A peer-reviewed version of this preprint was published in PeerJ on 3 November 2016.

Rall BC, Latz E. 2016. Analyzing pathogen suppressiveness in bioassays with natural soils using integrative maximum likelihood methods in R. PeerJ 4:e2615 https://doi.org/10.7717/peerj.2615

Assessing plant pathogen infection rates in natural soils using R

Björn C Rall Corresp., [1, 2, 3, 4] , **Ellen Latz** [1, 2, 5]

[1] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[2] Institute of Ecology, Friedrich-Schiller Universität Jena, Jena, Germany
[3] Department of Aquatic Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
[4] Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
[5] Department of Animal Ecology, J.F. Blumenbach Institute of Zoology and Anthropology, Georg-August Universität Göttingen, Germany

Corresponding Author: Björn C Rall
Email address: bjoern.rall@idiv.de

The potential of soils to naturally suppress inherent plant pathogens is an important ecosystem function. Usually, pathogen infection assays are used for estimating the suppressive potential of soils. In natural soils, however, co-occurring pathogens might simultaneously infect plants complicating the estimation of a focal pathogen's infection rate as a measure of soil suppressiveness. Here, we present a method in **R** correcting for these unwanted effects by developing a two pathogen mono-molecular infection model. We fit the two pathogen mono-molecular infection model to data by using an integrative approach combining a numerical simulation of the model with an iterative maximum likelihood fit. We show that in presence of co-occurring pathogens uncorrected data critically under- respectively overestimate soil suppressiveness measures. In contrast, our new approach enables to precisely estimate soil suppressiveness measures such as plant infection rate and plant resistance time. Our method allows a correction of measured infection parameters that is necessary in case different pathogens are present. We propose our method to be particularly useful for exploring soil suppressiveness of natural soils from different sites (e.g., in biodiversity experiments).

Assessing Plant Pathogen Infection Rates in Natural Soils using R

Björn C. Rall[1,2,3,4] **and Ellen Latz**[1,2,5]

[1]**German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany**
[2]**Institute of Ecology, Friedrich Schiller University Jena**
[3]**Department of Aquatic Ecology, Netherlands Institute of Ecology (NIOO-KNAW)**
[4]**Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW)**
[5]**Department of Animal Ecology, J.F. Blumenbach Institute of Zoology and Anthropology, Georg-August-Universität Göttingen**

ABSTRACT

The potential of soils to naturally suppress inherent plant pathogens is an important ecosystem function. Usually, pathogen infection assays are used for estimating the suppressive potential of soils. In natural soils, however, co-occurring pathogens might simultaneously infect plants complicating the estimation of a focal pathogen's infection rate as a measure of soil suppressiveness. Here, we present a method in R correcting for these unwanted effects by developing a two pathogen mono-molecular infection model. We fit the two pathogen mono-molecular infection model to data by using an integrative approach combining a numerical simulation of the model with an iterative maximum likelihood fit. We show that in presence of co-occurring pathogens uncorrected data critically under- respectively overestimate soil suppressiveness measures. In contrast, our new approach enables to precisely estimate soil suppressiveness measures such as plant infection rate and plant resistance time. Our method allows a correction of measured infection parameters that is necessary in case different pathogens are present. We propose our method to be particularly useful for exploring soil suppressiveness of natural soils from different sites (e.g., in biodiversity experiments).

Keywords: infected control treatments, maximum likelihood estimation, ordinary differential equation, numerical simulation, biodiversity, soil resistance, bioassay, R, bbmle, deSolve, Manual

INTRODUCTION

Pathogen infection assays are a standard method for estimating plant resistance to pathogens, induced systemic resistance in plants, the effect of artificial or natural plant protectants (e.g. plant beneficial bacteria), and a soil's suppressive potential. Such bioassays compose of a soil or substrate inoculated with a pathogen and a pathogen sensitive plant, and data is collected at just a single point in time (Maurhofer et al., 1994; Pierson and Weller, 1994; Postma et al., 2008) or at multiple points in time (e.g. Postma et al., 2008; Hanse et al., 2011; Latz et al., 2012, 2016). Remarkably, in the latter case often only one single point in time is chosen for evaluation (e.g. Postma et al., 2008; Hanse et al., 2011; Latz et al., 2012), or the increase from one to the next point in time is evaluated (Kushalappa and Ludwig, 1982). However, disease progression is more precisely described by classical growth curve models (Neher and Campbell, 1992). Out of the plethora of growth models (Paine et al., 2012), the mono-molecular model has often been used to describe bioassays with soil-borne pathogens (Stanghellini et al., 2004; Wilson et al., 2008). The mono-molecular infection model describes the disease progression (the change of infections over time) with an initial linear increase of infections (the infection rate), followed by a saturation (given by the maximum number of infectable plants, also known as carrying capacity or asymptotic growth).

The infection rate was suggested to be the most important parameter for determining pathogenicity (Raaijmakers et al., 2009). However, when estimating a soil's suppressive potential, the time until infections occur (resistance time) might be even more important since pathogen inhibition occurs largely during pathogen growth. Actually, only a few experimental setups allow the investigation of both, infection rate and resistance time. To measure an infection rate it is necessary to use a system with multiple plant

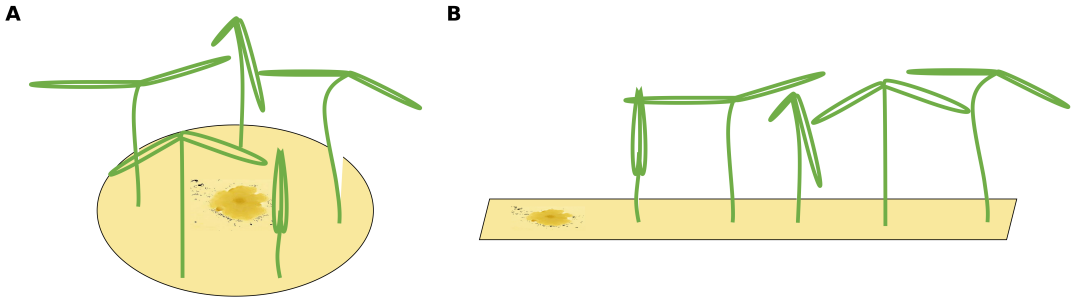

Figure 1. Two different possible setups for infection treatments. The circular setup with a centered pathogen surrounded by plants (a) may lead to a steep linear infection scenario as all plants are probably infected by the source pathogen at more or less the same time. Only the linear spatial assembly (b) allows for a consecutive infection of plants resulting in a linear increase that can be modeled by the mono-molecular infection model.

individuals (Figure 1) where plants can be infected one after another (i.e. measuring a time-series). In such experiments, the pathogen inoculant can be applied in different ways: (i) equally distributed application, i.e. homogeneously mixed in the soil or growth-substrate, or (ii) single point application (where pathogen spread can be assessed; Figure 1). If a pathogen is homogeneously distributed in the plant growth substrate, it is possible to measure the number of infected plants over time. The measured infection rate, however, would not represent the infection rate per se but rather the resistance variance of the plant community to the pathogen. The same problem occurs if a pathogen is applied to one location in the substrate and plants are planted at equal distances around the inoculum (Figure 1A). Linear spatial designs (Figure 1B) have the potential to estimate the correct infection rate in addition to the resistance time, whereas the further mentioned approaches solely allow to estimate the resistance time. Hence, it is important to keep in mind that the design determines the hypothesis that can be tested. Another difficulty in performing pathogen infection assays occurs if natural field soils are used as substrate (e.g. Mendes et al., 2011; Latz et al., 2012, 2016). Here, in addition to the applied pathogen, other unknown pathogens may already exist in the soil and may increase infection in the plants. To cope with this problem, control treatments may be used to reveal the occurrence of natural soil inhabiting pathogens. If controls show infections, (i) these infections might be ignored if they are evaluated as statistically not relevant (Fig. 2A), (ii) the treatments where the corresponding controls showed infections may be excluded from further analyses (Fig. 2B), (iii) the treatments may be linearly corrected by simply subtracting the total amount of infectable plants by the infections that occurred in the control (Fig. 2C). The third approach may lead to erroneous results in non-linear analyses as shown for functional response models (McCoy et al., 2012). However, none of these approaches are desirable as they may lead to a bias in single infection rate measures (due to ignoring or wrongly correcting infections of a naturally occurring pathogen) and the loss of data (exclusion of treatments where the corresponding control was infected).

Here, we present an alternative approach that incorporates infections caused by any additional pathogens in the system by using a two pathogen mono-molecular infection model inspired by the competition model for logistic growth (Lotka, 1925; Volterra, 1926). This two pathogen mono-molecular model is an ordinary differential equation system with two equations. Systems with two equations are hardly analytically integrable to a single equation describing the progress of infections over time, therewith preventing the use of standard linear or non-linear fitting algorithms. To overcome this limitation, we applied a numerical integration routine (Soetaert et al., 2010) combined with a maximum likelihood optimizer (Bolker and Team, 2016) to fit our model to data. Our method allows for the use of natural soils (i) already contaminated with naturally occurring pathogens, and (ii) from different origins and habitats, while allowing for accurate evaluation of pathogenicity and plant resistance patterns in the field.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

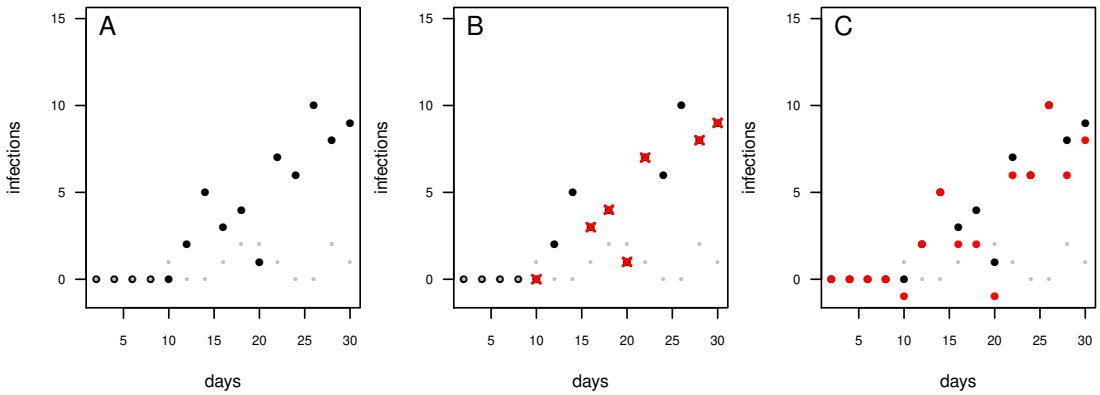

Figure 2. Known practices to deal with infections observed in control treatments. Number of infected plants at each time-point in a single, independent pot. Grey dots: control pots (without having added the pathogen); black dots: treatment pots (with having added the pathogen). (a) Infected controls are ignored and treatment data remains uncorrected. (b) In case the control pot showed an infection the respective treatment data is excluded (red crosses). (c) The treatment data is "corrected" by subtracting the number of infections in the control from the number of infections in the treatment (red dots, note that may lead to some negative infections).

METHODS

Simulations

We solved the differential equation systems (eqn. (2) & (3)) using the `lsoda()`-function (version 1.13; references: Soetaert and Herman, 2008; Soetaert et al., 2010) in R (R Core Team, 2016). The time-series length was set to 30 days with a temporal resolution of 0.01 days. I_{max} was fixed to 10 plants. We simulated two different scenarios; scenario 1: the natural pathogen has lower infection rates ($0.001 \leq r_{control} \leq 0.1$; $0.1 \leq r_{treatment} \leq 0.5$) and occurs earlier in the time series as the treatment pathogen ($1 \leq t0_{control} \leq 5$; $5 \leq t0_{treatment} \leq 10$); and scenario 2: the natural pathogen has comparable infection rates($0.01 \leq r_{control} \leq 0.1$; $0.01 \leq r_{treatment} \leq 0.1$) to the experimentally added pathogen and occurs later in the time series ($5 \leq t0_{control} \leq 10$; $1 \leq t0_{treatment} \leq 5$). We draw all infection rates, r, and time of first infections, t_0, from uniform distributions.

After simulating the time series, we sampled randomly four data-points for each full time-point (i.e. $t = 1, 2, \ldots, 30$) assuming a binomial distribution with a size of I_{max} and a probability of the simulated number of infections at time t divided by I_{max} resulting in 120 independent data points for each simulated infection assay. Additionally simulated one or four consecutive time series resulting in 30 data points of one experimental unit (temporal autocorrelated) and 120 of four experimental units (each time series contains 30 temporal autocorrelated data points). We repeated this simulation of data 1100 times for each scenario. We excluded model fits for both, the one-pathogen model and the two-pathogen model, if the fitting of one or the other failed and used the first 1000 results of the cleaned data set.

Statistical analyses

We analyzed the simulated data using an iterative maximum likelihood algorithm (function `mle2()` from the package **bbmle** version 1.0.18; references: Bolker, 2008; Bolker and Team, 2016) to fit equations (2) & (3) to the data using R (R version 3.3.0; reference: R Core Team, 2016). See the supplemental manual for an in-depth description of the methodology.

We saved all results for the one-pathogen model fitting and the two-pathogen model fitting for each scenario and each setting (independent, one time series and four time series). Subsequently, we analyzed the log_{10}-ratio of the fitted parameters to the initially simulated values. The starting values for infection rates where set to 50% of the simulated value and for resistance time to 75% of the simulated value. In scenario 2, the starting values for infection rates where set to 50% of the simulated value and for resistance time to 0.5 days of the simulated value.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

Additional Files

Additional file 1 — MainSources.zip

This compressed folder contains the sub-folders "data", "script" and "source". The folder "data" contains the data to reproduce figure 4, "data_scenario01.csv" & "data_scenario02.csv" shown in figure 4. The folder "script" contains the script files "scenario01.r" to "scenario06.r" that allow for reproducing the data shown in figure 4. The folder "source" contains the R-source files "infections.models.r" and "infection.nll.r" that are required to run the script files.

Additional file 2 — manual-assessing-plant.pdf

This document includes an in-depth description on how to apply the method presented in this study in R. Including how to create regression lines, trouble shooting, how to use the functions if there are different Imax, and an in-depth description of the source files.

Additional file 3 — ManualSources.zip

This compressed folder includes all necessary data, scripts and source files to reproduce the statistics and plots from the manual.

RESULTS AND DISCUSSION

The Model

The mono-molecular infection model (Raaijmakers et al., 2009; Paine et al., 2012) describes the increase of infections in a (plant) community over time, $dIdt^{-1}$, by:

$$\frac{dI}{dt} = r\,(I_{max} - I) \tag{1}$$

with r [time^{-1}] being the infection rate and I_{max} [Infected (Plants) Area^{-1}] being the maximum number of potentially infectable plants.

The infection of the first plant is not necessarily instantaneous, but depends on the resistance of the soil and the plants to the pathogen, leading to a lag phase at the beginning of the experiment. To account for this mechanism, we extend the mono-molecular infection model by the resistance time, t_0:

$$\frac{dI}{dt} = \begin{cases} 0 & \text{if } t < t_0 \\ r\,(I_{max} - I) & \text{if } t \geq t_0. \end{cases} \tag{2}$$

Below t_0 new infections are zero whilst above, the occurrence of new infections follows the mono-molecular infection model. We will refer to this model as one-pathogen model (Fig. 3A, B).

In experiments using natural soils, natural occurring pathogens may be responsible for additional infections during the experimental trial. To correct for those infections, we extend the one-pathogen model to a two-species mono-molecular infection model, inspired by the two-species competition growth model (Lotka, 1925; Volterra, 1926):

$$\begin{aligned} \frac{dI_p}{dt} &= \begin{cases} 0 & \text{if } t < t_{0_p} \\ r_p\,(I_{max} - (I_p + I_c)) & \text{if } t \geq t_{0_p}, \end{cases} \\ \frac{dI_c}{dt} &= \begin{cases} 0 & \text{if } t < t_{0_c} \\ r_c\,(I_{max} - (I_p + I_c)) & \text{if } t \geq t_{0_c}, \end{cases} \end{aligned} \tag{3}$$

where I_p is the number of infected plants due to the pathogen, I_c is the number of infected plants in the control; r_p and r_c are the infection rates of the pathogen and the control treatment, respectively; and t_{0_p} and t_{0_c} are the resistance times of the pathogen and the control treatment, respectively. We will refer to this model as two-pathogen model.

Below, we will give two examples of different model-parameter combinations, based on two different biological examples that might lead to two different misleading fitting results if the one-pathogen model is used in case of contaminated pots.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

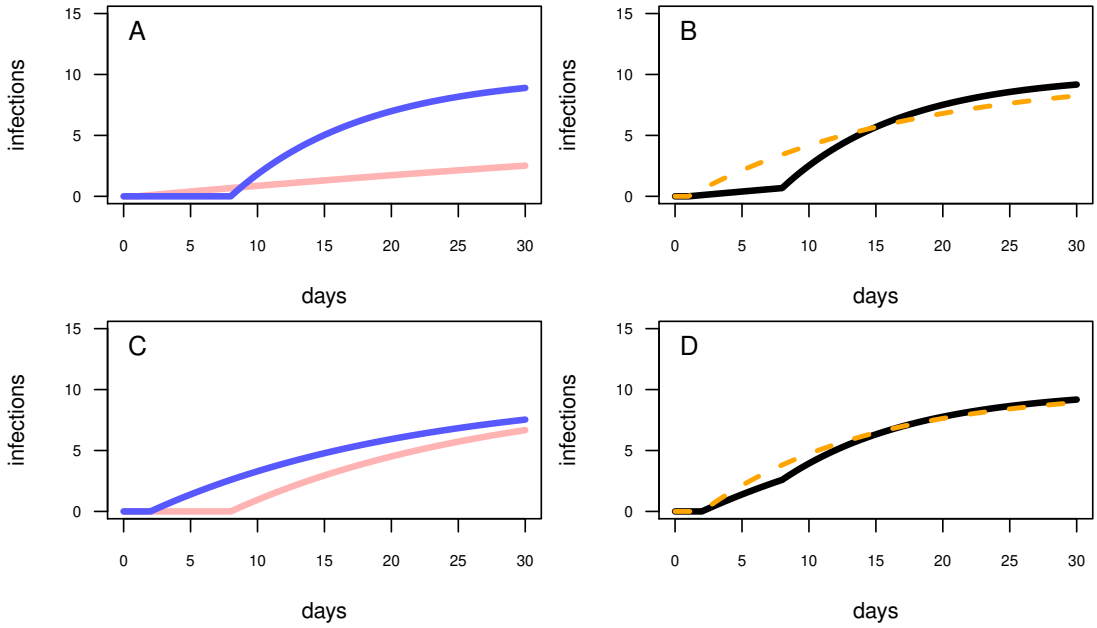

Figure 3. Different model configurations. (a) The one-pathogen model with 2 different settings of parameter values (light red line: $r = 0.01$ & $t_0 = 1$; blue line: $r = 0.1$ & $t_0 = 8$). (b) The two-pathogen model (black line) incorporates the parameter values of (a) and lies slightly above the one-pathogen model (blue line in (a)); hypothetically, using a one-pathogen model to fit the black line will result in a different parameter estimation (dashed orange line). (c) The one-pathogen model with 2 different settings of parameter values (light red line: $r = 0.05$ & $t_0 = 8$; blue line: $r = 0.05$ & $t_0 = 2$). (d) The two-pathogen model (black line) incorporates the parameter values of (c) and lies only above the one-pathogen model (blue line in (c)) at a late stage of the experiment, hypothetically using a one-pathogen model to fit may result in the orange model fit.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

First, we assume a high infection rate r, and an experimental pot showing a high resistance time, t_0. This will result in a first half of the experiment without any infections while in the second half of the experiment the plants will become infected rapidly (Fig. 3A, blue line). We interpret in this case an experimentally added pathogen (treatment pathogen) being inoculated in a defined distance to the seedlings, a soil showing high suppressivess and/or highly resistant plants (high resistance time), but the pathogen being highly abundant and able to infect plants rapidly after the first infection (high infection rate). However, this scenario presumes sterile soil previous to having added a treatment pathogen, whereas natural soils might be contaminated by already naturally occurring pathogens. A contaminated control pot without an experimentally added pathogen may then show early infections followed by a shallow increase of infections over time (Fig. 3A, light red line). The combined progression of the infections over time in a contaminated treatment pot is more complex than that of assuming only a treatment pathogen being present, with showing a shallow increase of infections at low densities and a steep increase of infections in the second half of the experiment (Fig. 3B, black line). Applying the one-pathogen model to estimate the resistance time and infection rate would lead to a misleading fit (Fig. 3B, dashed orange line).

Second, we assume the plants having a rather small resistance time, t_0, and the pathogen being less aggressive (low infection rate,r; Fig. 3C, blue line. Here, we assume a perfectly sterile experiment for both, the treatment and the control. In this example, the control treatments should not show any infections over time. However, pathogens could also disperse into the experimental pots during the experimental trial, leading to late infections of the control (Fig. 3C, red line). This might be the case when experimental pots can not be isolated from the environment, e.g. partially open mesocosms, resulting in more than the treatment pathogen being responsible for infections (Fig. 3D, black line vs . Fig. 3C, blue line). Applying here the one-pathogen model to estimate the infection parameters may lead to the correct estimation of the resistance time but to a underestimation of the infection rate of the treatment pathogen (Fig. 3D, dashed orange line).

In both scenarios, the use of the one-pathogen model would lead to misleading parameter estimations. To overcome this issue the two-pathogen model should be fitted to the data.

Statistical model evaluation

Independent data

We tested our model framework by simulating two separate scenarios (subsequently called scenario 1 and scenario 2). In scenario 1, naturally occurring pathogens infect seedlings earlier than the treatment pathogen, but the naturally occurring pathogens are less infectious (i.e. a lower infection rate, r; Fig. 3A,B). In scenario 2, the naturally occurring pathogens infect the seedlings later than the treatment pathogen but are similar infectious (Fig. 3C,D). We simulated 1000 data sets where each simulated data point represents an independent measure (i.e. the end point of a single time series) for each scenario and fitted (i) the one-pathogen model to each data set (equation (2)) and (ii) the two-pathogen model (equation (3)) to each data set. We compared the fitted parameter values (i.e. the infection rate, r, and the time of first infection, t_0) by taking the log-ratio. See methods for a detailed description of the procedure.

Using the one-pathogen model leads to a systematic underestimation of infection rates, r, (Fig. 4A, lower row) whereas the two-pathogen model performs well (Fig. 4A, upper row). Also, the resistance times, t_0, are underestimated by the one-pathogen model (Fig. 4B, lower row) whereas the two-pathogen model predicts the resistance time very precisely (Fig. 4B, upper row).

The underestimation of both the resistance times and the infection rates nicely reflect our assumptions when fitting the one-pathogen model to the treatment (Fig. 3C). The real increase in infection is rather strong, and coupled to a late first occurrence of infections (Fig. 5A, gray dashed line). But the one-pathogen model estimates a mixed increase of both, the infections caused by the control and the treatment pathogens. This means that the resistance time is driven by the control pathogen leading to an underestimation of infection rates (Fig. 5A, orange dashed line). The two-pathogen model, however, resolves the strong non-linear interactions between the model paraemters and leads to a infection curve with the correct infection rates and resistance times (Fig. 5A, blue line) that lies slightly beneath the total infection (Fig. 5A, black line).

In the second scenario (higher resistance time for the control pathogen with similar infection rates for both) the one-pathogen model overestimates the infection rates systematically (Fig. 4C, lower row). Surprisingly, also the resistance times are overestimated (Fig. 4D, lower row) contrasting our expectations. In contrast, the two-pathogen model predicts the simulated parameter values precisely and outperforms

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

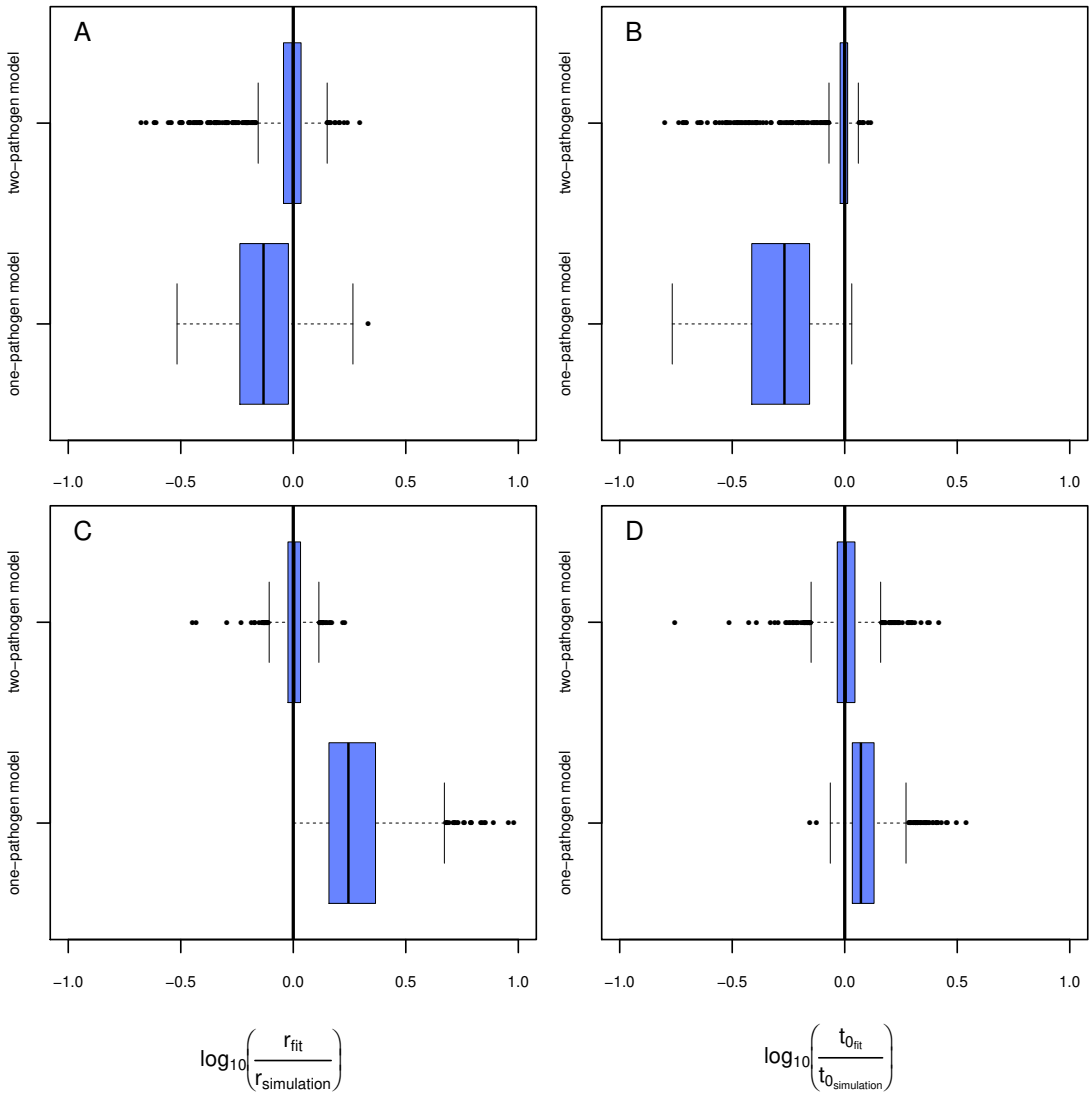

Figure 4. Results of the model evaluation of the one pathogen model versus the two pathogen model. The results of scenario 1 (a & b) and scenario 2 (c & d) for infections rate, r_p, (a & c) and resistance time, t_{0_p} (b & d). The log_{10}-ratio of the parameter fit to the real parameter used for simulating is given on the x-axis. If zero, the fit is perfectly reflecting the simulation, if larger than zero, the fit overestimates the real value, if smaller, the fit underestimates the real value.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

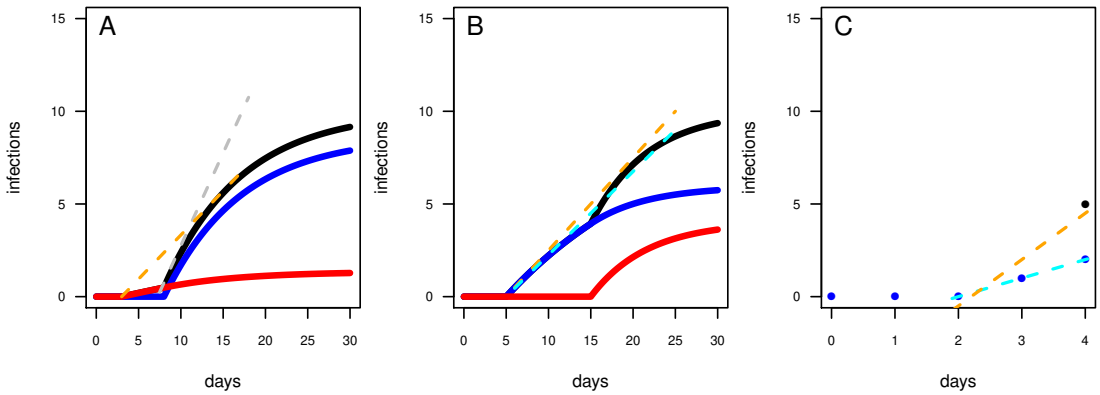

Figure 5. Mechanics of fitted results. (a) The two-pathogen model shows a steep increase (dashed gray line) in infections (black line) when the treatment pathogen enters the system (blue line). The controls, however, lead to infections earlier (red line) leading to a decreased increase in infections using a one-pathogen model for fitting to contaminated data (dashed orange line). (b) The infection rates are overestimate by using a one-pathogen model (orange line) to a treatment with two pathogens (black line). The real treatment infections must be lower (dashed cyan line) as only a part of the infections are caused by the treatment pathogen (blue line), the rest is caused by the control pathogen (red line). (c) The resistance time is mainly inferred by the fit using knowledge on the correct infection rates (dashed cyan line), if the infection rate is overestimated due to additional late occurring control infections (black dot) the resistance time is also overestimated (dashed orange line).

the one-pathogen model dramatically (Fig. 4C,D, upper rows). The overestimation of the infection rates by the one-pathogen model can be explained by an additional boost of infections later in the experiment (Fig. 5B, black line) by additional infections of the control pathogen (Fig. 5B, red line) additionally to the infections of the treatment pathogen (Fig. 5B, blue line). This additional infections lead to an increase in estimated infection rates (Fig. 5B, orange line) compared to the prediction of the isolated infections of the treatment pathogen (Fig. 5B, cyan line). Interestingly, the resistance times are also overestimated. This is a rather small effect and may caused by the fact that, if the correct resistance time lies between two time steps (e.g. $t_0 = 2.1$), the next full time step (e.g. $t_0 = 3$) may show the first infection and the third time step the second infection we expect a rather linear increase from zero to two from time step 2 to 4 (Fig. 5C, cyan line). If a control pathogen also causes an infection at the third time step, the fitting algorithm will estimate a steeper increase to the cost of a higher estimates resistance time (that must still be below 3 in this example, Fig. 5C, orange line).

Consecutive time-series data

For the above described model comparison we used data that consisted of independent measures. This means each data point was derived from a single experimental pot that has been destructively sampled. If applying this approach to an experiment running 30 days with a resolution of one measurement per day and 4 replicates the total amount of pots that must be maintained is 120 (as in our above described analyses). Applying an additional gradient (e.g. biodiversity) would lead to a not feasible amount of experimental units. To avoid such a laborious approach, most studies measure consecutive time series where data for each temporal replicate originates from the same experimental unit. To test if our model approach is also able to fit such data adequately we simulated: (1) data of a single time series resulting in 30 measures from one experimental pot; (2) data of four time series resulting in 120 measures from four experimental pots. We only applied the two-pathogen model to the simulated data. Subsequently, we compared the deviations of the model fits to the original simulated parameter values and we cross-compared the quality of the fits using independent data (120 measures from 120 experimental pots).

Fitting the model to data from a single time series in scenario 1 (Fig. 6A,B, topmost rows, naturally occurring pathogens infect the plants earlier but less strong) leads to a slight overestimation of infection rates but in average correctly estimated resistance times. Using data from four consecutive time series (Fig. 6A,B, middle rows) results in a very precise fit that is not distinguishable from the fit using independent data (Fig. 6A,B, lowermost rows). In scenario 2 (Fig. 6C,D, topmost rows, naturally occurring pathogens

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

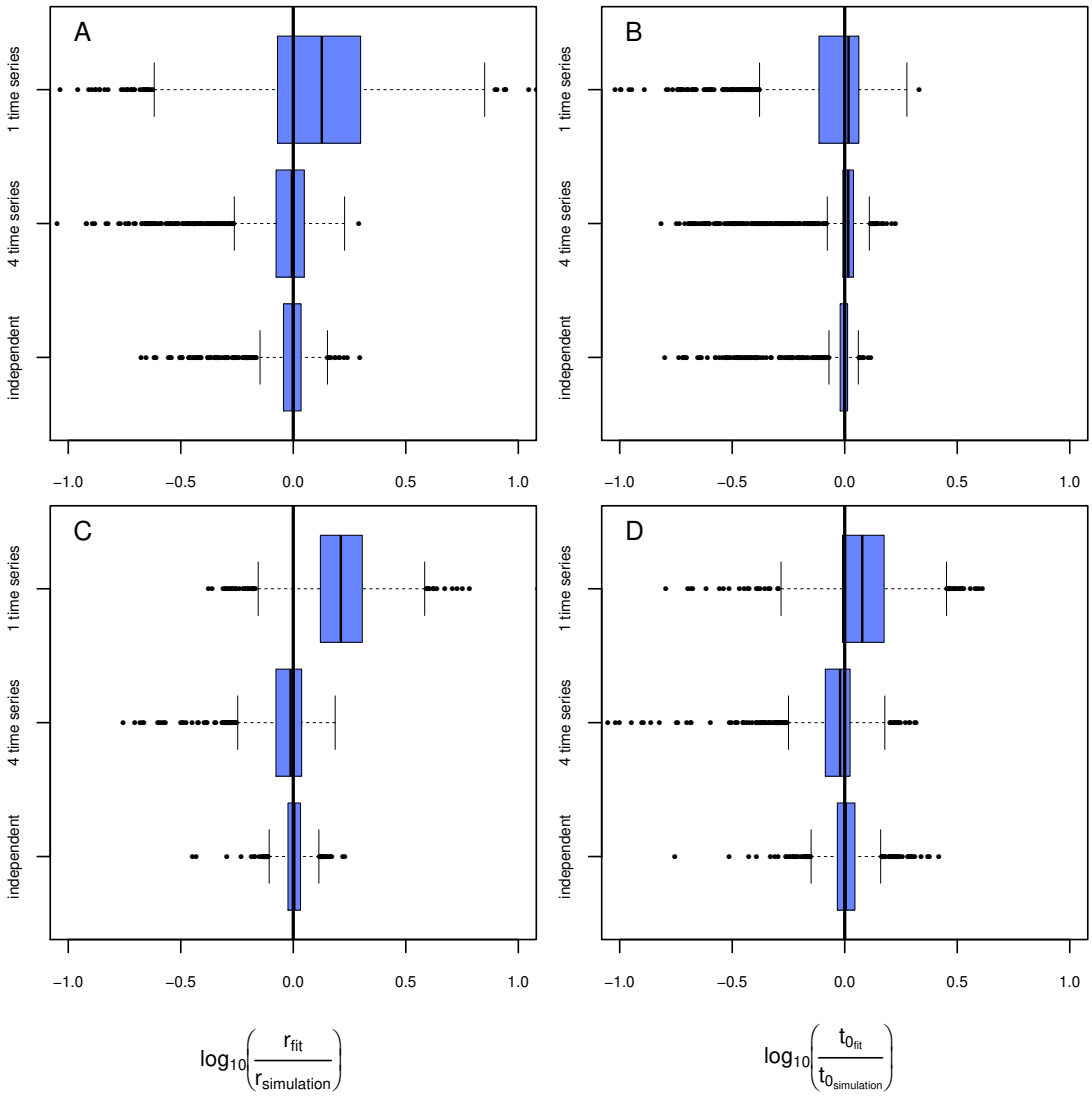

Figure 6. Results of the model evaluation comparing independent measures and consecutive time-series data. The results of scenario 1 (a & b) and scenario 2 (c & d) for infections rate, r_p, (a & c) and resistance time, t_{0_p} (b & d). The log_{10}-ratio of the parameter fit to the real parameter used for simulating is given on the x-axis. If zero, the fit is perfectly reflecting the simulation, if larger than zero, the fit overestimates the real value, if smaller, the fit underestimates the real value.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

infect the plants later but equally strong) both, the infection rate and the resistance time, are systematically overestimated. Using data from four time series to estimate the parameter values statistically increases the preciseness of the fit dramatically and the results do not differ significantly from the expected simulated values (Fig. 6C,D, middle rows) and are only marginally worse than the results from the fit using independent data (Fig. 6C,D, lowermost rows). The systematic overestimation of infection rates in both, scenario 1 and scenario 2, might be reasoned by the fact that in consecutive time series the number of infected plants can only increase opposing the independent measures where infection can also decrease as they are results from independent time series (e.g. Fig. 2A).

General discussion

In both scenarios, the two-pathogen model outclasses the one-pathogen model in predicting both, resistance time and infection rates. Moreover, our approach allows to use data from just a few (in our case: four) consecutive time series reducing the number of pots to be maintained dramatically (in our example 4 versus 120 pots). This reduction of experimental units also allows to investigate the suppressive potential of soils in dependence of other independent variables such as biodiversity, environmental changes (e.g. a nutrient or temperature gradient), diversity and abundance of plant beneficial bacteria or pesticides (see reference (Latz et al., 2016) as an example). To provide a relatively simple entry into our statistical method, we provide the R-code to reproduce all data and statistics presented above. Moreover we provide an in-depth manual as additional online file (see section additional files below for further information). Our model approach should be easily extendable to other kinds of growth or infection models (find other growth models in reference Paine et al., 2012) to e.g. describe pathogen dispersion in larger plant communities or to include more than one treatment pathogen to estimate the competition ability of different pathogens when used together. The statistical method presented here is also superior to classical analytic approaches such as the linearization of the growth model (Neher and Campbell, 1992), the estimation of infection rates by analyzing the initial increase in infections (Kushalappa and Ludwig, 1982), or the arbitrary selection of a single point in time (Maurhofer et al., 1994; Pierson and Weller, 1994; Postma et al., 2008; Hanse et al., 2011; Latz et al., 2012) as it allows (1) to analyze the complete disease progression over time and (2) it allows to correct for naturally occurring pathogens.

CONCLUSIONS

Keystone plants as well as diverse plant communities have shown to increase the pathogen suppressive potential of soils, an effect that would vanish if soils would be sterilized. However, if standard approaches or the one-pathogen infection model is applied, a sterile soil is required to prevent infections by non-treatment pathogens and non-sterile soils consequently prevent the correct estimation of the pathogen suppressive potential of natural soils. This problem can be overcome by using the two-pathogen model presented in this study as it allows for the correct estimation of infection rates and resistance times using natural soils. Our method will thus enable to estimate the natural suppressive potential of soils allowing to investigate how e.g. keystone plants or specifically mixed plant communities naturally contribute to a soil resistance against pathogens.

ACKNOWLEDGMENTS

We thank Myriam Hirt, Amrei Binzer, Andrew Barnes and Ulrich Brose for proof reading our manuscript and beta-testing the R-Code in the manual.

REFERENCES

Bolker, B. (2008). *Ecological Models and Data in R*. Princeton University Press, Princeton, N.J.

Bolker, B. and Team, R. D. C. (2016). ***bbmle**: Tools for general maximum likelihood estimation*. R package version 1.0.18.

Hanse, B., Schneider, J. H. M., Termorshuizen, A. J., and Varrelmann, M. (2011). Pests and diseases contribute to sugar beet yield difference between top and averagely managed farms. *Crop Protection*, 30(6):671–678.

Kushalappa, A. C. and Ludwig, A. (1982). Calculation of apparent infection rate in plant diseases: development of a method to correct for host growth. *Phytopathology*, 72(10):1373–1377.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016

Latz, E., Eisenhauer, N., Rall, B. C., Allan, E., Roscher, C., Scheu, S., and Jousset, A. (2012). Plant diversity improves protection against soil-borne pathogens by fostering antagonistic bacterial communities. *Journal of Ecology*, 100(3):597–604.

Latz, E., Eisenhauer, N., Rall, B. C., Scheu, S., and Jousset, A. (2016). Unravelling linkages between plant community composition and the pathogen-suppressive potential of soils. *Scientific Reports*, 6:23584.

Lotka, A. J. (1925). *Elements of Physical Biology*. Williams & Wilkins Company, Baltimore.

Maurhofer, M., Hase, C., Meuwly, P., Metraux, J., and Defago, G. (1994). Induction of systemic resistance of tobacco to tobacco necrosis virus by the root-colonizing *Pseudomonas fluorescens* strain CHA0: Influence of the gacA gene and of pyoverdine production. *Phytopathology*, 84(2):139–146.

McCoy, M. W., Stier, A. C., and Osenberg, C. W. (2012). Emergent effects of multiple predators on prey survival: the importance of depletion and the functional response. *Ecology Letters*, 15(12):1449–1456.

Mendes, R., Kruijt, M., Bruijn, I. d., Dekkers, E., Voort, M. v. d., Schneider, J. H. M., Piceno, Y. M., DeSantis, T. Z., Andersen, G. L., Bakker, P. A. H. M., and Raaijmakers, J. M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, 332(6033):1097–1100.

Neher, D. and Campbell, C. (1992). Underestimation of disease progress rates with the Logistic, Monomolecular, and Gompertz models when maximum disease intensity is less than 100-percent. *Phytopathology*, 82(8):811–814.

Paine, C. E. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., and Turnbull, L. A. (2012). How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods in Ecology and Evolution*, 3(2):245–256.

Pierson, E. and Weller, D. (1994). Use of mixtures of fluorescent pseudomonads to suppress take-all and improve the growth of wheat. *Phytopathology*, 84(9):940–947.

Postma, J., Schilder, M. T., Bloem, J., and van Leeuwen-Haagsma, W. K. (2008). Soil suppressiveness and functional diversity of the soil microflora in organic farming systems. *Soil Biology and Biochemistry*, 40(9):2394–2406.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raaijmakers, J. M., Paulitz, T. C., Steinberg, C., Alabouvette, C., and Moënne-Loccoz, Y. (2009). The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil*, 321(1-2):341–361.

Soetaert, K. and Herman, P. M. J. (2008). *A Practical Guide to Ecological Modelling: Using R as a Simulation Platform*. Springer, 1 edition.

Soetaert, K., Petzoldt, T., and Setzer, R. (2010). Solving differential equations in R: package **deSolve**. *Journal of Statistical Software*, 33(9):1–25.

Stanghellini, M. E., Kim, D. H., Waugh, M. M., Ferrin, D. M., Alcantara, T., and Rasmussen, S. L. (2004). Infection and colonization of melon roots by *Monosporascus cannonballus* in two cropping seasons in Arizona and California. *Plant Pathology*, 53(1):54–57.

Volterra, V. (1926). Variations and fluctuations of the number of individuals in animal species living together. In Chapman, R. N., editor, *reprinted in: Animal Ecology (1931)*, pages 412–414. McGraw Hill, New York.

Wilson, P. S., Ketola, E. O., Ahvenniemi, P. M., Lehtonen, M. J., and Valkonen, J. P. T. (2008). Dynamics of soilborne *Rhizoctonia solani* in the presence of *Trichoderma harzianum*: effects on stem canker, black scurf and progeny tubers of potato. *Plant Pathology*, 57(1):152–161.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2156v1 | CC BY 4.0 Open Access | rec: 24 Jun 2016, publ: 24 Jun 2016