KAIQUE DOS SANTOS ALVES

ESTIMATION OF THE TIME-VARYING APPARENT INFECTION RATE FROM PLANT DISEASE PROGRESS CURVES: A PARTICLE FILTER APPROACH

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de *Magister Scientiae*.

Orientador: Emerson Medeiros Del Ponte

VIÇOSA - MINAS GERAIS

2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T

Alves, Kaique dos Santos, 1995-

A474e
2019

Estimation of the time-varying apparent infection rate from plant disease progress curves : a particle filter approach / Kaique dos Santos Alves. – Viçosa, MG, 2019.

26f. : il. (algumas color.) ; 29 cm.

Texto em inglês.

Inclui anexos.

Inclui apêndices.

Orientador: Emerson Medeiros Del Ponte.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.19-22.

1. Plantas - Doenças e pragas - Modelos matemáticos. 2. Problemas inversos (Equações diferenciais). 3. Método Monte Carlo. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22 ed. 632.5

KAIQUE DOS SANTOS ALVES

ESTIMATION OF THE TIME-VARYING APPARENT INFECTION RATE FROM PLANT DISEASE PROGRESS CURVES: A PARTICLE FILTER APPROACH

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de *Magister Scientiae*.

APROVADA: 12 de agosto de 2019.

Assentimento:

Kaique dos Santos Alves

Kaique dos Santos Alves
Autor

Emerson Medeiros Del Ponte
Orientador

AGRADECIMENTOS

Aos meus pais Cidileia dos Santos Alves e João Alves Neto por todo apoio em todas as minhas decisões. Ao professor Emerson Medeiros Del Ponte pela orientação e confiança; aos professores Dr. Willian Bucker Moraes, Dr. Wellington Betencurte da Silva e Dr. Nik Cunniffe pela colaboração no desenvolvimento da pesquisa; à banca examinadora pela disponibilidade de participar e pelas contribuições acerca da dissertação; aos membros do laboratório de Epidemiologia pelo companheirismo e paciência; à Universidade Federal de Viçosa e ao Departamento de Fitopatologia e ao Programa de Pós-graduação em Fitopatologia; ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

EPÍGRAFE

"Chemical industry and plant breeders have forged fine tactical weapons; but only epidemiology sets the strategy." **(J. E. Vanderplank)**

ABSTRACT

ALVES, Kaique dos Santos, M.Sc., Universidade Federal de Viçosa, August, 2019. **Estimation of the time-varying apparent infection rate from plant disease progress curves: a particle filter approach**. Adviser: Emerson Medeiros Del Ponte.

The parameters of the simplest (two-parameter) epidemiological models that best fit plant disease progress curve (DPC) data are biologically meaningful: one is the surrogate for initial inoculum (y_0) and the other is the (constant) apparent infection rate (r), both being useful for understanding, predicting and comparing epidemics. The assumption that r is constant is not reasonable and fluctuations are expected due to systematic changes in factors affecting infection (e.g. weather, host susceptibility, etc.), thus leading to a time-varying r, or r_k, being $k = 1,2,...,N$ and $N$the final epidemic time. A rearrangement in formulation of these models (e.g. logistic, monomolecular, etc.) can be used to obtain r between two time points, given the disease (y) data are available. We evaluated one of the several data assimilation techniques, the Particle Filter (PF), as an alternative method for estimating r_k. Synthetic DPC data for hypothetical polycyclic epidemics were simulated using the logistic differential equation for scenarios that combined five patterns of r_k (constant, increasing, decreasing, random or sinusoidal); five increasing time assessment interval (Δt = 1, 3, 5, 7 or 9 time units - t.u.); and two levels of noise (0.1 or 0.25) assigned to y_k. The analyses of 50 simulated 60-t.u. DPCs showed that the errors of PF-derived$\hat{r}_k$were lower (RMSE < 0.05) for Δt < 5 t.u. and less affected by the presence of noise in the measure compared with the logit-derived r_k. The ability to more accurately estimate r_k may be useful to increase knowledge of field epidemics and identify within-season drivers of r_k behaviour.

Keywords: Data assimilation. Inverse problems. Sequential Monte Carlo.

RESUMO

ALVES, Kaique dos Santos, M.Sc., Universidade Federal de Viçosa, agosto de 2019. **Estimação da taxa aparente de infecção de variação temporal a partir de curvas de progresso de doença: uma abordagem usando filtro de partículas.** Orientador: Emerson Medeiros Del Ponte.

Os dois parâmetros dos modelos (não flexíveis) de epidemias de doenças de plantas que são ajustados às curvas de progresso de doenças (CPD), representam o inóculo inicial (y_0) e a taxa (fixa) de infecção aparente (r), sendo ambos úteis para entendimento, predição e comparação de epidemias. O pressuposto de que r é constante não é apropriado visto que flutuações são esperadas devidos à mudanças nos fatores que afetam o processo de infecção (ambiente, suscetibilidade do hospedeiro etc.) levando assim a uma r com variação temporal ou r_k, sendo $k = 1,2,...,N$, e No tempo final. O rearranjo das equações desses modelos simples (logístico, monomolecular etc.), pode ser utilizado para obter r entre dois tempos, uma vez que a medida de doença (y) esteja disponível. Nesse estudo, foi avaliado uma das várias técnicas de assimilação de dados, o Filtro de Partículas (FP), como um método alternativo para estimar r_k. Dados de CPD sintéticas de epidemias policíclicas hipotéticas foram simuladas utilizando o modelo logístico em cenários em que foram combinados cinco padrões temporais, simulados, de r_k (constante, aumentando, diminuindo, aleatório ou sinusoidal), cinco intervalos de tempo crescentes (Δt = 1, 3, 5, 7 ou 9 unidades de tempo - ut), e dois níveis de ruído (0.1 ou 0.25) para y_k. A análise de 50 CPD simuladas com 60 ut mostraram que os erros$\hat{r}_k$provenientes do FP foram menores para Δt < 5 ut e menos afetados pela presença de ruídos nas medidas quando comparado com resultado do método de rearranjo do modelo. A habilidade de estimar r_k mais acuradamente pode ser útil para aumentar o conhecimento de epidemias de campo e identificar fatores sazonais que possam influenciar no comportamento de r_k.

Palavras-chave: Assimilação de dados. Problemas inversos. Monte Carlo sequencial.

SUMÁRIO

1 Introduction

Mathematical models are powerful tools for describing, analysing and comparing plant disease progress curve (DPC) data. In general, these models are derived from two approaches: 1) statistically, or data-driven, when mathematical (population dynamics) models are fitted to observed or simulated DPC data; and 2) fundamentally, or concept-driven, by linking ordinary differential equations (ODE) for describing system behavior and predicting DPC data or fitting ODEs to estimate model parameters (Madden et al. 2007). The fit of more complex multiple parameter (infection rate, incubation period, host growth rate, etc) epidemiological models is challenging. Hence, simpler models with fewer, yet biologically-relevant, parameters have been preferred and have been used extensively for describing and comparing plant disease epidemics (Kranz 2003; Madden et al. 2007; Jeger et al. 2011a; 2011b).

Traditionally, and most commonly, two-parameter non-flexible population dynamics models (monomolecular, logistic or Gompertz) are fitted to DPC data, and although more flexible ones (with additional parameters) are available, they are seldom used (Berger 1981; Campbell and Madden 1990; Madden et al. 2007). These models describe and allow to predict and compare epidemics: the intercept parameter, representing the disease intensity (y) level at time (k) zero (y_0) (being $k = 0,1,2,\dots,N$, and N the final time), is a surrogate for initial inoculum, and the slope parameter, representing the "apparent infection rate" (r), describe the increase in disease intensity per unit of disease intensity per unit of time (Vanderplank 1963; Campbell and Madden 1990). The rate was called "apparent" due to the appearance of symptoms lagging behind actual infections. Another kind of rate obtained directly from the DPCs, without model fitting, is the absolute rate of disease increase, or the difference (Δ) between disease intensity at two time-points, given by $\frac{dy}{dt}$. While Δ is inherently variable over time according to changes in disease intensity between times, r behaves differently: if the epidemics are not disturbed (by a driver), r remains constant over time.

The (constant) rate parameter is obtained from fitting models to DPC data over the course of the epidemics. However, as discussed by Madden et al. (2007), it is not reasonable to assume a constant behavior for r and fluctuations are expected due to temporal changes in the drivers of infection (e.g. weather, host susceptibility, etc.), thus leading to a time-varying r. In fact, once it is known how r behaves (increasing, decreasing, oscillating, etc.), modifications of the population dynamics models (logistic, monomolecular, etc.) are possible to account for r_k depending on its actual form, which may vary considerably across pathosystems, thus limiting generalizations (Madden et al. 2007).

The traditional way of obtaining r_k during epidemics was presented early by Vanderplank (1963) and further modifications were proposed 20 years later to correct it for host growth (Kushalappa 1980; 1982). However, the uncertainty of the measurements of disease intensity at each time point is not handled directly and it is not known how accurately they represent actualr_k, which in fact cannot be measured like disease intensity.

In this study, we evaluated an alternative method to estimate a time-varying r during the course of the epidemics using one of the various data assimilation modeling techniques. Conceptually, data assimilation methods are those which allow performing space-state estimations in a dynamic system by combining observation and model information, assuming that both have errors (Zhang and Moore, 2015). Specifically, we tested Particle Filter (PF), a Bayesian method that allows parameters to be estimated using the measurement data with a considerable level of uncertainty (Turner and Sherlock 2013). In brief, the PF algorithm uses a measured state-variable (e.g. y) to estimate a parameter/variable (e.g. r) at the same time point, which cannot be measured directly. The basic idea behind the PF is to approximate the posterior probability of the states using a large number of samples (particles) with their associated weights. These particles and weights are then updated sequentially along with the state evolution when new measurements become available (Chen 2003). Primarily used by engineers for solving problems in physics and chemistry, particle filter has been used in human epidemiology to forecast vector-borne epidemics such as influenza (Sheinson et al. 2014; Yang et al. 2014; Dawson et al. 2015; Moss et al., 2016; Ristic and Dawson 2016).

The first application of a filter technique in plant disease epidemiology was the Kalman filter, a predecessor of the PF, used a decade ago for modeling the spatial relationships between grain yield and reflectance measurements of wheat streak intensity estimated across the field (Workneh et al. 2009). Authors of that study showed that filters assuming stochastic trends without slopes or deterministic trends with or without slopes, best described yield-reflectance relationships across the field. Nonetheless, no further study in plant pathology has used particle filters as far as known. Our main objective was to evaluate the particle filter as a method for estimating r_k from disease progress curve data. We did so by creating and analyzing synthetic epidemic data using the differential form of the logistic model for various scenarios that combined different patterns of simulated r_k, five interval lengths in time units between two disease assessment times and two levels of noise in disease data.

2 Material and Methods

A common practice when evaluating the performance of a Particle Filter for parameter estimation is the generation of synthetic parameter data (da Costa et al. 2018; Salardani et al. 2018; Dias et al. 2017). Because r is never known due to impossibility to assess it in plant disease epidemics, and assuming that several different forms of r_k exist for all sorts of plant disease epidemics, we generated synthetic r_k assuming five different patterns: constant, increasing, decreasing, random and sinusoidal. DPCs were then constructed using the differential form of the logistic model, which describes the behavior of polycyclic epidemics (Berger 1981).

The use of a known parameter to predict the measurements is called “direct problem”, while the “inverse problem” is to estimate the parameter based on the known (observed) measurement (Colaço et al. 2012; Kaipio and Somersalo 2005). The inverse problem approach may suffer from inverse crime, or when unrealistic results arise from using the same mathematical model to obtain the predicted and estimated data (Kaipio and Somersalo 2005). To alleviate the issue, a few discrepancies may be introduced, including 1) adding noise to the simulated data; 2) using different time intervals in the forward and the inverse procedures, and 3) using a different mathematical model for the forward and the inverse procedures (Kaipio and Somersalo 2005). In our study, the two first two discrepancies were used by a) assigning a randomly distributed noise to disease intensity value at each time point and b) extracting the values at different time points, according to the different interval length. The third discrepancy was considered not appropriate for our objective of comparing our estimates of r_k with the one used to construct the DPCs, which should be based on the same model.

2.1 Generation of synthetic r_k data

We used the differential logistic model to generate DPC data for a hypothetical polycyclic epidemics. Five different patterns of r_k were simulated. Initial values and equations varied according to the desired pattern (Table 1). The initial inoculum parameter at time 0 was set to 0.1% disease intensity ($y_0 = 0.001$ and r_k values specific to a time-varying pattern (Equation 1).

$$\frac{dy}{dt} = yr_k(1 - y) \quad (1)$$

2.2 Construction of disease progress curves

We produced synthetic DPCs with five increasing fixed time interval length between two disease assessments (Δt = 1, 3, 5, 7 or 9 time units), meaning that k varied from 1 to N by Δt, Nbeing the duration of the epidemics set to 60 time units (t.u.). A normally distributed random noise was assigned to predictions of y_k by:

$$y_k = N(y_k, \delta_k), \tag{2}$$

where $N(y_k, \delta_k)$ is normally distributed with mean y_k and non-linear standard deviation $\delta_k = \alpha y_k(1 - y_k)$, with normal approximation to the binomial, during the epidemic and maximal when disease intensity values reach 50%. We disturbed the epidemics by adding two values of noise (α = 0.1 or 0.25) to y_k. The ordinary differential equations (ODEs) were solved using the ode() function of the deSolve package (Soetaert et al. 2010) of R (R Core Team 2018).

Table 1: Initial values of the apparent infection rate (r) at time 0 (r_0) and the respective equations for updating r (i.e. r_k) following different temporal patterns, over the course of a simulated polycyclic epidemics of 60 time units using the differential logistic model.

Time-varying r	Initial value ($r0$)	Updated value ($rk + \Delta t$)
Non variable (constant)	0.2	$r_{k+1} = r_k$
Increasing	0.05	$r_{k+1} = r_k + 0.0035$
Decreasing	0.3	$r_{k+1} = r_k - 0.0035$
Sinusoidal	0.2	$r_{k+1} = r_k + \frac{\pi\cos(\pi\frac{k-1}{30})}{360}$
Random[1]	0.2	$r_{k+1} = r_k + 0.05v_k$

[1] v_k is a noise added to rate based on a uniform distribution bounded between -1 and 1.

2.3 Particle filter estimation

Particle Filter (PF) is a Monte Carlo procedure also known as Sequential Monte Carlo (SMC), used here to estimate the parameter of interest, based on the measurements. Among a few different particle filters, we opted test one that has proven superior compared with the first filter algorithms, the Sequential Importance Resampling PF (SIR-PF) (More details in Gordon et al. 1993, Chen 2003, Colaço et al. 2012, Dias et al. 2017).

Synthetic measurements of y were produced and the target parameter to estimate was $\hat{r}$ which are both linked by an epidemiological model. In the algorithm, both $\hat{y}_{k+1}$ and $\hat{r}_{k+1}$are estimated simultaneously from an independent randomly-generated samples (the particles) of $y_{k+1}$and r_{k+1} obtained from the *priori* $\hat{y}_k$and $\hat{r}_k$.

$$r_{i,k+1} = f(\hat{r}_k, v_i) \tag{3}$$

$$y_{i,k+1} = h(\hat{y}_k, \hat{r}_k, \eta_i) \tag{4}$$

Where $i = 1,2,\dots,P$, and P is the number of particles in the sample, v_i is the uniformly distributed noise and η_i is a normally distributed noise. The particles are weighted by their likelihood to the measurement z_{k+1}(the actual value of disease intensity).

Equation 5 provides the weight $w_{i,k}$ for the respective particle. In Equation 6, the likelihood function, where W is the inverse of the covariance matrix, in that, $W = \sigma_0 Inv_{n,n}$, been σ_0a measurement of uncertainty (Ozisik and Orlande 2000).

$$w_{i,k} = \pi(z_k \vee y_{i,k}) \tag{5}$$

$$\pi(z_{k+1} \vee y_{i,k+1}) = (2\pi)^{-n/2}|W|^{-1/2}\exp\{\tfrac{-1}{2}[z_{k+1} - y_{i,k+1}]^T W^{-1}[z_{k+1} - y_{i,k+1}]\} \tag{6}$$

Note that in the likelihood function only y is assigned, meaning that the weights of the particles of y and r are the same. Additionally, Equation 7 provides the weights standardized between zero and one ($0 \leq ws_{i,k} \leq 1$). Then, the *posteriori* estimates $\pi(\hat{y}_{k+1}^{\square} \vee z_{k+1})$ and $\pi(\hat{r}_{k+1}^{\square} \vee z_{k+1})$ are calculated by a weighted mean (see Fig. 1) of the particles and its respective particle as follows:

$$ws_{i,k} = \frac{w_{i,k}}{\sum_{i=1}^{P} w_{i,k}} \tag{7}$$

$$\pi(\hat{y}_{k+1}^{\square} \vee z_{k+1}) = \sum_{i=1}^{P}(ws_{i,k+1}y_{i,k+1}) \tag{8}$$

$$\pi(\hat{r}_{k+1}^{\square} \vee z_{k+1}) = \sum_{i=1}^{P}(ws_{i,k+1}r_{i,k+1}) \tag{9}$$

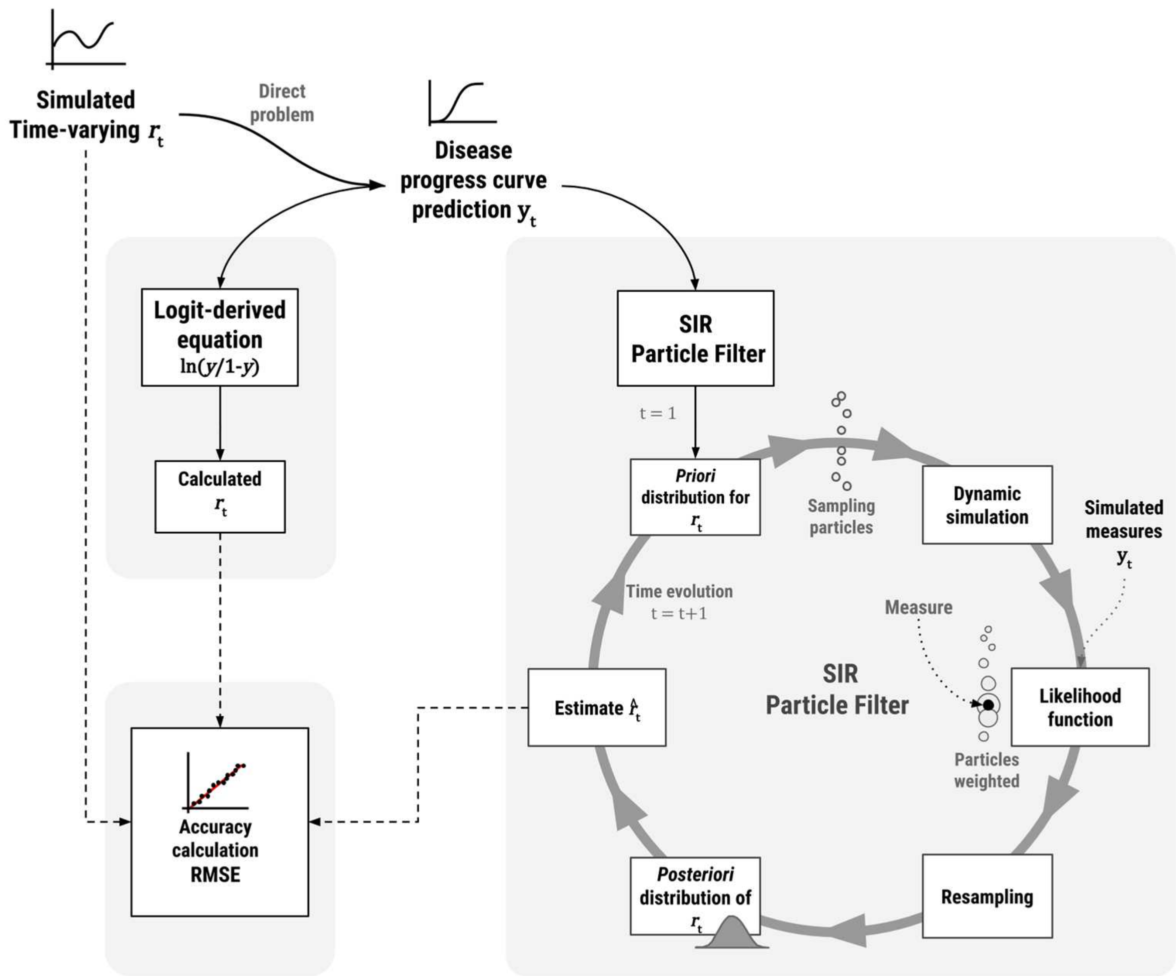

Fig. 1. Workflow of the approach for generating synthetic epidemics by the differential logistic model with different patterns of time-varying apparent infection rate (r_k), and evaluating the performance of a Sequential Importance Resampling Particle Filter (SIR-PF) algorithm to obtain/update $\hat{r}_k$ over time. The error of the PF-derived estimates and logit-derived calculated rates were evaluated by the root mean square error (RMSE) as a measure of accuracy. SIF-PF scheme adapted from Dias et al. (2017).

2.4 Setup for obtaining PF-derived r_k

For generating the particles, a 0.5% uncertainty in *y* was assumed (Equation 10). Such small value was chosen because the variance of the particles should result from the uncertainty in the parameter, i.e the next particles of the next state will be close to the predicted by the model.

$$y_{i,k+1} = \hat{y}_k r_{i,k+1}(1 - \hat{y}_k) + N(\hat{y}_k, 0.005\hat{y}_k)_i \tag{10}$$

To calculate the weights using the likelihood equation, an uncertainty of 25% (maximum noises added in the measurements) was assigned to the measurements at each time step of the epidemics, calculated using the same equation for adding noise in the constructed DPC. In the synthetic epidemic, the level of uncertainty assigned to update the parameter (particle generation for the parameter) was 15% of the previous r_k (Equation 11), assuming that the next state of the parameter will have a value around ±15% of the previous state. Therefore, the evolution model for particle generation of y_k and r_k depends on the uncertainty of both in the model and in the parameter, respectively and, is given by the following where $v_i$is a uniformly distributed noise.

$$r_{i,k+1} = \hat{r}_k + 0.15\hat{r}_k v_i. \tag{11}$$

Eventually, the parameter can be updated at each time step based on estimates obtained using another model (weather-based model, for example), but in this work, the updates in r were assumed to originate from a random process. All estimations were obtained from generating 1,000 particles each time step and 99% credible interval of the normal distribution was obtained. The error of the estimates of r_k was summarized by the root mean square error (RMSE) for simulated values until the epidemics reached 100% severity.

2.5 Apparent infection rate calculation

A reference method for calculating r_k was based on the difference in the logits (assuming the logistic model) of two consecutive disease intensity, for each Δt defined previously, divided by the respective length in time units (Equation 15) (Madden et al. 2007).

$$r_{k+1} = [\ln(\frac{y_{k+1}}{1 - y_{k+1}}) - \ln(\frac{y_k}{1 - y_k})]/(t_{k+1} - t_k) \tag{12}$$

The index $k + 1$ presented in the equations above is valid for a time step $\Delta t = 1$, for the remaining time steps the index may be read as $k + \Delta t$.

2.6 Data availability and reproducibility

All data from our simulations, analysis and graphical work were conducted using R version 3.5.2 (R Core Team 2018). To fully reproduce our work, a research compendium was organized and all R scripts are available for download (https://osf.io/7nyuj/) (Alves and Del Ponte 2019).

3 Results

Fifty DPCs in total (5 x 5 x 2) were generated from the combination of five temporal patterns of r_k, five Δt (time interval) and twoα(noise) (Appendices: Fig. S1). The typical sigmoid-shaped curves were produced for most epidemics. Epidemic curves for $\Delta t = 3$ t.u., chosen as representative of all patterns produced for the other Δt (Fig S1) are depicted in Fig. 2. All but the DPCs resulted from decreasing r_k reached the upper limit (y = 1.0) within 60 t.u., but varied in their shapes and, consequently, in the area under the disease progress curves (data not shown). For example, DPCs resulting from increasing r_k lagged behind the other DPCs (Fig. 2). The ones resulting from random r_k lagged at the beginning of the epidemics but peaked up later in time, reaching maximum earlier than the constant r_k due the parameter has assumed lower values in the beginning and increased fast later in the epidemic. The effect of the two values of α on the shape of the observed data was more apparent for the shorter rather than the longer Δt (Appendices: Fig. S1).

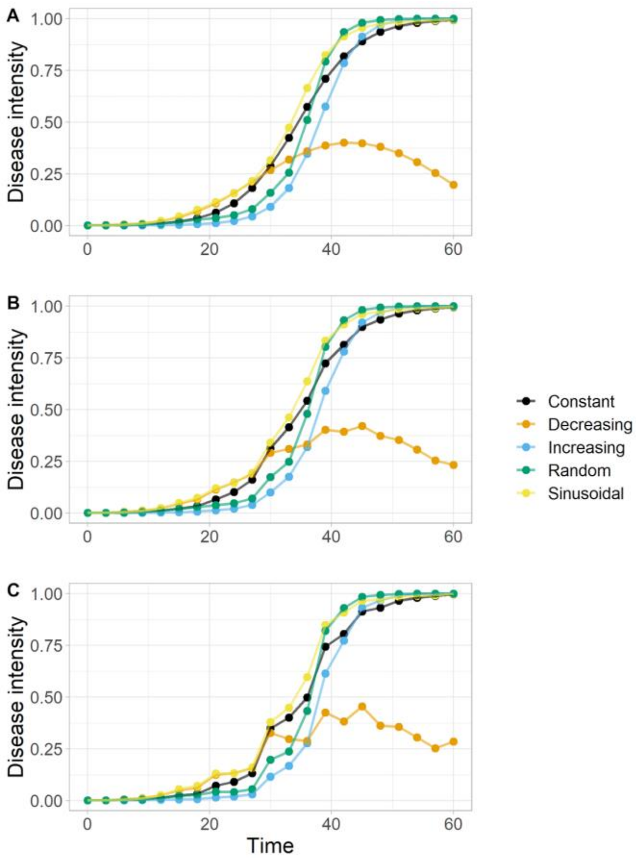

Fig. 2. Generation of synthetic disease progress curves (DPCs) data using a differential logistic model with five different temporal patterns of r_k(colored dotted lines - see equations in Table 1). (**A**) No noise in disease intensity; (**B**) 10% noise ($\alpha = 0.1$); and (**C**) 25% noise ($\alpha = 0.25$).

The $\hat{r}_k$- particle-filter estimates of time-varying rate were generally closer to the synthetic r_k than the logit-derived r_k, regardless of the temporal pattern. However, the superiority of the particle filter compared with logit-derived r_k was more evident for scenarios of shortest Δt and largest α than for scenarios of largest-intervals and lowest-noise (Appendices:: Fig. 3, Fig. 4, Fig. S2 and Fig. S3). The two approaches performed similarly when Δt > 5 t.u., especially for the lowest α(10% noise), and > 7 t.u. for the high α (25% noise) (Fig. 3). The SIR-PF allowed obtaining estimates of disease intensity at values that matched quite well the simulated disease intensity which fell within the 99% confidence intervals of the estimates (Appendices: Fig. S4).

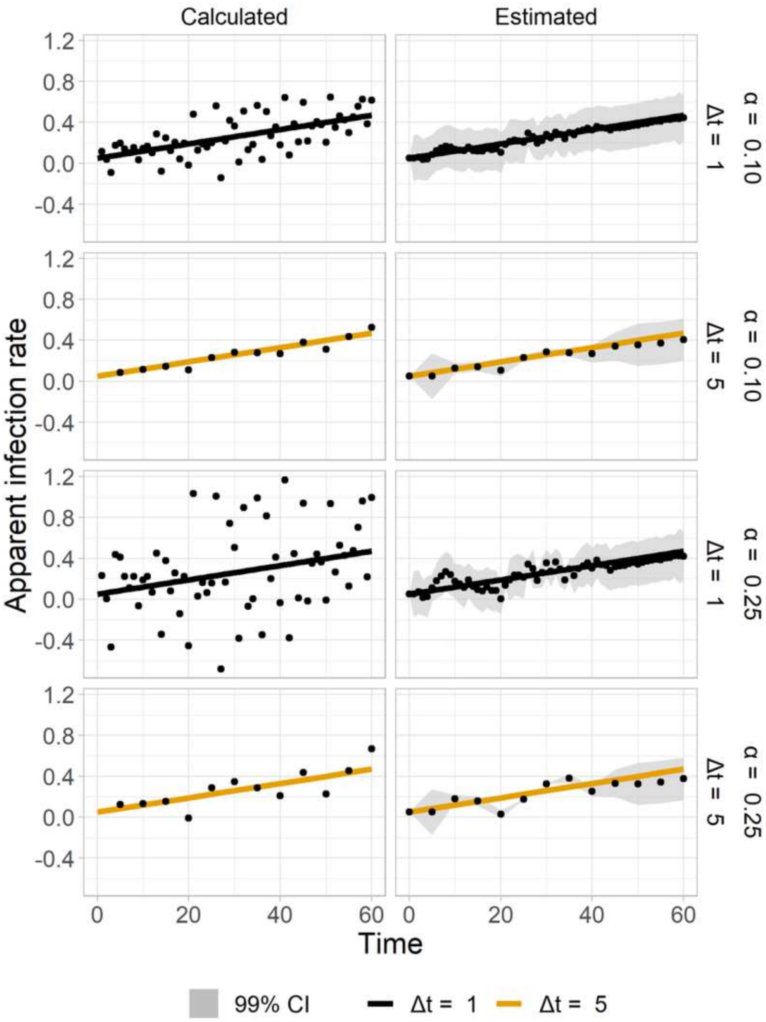

Fig. 3. Increasing apparent infection rate r_k obtained by a rearrangement of the logistic model (left panel) or an estimation method based on the sequential importance resampling particle-filter approach (SIR-PF) (right panel) for scenarios of two intervals between two assessments ($\Delta t = 1$time unit and $\Delta t = 5$time units) and two levels of noise ($\alpha = 0.10$ and $\alpha = 0.25$assigned to disease intensity data. Solid line represents original simulated increasing r values and particle filter- and logit-derived r_k are represented by the dots at the respective panel.

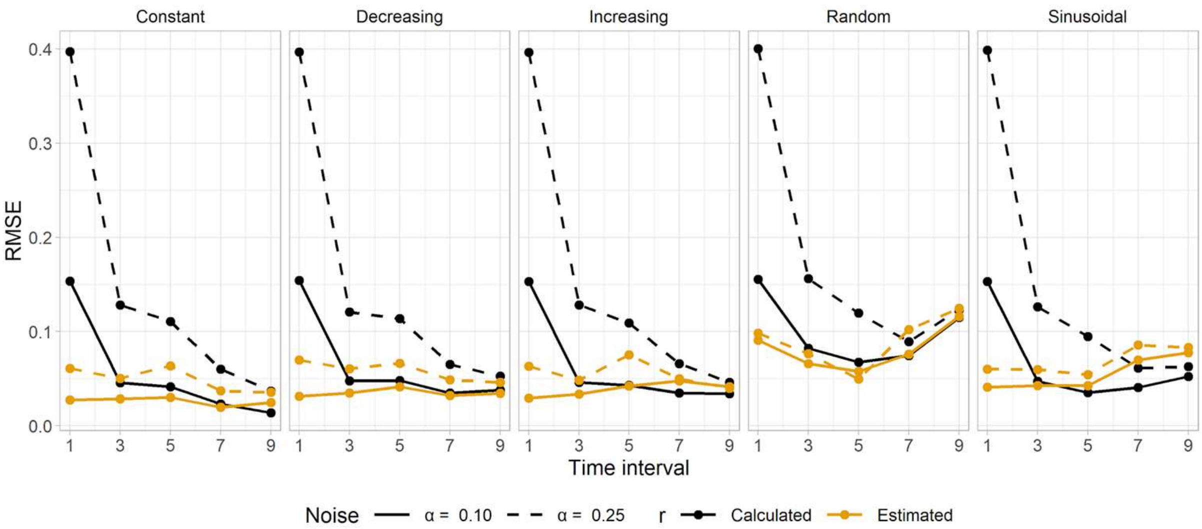

Fig. 4. Accuracy evaluation based on the root mean square error (RMSE) of estimation/calculation of the apparent infection rate r_k using two methods for recovering it: the first, a sequential importance resampling particle-filter (SIR-PF), and the second based on the rearrangement of the logistic model, both compared with simulated rates to account for five systematic r_k patterns (each plot in the panel). The methods were evaluated for different time length (in time units) between two assessments and two levels of noise level assigned to simulated disease intensity simulated by a differential logistic model.

4 Discussion

In this study, an alternative method was proposed and evaluated using a simulation approach for estimating the time-varying apparent infection rate r_k over the course of synthetic epidemics simulated with variable r_k following systematic changes. We showed that the Particle Filter method allowed to obtain estimates of r_k with lower error than a simple calculation when interval between time assessment was the shortest among evaluated here and the level of uncertainty of the measurements was as high as 25%. State-space modelling, including the particle filter, is a well developed technique that has found practical application in other fields and could be useful to tackle other plant pathology problems, as shown previously for a spatial analysis study (Workneh et al. 2009). Here, an unmeasured parameter (infection rate) was estimated (recovered) from measured ones (disease intensity), the inverse problem.

The need to account for time-varying rate in epidemics has been highlighted as one of the challenges in modeling plant diseases (Cunniffe et al., 2015) and formulations to account for a time-varying rate in population dynamics models have been proposed (see page 115 in Madden et al. 2007). Hence, the ability to obtain more accurate estimates of r_k, especially when disease data are available at relatively short intervals (monthly assessment for diseases of perennial crops, for example) between two disease assessments may help to better describe r_k behavior and investigate factors that drive its behavior.

It is well known that the dynamics of field epidemics are disturbed by within-season continuous variations in weather at short (hourly) time scales which affect processes of the disease cycle, mainly the apparent infection rate. For instance, a decline in r_k obtained from three assessments in time in Cercospora leaf spot (*Cercospora beticola*) epidemics in sugar beet was associated with variations of temperature that turned out less favorable for new infections later in the season (Pundhir and Mukhopadhyay, 1987). Waggoner (1986) assembled mathematical equations to account for parabolic and sinusoidal patterns around an optimum r_k in order to reflect the effects of climatic variations on late blight epidemics. Variations in host susceptibility are also known to affect r_k and consequently the patterns of the DPCs (Fillipe and Gilligan, 2003). An exponential decrease in infection efficiency was observed with leaf aging of *Vitis vinifera* affected by powdery mildew (*Erysiphe necator*) (Calonnec et al., 2018). The same leaf aging effect slowing down epidemics has been reported for powdery mildew in strawberry (Carisse and Bouchard 2010) and myrtle rust in *Eucalyptus grandis* (Xavier et al. 2015). Once the drivers are identified, the infection rate can be corrected/updated at each time point. For example, modifications of a generic rice disease simulation model to account for decreasing or increasing the susceptibility of the crop to the disease over time, among other disease-specific parameters, successfully reproduced the typical epidemic curves of rice blast and rice brown spot, respectively (Savary et al. 2012). These curves are somewhat similar to those we produced using the increasing and decreasing r_k (Appendices: Fig. S1). A classical SEI model was modified to simulate slow down of epidemics due to decay in host susceptibility described by time-varying rates of primary and secondary infections (Fillipe and Gilligan, 2003).

Our results showed that the accuracy of the logit-derived r_k values was the poorest for scenarios of high noise (uncertainty) in the measurements and time intervals shorter than 5 t.u., a situation in which the particle filter performed best. In fact, particle filter has shown to perform well when recovering information from data even with large signal-to-noise ratio (Dias et al. 2017; Leung et al. 2016; Gordon et al. 1993). These results may be generalized to any time scale (day, week, month, year) and so the advantage of PF becomes more evident. While epidemics in annual crops are usually assessed at a 7 to 10-day interval, epidemics in perennial crops are usually assessed monthly or weekly.

In some situations, a transient state in disease may occur, i.e. momentaneous increase followed by a sudden decrease or vice versa. This is true for some pathosystems where the host is susceptible only for a short time such as Fusarium head blight in small grains - the host is largely vulnerable to infection during a relatively short period, restricted to flowering, becoming much less susceptible afterwards (Shah et al. 2018; Del Ponte et al. 2005). In managed epidemics using fungicide spray programs, r_k can be greatly reduced, but for a relatively short time. A recent study reported an effective period of fungicides (inferred from reduction of infection rate) for three sole fungicides and a pre-mix for controlling septoria leaf spot in winter wheat, ranged from 16 to 22 days (Greiner et al. 2019). However, differences between two absolute rates (treatment and control) were calculated over time to determine when they were equivalent (end of effective protection). The approach developed here could be used to obtain estimates of r_k with the advantage of accounting for measurements of uncertainty.

The possibility to obtain more realistic estimates of r_k may help to increase our understanding of the epidemics and identify within-season drivers of the r_k behaviour, which could incorporate disease prediction models. Further studies on this specific problem should focus on using data from actual field epidemics. If an informative model (e.g. weather-driven) for r_k is available, future disease levels can be predicted using real-time or forecast weather. Further, besides the estimation of an epidemiological parameter, an unmeasured state variable could be estimated, such as the number of plants at the asymptomatic stage for diseases of relatively long incubation period and for which an informative weather-driven model (oscillating temperature) could be useful as well.

Literature cited

Alves, K. S., and Del Ponte, E. M. (2019, May 2). Research compendium - Estimation of a time-varying apparent infection rate from plant disease progress curves: a particle filter approach. https://doi.org/10.17605/OSF.IO/7NYUJ

Berger, R. D. 1981. Comparison of the Gompertz and logistic equations to describe plant disease progress. Phytopathology. 71:716–719.

Calonnec, A., Jolivet, J., Vivin, P., and Schnee, S. 2018. Pathogenicity traits correlate with the susceptible *Vitis vinifera* leaf physiology transition in the biotroph fungus erysiphe necator: an adaptation to plant ontogenic resistance. Front Plant Sci. 9:1808.

Campbell, C. L. and Madden, L.V. 1990. Introduction to plant disease epidemiology, Wiley, New York, 532 pp.

Carisse, O., and Bouchard, J. 2010. Age-related susceptibility of strawberry leaves and berries to infection by *Podosphaera aphanis*. Crop Prot. 29:969–978.

Chen, Z. 2003. Bayesian filtering: from Kalman filters to particle filters, and beyond. Statistics, 182:1–69.

Colaço, M. J., Orlande, H. R. B., da Silva W. B., Dulikravich, G. S. 2012. Application of two Bayesian filters to estimate unknown heat fluxes in a natural convection problem. J. Heat Transfer. 134:092501(1-9).

Cunniffe, N. J., Koskella, B., E. Metcalf, C. J., Parnell, S., Gottwald, T. R., and Gilligan, C. A. 2015. Thirteen challenges in modelling plant diseases. Epidemics. 10:6–10.

da Costa, J. M. J., Orlande, H. R. B., da Silva, W. B. 2018. Model selection and parameter estimation in tumor growth models using approximate Bayesian computation-ABC. J. Comput. Appl. Math. 37:2795-2815.

Dawson, P., Gailis, R., and Meehan, A. 2015. Detecting disease outbreaks using a combined Bayesian network and particle filter approach. J. Theor. Biol. 370:171–183.

Del Ponte, E. M., Fernandes, J. M. C., and Pavan, W. 2005. A risk infection simulation model for fusarium head blight of wheat. Fitopatol. Bras. 30:634–642.

Dias, A. C. S. R., Silva, W. B., Dutra, J. C. S. 2017. Propylene polymerization reactor control and estimation using a particle filter and neural network. Macromol. React. Eng. 11:1–20.

Filipe, J. A. N., and C. A. Gilligan. 2003. Solution of epidemic models with quenched transients. Physical Review E 67:021906.

Gordon, N. J., Salmond, D. J., Smith, A. F. M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc-F. 140:107.

Greiner, S. D., Racca, P., Jung, J., and von Tiedemann, A. 2019. Determining and modelling the effective period of fungicides against Septoria leaf blotch in winter wheat. Crop Prot. 117:45–51.

Jeger, M. J., Chen, Z., Powell, G., Hodge, S., and van den Bosch, F. 2011. Interactions in a host plant-virus–vector–parasitoid system: Modelling the consequences for virus transmission and disease dynamics. Virus Res. 159:183–193.

Jeger, M. J., van den Bosch, F., and Madden, L. V. 2011. Modelling virus- and host-limitation in vectored plant disease epidemics. Virus Res. 159:215–222.

Kaipio, J., and Somersalo, E. 2005. *Statistical and Computational Inverse Problems*. New York: Springer-Verlag.

Kranz, J. 2003. Comparative Epidemiology of Plant Diseases. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kushalappa, A. C.1980. Um método para calcular a taxa de infecção aparente corrigida para queda e formação de novas folhas. Fitopatol. Bras. 5:411.

Kushalappa, A. C. 1982. Calculation of apparent infection rate in plant diseases: development of a method to correct for host growth. Phytopathology. 72:1373.

Leung, C., Forbes, C. S., Martin, G. M., and McCabe, B. P. M. 2016. Data-driven particle filters for particle Markov chain Monte Carlo. Monash Econometrics and Business Statistics.

Madden, L. V., Hughes, G., Bosch, F. 2007. The study of plant disease epidemics. American Phytopathological Society (APS Press). St. Paul, MN.

Moss, R., Zarebski, A., Dawson, P., and McCaw, J. M. 2016. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. Influenza Other Respir. Viruses. 10:314–323.

Ozisik, M. N. and Orlande, H. R. B. 2000. Inverse heat transfer fundamentals and applications. Taylor & Francis, New York.

Pundhir, V. S., and Mukhopadhyay, A. N. 1987. Epidemiological studies on Cercospora leaf spot of sugar beet. Plant Pathol. 36:185–191.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ristic, B., and Dawson, P. 2016. Real-time forecasting of an epidemic outbreak: Ebola 2014/2015 case study. In *2016 19th International Conference on Information Fusion (FUSION)*, p. 1983–1990.

Salardani, L. S. F., Albuquerque, L. P., da Costa, J. M. J., da Silva, W. B., and Dutra, J. C. S. 2018. Particle filter-based monitoring scheme for simulated bio-ethylene production process. Inverse Probl. Sci. Eng. 1:1–21.

Savary, S., Nelson, A., Willocquet, L., Pangga, I., and Aunario, J. 2012. Modeling and mapping potential epidemics of rice diseases globally. Crop Prot. 34:6–17.

Shah, L., Ali, A., Yahya, M., Zhu, Y., Wang, S., Si, H., Rahman, H. and Ma. C. 2018. Integrated control of fusarium head blight and deoxynivalenol mycotoxin in wheat. Plant Pathol. 67:532–548.

Sheinson, D. M., Niemi, J., and Meiring, W. 2014. Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. Math. Biosci. 255:21–32.

Soetaert, K., Petzoldt, T., and Setzer, R. W. 2010. Solving differential equations in *R*: Package deSolve. J. Stat. Soft. 33 Available at: http://www.jstatsoft.org/v33/i09/.

Turner, L. and Sherlock, C. 2013. An introduction to particle filtering. Lancaster University, Lancaster.

Vanderplank, J. E. 1963. Plant diseases: epidemics and control. Academic Press.

Waggoner, P. E. 1986. Progress curves of foliar diseases: Interpretation and use. Pages 3-37 in: Plant Disease Epidemiology. K. J. Leonard and W. E. Fry, eds. MacMillan Publishing Co., New York.

Workneh, F., Jones, D. C., and Rush, C. M. 2009. Quantifying wheat yield across the field as a function of wheat streak mosaic intensity: A state space approach. Phytopathology 99:432-440.

Xavier, A. A., da Silva, A. C., da Silva Guimarães, L. M., Matsuoka, K., Hodges, C. S., and Alfenas, A. C. 2015. Infection process of *Puccinia psidii* in *Eucalyptus grandis* leaves of different ages. Trop. Plant Pathol. 40:318–325.

Yang, W., Karspeck, A., and Shaman, J. 2014. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics ed. Neil M. Ferguson. PLoS Comput Biol. 10:e1003583.

Appendices

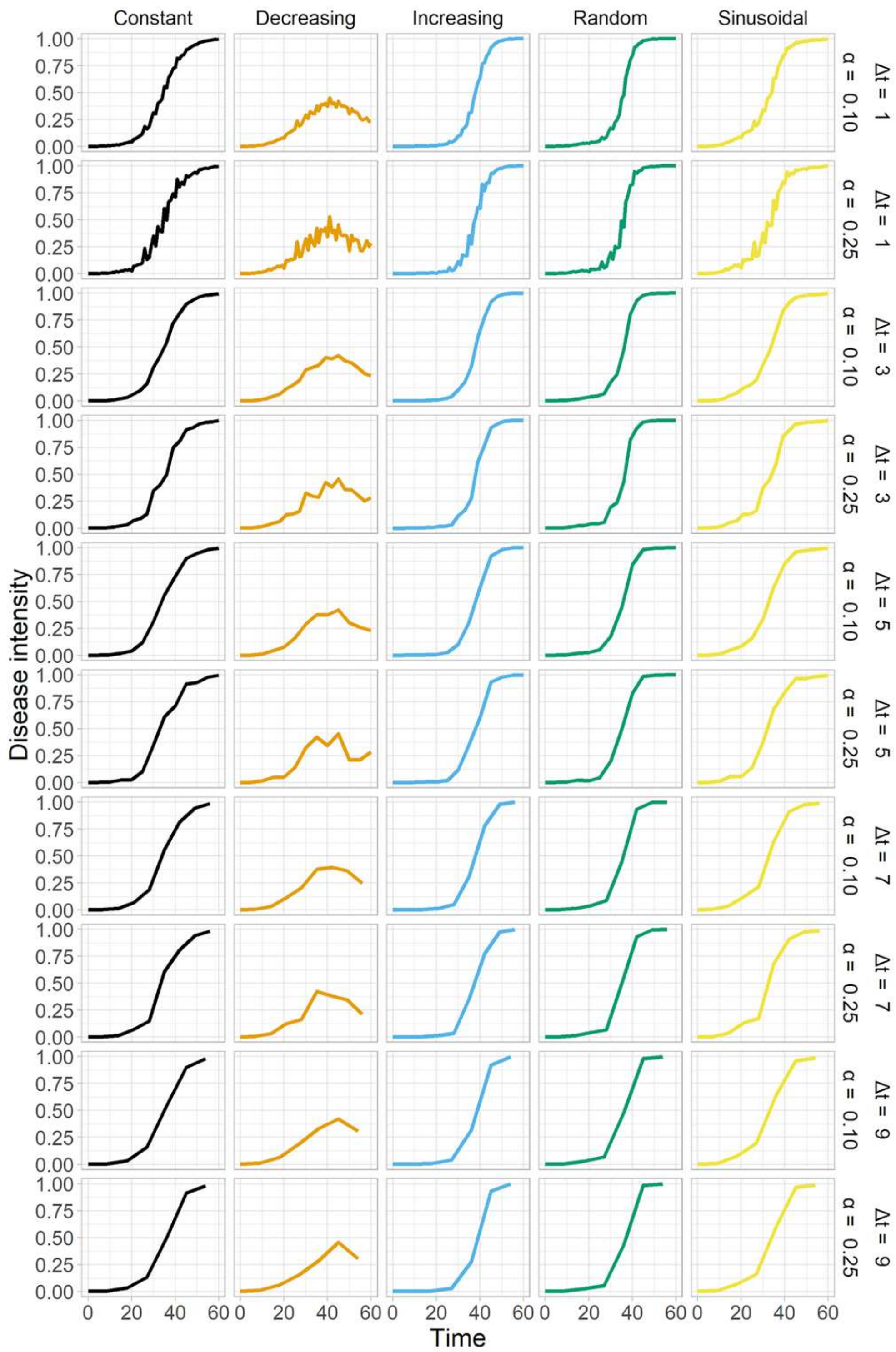

Fig. S1. Disease progress curves of 60-day epidemics simulated by a differential form of the logistic model for five patterns of time-varying apparent infection rate r_k (constant, decreasing, increasing, random and sinusoidal), five intervals in time units ($\Delta t = 1$, 3, 5, 7 and 9) and two levels of noise ($\alpha = 0.1$and 0.25) assigned to disease intensity data.

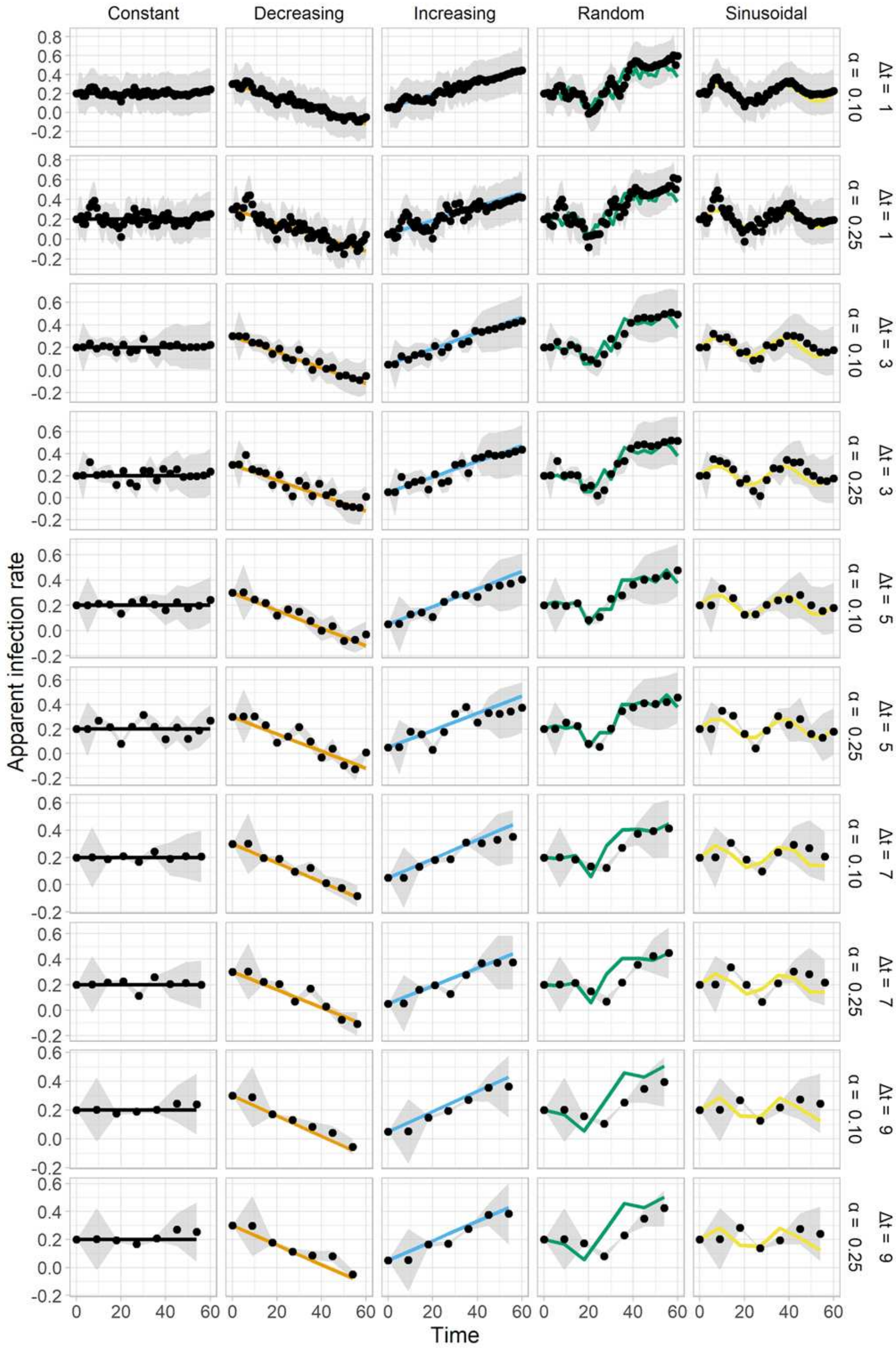

Fig. S2. Particle filter estimates of a time-varying apparent infection rate in synthetic disease progress curves simulated by a logistic model with five predefined patterns of time-varying apparent infection rate r_k (constant, decreasing, increasing, random and sinusoidal), five intervals in time units (Δt = 1, 3, 5, 7 or 9) and two levels of noise ($\alpha = 0.1$or 0.25) assigned to disease intensity data. Solid dots and triangles represent the estimates for the respective level of noise and solid lines represents simulated values of r_k. The shaded area represents the 99% confidence interval of the particle-filter estimates.

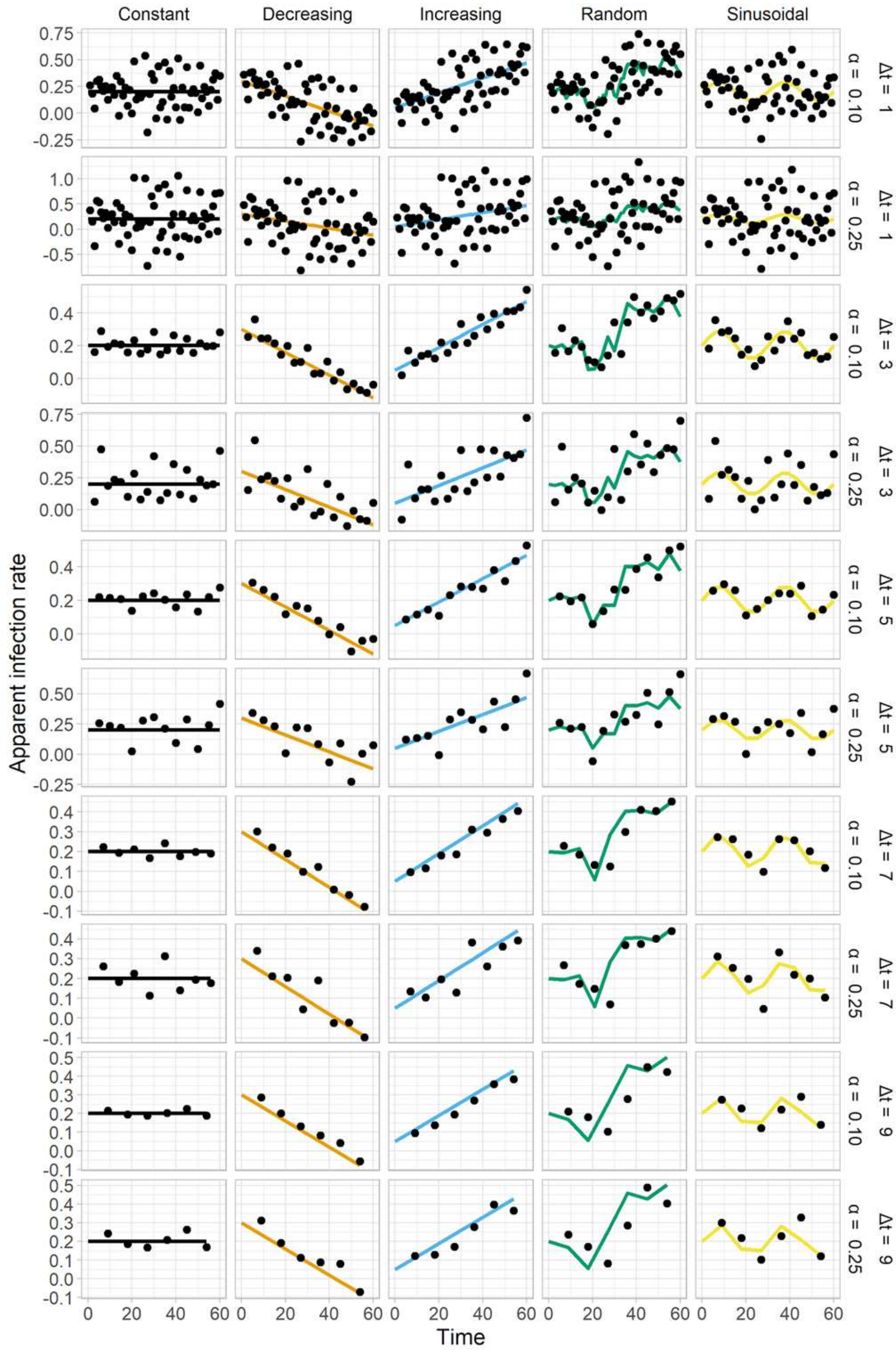

Fig. S3. Calculation of the time-varying apparent infection rate, based on the rearrangement of the logistic equation, in synthetic disease progress curves simulated by a logistic model with five predefined patterns of time-varying apparent infection rate r_k (constant, decreasing, increasing, random and sinusoidal), five intervals in time units (Δt = 1, 3, 5, 7 or 9) and two levels of noise ($\alpha = 0.1$or 0.25) assigned to disease intensity data. Solid dots and triangles represent the calculated r for the respective level of noise and solid lines represents simulated values of r_k.

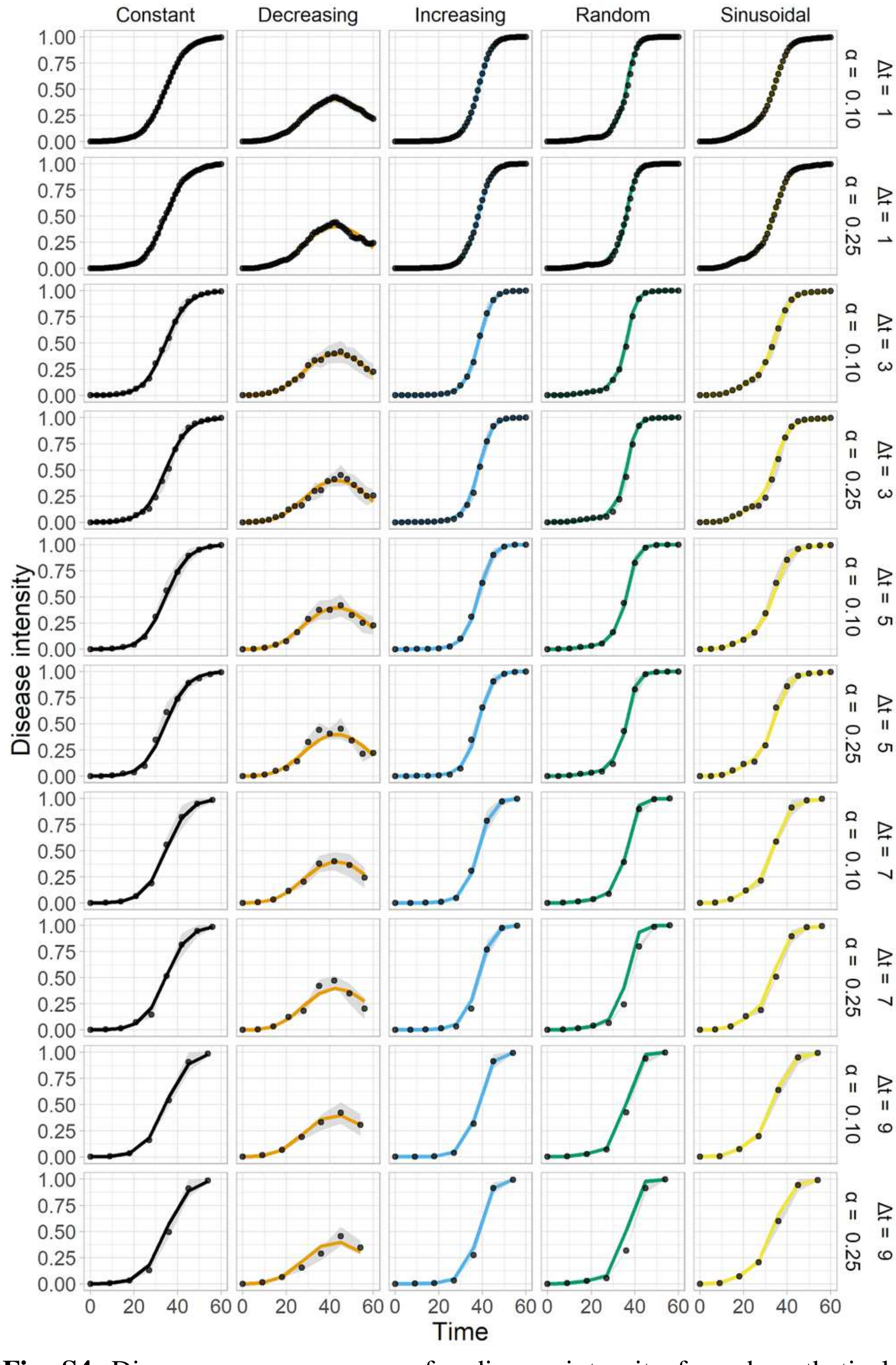

Fig. S4. Disease progress curves of a disease intensity for a hypothetical polycyclic plant disease by a logistic model (solid line) and estimates of disease intensity by a particle filter approach for five patterns of time-varying apparent infection rate r_k (constant, decreasing, increasing, random and sinusoidal), five intervals in time units (Δt = 1, 3, 5, 7 or 9) and two levels of noise ($\alpha = 0.1$or 0.25) assigned to disease intensity data. Dots represent the state estimation by the SIR-PF and shaded area represents the 99% confidence interval.