *et al.*, 2000).

An approximate vector sum was calculated to characterize the wind into one variable: the component vectors for different wind directions were weighted based on the cumulative amounts of time the wind direction was in each direction and summed as vectors to approximate the path that the carried sporangia might have taken. This cumulative vector sum accounted for the changing direction of the wind during these critical morning hours, as well as the amount of time in each direction. This wind vector takes into account only the times of day when the solar radiation was not high enough to kill the sporangia, usually just a few hours in the morning. Because it is unknown for how long a given sporangium is airborne, the cumulative wind vector must account for the sporangia released at the latest possible time (right before the solar radiation is high enough to kill them), which is why the total time included is not dependent on the exposure period. The program that mined the weather data for viable hours checked to make sure the hours of spread were

after sufficiently high relative humidity in the night, before high relative humidity the day after, and low during the spread. These are the best-known factors that are conducive to the production of viable sporangia spores.

To simplify the model, effects of abiotic factors on sporangia viability during their aerial transport such as ambient relative humidity and radiation intensity, were not taken into account. Wind speed was also not taken into account for several reasons: first, it is unknown if the sporangia are travelling exactly with the wind throughout the duration. If this were the case, then late blight would be spreading over long distances (as the wind would travel these large leaps), when in reality late blight is documented to spread a maximum of only a few kilometres. Secondly, the vertical speed and direction of the wind is not known either. If the wind is moving very quickly, but blowing the sporangia towards the ground, they would travel shorter distances than a much slower wind pushing the sporangia up into the air. Using this cumulative wind vector approximates the most likely path a particle would travel in the wind if it were released into the wind at a random time and fell onto the ground during the duration of the wind. The wind vector function only examined times in which spread was viable, and thus the model is based on these periods of time. This narrows down the weather data to only the relevant hours, and more accurately models the disease spread. It should be noted that, as the dispersal distance was modelled using empirical data (2005–2006 data), effects of abiotic factors on sporangia viability and wind speed were represented in the wind vector function. In that regard, it should be noted that the model was developed to simulate pathogen spread over short distances (a few kilometres), between fields within a region.

For each subsequent infection of the 2005–2006 season, the angle between this cumulative wind vector and the direction from the point of origin was computed. Like the distance-based probabilities, the probability of infection at a given angle to the wind vector was estimated as the infected fraction of that pie-shaped octant of a circle centred at the first site. Note that neither of these calculations required estimation of absolute probabilities, but only relative probabilities that could be used to compare the likelihoods of infection among different locations.

Model development

The model parameterization (of wind- and distance-based probabilities) technique used was a qualitative analysis of which parameterization of the two probabilities best described known disease dynamics. Using data from the 2004–2005 season, discrete parameterizations of wind- and distance-based probabilities were created and analysed to see which one best modelled the spread of the disease. This model assumed each field (with exceptions noted below) was infected from the previously infected field that had the highest probability of infecting it.

Each of the probabilities estimated represents the likelihood of a specific location being a source of inoculum to cause disease in another specific location. To model the spread of the disease, the probabilities of every field being infected from each of the previously diseased fields was calculated based on the distance- and wind-related probabilities. The location from the previous week having the highest relative probability was deemed to be the source of the new infection for analysis purposes.

For each new diseased field, the distances to all of the prior diseased fields were taken and plugged into the hyperbolic approximation to compute probabilities. These were then normalized to a scale from 0 to 1, by dividing by their sum. In addition, for each prior diseased field, the angle between the vector to the new infection and the cumulative wind vector was calculated. These angles were each normalized by dividing by their sum, inverted by taking 1 minus the normalized angle. This was done because a larger angle corresponds to a smaller probability because the force of the wind is directed further away from the direction of the sporangia travel. Next, the angles were renormalized to sum to 1. The probabilities for the wind direction had to be inverted so that a larger difference would denote a lower probability. For example, if the original angles were A_1, …, A_n, then A_i was normalized to

$$[1 - A_i / \sum A_i]/(n-1)$$

If the previously diseased fields all had similar relative probabilities within a narrow range (2.22% as explained below), then the new infection was assumed to have come from an external source, such as an unknown infected tomato field, infected seed tubers, or an infected volunteer plant.

By normalizing the probabilities, the model used the assumption that every diseased field came from a prior diseased field regardless of conditions. Even if the wind was going in the opposite direction and the fields were far away from other possible sporangia sources, all the probabilities of infecting a given field would sum to one. Thus, it could not account for a new source of inoculum. The 2.22% rule was used to account for this. The value 2.22% was chosen as 100% divided by the total number of diseased locations in 2004–2005 (45 in this case). The normalized probabilities from the wind and the distance were combined in a weighted average, to yield a final relative probability between a newly infected field and a previously diseased field. The spatiotemporal dynamic model was run with different sets of weightings to simulate the relative effects of wind versus distance. Graphs were created with wind-to-distance weightings on a discrete spectrum with a full range of integral ratios (e.g. 1:1, 6:1 and 10:1). These ratios had to be simple to account for uncertainty. Within any range, the actual model might not shift until a critical value is reached. Because there is no way to say which value within the most accurate range is correct, the lowest level of uncertainty was taken, in this case integer ratios. For each newly infected field, the previously diseased location with the highest calculated probability of infecting it was designated as its source. This assumption was used to effectively compare various parameterizations of the wind- and distance-based factors. Creating charts to show the paths of the disease best predicted by these parameters can show how these different factors interact and which parameterization makes most biological and probabilistic sense.

Risk maps

Sample risk maps were created based on the 2014–2015 data by assuming that nondiseased (healthy) fields could be infected by any existing diseased field from previous weeks, to demonstrate the tools this model offers for farmers. By mathematically aggregating the relative probabilities of infection from every diseased field, a relative probability of infection was calculated for each healthy location. Thus, the risk map shows the relative likelihood of infection for each location for a single week during the season. For clarity, in this paper, the probabilities described below are relative probabilities and will be different in each week, because they depend on the numbers of previously infected fields.

To create the risk map, the wind-based probability approximation was refined, using a log-linear, best-fit regression, as was

originally done for the distance-based probability, as this yielded the highest correlation to the data and matched current hypotheses on the spread of pathogens. This also yielded a hyperbolic approximation of the form $Y = kX^{-j}$. The risk map was created by running the model (at a 6:1 wind-to-distance weighting) to calculate for each uninfected location, the probability of infection of that point from each of the diseased locations. The total probability of infection at that location is the mathematical union of all of these calculated probabilities.

Mathematically, separate probabilities of infection cannot just be summed, generally speaking, because for each subset of probabilities, this would double-count the probabilities that infection came from more than one of the previously diseased locations. The probabilities must be combined using the inclusion–exclusion principle, which nominally requires calculating and summing a number of terms that is exponential in the number of infected locations. To estimate this efficiently, the individual probabilities can be combined step by step, so as to only take the union of two probabilities at a time. Each future point's probability is calculated, and the algorithm to take the union of two probabilities is run. This is mathematically proven to be equal to the total union in this scenario.

To create the risk map, discrete points (x, y) are chosen according to the desired precision. Specifically, for this case study, they are spaced every 300 m forming a grid. Each point is then compared to all previously diseased fields, taking into account the distance (γ) and angle (θ). Given the current wind vector (φ) and parameterization of wind- and distance-based probabilities, a custom heuristic generates the probability of infection. Because certain points would lie on fields if they were close enough to 1, if the location of (x, y) was on an already diseased field ($\gamma \leq 300$ m), that field's value was set to 0. Points within this range were reasonably deemed to be the same field, and to try to find the probability of infection of the average uninfected field, this value must be removed and set to 0. The overall probability would not be changed drastically once it was added to the other values in the union of the set. If the value of the wind-based probability was >1 (although this never occurred), it would be replaced with 0.999. This formula represented the weighted average of the distance- and wind-based regressions, which were designed to fall between 0 and 1 for relevant fields. Therefore, fields that were not related to a diseased field (not the same or very close), would have realistic estimated probabilities.

Model validation

The validation method reverses the inputs and the outputs of the model. Feeding backwards, the validation method took inputs of the currently infected fields and which fields were to be infected in the coming week. Given these inputs, its goal was to predict the conditions (in this case, the wind direction because that is the input data in the model) that would lead to some fields getting infected and others remaining uninfected based on their associated risk values. These risk values would be calculated using the model's formula and algorithm without knowing the wind direction. It would thus be 'guessing' the wind direction and the validation would test to see if it was correct in isolating the direction (or multiple directions in some weeks, if conditions were viable on multiple nights) of the wind. If it were able to consistently do this over multiple weeks in multiple seasons, then the model predicting the relative risk values throughout the season was clearly differentiating high risk from low risk areas.

For the validation of the developed spatiotemporal dynamic model, data sets from the years 2006–2007, 2014–2015 and 2015–2016 were used. Ten independent weeks were used for validation. The infection events were aggregated into weeks for methodological reasons, as a minimum number of infection events were needed in order to make the test. 'Weeks' were used because the infection cycle of late blight takes 5–8 days, so weeks are a good representative of the time range. Standard validations of similar models were not feasible, because the data were limited (as only a few fields were diseased each week) compared to other studies of disease modelling, and because the data were presence-only. Although the disease was not found in many fields, it was never fully known if the disease was not there or just was not found by the scouts. There were also different areas of data not collected in different years, so it was impossible to assume that just because an area showed no late blight, it was not present there. Additionally, as all fields were sprayed with fungicides, the fact that disease was not observed in a specific field may have resulted from effective control of sporangia and not from lack of sporangia availability. Accordingly, the focus must be on the diseased, rather on the healthy fields. Furthermore, it is better to have a false positive (and overspray) than a false negative (and get infected).

The model was validated using only the field locations and 70 randomly selected uninfected fields within the region (equal to the average number of uninfected fields) as the inputs. Then a program was compiled that tested all possible wind directions from 0 to 360° in 2° intervals, and for each wind direction, took the cumulative average of the diseased fields and the randomly assigned points from all the infected sites at that time. Because there was no way to know exactly where the disease was absent, the randomly assigned points represented uninfected fields. Even if some of them were infected and it was unknown, the average probability would still represent that of uninfected fields. The wind direction values that yielded the largest positive difference in probability between the averages of the actual diseased locations and the randomly chosen locations represented the wind directions that the model claimed would best support the data. These 'predicted wind directions' show how statistically accurate the model is at creating probability differences between areas that will get infected compared to those that will not. These wind direction values were compared to the observed wind directions on the days viable for spreading infection. In two cases, a day with RH > 88% for 8 h in the preceding night (instead of RH > 90% for 10 h) was chosen because no suitable days were found according to the stricter conditions, so the disease must have spread on a slightly less ideal day. These two cases occurred late in the season (week 3), so if the infection in these two fields were internal (e.g. infected tubers) it would be expected that symptoms would be evident earlier. The loosening of the conditions was done to gain more power for the validation as the data was limited. If the predicted wind direction matches the actual wind direction to a close degree, it determines that the model is showing a clear difference between infected fields and uninfected fields.

Results

Using the 2005–2006 data and the associated meteorological records, the probabilities of disease spread at any given distance (Fig. 1a), and the difference in angle from the wind vector were calculated (Fig. 1b). In both analyses, the estimated probability of infection varied inversely

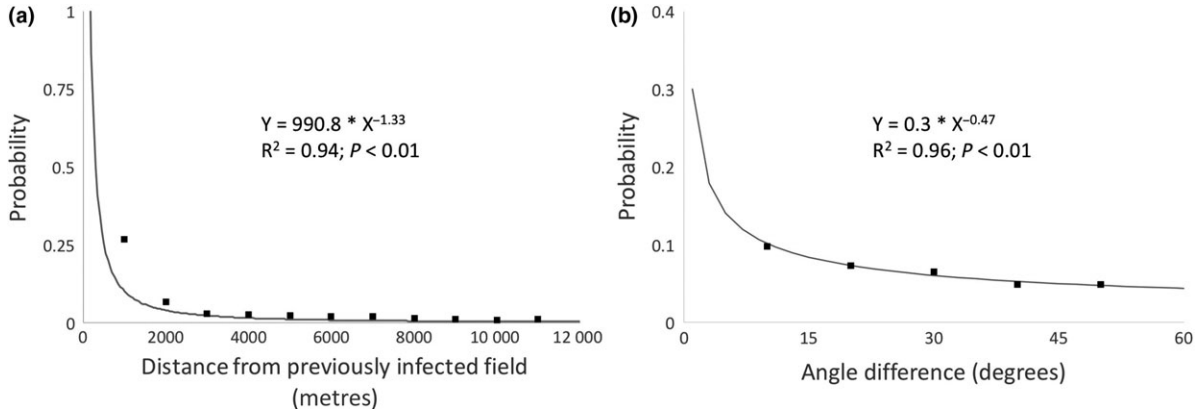

Figure 1 The probability that disease at a certain potato field (the recipient field) originated from a field that was diseased by late blight 1 week earlier (the source field), as a function of: (a) the distance between the fields, and (b) the angle difference between the fields and the cumulative wind vector. Regressions were derived from the 2005–2006 infected fields and weather data.

with a power of the distance, yielding a hyperbolic estimation function that matched the hypothesis. The 'critical value', where it reaches a probability of 1, was 178 m for the distance-based probability (below the average radius of a field) and 0.078° for the wind. Although the wind-probability graph may appear linear due to the scale (Fig. 1b), its correlation was lower than that of the hyperbolic function (R^2 values of 0.90 and 0.96, respectively). Because the hyperbolic fit had a better correlation and matched the hypothesis, that function was used in the model.

Various arrow infection diagrams were created to attempt to find a plausible parameterization for the final model (Fig. 2). When examined, the 6:1 (Fig. 2b) wind-to-distance ratio was deemed to be the most appropriate for two reasons. First, Figure 2a (1:1 ratio) has some aspects that logically should not have occurred based on the biology of the disease. Field 22 was closer to field 16 than field 11, but is very similar to field 21 in angle from both. Because they were infected at the same time and were in close vicinity to each other, it is reasonable to assume that they were either infected from the same source, or at least from very close sources. However, in this parameterization, the algorithm assigned field 22 as having been infected from field 16 because it is closer. Because the distance-based probability had such weight, this affected the average too much and assigned field 22 as having been infected from field 16. The same principle was applied in field 2 spreading to field 3. Fields 3 and 4 have similar locations and should have been infected from the same source. Because of this, it is believed that this model overvalued distance. Secondly, in Figure 2c (10:1 ratio), the assignment of field 22 as having been infected from field 11 seems reasonable, but there are some other logical inaccuracies. The infection of field 7 by field 3 does not seem reasonable. Field 5 is a lot closer, and the difference in the wind direction should not be enough to switch it over to field 3. This set of wind-to-distance weightings may have overvalued the importance of wind direction. The 6:1 (Fig. 2b) ratio did not have the same fallacies as the other parameterizations and was therefore chosen to be in the model. The resultant combined formula for the probability of infection based on the distance (γ), the angle (θ) between the fields, and the current wind vector (φ) was:

$$P(x,y) = \frac{6 * \frac{0.3}{|-\varphi|^{0.472}} + \frac{990.8}{\sqrt[3]{\gamma^4}}}{7}$$

The risk maps demonstrate how the spatiotemporal dynamic model would use previously diseased fields and the weather data to determine relative probabilities for the nearby area. As an example, a risk map for the 2014–2015 season is presented in Figure 3. The disease was first observed in the first week of November 2014 in two fields and spread to the southwest. As the season progressed and more fields got infected, the high-risk area values increased.

The model was validated with independent data sets, recorded in the 2006–2007, 2014–2015 and 2015–2016 seasons. When the predicted values for wind direction were compared to the observed wind directions, the average difference was 7.7° (RMS of 9.66), and the total difference in arithmetic means was 3.1°. Next, the predicted versus expected wind directions were plotted (Fig. 4). The slope of the trend line was 0.95 with an intercept of 8.88. The Pearson correlation coefficient was 0.98 (Fig. 4).

Discussion

Using mathematical modelling to better understand biological events or disease spread has been shown to be effective in many studies and for various diseases (Mollentze *et al.*, 2014). In this work, a methodology is offered to track diseases that spread in time and space, and a spatiotemporal dynamic model developed for

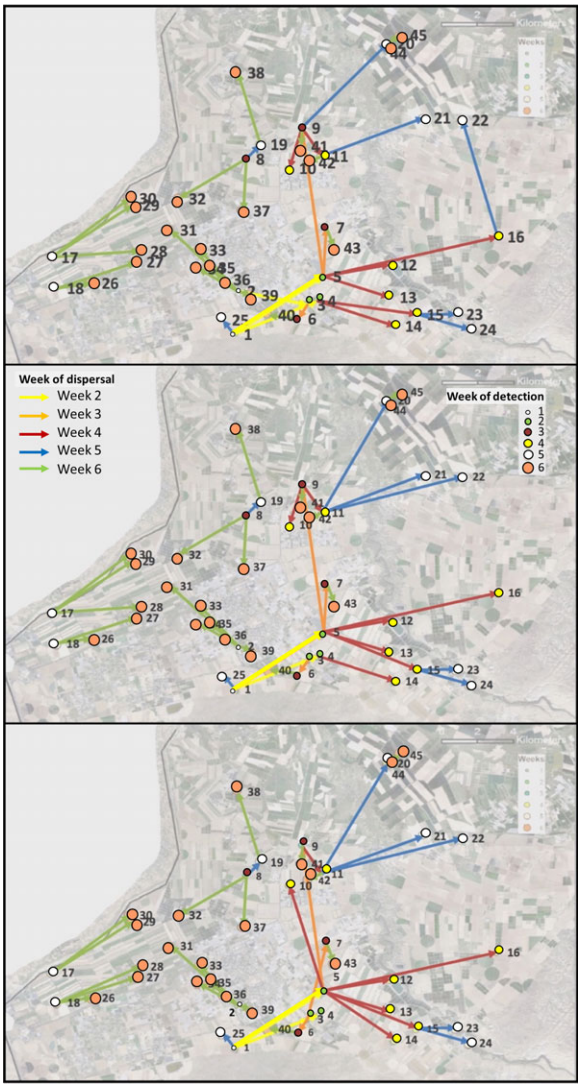

Figure 2 Verification of the spatiotemporal dynamic model for *Phytophthora infestans* using the data set recorded in 2004–2005. Outputs of three representative models with different wind-to-distance weightings are presented. These show how the 6:1 ratio was chosen by exemplifying the reasons why when scaled in other directions, either the wind component is under-represented (a) or over-represented (c), and thus the models do not accurately model disease dynamics. (a) wind-to-distance weighting of 1:1; (b) wind-to-distance weighting of 6:1; and (c) wind-to-distance weighting of 10:1. The circles and the adjacent numbers represent the diseased potato fields, and the time of disease onset is marked by the colour and size of the circles. The arrows show how the disease spread from field to field, each arrow denoting the most likely source of infection to that field, based on the model. The coloured circles and arrows show the week that the infection took place. Fields that do not have an arrow leading to them are either fields infected in week 1 (fields 1 and 2) or determined by the model to have been infected by external sources and declared new inoculum sources (fields 8, 17 and 18).

P. infestans at a regional scale. Specifically tracking which field infected another poses a difficult problem, but offers powerful information for future research about aspects of the disease spread, such as maximal distance, directionality and likelihood of random appearance. For late blight, none of these are known accurately, but the research includes various hypotheses (Fry *et al.*, 2012).

Another problem with many disease data sets is that they record only where the disease was observed (presence-only data), but do not confirm where it was not. Scientists in the field have looked for ways to overcome such limitations (Sutrave *et al.*, 2012). Many model validation techniques test for accuracy by comparing the successes and failures of the model using simulated versus real data (Rykiel, 1996), but when just presence-only data is available, these techniques cannot be used, especially when the data is complicated by treatments such as fungicides. This creates a problem for some disease studies. In this study, a method of model validation is proposed that can use presence-only data and does not require large amounts of data for each infection step, but can still effectively evaluate the accuracy of a model.

This approach validates the relative probabilities of the areas, meaning that the calculated value may not be the true probability of infection, but relative to the area as a whole. The model can examine the region as a whole and determine areas of relative high risk and low risk. A strong difference in probabilities indicates that the farmers in one area should spray fungicide, and in another area they can decide not to do so.

The developed spatiotemporal dynamic model performed reasonably well at all points throughout the season. Due to the nature of presence-only data, the only quantifiable measure to evaluate the model was the wind direction. Compared to the observed wind direction, the model was able to predict the wind direction within a degree in some weeks, empirically, by calculating the probabilities of infected versus the random uninfected fields. Additionally, within each week of spread, the model not only predicted exactly how many days of spread there were in that week (based on favourable weather for inoculation), but could accurately predict all of them. Stemming from the fact that in any given week, there might only be 1 or 2 viable days to spread, and in those days only a few hours will have the suitable conditions, the actual direction of the wind is often limited and directed over small time spans. This shows that the model can successfully represent the complexity of the spread via wind.

Based mostly on previously diseased fields, it might be expected that the performance of the model would improve as the season continues and more fields get infected. However, in the first week (14–21 October) of the 2014–2015 data, with only two previously diseased fields, the model was only 1° off of the observed wind. In the following week, with only six infected fields, the two predicted wind values were 6 and 7° off, respectively. Because the model was able to perform at this level of accuracy so early in the season, it most likely means that the diseased fields after week 1 were infected from the originally diseased fields and not from

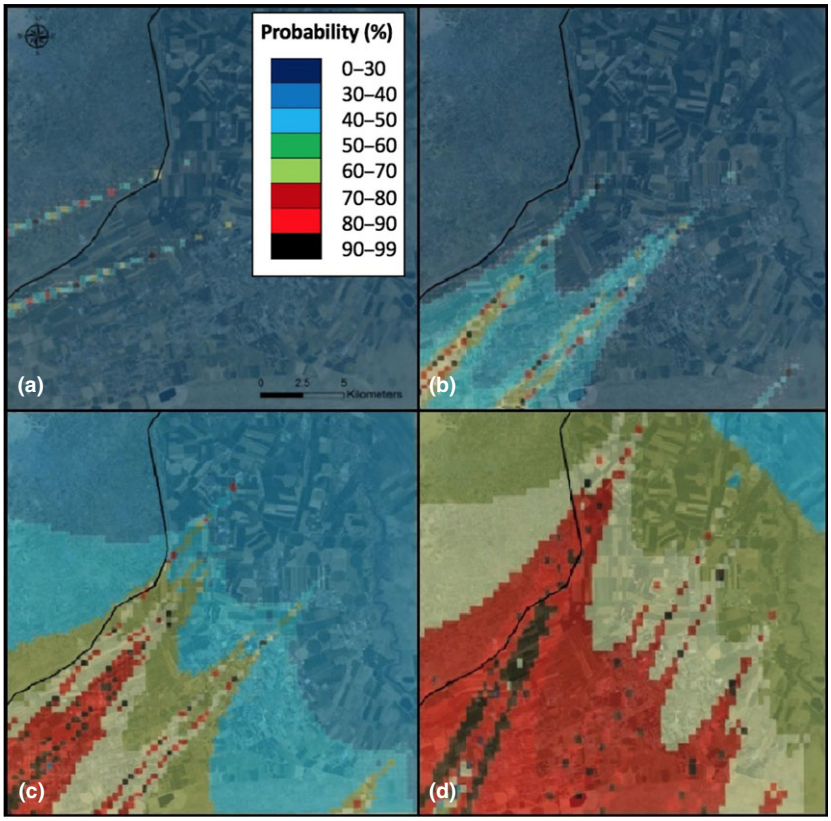

Figure 3 Risk maps for late blight using the data set recorded in 2014–2015. The coloured scale represents the probability of disease spread from the diseased potato fields to the surroundings, in the northwestern Negev region. Each map represents the probability of disease spread in a certain week: (a) 1–6 November 2014. The cumulative wind vector direction was 249°; (b) 7–13 November 2014. The cumulative wind vector direction was 226°; (c) 14–20 November 2014. The cumulative wind vector direction was 220°; and (d) 21–26 November 2014. The cumulative wind vector direction was 220°.

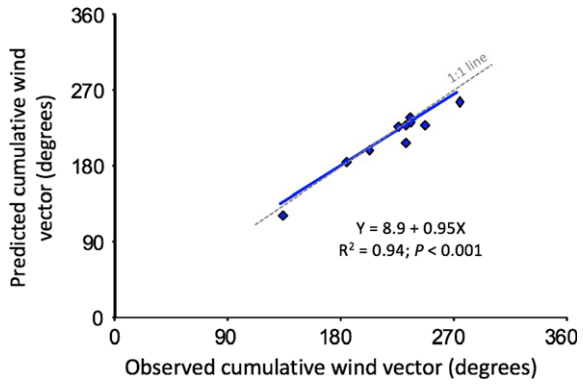

Figure 4 Validation of the spatiotemporal dynamic model for *Phytophthora infestans* using data sets recorded in 2006–2007, 2014–2015 and 2015–2016 (which were not used to quantify the model's parameters). Data points represent predicted and observed cumulative wind vector directions. The regression line represents the correlation between predicted and observed data, and the dashed grey line (the 1:1 line) represents a theoretical perfect fit between the two variables.

random inoculation from infected tubers or volunteer plants. This fully supports the original assumption that after disease onset in the first fields in the region, the fields that were infected subsequently were infected from the previously infected fields as opposed to infected tubers. If this were not the case, the model could have never been so accurate at predicting these early weeks of infection. It should be noted that potato fields in the studied area are planted over a short period of time. Thus their emergence and subsequent development occurs almost simultaneously.

The methodology used allows this model to cater specifically to late blight. The algorithms used were based on how *P. infestans* spreads, both spatially and temporally. The needed input is minimal, and includes the locations of infected fields during the season and the proportion of potato fields in the area. Therefore, the methodology presented in the paper could be replicated in other regions and implemented in new areas if their data were available. This makes the model available to potato farms worldwide. All that is necessary to operate the model is the establishment of the probabilities and the parameterization of the wind- and distance-based probabilities. Both of these can be made from a single year's data of late blight occurrences, marking the date and location of infected fields. Creating a specific model for tracking disease spread allows for more research to be done on late blight spatial dynamics, and for example, to quantify the maximal spread of the disease and study what can facilitate *P. infestans* movement.

This model showed that *P. infestans* spread has a local directionality to it. However, within the ideal parameterization, multidirectional spread still occurred. Thus, within 1 week, late blight could spread in different

directions if there were multiple days that allowed for viable spreading. This was corroborated by the model being able to predict multiple wind directions in a week during the validation process.

Assessing the risk and predicting late blight can be a powerful tool to combat late blight and help prevent excessive fungicide treatment. Isolating areas that are at risk can help growers better protect themselves with proper fungicide treatment, but arguably more important is the model's ability to locate areas that are at low risk. Those farms can afford to spray considerably less fungicide or more environmentally friendly ones even if they are less effective. Knowing a proper risk assessment can help optimize late blight research and data collection. Farmers can use these models to better monitor fields at risk and to focus their resources on the most relevant areas.

The methodology used to create this model could be effective at predicting the spatial spread of other pathogens. Probability functions for pathogen spread similar to the ones developed in this study could be created from existing data and used to help study and better manage those diseases.

Acknowledgements

The authors would like to thank the farmers and Ma'on Enterprises for their help in collecting the data. The work is a contribution of the Agricultural Research Organization, Volcani Center, Israel, no. 582/18. This work was supported by a grant from the Chief Scientist of the Israeli Ministry of Agriculture and Rural Development awarded to L.B. and D.S. (grant no. 132-1830).

References

Andrade-Piedra JL, Hijmans RJ, Juárez HS, Forbes GA, Shtienberg D, Fry WE, 2005. Simulation of potato late blight in the Andes. II: validation of the LATEBLIGHT model. *Phytopathology* **95**, 1200–8.

Aylor DE, 2005. Quantifying maize pollen movement in a maize canopy. *Agricultural and Forest Meteorology* **131**, 247–56.

Aylor DE, Fry WE, Mayton H, Andrade-Piedra JL, 2001. Quantifying the rate of release and escape of *Phytophthora infestans* sporangia from a potato canopy. *Phytopathology* **91**, 1189–96.

Aylor DE, Boehm MT, Shields EJ, 2006. Quantifying aerial concentrations of maize pollen in the atmospheric surface layer using remote-piloted airplanes and Lagrangian stochastic modeling. *Journal of Applied Meteorology and Climatology* **45**, 1003–15.

Aylor DE, Schmale DG, Shields EJ, Newcomb M, Nappo CJ, 2011. Tracking the potato late blight pathogen in the atmosphere using unmanned aerial vehicles and Lagrangian modeling. *Agricultural and Forest Meteorology* **151**, 251–60.

Blank L, Cohen Y, Borenstein M *et al.*, 2016. Variables associated with severity of bacterial canker and wilt caused by *Clavibacter michiganensis* subsp. *michiganensis* in tomato greenhouses. *Phytopathology* **106**, 254–61.

Brooks CP, Antonovics J, Keitt TH, 2008. Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *The American Naturalist* **172**, 149–59.

Brown JKM, Hovmøller MS, 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–41.

Fry WE, 2016. *Phytophthora infestans*: new tools (and old ones) lead to new understanding and precision management. *Annual Review of Phytopathology* **54**, 529–47.

Fry WE, McGrath MT, Seaman A *et al.*, 2012. The 2009 late blight pandemic in the eastern United States – causes and results. *Plant Disease* **97**, 296–306.

Haverkort AJ, Boonekamp PM, Hutten R *et al.*, 2008. Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. *Potato Research* **51**, 47–57.

Hoyle M, Cresswell JE, 2007. The effect of wind direction on cross-pollination in wind-pollinated GM crops. *Ecological Applications* **17**, 1234–43.

Hu C-H, Perez FG, Donahoo R *et al.*, 2012. Recent genotypes of *Phytophthora infestans* in the eastern United States reveal clonal populations and reappearance of mefenoxam sensitivity. *Plant Disease* **96**, 1323–30.

Jeger MJ, Pautasso M, 2008. Epidemic threshold and network structure: the interplay of probability of transmission and of persistence in small-size directed networks. *Ecological Complexity* **5**, 1–8.

Jeger MJ, Pautasso M, Holdenrieder O, Shaw MW, 2007. Modelling disease spread and control in networks: implications for plant sciences. *New Phytologist* **174**, 279–97.

Kirk WW, Felcher KJ, Douches DS *et al.*, 2001. Effect of host plant resistance and reduced rates and frequencies of fungicide application to control potato late blight. *Plant Disease* **85**, 1113–8.

Krause RA, Massie LB, Hyre RA, 1975. BLITECAST, a computerized forecast of potato late blight. *Plant Disease Reporter* **59**, 95–8.

Levin SA, 1992. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–67.

Metz MR, Frangioso KM, Wickland AC, Meentemeyer RK, Rizzo DM, 2012. An emergent disease causes directional changes in forest species composition in coastal California. *Ecosphere* **3**, 1–23.

Meyer M, Burgin L, Hort MC, Hodson DP, Gilligan CA, 2017. Large-scale atmospheric dispersal simulations identify likely airborne incursion routes of wheat stem rust into Ethiopia. *Phytopathology* **107**, 1175–86.

Mizubuti ESG, Aylor DE, Fry WE, 2000. Survival of *Phytophthora infestans* sporangia exposed to solar radiation. *Phytopathology* **90**, 78–84.

Mollentze N, Nel LH, Townsend S *et al.*, 2014. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proceedings of the Royal Society B* **281**, 20133251.

Rykiel EJ, 1996. Testing ecological models: the meaning of validation. *Ecological Modelling* **90**, 229–44.

Shaw MW, Pautasso M, 2014. Networks and plant disease management: concepts and applications. *Annual Review of Phytopathology* **52**, 477–93.

Shtienberg D, 2013. Will decision-support systems be widely used for the management of plant diseases? *Annual Review of Phytopathology* **51**, 1–16.

Skelsey P, Holtslag AAM, van der Werf W, 2008. Development and validation of a quasi-Gaussian plume model for the transport of botanical spores. *Agricultural and Forest Meteorology* **148**, 1383–94.

Skelsey P, Kessel GJT, Holtslag AAM, Moene AF, van der Werf W, 2009a. Regional spore dispersal as a factor in disease risk warnings for potato late blight: a proof of concept. *Agricultural and Forest Meteorology* **149**, 419–30.

Skelsey P, Kessel GJT, Rossing WAH, van der Werf W, 2009b. Parameterization and evaluation of a spatiotemporal model of the potato late blight pathosystem. *Phytopathology* **99**, 290–300.

Small IM, Joseph L, Fry WE, 2015. Evaluation of the BlightPro decision support system for management of potato late blight using computer simulation and field validation. *Phytopathology* **105**, 1545–54.

Sutrave S, Scoglio C, Isard SA, Hutchinson JMS, Garrett KA, 2012. Identifying highly connected counties compensates for resource

limitations when evaluating national spread of an invasive pathogen. *PLoS ONE* **7**, e37793.

Thébaud G, Sauvion N, Chadœuf J, Dufils A, Labonne G, 2006. Identifying risk factors for European stone fruit yellows from a survey. *Phytopathology* **96**, 890–9.

Zwankhuizen MJ, Govers F, Zadoks JC, 1998. Development of potato late blight epidemics: disease foci, disease gradients, and infection sources. *Phytopathology* **88**, 754–63.