RESEARCH ARTICLE

Statistical modelling of conidial discharge of entomophthoralean fungi using a newly discovered *Pandora* species

Niels Lundtorp Olsen[ID][1]*, Pascal Herren[2], Bo Markussen[1], Annette Bruun Jensen[2], Jørgen Eilenberg[2]

1 Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark, **2** Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark

* niels.olsen@math.ku.dk

Abstract

Entomophthoralean fungi are insect pathogenic fungi and are characterized by their active discharge of infective conidia that infect insects. Our aim was to study the effects of temperature on the discharge and to characterize the variation in the associated temporal pattern of a newly discovered *Pandora* species with focus on peak location and shape of the discharge. Mycelia were incubated at various temperatures in darkness, and conidial discharge was measured over time. We used a novel modification of a statistical model (pavpop), that simultaneously estimates phase and amplitude effects, into a setting of generalized linear models. This model is used to test hypotheses of peak location and discharge of conidia. The statistical analysis showed that high temperature leads to an early and fast decreasing peak, whereas there were no significant differences in total number of discharged conidia. Using the proposed model we also quantified the biological variation in the timing of the peak location at a fixed temperature.

Introduction

Fungi are important as biological control agents, and their effect is due to their infective spores [1]. The mechanisms for spore releases differ among fungal taxons, one mechanism is shooting off spores as found in fungus order Entomophthorales, which are insect and mite pathogens. Conidia are the infective units of entomophthoralean fungi, and for the majority of species they are actively discharged [2]. The large conidia (mostly between 15 and 40 microns in length) in Entomophthorales demands high energy to be discharged. The spore discharge mechanism for entomophthoralean fungi allow fungi to convert elastic energy into kinetic energy, ensuring that spores are discharged at sufficient speeds.

The infection success depends, among other things, on the attachment of the discharged conidium after landing on host cuticle [3]. The conidia of entomophthoralean fungi are discharged with fluid from the conidiophore, which further assist the conidium to stick to host cuticle after landing [4, 5]. Once the conidia of entomophthoralean fungi are discharged from
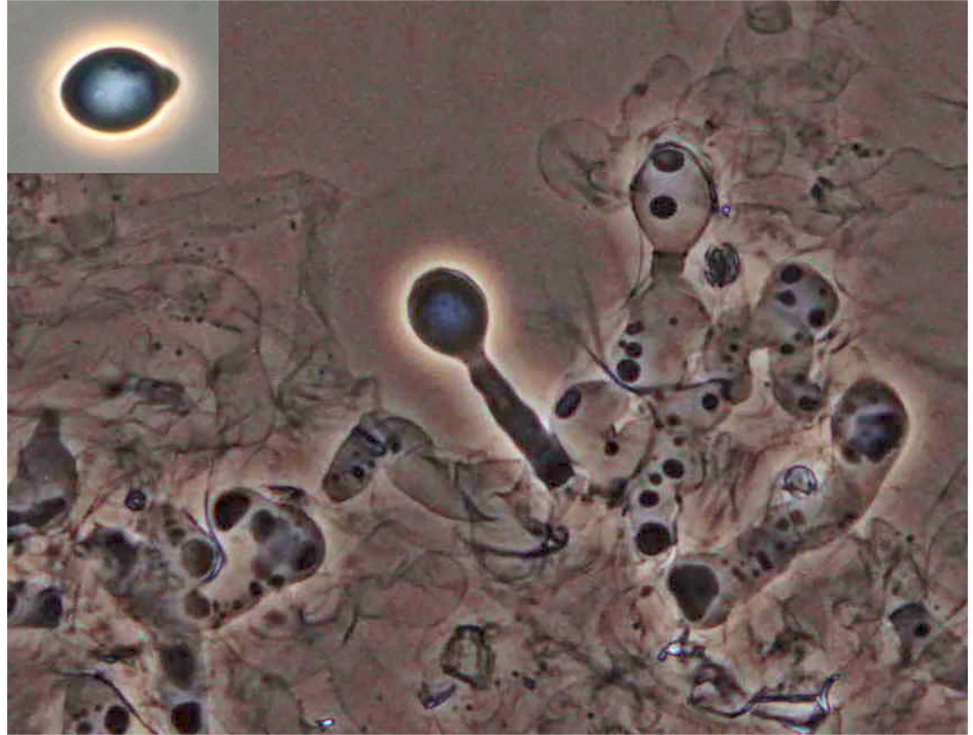
Fig 1. The mycelium of *Pandora* sp. from *Cacopsylla* sp. with a primary conidium on top of a conidiophore and a discharged primary conidium (insert in upper left corner). The length of conidia of *Pandora* sp. is 15.6-23.3 μm [8].

https://doi.org/10.1371/journal.pone.0215914.g001

the conidiophores they have a short longevity [6, 7]. Fig 1 shows a conidium of *Pandora* sp., a species from Entomophthorales isolated from an infected Psyllid in 2016 [8].

The temporal pattern of conidial discharge from infected and dead hosts have been studied for several species of Entomophthorales belonging to the genera *Entomophthora*, *Entomophaga*, *Pandora* and *Zoophthora* [5, 9–12]. The studies show the same overall pattern: after a lag phase of a few hours after the death of the host, conidial discharge is initiated. Depending on host species, fungus species, and temperature, the peak in discharge intensity will be reached within one or two days, thereafter the intensity drops although conidia may still be produced and discharged several days after death of the host. In principle the same pattern appears when conidia are discharged from in vitro cultures. Here the starting point will be when a mycelium mat, which has been grown on nutritious solid medium, is transferred onto for example moist filter paper or water agar, from where conidial discharge will be initiated. The conidial discharge of different species of Entomophthorales is affected by temperature in relation to the intensity and the total number of discharged conidia [10, 13–15].

It is essential to study mechanisms of spore discharge in insect pathogenic fungi [16], effects of environmental factors on spore discharge and also, modelling of spore dispersal in the field [17]. As pointed out by [17], there is a need for more studies on spore discharge and the modelling of dispersal in order to understand natural ecosystem functioning and in order to develop more biological control based on fungi. It is, however, a methodological challenge to study patterns at a quantitative level over time of conidial discharge in Entomophthorales since the system is very dynamic and conidia are sticky. People have therefore used various methods to collect and count discharged conidia. In [18] different methodologies applied to entomophthoralean fungi are reviewed, and a common trait is that the setup should as much

as possible reflect the natural condition, where insects are killed and thereafter initiate discharge of conidia. Different laboratory setups have been used for obtaining discharged conidia counted on glass slides referring to specific time intervals and/or different distances [5, 10, 19]. The data treatments in studies on conidial discharge are mostly rather simple and include for example calculations of mean and standard deviation for replicates, pairwise comparisons or analysis of variance, and a description in words about peak of intensity and length of period with conidial discharge. While these methods are valid and may offer a fair background for conclusions, they nevertheless do not make use of the total biological information in the study.

Statistical modelling of temporal variation in biological systems using functional data

For biological processes that progress over time, we may like to think in terms of idealised systems with a clear time-dependent profile. However, it is often the case that different instances/replications of such processes show some variation in timing. Within statistics, this variation is commonly referred to as *temporal variation* or *phase variation*.

The usual interpretation of temporal variation is *biological time*; that is, the clock of the underlying biological system is out of synchronization with the idealised system (this may be for various reasons), but it is the same underlying processes that are taking place. A common example is puberty for boys: healthy boys enter the pubertal stage, which has some common characteristics for all boys, at some point, but when that happens more exactly varies considerably between individuals.

Inferring the effect of biological time obviously requires replications of the same experiment, and when the underlying structure is a continuous process, such data is naturally handled within the framework of functional data analysis (FDA) [20]. The use of FDA will allow us to estimate sharply defined curves and to estimate and adjust for the variation in peak position [21]. We believe that the effects of temporal variation are sometimes neglected in the biological sciences, something which can lead to weak or in worst cases even misleading conclusions [21].

There are various approaches to modelling functional data with temporal variation (misaligned functional data). We intend to follow the novel methodology of [22], which we will refer to as the *pavpop* model (*Phase and Amplitude Variation of POPulation means*). This methodology has been used in different applications with great success [23, 24]. The main idea of [22] is simultaneous modelling of amplitude and temporal variation, where temporal variation is modelled as a spline interpolation of a latent Gaussian variable that represents temporal deviation from the idealised system. For a review of methods for handling misaligned functional data, we refer to [22, 23].

Whereas classification is often part of papers on misaligned functional data, inference in form of hypothesis testing has got little attention in misaligned functional data. In general, inference in functional data is not easy and requires either strong parametric assumptions, which can be wrong, or the use of non-parametric tests, which can be computationally difficult.

Purpose and content of this study

Overall, we aimed at getting a better understanding of the temporal progression of conidial discharge in entomophthoralean fungi by applying dedicated methods from functional data analysis (FDA). To the authors' knowledge, FDA has not been applied to this kind of studies before, this is a secondary aim of this study. The novelty of this work is also the extension and application of the pavpop model to discrete data, which are generated from an unobserved

biological system with temporal progression. We consider inferential questions, which is new to this methodology as well. In a more broad context, this can be seen as combining the pav-pop model with generalized linear models. In this application we use a negative binomial response model; the appendix describes various other response models.

In this study, discharge of *Pandora* conidia (Fig 1) as a function of time was studied at different temperatures. We hypothesize that high temperature leads to an early and fast decreasing peak when looking at (the intensity of) conidial discharge, whereas a low temperature leads to a late and more slowly decreasing peak, and that a low temperature leads to a higher total production of conidia in the first 120 hours, as compared with higher temperatures.

Methods

Experiment and data collection

A detailed description of the data collection can be found in [25].

We used an isolate of *Pandora* sp. (KVL16-44). It is an insect pathogenic fungus found first time in 2016 in Denmark on *Cacopsylla* sp. (Hemiptera). The conidia have typical *Pandora* morphology, including mononucleate conidia [8]. The fungus is however a new, yet undescribed species based on molecular sequencing [8], and we will therefore in our paper use the name *Pandora* sp.

Mycelium production. The fungus was grown on Sabouraud Dextrose Agar (SDA) supplemented with egg yolk and milk [18] in Petri dishes (55 mm diameter). To produce fresh material mycelium mats were transferred to new petri dishes and incubated at 18.5˚C in dark conditions for 20 days. Using mycelial mats has the advantage, compared to use fungus killed insects, that conidia production can be syncronized more precisely.

Conidia production. Filter papers of 18 x 18 mm, moistened with 0.75 ml of autoclaved water, were placed in the center of lids of petri dishes (34 mm diameter). Four squares of 5 x 5 mm mycelium mats were cut from the same mycelium mat 20 mm away from the centre of the Petri dish and put upside down in the edges of the moist filter papers. All lids were put on the counter parts of the empty Petri dishes and they were kept in three different temperatures (12.0, 18.5 and 25.0˚C) in complete darkness at 100% RH. The mycelium mats were facing downwards. For each temperature, five replicates were made.

Conidia discharge over time. To measure the conidial discharge a small stripe of Parafilm with a cover slip (18 x 18 mm) on top was placed inside the lower part of each petri dish. The dishes were then placed in incubators (12.0, 18.5 and 25.0˚C) for 30 min. The cover slips were placed underneath the four mycelium mats. The cover slips were removed immediately afterwards, and the conidia laying on the slip were stained with lactic acid (95%). This procedure was repeated every eight hours for 120 h, which meant that in total, we obtained observations from 16 time points. The lower parts of the Petri dishes were cleaned with ethanol (70%) and demineralised water every eight hours to ensure that primary conidia did not discharge secondary conidia on the cover slips. The conidia were counted in each of the four corners of the cover slips. In total, we got four observations per time-point, replicate and temperature (4 counts * 5 replicates * 16 time-points * 3 temperatures = 960 observations). Conidia on cover slips were counted with the aid of a light microscope (Olympus Provis) at x 400 magnification.

Statistical modelling

We consider a set of $N = 15$ latent *mean curves*, $u_1, \ldots, u_N : [0, 1] \to \mathbb{R}$ from $J = 3$ treatment groups. The mean curves are assumed to be independently generated according to the

following model

$$u_n(t) = \theta_{f(n)}(v_n(t)) + x_n(t), \quad n = 1, \dots, N \tag{1}$$

where f maps curves into treatment groups. That is, to each subject corresponds a fixed effect θ_j, which is perturbed in time by v_n and in amplitude by x_n, both assumed to be random. The temporal perturbation v_n is usually referred to as a *warping function*.

To each curve corresponds a set of discrete observations $(t_{n1}, y_{n1}), \dots, (t_{nm_n}, y_{nm_n}) \in [0, 1] \times \mathcal{Y}$ where $(t_{n1}, \dots, t_{nm_n})$ are m_n pre-specified time points and $\mathcal{Y} \subseteq \mathbb{R}$ is the sample space for the observations.

We assume that the observations conditionally on the latent mean curves are independently generated from an exponential family with probability density function

$$p(y|\eta) = b(y) \exp(\eta \cdot y - A(\eta, y)), \quad \eta \in \mathbb{R}, y \in \mathcal{Y} \tag{2}$$

where η is the value of the latent mean curve at a given time, and y is the canonical statistic for the observations. A and b are functions defining the exponential family. We assume that $A(\eta, y)$ is two times continuous differentiable in η with the property that $A_\eta''(\eta, y) > 0$ for all η and y, and we assume that all hyperparameters describing A and b are known and fixed beforehand. More details on response models can be found in the appendix.

The amplitude variation x_n is assumed to be a zero-mean Gaussian process. The fixed effects θ_n are modelled using an appropriate spline basis, and the warping functions v_n are parametrised by Gaussian variables $w_n \in \mathbb{R}^{m_w}$ such that $w_n = 0$ corresponds to the identity function on $[0, 1]$. More details on fixed effects and phase variation can be found in the appendix.

Estimation in this model is presented in the appendix. Estimation is a major challenge, as direct estimation is not feasible due to the large number of latent variables. Furthermore, unlike [23], the response is not Gaussian, which require additional considerations. We propose to use a twofold Laplace approximation for doing approximate maximum likelihood estimation; details on the Laplace approximation are found in the appendix.

Data analysis

As described in the data collection section, data consist of 960 observations (4 counts * 5 replicates * 16 time-points * 3 temperatures) in $\mathbb{N}_0$. The largest count was 211, and a large fraction of the counts was zero.

Samples no. 7 and no. 13 effectively terminated the discharge of conidia after 48 and 40 hours, respectively, and therefore measurements from these samples were cancelled after these hours.

Response model. A popular choice for modelling count data from biological experiments are Poisson models. This is backed by a strong theoretical reasoning; using our data as an example, one would expect that while the fungi are placed in the incubators, they would independently discharge conidia at random and at a constant rate. This would be a typical example of a Poisson process.

However, a unique feature of the data set was the four samples taken from each batch used for conidia count, which can reasonably be assumed to be independent and identically distributed conditionally on the latent curve u. By comparing sample means and sample variances across the 240 measurements, this allowed us to assess if data was in reasonable agreement with a Poisson model or overdispersed relative to this. As indicated in Fig 2, data was clearly overdispersed; the Poisson model corresponds to the dotted line.
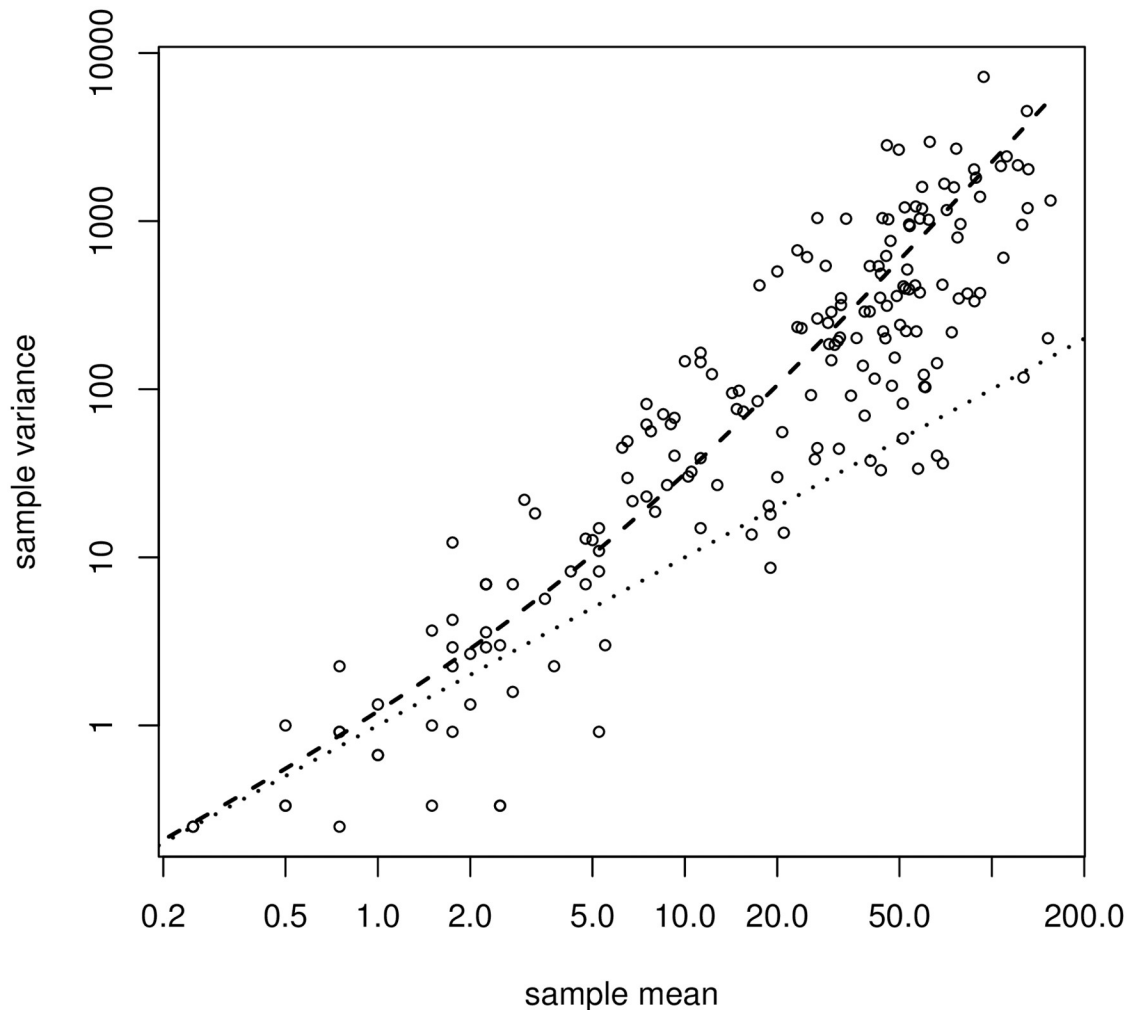
Fig 2. Sample variance as a function of sample means across measurements. Dashed line is fit using a NB(4.66)-model; dotted line is fit using a Poisson model Each sample consisted of four observations.

Because data was overdispersed, we instead fitted an unstructured negative binomial regression model with common rate r to the data instead. This was in good concordance with data: the estimated rate was $r_0 = 4.658$, and the dashed line in Fig 2 indicates the corresponding mean/variance relation. This value was fixed and used in the subsequent analysis.

Having estimated the dispersion, the counts at individual measurements were added for the subsequent analysis as the sum of counts is a sufficient statistic for our model. The sum of independent and identically distributed negative binomial random variables is again negatively binomially distributed; the rate parameter r is multiplied by the number of counts; thus we got $r = 4 \cdot r_0 = 18.63$. The summed counts are displayed in Fig 3.

Model for mean curves. Time was rescaled to the unit interval such that $t = 0$ corresponded to 0 hours and $t = 1$ corresponded to 120 hours. Warping functions were modelled as increasing cubic (Hyman filtered) splines with $m_w = 7$ equidistant internal anchor points with extrapolation at the right boundary point. The latent variables w_n were modelled using a Matérn covariance function with smoothness parameter $\alpha = 3/2$ and unknown range and scale
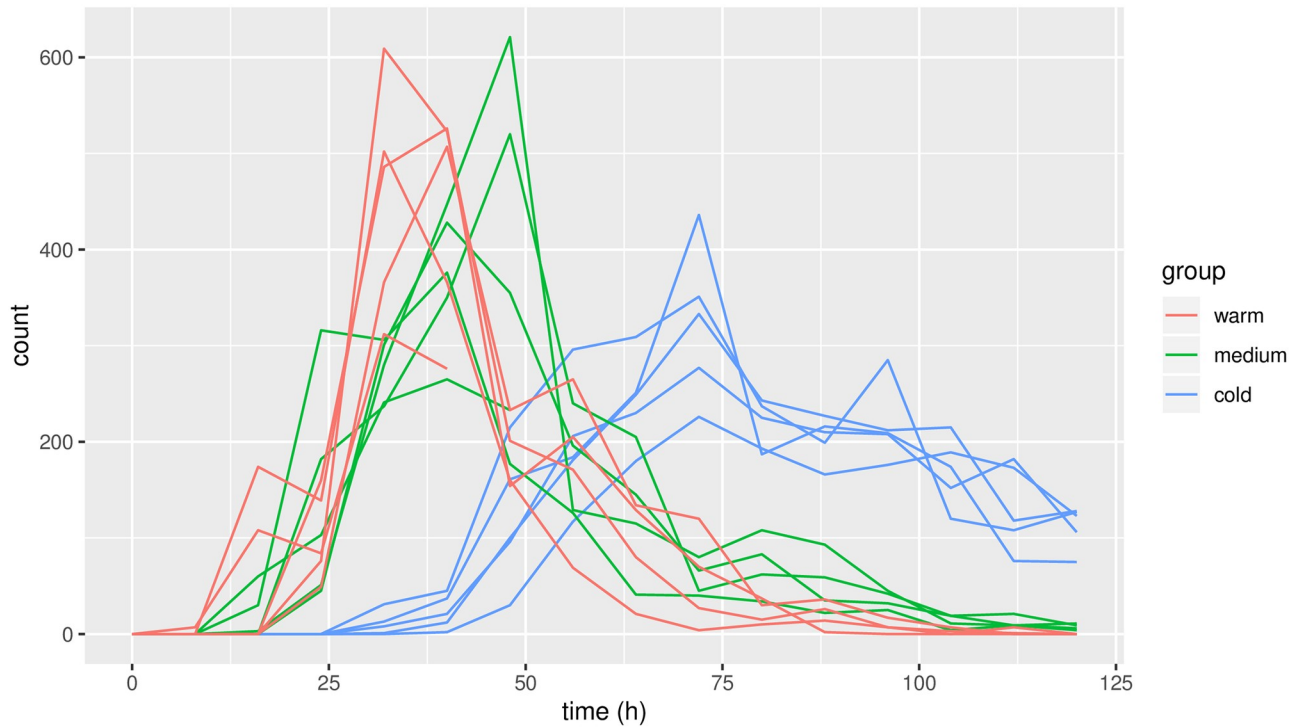
Fig 3. Summed counts of conidia for the individual fungi as functions of time, color-coded according to temperature.

parameters. This corresponds to discrete observations of an integrated Ornstein-Uhlenbeck process. This gave a flexible, yet smooth, class of possible warping functions which also take into account that the internal clocks of individual fungi could be different at the end of the experiment.

Population means θ_{cold}, θ_{medium}, θ_{warm} were modelled using natural cubic splines with 11 basis functions and equidistant knots in the interval [0, 1]. Natural cubic splines are more regular near boundary points than b-splines which reduce the effect of warping on estimation of spline coefficients.

Amplitude covariance was modelled using a Matérn covariance function with unknown range, smoothness and scale parameters; details are provided in the appendix.

Hypotheses. We define 'peak location' as the time with maximal condidial discharge, and 'peak decrease' as the average decrease in discharge between 'peak location' and end of the experiment:

$$\text{peak location}_j = \arg \max \ \theta_j, \quad \text{peak decrease}_j = \frac{\max(\theta_j) - \theta_j(120h)}{120h - \text{peak location}_j}$$

Note that this is defined on log-scale, so peak decrease should be interpreted as a relative decrease of conidial discharge.

One can qualitatively assess the hypotheses without strict definitions, but in order to do statistical inference, a mathematical definition is needed. We remark that here we consider population means; temporal variation may also affect peak location for individual fungi.

Results

Predicted mean trajectories for u, evaluated at observed time points, along with population means are displayed in Fig 4. We observe a slightly odd behaviour around $t = 0$. This is an artifice; most observations around $t = 0$ are zero. When the predicted values of u are exp-transformed, these are mapped into almost-zero values. In concordance with our hypothesis, the three population means are clearly separated and fit well into what we expected: θ_{warm} peaks first and has the highest peak; θ_{middle} is in-between and θ_{cold} peaks latest and has a smaller and more slowly decreasing peak.

Predicted warping functions are displayed in Fig 5. The scale parameter for the warp covariance was estimated to be 0.026; this corresponds to a standard deviation of around 3.1 hours on temporal displacement, or a 95% prediction interval of roughly 6 hours.

The results in Fig 5 are closely connected with those in Fig 4: a vertical change in Fig 5 corresponds to a horisontal change in Fig 4. One may interpret the trajectories in Figs 4 and 5 as *smoothing* of the data: Fig 3 shows the raw data counts; Fig 4 displays the smoothed curves, which are our predictions of the intensity of conidial discharge (the underlying biological quantity of interest) for individual fungi; and finally Fig 5 displays the corresponding predictions of the biological times.

The trajectory for an individual fungus is of little interest by itself as that fungus is confined to this experiment. However, when the trajectories are viewed together, they illustrate the variation on population level allowing us to assess variation between individual fungi from the same treatment group, and also to compare this to fungi from other treatment groups.

Discharge of conidia above certain levels

For practical applications it is relevant to know when the intensity of conidia discharges reaches a given level and for how long this happens. Although one conidium is enough to infect an insect [26], the chance of a conidium landing on an insect is small. Therefore we chose a range from low to very high discharge of conidia. The lowest threshold was 0.5 and the
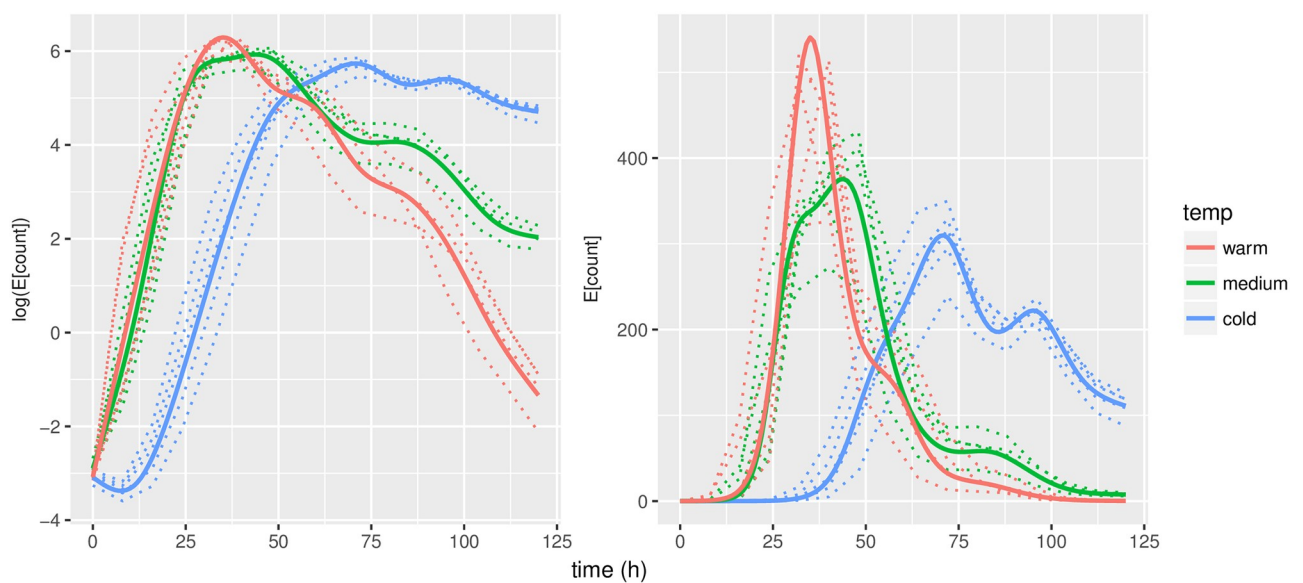

Fig 4. Predicted trajectories for *u* (dashed lines). Left is on model scale, right is exp-transformed (same scale as the observations). Thick lines indicate estimated population means.

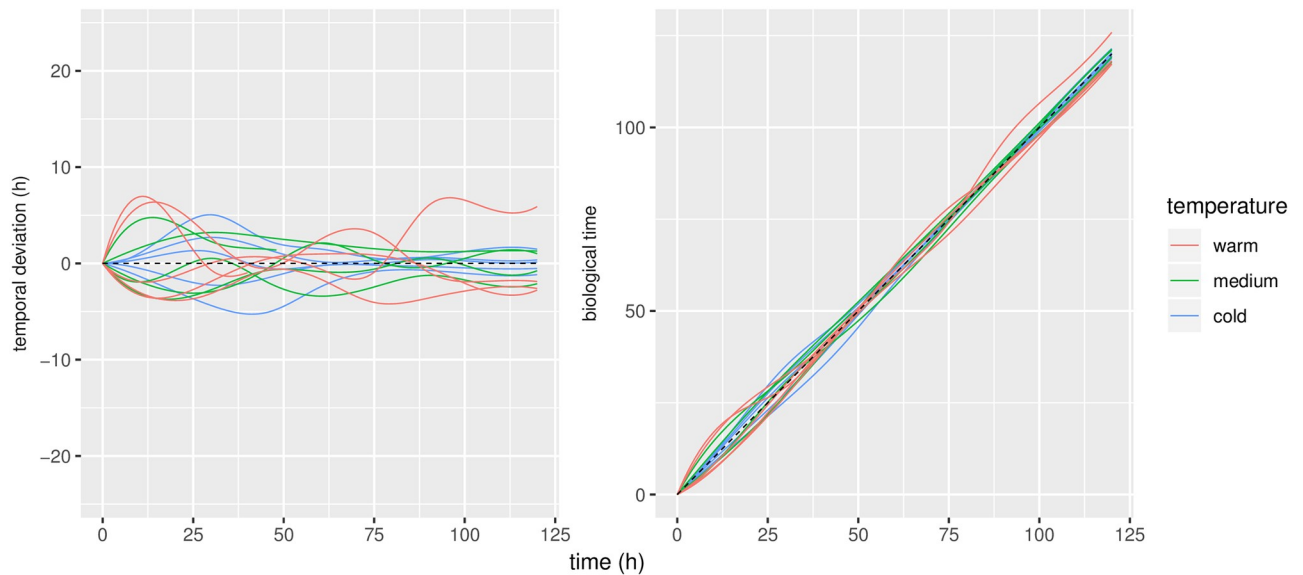https://doi.org/10.1371/journal.pone.0215914.g004

Fig 5. Predicted warp functions. Left panel: deviations from the identity. Right panel: resulting warping functions. Black line indicates the identity, ie. no temporal deviation.

largest threshold was 5.5 with a step size of 0.5. One step corresponds to an increase in conidia discharge of $\approx 65\%$. Using the results of the analysis, we simulated trajectories of u from the model. For a given trajectory and threshold, we measured the first time this threshold was reached, and for how long u remained above this level.

The results are seen in Fig 6. There are generally large variations, but we see that fungi at low temperatures are consistently slower at reaching the threshold. It should be noted that the duration is only counted until end of experiment (120 h) so the actual duration values for cold fungi could be larger when viewed over a longer time span.

Total conidia discharge

The total number of discharged conidia by individual fungi is displayed in Table 1. Looking at the numbers, there is a decrease in total conidia count towards higher temperatures, also when discarding samples 7 and 13, which terminated discharge of conidia during the experiment.

However, a one-way anova test gave a *p*-value of 0.075 (excluding samples 7 and 13), and pairwise Wilcoxon tests and a Kruskal-Wallis test gave even larger p-values. So while it is evident that temperature has an effect on conidia discharge as a function of time, we are not able to detect a significant effect of temperature on the total amount of conidia discharged within the first 120 h.

Inference for population means

Following the approach outlined in the appendix, we estimated the information matrices for the spline coefficients, I_{cold}, I_{medium}, I_{warm}. The information matrices themselves are of little interest, but following Berstein-von Mises theorem, the information matrix can be used to quantify uncertainty and standard error for any given value of θ, see Fig 7. We have much more uncertainty for small values of θ. This is as expected; small values of θ corresponds to few conidia counts and thus only little data to estimate from. The pointwise standard errors for θ in regions with large counts are around 0.20-0.25 or 20-25% when exp-transformed.

Initial reach of threshold

Duration

Initial reach of threshold, averages
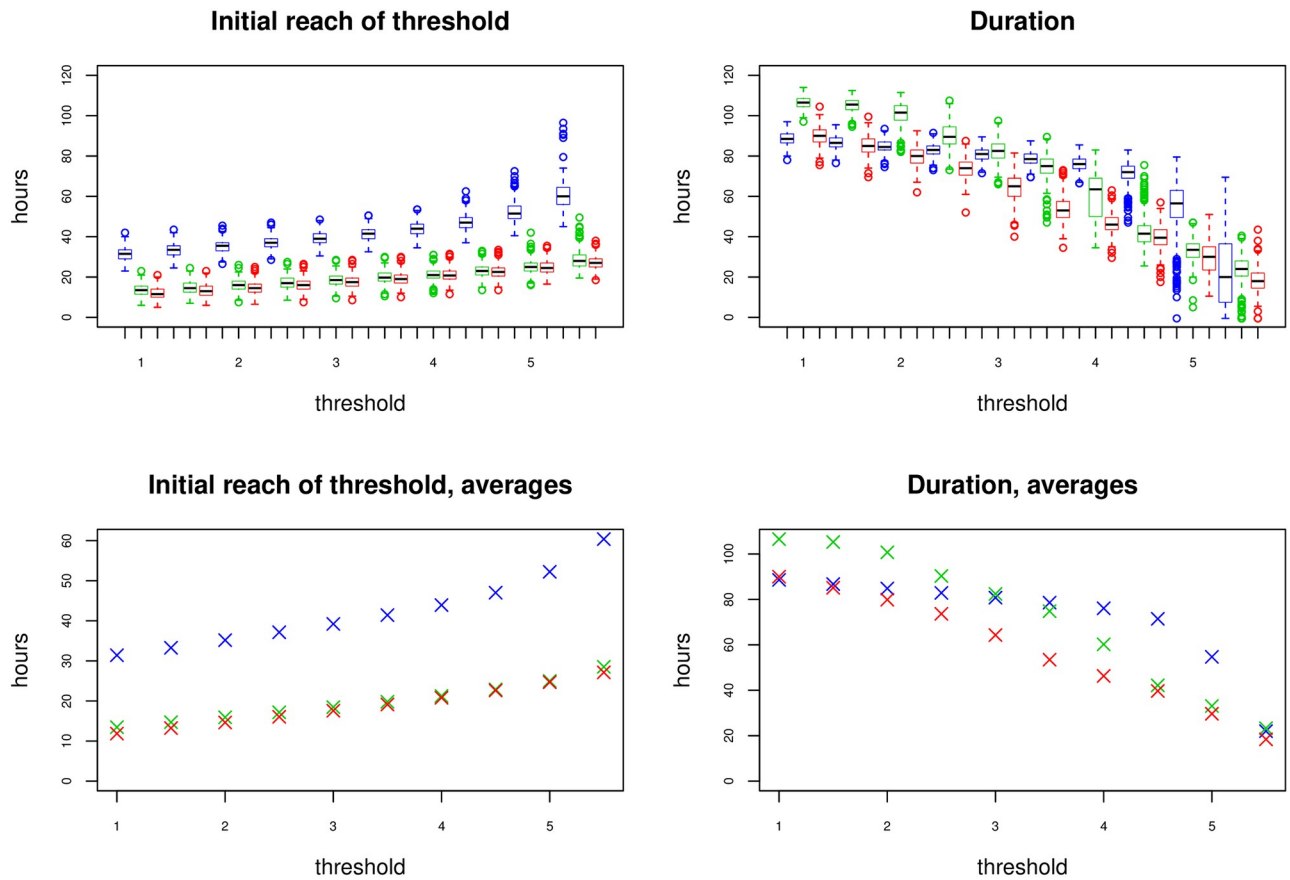
Duration, averages

Fig 6. Left: First time conidial discharge intensity reaches given threshold according to the model, for different thresholds. Some trajectories did not reach given thresholds and have been omitted from the corresponding boxplots. Right: Duration that conidial discharge intensity is above given threshold. Blue: 12.0˚C, Green: 18.5˚C, Red: 25˚C.

https://doi.org/10.1371/journal.pone.0215914.g006

Peak location and decrease

Using the standard error estimates from the previous section, we made inference on the location and decrease of peaks. This was done by simulating from the approximate distributions of the estimators. 1000 simulations were used, results are in Table 2. As we expected, θ_{cold} peaked late, around 70h after start, while the fungi stored at higher temperatures peaked much earlier. We observed a large and skewed 95% confidence interval for peak location of θ_{medium}, even containing the similar confidence interval for θ_{warm}. Regarding the second element, peak decrease, we saw a roughly linear relationship between temperature and decrease. The confidence interval for θ_{warm} is broader than the other confidence intervals; this is due to the lack of

Table 1. Sums of discharged conidia. * indicate fungi that terminated discharge of conidia during the experiment.

cold	medium	warm
1575	2003	1742
2019	902*	1764
1921	1510	787*
2019	1991	1470
2323	1720	1769

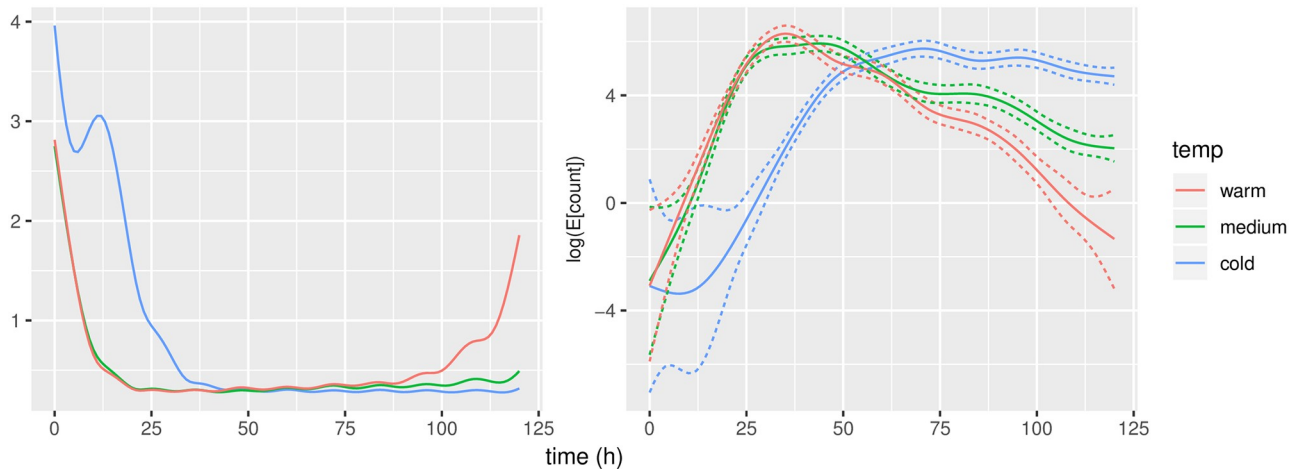https://doi.org/10.1371/journal.pone.0215914.t001

Fig 7. Left: Pointwise evaluations of $1.96 \cdot I^{-1}$, where I is the information matrix. Right: Corresponding pointwise confidence intervals.

https://doi.org/10.1371/journal.pone.0215914.g007

Table 2. Approximate 95% confidence intervals for peak location (left) and peak decrease (right). Units are hours after start of experiment and %/h, respectively.

	2.5%	Estimate	97.5%		2.5%	Estimate	97.5%
cold	66.0	70.7	73.7	cold	1.29	2.10	2.99
medium	32.7	43.8	46.6	medium	4.14	5.07	5.98
warm	33.7	35.1	36.1	warm	6.78	8.94	11.42

https://doi.org/10.1371/journal.pone.0215914.t002

data for small values of θ, cf. Fig 7. However, all confidence intervals are clearly separated at a 95% level, and we can firmly conclude that lower temperatures leads to a more slowly decreasing peak, with the consequence of increasing the duration of high conidial discharge.

Credibility of hypotheses. By comparing the approximate distributions of the estimators, we assessed the credibility of the hypotheses stated in the data analysis section. This was done by pairwise comparison of estimators using $q = P(f(\hat{X}) < f(\hat{Y}))$, where $f(\hat{X})$ and $f(\hat{Y})$ were sampled independently under the posterior distributions of the parameter functions, e.g. $f(X) = peak(\theta_{cold})$ and $f(Y) = peak(\theta_{middle})$.

Identical posterior distributions of $f(\hat{X})$ and $f(\hat{Y})$ implies $q = 0.5$, so small or large values of q are evidence against the hypothesis $f(X) = f(Y)$. Results are shown in Table 3. Apart from $peak(\theta_{middle}) = peak(\theta_{warm})$, all q-values are very close to one. As a result, our analysis very strongly supports that higher temperatures lead to faster decreasing peaks, and that a low temperature gives a late peak in comparison to middle and high temperatures.

Table 3. Pairwise comparisons of hypotheses with credibility values.

hypothesis	q
$peak(\theta_{cold}) = peak(\theta_{middle})$	1.00
$peak(\theta_{cold}) = peak(\theta_{warm})$	1.00
$peak(\theta_{middle}) = peak(\theta_{warm})$	0.108
$slope(\theta_{cold}) = slope(\theta_{middle})$	0.9999
$slope(\theta_{cold}) = slope(\theta_{warm})$	1.00
$slope(\theta_{middle}) = slope(\theta_{warm})$	0.9993

https://doi.org/10.1371/journal.pone.0215914.t003

Table 4. Parameter estimates and leave-one-out results. Note: An upper bound of 10 for the Matérn-smoothness was used in the analysis.

Parameter	nb-dispersion	range$_{amp}$	smoothness$_{amp}$	scale$_{amp}$	range$_{warp}$	scale$_{warp}$
Lower bound	4.40	0.314	2.30	0.0066	0.072	0.023
Estimate	4.66	0.458	7.21	0.072	0.083	0.026
Upper bound	5.21	0.523	10.0	0.084	0.691	0.034

Robustness of statistical analysis

Leave-one-out-analysis. To assess the uncertainty and robustness of the parameter estimates, a leave-one-out analysis was performed: One observation (or in our case, one curve) is removed from the data set, and the model is fitted to the reduced data set. This is done for all *N* observations in turn, and the results are compared in the end. These estimates should preferably not differ by much; this is called *robustness*; lack of robustness is an indication of overfitting, that is too many features or variables are included in the model. Robustness is related to *generalised cross-validation*; see e.g. [27] for a reference.

As our model is highly non-linear and consists of several layers, each with different parameters, it was of interest to study the robustness. As seen in Table 4 we got a fairly large spread on the amplitude covariance parameters. However, this can be explained by the many kinds of variation in data; it is more relevant that the mean curves are very robust (see Fig 8), as the population means are main interest of this study. The explanation behind the large spreads observed in beginning is that large negative values are mapped into almost-zero values.
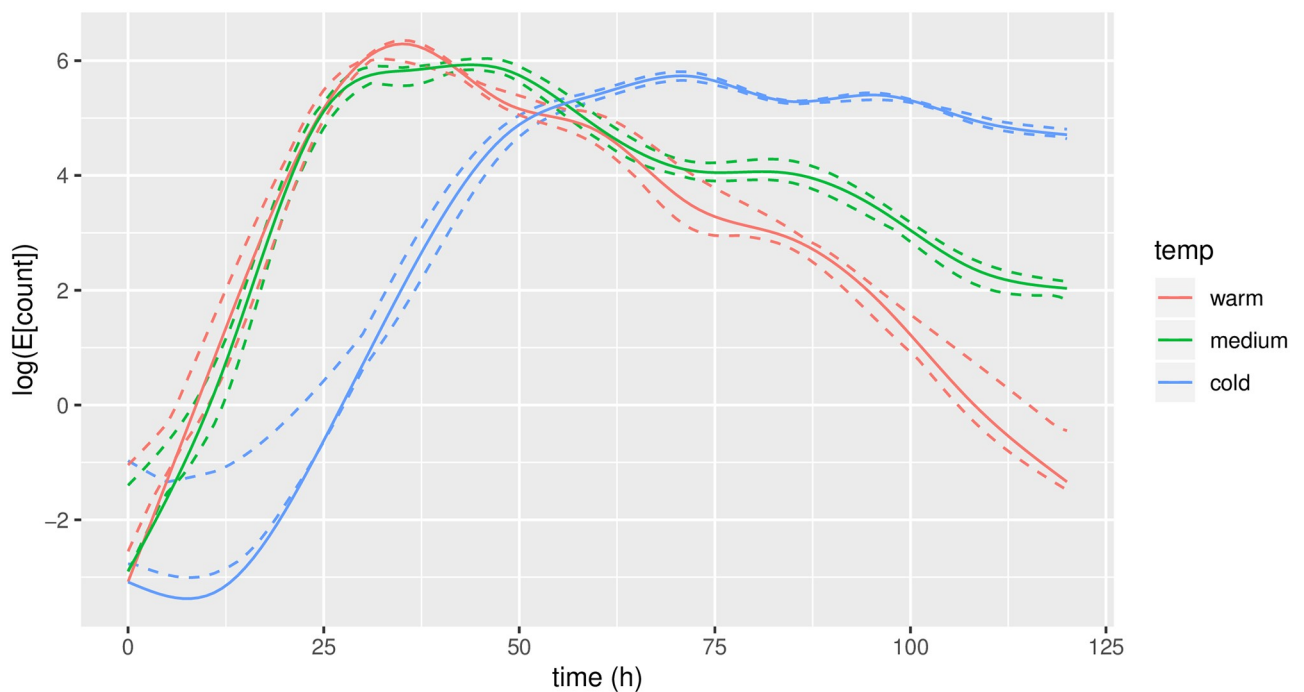

Fig 8. Pointwise estimates, and upper and lower bounds for leave-one-out analysis.

Discussion

With the applied statistical methods, we were able to characterize the temporal patterns of conidial discharge to a much better degree than previous studies, and we characterized the variation between individual fungi at the same temperature (i.e. of the same population). With a 95% prediction interval of roughly 6 hours, the temporal variation is too little for changing the overall shapes, but still large enough to be important for the analysis and to shift the peaks for individual fungi significantly.

Good statistical methods are essential when analysing biological systems with a temporal pattern, and allow researchers to get a better interpretation of data. Advanced statistical methods are not always better than simple ones, but the applied methods should be able to capture all essential variations in data. The presented model accounts for all these variations, which is a major advantage to previous methods. We believe this model to be more realistic compared to other models used in similar studies.

Examples of statistical analyses (some using the pavpop model) of other biological systems, where a model of the temporal variation was essential for the data analysis and interpretation of results, include electrophoretic spectra of cheese [28], growth of boys [23] and hand movements [23, 24].

In this study we demonstrated the flexibility of the pavpop model by successfully fitting to a complete different kind of data: namely discrete data with many zeros, where a Gaussian approximation would be unreasonable. With this success, there is reason to believe that this framework would work well in applications with other commonly used response models, for example binary response models (logistic regression).

Having several counts per measurement allowed us to look into response models. The Poisson model was invalidated, so we applied a negative binomial model instead. This is also relevant for similar/future studies: a negative binomial distribution gives rise to larger standard errors on estimates than a corresponding Poisson model. Thus, if one naïvely applies a Poisson model, where a negative binomial model is correct, this increases the risk of making type I errors.

There were some non-robustnesses in the estimation, but given the comparatively small amount of data, this is adequate. The robustness analysis can be used to asses which parameters are identifiable in practice. Although some of the variance parameters were not well identified, the dispersion parameter, average temporal deviation and population means were found to be robust.

We were not able to detect significant differences in total number of discharged conidia in this study. However, the fungi stored at 12˚C were still discharging many conidia at $t = 120h$, so there is good reason to believe that there would have been significant differences if a longer time span had been used; the authors have data that supports this, too. In a study conducted on mycelial mats of *Pandora neoaphidis* over 168 h this could be observed: At 25˚C the mycelium mats produced less primary conidia compared to mycelium mats incubated at 10, 15 or 20˚C [14]. Aphid cadavers infected by *P. neoaphidis* discharged similar numbers of primary conidia at temperatures between 5 and 25˚C in the first 24 hours [13].

On the other hand, we detected significant differences in peak location and shape: high temperature leads to an early peak but fast decreasing intensity of conidial discharge compared to low temperature. Other authors also found an earlier peak and faster decreasing intensity of conidial discharge at 25˚C compared to lower temperatures in other species of fungi, but the position of the peak and decrease of conidial discharge intensity was not statistically analysed [15, 19]. Our findings agree with those of [15]; lower temperature leads to longer durations of

conidial discharge. When the host population is large, the chance of a conidium landing on a host is larger and there is no advantage of prolonging the conidial discharge [15].

Our work can also be seen in the perspective of disease forecasting and fungal pathogen modelling [29, 30], where good models for spore discharge are an important ingredient, and we believe there is an important potential (in this direction) for future research.

Biological control

There is an important plant protection perspective in this study as the fungus considered in this study has a high virulence against insects from the genus *Cacopsylla* (Hemiptera, Psyllidae). Psyllids harm fruit trees by sucking from the leaves, and furthermore, the species *C. picta* can transmit plant diseases Candidatus (Ca.) Phytoplasma pyri to pear trees and Ca. Phytoplasma mali to apple trees, causing large economic losses [31]. There is an ongoing effort to reduce the usage of chemical pesticides for controlling psyllids in fruit and to switch to alternative control methods [25, 32].

The effects of temperature on temporal pattern of conidial discharge are important in practical applications and for the potential of this species as a biocontrol agent. The most important factor is the duration of intense conidial discharge, thus we believe the biocontrol potential to be largest at cold temperatures; the effect is illustrated in Fig 7. To get a better understanding of the environmental tolerance of a fungus regarding conidial discharge, experiments including fluctuating temperature, different relative humidity and light levels need to be conducted. Furthermore, the conidial discharge from insect cadavers can be measured to get a better understanding of the development of epizootics in the field. The presented statistical framework will likely be of great benefit for future data analysis of any experiments in which conidial discharge is measured over time.

Appendix

Statistical estimation

Direct estimation in the statistical model (1) and (2) is not feasible due to the large number of latent variables. Furthermore, unlike the setup in [23], the response is not Gaussian, which further complicate estimation. One solution would be to use MCMC methods, which are generally applicable. However, we propose to use a double Laplace approximation for doing approximate maximum likelihood estimation.

This actually consists of a linearisation around the warp variables w_n followed by a Laplace approximation on the discretised mean curves $\boldsymbol{u}$; $\boldsymbol{u}_n = \{u_n(t_{nk})\}_{k=1}^{m_n}$ for $n = 1, \ldots, N$. When these approximations are done at the maximum posterior values of $(w_n, \boldsymbol{u}_n)$, this is equivalent to a Laplace approximation jointly on $(w_n, \boldsymbol{u}_n)$.

The main difference from the estimation procedure of [23] is the non-trivial addition of a second layer of latent variables, u.

Posterior likelihood. To perform Laplace approximation, we need the mode of the joint density of observations and latent variables; this can be found by maximising the posterior likelihood for the latent variables.

The joint posterior negative log-likelihood for sample n is proportional to

$$L = \Big[\sum_{k=1}^{m_n} A(u_{nk}) - u_{nk}y_{nk}\Big] + \frac{1}{2}(\boldsymbol{\gamma}_{w_n} - \boldsymbol{u}_n)^\top S_n^{-1}(\boldsymbol{\gamma}_{w_n} - \boldsymbol{u}_n) + \frac{1}{2}w_n^\top C^{-1}w_n \tag{3}$$

where $\boldsymbol{\gamma}_{w_n}$ denote the vector $\{\theta_{f(n)}(v(t_{nk}, w_n))\}_{k=1}^{m_n}$. Spline coefficients for the fixed effects are indirectly present in the posterior likelihood through γ_{w_n}; more details follow below. It should

be noted that under relatively mild assumptions, minimizing (3) for a fixed w is a convex optimization problem.

Likelihood approximation. To approximate the likelihood, we firstly linearise around w^0 to approximate $p(u)$ with a Gaussian distribution and secondly we make a Laplace approximation of the joint linearised likelihood.

The linearization around w^0 to approximate the likelihood for density the mean curves, $p(u_n)$, is described in detail in [22, 23]. The result of doing this is a Gaussian approximation of the latent u, ie. $\boldsymbol{u}_n \overset{D}{\approx} \tilde{\boldsymbol{u}}_n$ where $\tilde{\boldsymbol{u}}_n \sim N(r_n, V_n)$. r_n and V_n are obtained from the Taylor approximation of u around the posterior maximum w_n^0; for details we refer to [22, 23].

In general, the Laplace approximation of an integral on the form $\int_{\mathbb{R}^d} e^{f(x)} \, \mathrm{d}x$ around the mode x_0 of f is given by

$$(2\pi)^{d/2} \mid -f''(x_0) \mid^{-1/2} e^{f(x_0)} \tag{4}$$

where $\mid -f'(x_0) \mid$ is the determinant of the Hessian of $-f$, evaluated in x_0. This approximation is exact if f is a second-order polynomial, and generally the approximation is directly related to the second-order Taylor approximation of f at x_0.

Up to some constants, which do not depend on the parameters, the likelihood for a single curve in the linearised model is given by the following integral, which we want to approximate:

$$L_n^{\mathrm{lin}} \propto \int_{\mathbb{R}^{m_n}} |V_n|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{u}_n - r_n)^\top V_n^{-1} (\boldsymbol{u}_n - r_n) + \sum_{k=1}^{m_n} y_{nk} \boldsymbol{u}_{nk}^0 - A(u_{nk}) \right) \mathrm{d}\boldsymbol{u}_n \tag{5}$$

Assuming $\boldsymbol{u}_n^0$ to be the maximum of the posterior likelihood (3), one can show that the negative logarithm of the Laplace approximation around $(\boldsymbol{u}_n^0, w_n^0)$ is given by:

$$1/2 \log |\tilde{\Sigma}_n| + \sum_{k=1}^{m_n} (A(u_{nk}^0) - y_{nk} \boldsymbol{u}_{nk}^0) + p(\boldsymbol{u}_n^0)$$

where $\tilde{\Sigma}_n = V_n^{-1} + 2 diag(A''(\boldsymbol{u}_n^0))$ and $p(\cdot)$ is the negative log-density for a $N(r_n^0, V_n)$-distribution. By assumption, $A''(u_{nk}^0) > 0$, so $|\tilde{\Sigma}_n| > |V_n|^{-1}$.

The total log-likelihood for all observations is then approximated by

$$\sum_{n=1}^{N} \left[\log |\tilde{\Sigma}_n| + \log |V_n| + (\boldsymbol{u}_n^0 - r_n)^\top V_n^{-1} (\boldsymbol{u}_n^0 - r_n) + 2\sum_{k=1}^{m_n} (A(u_{nk}^0) - y_{nk} \boldsymbol{u}_{nk}^0) \right] \tag{6}$$

Inference. We propose to use alternating steps of (a) estimating spline coefficients for the fixed effects and predicting the most likely warps and mean curves by minimizing the posterior log-likelihood (3) and (b) estimating variance parameters from minimizing the approximated log-likelihood (6).

Fixed effects and phase variation

Fixed effects are modelled using a spline basis that is assumed to be continuously differentiable, e.g. a Fourier basis or B-spline bases. A typical choice for non-periodic data would be B-splines; we have used natural cubic splines in the data application. Fixed effects are estimated using the posterior likelihood (3). For a fixed value of w_n, $\boldsymbol{\gamma}_{w_n}$ is a linear function of the spline coefficients, and thus the optimal value can be found using standard linear algebra tools.

Phase variation is modelled by random warping functions $v_n = v(\cdot, w_n):[0, 1] \to D$, parametrized by independent zero-mean Gaussian variables $w_n \in \mathbb{R}^{m_w}$. $v : [0, 1] \times \mathbb{R}^{m_w} \to D$ is a suitable spline interpolation function, such that $v(\cdot, 0)$ is the identity on $[0, 1]$.

The latent trajectories v_n are modelled as deviations from the identity function at pre-specified time points $(t_k)_{k=1}^{m_w}$, subject to a Hyman filtered, cubic spline interpolation for insuring monotonicity, $v_n(t_k) \approx t_k + w_{nk}$. A more detailed discussion of modelling phase variation using increasing spline functions can be found in [23].

Uncertainty for fixed effects. As our model is highly non-linear, we cannot expect to find closed-form expressions for the uncertainty of the parameter estimates. Furthermore, the latent variables complicate assessment of uncertainty as these are uncertain themselves.

A standard quantifier for assessing uncertainty in statistical models is the *information matrix*, which can be approximated by the second-order derivative of the log-likelihood at the MLE. However, directly using (6) would underestimate the information, as (6) depend on the optimal value of the posterior likelihood (3), which itself is a function of parameters.

Let c_j denote the spline coefficients which determine the population mean θ_j for treatment group j. c_j is determined from the posterior likelihood $L = L(c, u, w)$, given in Eq 3. As u and w are latent, it would be wrong to use the second derivative of L for the information matrix; instead we use the second derivative of $f(c) = L(c, u(c), w(c))$, where u and w are viewed as functions that map c into the max-posteriors of u and w given c.

This will more correctly ensure that the uncertainty of u and w is taken into account when estimating the information matrix. Furthermore, positive definiteness of L'' will imply positive definiteness of f''.

Response models

In the application presented in this paper we assume that $(y|u)$ follows a negative binomial distribution. There are various choices of response models, a list of important ones are stated below. Note that not all exponential families fits naturally with our methodology; $y|u$ must be well-defined for all $u \in \mathbb{R}$.

Binary response. For binary responses, the sample space is $\mathcal{Y} = \{0, 1\}$. If we define $p := P(Y = 1|\eta)$, and set $A(\eta) = \log(1 + e^{\eta})$, we get that $\eta = \log(p) - \log(1 - p)$, the canonical link function for regression models with binomial response.

Poisson model. For the Poisson model we have $Y \in \mathbb{N}_0$ where $A(\eta) = e^{\eta}$. The conditional mean satisfies $\mathrm{E}[Y|\eta] = e^{\eta}$, and by inverting this relation we get $\eta = \log \mathrm{E}[y|\eta]$, the canonical link function.

Negative binomial model. Negative binomial distributions are often viewed as overdispersed versions of Poisson models. Let the rate parameter $r > 0$ be given such that $V[Y|\eta] = E[Y|\eta] + E[Y|\eta]^2/r$; the limit $r \to \infty$ corresponds to the Poisson model.

We get $A(\eta, y) = (r + y) \log \left(1 + \frac{e^{\eta}}{r}\right)$ and $A''(\eta; y) = (y + r) \frac{re^{\eta}}{(r + e^{\eta})^2}$. Unlike the Poisson and binomial models, the link function A depends on y, but it is easily seen that $A(\eta, y)$ approximates e^{η} in the limit $r \to \infty$.

Normal distribution with known variance σ^2. For normal distributions we have $Y \in \mathbb{R}$. By setting $\tilde{Y} = Y/\sigma^2$, then $A(\eta) = \eta^2/2\sigma^2$, $\mathrm{E}[\tilde{Y}|\eta] = \eta/\sigma^2$, and $\mathrm{E}[Y|\eta] = \eta$. This is the most basic response model, and the one used in [22]. [22] use a different formulation and also treats σ^2 as an unknown parameter. The Laplace approximation becomes exact when using normal distributions, simplifying estimation to become the approach used in [23].

Matern covariance function

The Matérn covariance function is commonly used in functional data analysis and spatial statistics. It is given by

$$f_{\sigma,\alpha,\kappa}(s,t) = \sigma^2 \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{\alpha|s-t|}{\kappa}\right)^\alpha K_\alpha\left(\frac{\alpha|s-t|}{\kappa}\right), \quad s,t \in \mathbb{R} \tag{7}$$

where K_α is the modified Bessel function of the second kind. Here σ is the scale parameter, α is the smoothness parameter and κ is the range parameter.

Supporting information

S1 File. Zipped data file. The zip file contains the data used for the analysis. Please see Description.txt for use.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Niels Lundtorp Olsen, Pascal Herren, Bo Markussen, Annette Bruun Jensen, Jørgen Eilenberg.

Data curation: Pascal Herren, Annette Bruun Jensen, Jørgen Eilenberg.

Formal analysis: Niels Lundtorp Olsen, Pascal Herren.

Investigation: Pascal Herren.

Methodology: Niels Lundtorp Olsen, Pascal Herren, Bo Markussen, Annette Bruun Jensen, Jørgen Eilenberg.

Software: Niels Lundtorp Olsen.

Supervision: Bo Markussen, Annette Bruun Jensen, Jørgen Eilenberg.

Validation: Niels Lundtorp Olsen, Pascal Herren, Bo Markussen, Annette Bruun Jensen, Jørgen Eilenberg.

Visualization: Niels Lundtorp Olsen, Pascal Herren.

Writing – original draft: Niels Lundtorp Olsen, Pascal Herren, Bo Markussen, Jørgen Eilenberg.

Writing – review & editing: Niels Lundtorp Olsen, Bo Markussen, Annette Bruun Jensen, Jørgen Eilenberg.

References

1. Hajek AE, Eilenberg J. Natural Enemies—an introduction to biological control. Cambridge University Press; 2018.

2. Shah PA, Pell JK. Entomopathogenic fungi as biological control agents. Applied Microbiology and Biotechnology. 2003; 61(5-6):413–423. https://doi.org/10.1007/s00253-003-1240-8 PMID: 12764556

3. Boomsma JJ, Jensen AB, Meyling NV, Eilenberg J. Evolutionary interaction networks of insect pathogenic fungi. Annual Review of Entomology. 2014; 59:467–485. https://doi.org/10.1146/annurev-ento-011613-162054 PMID: 24160418

4. Money NP. Spore production, discharge, and dispersal. In: Waktinson SC, Boddy L, Money NP, editors. The Fungi (Third Edition). Academic Press; 2016. p. 67–97.

5. Eilenberg J. The culture of *Entomophthora muscae* (C) Fres. in carrot flies (*Psila rosae* F.) and the effect of temperature on the pathology of the fungus. Entomophaga. 1987; 32(4):425–435. https://doi.org/10.1007/BF02372452

6. Hajek AE, Meyling NV. Fungi. In: Hajek AE, Shapiro-Ilan DI, editors. Ecology of Invertebrate Diseases. John Wiley & Sons; 2018.

7. Furlong MJ, Pell JK. The influence of environmental factors on the persistence of *Zoophthora radicans* conidia. Journal of Invertebrate Pathology. 1997; 69(3):223–233. https://doi.org/10.1006/jipa.1996.4649

8. Jensen AH. A new insect pathogenic fungus from Entomophthorales with potential for psyllid control. MSc thesis. Department of Plant and Environmental Sciences, University of Copenhagen; 2017.

9. Aoki J. Pattern of conidial discharge of an *Entomophthora* species ("grylli" type) (Entomophthorales: Entomophthoraceae) from infected cadavers of *Mamestra brassicae L.* (Lepidoptera: Noctuidae). Applied Entomology and Zoology. 1981; 16(3):216–224. https://doi.org/10.1303/aez.16.216

10. Hemmati F, Mccartney HA, Clark SJ, Deadman ML, et al. Conidial discharge in the aphid pathogen *Erynia neoaphidis*. Mycological research. 2001; 105(6):715–722. https://doi.org/10.1017/S0953756201004014

11. Hajek AE, Davis CI, Eastburn CC, Vermeylen FM. Deposition and germination of conidia of the entomopathogen *Entomophaga maimaiga* infecting larvae of gypsy moth, *Lymantria dispar*. Journal of invertebrate pathology. 2002; 79(1):37–43. https://doi.org/10.1016/S0022-2011(02)00010-1 PMID: 12054785

12. Wraight S, Galaini-Wraight S, Carruthers R, Roberts DW. *Zoophthora radicans* (Zygomycetes: Entomophthorales) conidia production from naturally infected *Empoasca kraemeri* and dry-formulated mycelium under laboratory and field conditions. Biological Control. 2003; 28(1):60–77. https://doi.org/10.1016/S1049-9644(03)00035-5

13. Yu Z, Nordin G, Brown G, Jackson D. Studies on *Pandora neoaphidis* (Entomophthorales: Entomophthoraceae) infectious to the red morph of tobacco aphid (Homoptera: Aphididae). Environmental Entomology. 1995; 24(4):962–966. https://doi.org/10.1093/ee/24.4.962

14. Shah PA, Aebi M, Tuor U. Effects of constant and fluctuating temperatures on sporulation and infection by the aphid-pathogenic fungus *Pandora neoaphidis*. Entomologia Experimentalis et Applicata. 2002; 103(3):257–266. https://doi.org/10.1046/j.1570-7458.2002.00980.x

15. Li W, Xu WA, Sheng CF, Wang HT, Xuan WJ. Factors affecting sporulation and germination of Pandora nouryi (Entomophthorales: Entomophthoraecae), a pathogen of *Myzus persicae* (Homoptera: Aphididae). Biocontrol Science and Technology. 2006; 16(6):647–652. https://doi.org/10.1080/09583150500532840

16. Eilenberg J, Bresciani J, Latgé JP. Ultrastructural studies of primary spore formation and discharge in the genus *Entomophthora*. Journal of Invertebrate Pathology. 19

17. Hesketh H, Roy HE, Eilenberg J, Pell JK, Hails RS. Challenges in modelling complexity of fungal entomopathogens in semi-natural populations of insects. BioControl. 2010; 55:55–73. https://doi.org/10.1007/s10526-009-9249-2

18. Hajek AE, Papierok B, Eilenberg J. Methods for study of the Entomophthorales. In: Lacey LA, editor. Manual of Techniques in Invertebrate Pathology (Second Edition). Elsevier; 2012. p. 285–316.

19. Kalsbeek V, Pell JK, Steenberg T. Sporulation by *Entomophthora schizophorae* (Zygomycetes: Entomophthorales) from housefly cadavers and the persistence of primary conidia at constant temperatures and relative humidities. Journal of Invertebrate Pathology. 2001; 77(3):149–157. https://doi.org/10.1006/jipa.2001.5012 PMID: 11356049

20. Ramsay JO. When the data are functions. Psychometrika. 1982; 47(4):379–396. https://doi.org/10.1007/BF02293704

21. Ramsay J, Silverman BW. Functional Data Analysis. Springer; 2005.

22. Raket LL, Sommer S, Markussen B. A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data. Pattern Recognition Letters. 2014; 38:1–7. https://doi.org/10.1016/j.patrec.2013.10.018

23. Olsen NL, Markussen B, Raket LL. Simultaneous inference for misaligned multivariate functional data. Journal of the Royal Statistical Society: Series C (Applied Statistics); 67(5):1147–1176. https://doi.org/10.1111/rssc.12276

24. Raket LL, Grimme B, Schöner G, Igel C, Markussen B. Separating timing, movement conditions and individual differences in the analysis of human movement. PLoS Computational Biology. 2016; 12(9): e1005092. https://doi.org/10.1371/journal.pcbi.1005092 PMID: 27657545

25. Herren P. Conidial discharge of *Pandora* sp., a potential biocontrol agent against *Cacopsylla* spp. in apple and pear orchards. MSc thesis. Department of Plant and Environmental Sciences, University of Copenhagen; 2018.

26. Yeo H, Pell J, Walter M, Boyd-Wilson K, Snelling C, Suckling D. Susceptibility of diamondback moth (*Plutella xylostella* (L.)) larvae to the entomopathogenic fungus, *Zoophthora radicans* (Brefeld) Batko. In: Proceedings of the New Zealand Plant Protection Conference. New Zealand Plant Protection Society; 1998; 2001. p. 47–50.

27. Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning. 2nd ed. Springer; 2009.

28. Rønn BB. Nonparametric maximum likelihood estimation for shifted curves. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001; 63(2):243–259. https://doi.org/10.1111/1467-9868.00283

29. Holtslag QA, Remphrey WR, St-Pierre RG, Ash GHB. The development of a dynamic disease forecasting model to control *Entomosporium mespili* on *Amelanchier alnifolia*. Canadian Journal of Plant Pathology. 2004; 26(3):304–313. https://doi.org/10.1080/07060660409507148

30. Zhou X, Gou K, Sheng-Feng M, He W, Zhu YY. Modeling analysis on sporulation capacity, storage and infectivity of the aphid-specific pathogen *Conidiobolus obscurus* (Entomophthoromycota: Entomophthorales) Mycoscience. 2014; 55(1):21–26.

31. Strauss E. Phytoplasma research begins to bloom. Science. 2009; 325(5939):388–390. https://doi.org/10.1126/science.325_388 PMID: 19628836

32. Patel A. Entwicklung neuartiger Formulierungen für verhaltensmanipulierende Strategien zur biologischen Bekämpfung von *Cacopsylla picta*, dem Überträger der Apfeltriebsucht.; 2015.