INFERRING STRATIFIED PARASITOID DISPERSAL MECHANISMS AND PARAMETERS FROM COARSE DATA USING MATHEMATICAL AND BAYESIAN METHODS

CHRISTOPHER STRICKLAND[1,2], NADIAH P. KRISTENSEN[3], AND LAURA MILLER[1,4]

Abstract. Biological invasions have movement at the core of their success. However, due to difficulties in collecting data, medium- and long-distance dispersal of small insects has long been poorly understood and likely underestimated. The agricultural release of parasitic hymenoptera, a group of wasps that are critical for biological pest control, represents a rare opportunity to study the spread of insects on multiple spatial scales. Since these insects are typically less than 1 mm in size and are challenging to track individually, a first-time biocontrol release will provide a known spatial position and time of initial release for all individuals that are subsequently collected. In this paper, we develop and validate a new mathematical model for parasitoid wasp dispersal from point release, as in the case of biocontrol. The model is derived from underlying stochastic processes but is fully deterministic and admits an analytical solution. Using a Bayesian framework, we then fit the model to an Australian dataset describing the multi-scale wind-borne dispersal pattern of *Eretmocerus hayati* Zolnerowich & Rose (Hymenoptera: Aphelinidae). Our results confirm that both local movements and long distance wind dispersal are significant to the movement of parasitoids. The model results also suggest that low velocity winds are the primary indicator of dispersal direction on the field scale shortly after release, and that average wind data may be insufficient to resolve long distance movement given inherent nonlinearities and heterogeneities in atmospheric flows. The results highlight the importance of collecting wind data when developing models to predict the spread of parasitoids and other tiny organisms.

1. Background

Movement is a fundamental characteristic of living organisms, and understanding movement is crucial for understanding the invasion [1] and persistence [2] of populations in ecosystems. A particular challenge is characterising and predicting movement and its consequences when the organism both: (1) is difficult to detect, and (2) has the potential for stratified movement over different spatial scales using different modes of movement. Many organisms share both of these characteristics, such as species with dispersal polymorphism [3] or organisms that undergo long-distance anthropogenic dispersal [4, 5]. In this paper, we used the first-release of a small parasitoid wasp, *Eretmocerus hayati* Zolnerowich & Rose (Hymenoptera: Aphelinidae), as a case study to quantify stratified dispersal on multiple scales. Such small entomophagous insects (< 1 mm) are significant for biocontrol of crop pests in agro-ecosystems [6], and *E. hayati* undergoes a form of stratified dispersal

Key words and phrases. Biological invasions, Biocontrol, Long-distance dispersal, Multi-scale dispersal, Active and passive spread, Fokker-Planck model.

[7] that is common in small insects: a combination of wind-borne dispersal over long distances with short-range dispersal governed by different mechanisms [8].

The challenge presented by stratified movement is both theoretical and empirical. The theoretical challenge is that models are required that can integrate different modes and scales of movement, such as individuals' response to local cues that trigger long-distance dispersal as well as aerodynamics at both the local and geophysical scales. The empirical difficulty lies in the challenge of obtaining multi-scale data to verify such models. The size of a study area is often constrained by research feasibility, and consequently there exists a bias towards measurement of routine small-scale movements compared to long-distance emigration and immigration movements [9]. Furthermore, wind data, if collected, is often temporally averaged and only collected at one location. The resolution of data collection is chosen to match a particular mode of movement. As a consequence, integrating multiple spatial scales is typically done by only including movement behaviours relevant to the focal scale and neglecting movement occurring on larger spatial scales. As a consequence of the latter, rates of spread tend to be underestimated due to study design [9], and the neglect of infrequent long-distance dispersal events has been identified as a cause of spread-rate underestimation in the literature [1].

Accurate descriptions and predictive models of long-distance dispersal events are important given their significance to population dynamics. For example, the agro-ecosystems that insect pests and biocontrol agents inhabit are typically fragmented, such that the functional connectivity of the landscape – which incorporates both distance between suitable habitat and dispersal ability – is a key determinant of spread [10]. Long-distance dispersal capacity may reduce colonisation lag leading to greater biocontrol efficacy [11], while local-scale search ability is necessary for effective pest suppression [6].

Local-scale and individual-level processes are also important to dispersal dynamics since these factors are often the trigger for shifts between different modes of dispersal. For example, the physiological internal state of the individual [12] and the organism's reaction to environmental cues (e.g. [13, 14]) may initiate long distance dispersal. Even during so-called 'passive' dispersal (long-distance dispersal primarily by wind or water advection), the individual controls entry into, exit from, and movement within the advective stream (examples include spider ballooning [15, 16], and the vertical migration of plankton [17]). These particulars of dispersal control can lead to biologically significant differences in dispersal distance [18]. Therefore, a fundamental challenge in the study of long distance dispersal is how to effectively integrate processes on multiple spatial scales and meaningfully confront hypotheses (i.e. models) with data.

Even rudimentary attempts to model dispersal on multiple spatial scales can reveal new insights into the fundamental biological and physical processes. Particularly in the case of simple models, mismatches between the theoretical distributions and data are especially useful for falsifying hypotheses about movement processes and identifying gaps in our understanding [1]. The subsequent challenge is then to obtain the data necessary to refine and validate more sophisticated models, and to develop the modelling techniques necessary to overcome data limitations.

For example, fast moving individuals dispersing across large spatial scales are unlikely to be detected, but it may be feasible to infer the presence of an organism from their impact – e.g. infections, parasitisations, or evidence of predation. With

an additional modelling layer inferring presence from impact, a link can be made between the question of interest and the type of data that is feasible to obtain. Such complex models can be fitted using Bayesian frameworks, which have been used to answer a variety of ecological hypotheses in recent years [19]. For example, Ovaskainen et al. [20] used Bayesian methods to discern the relationship between emigration rate and age, and Bayesian methods have also been used to estimate extinction risks [21], parameters for demographic models [22], epidemiological parameters [23], and gene frequencies [24]. In contrast to mechanistic mathematical simulation models, which are typically validated with data using ad hoc approaches combined with parameter sensitivity analysis, the key benefit of the Bayesian approach is that it can connect model outputs to proxy data in a systematic way. The kinds of complex and multi-scale spatial models that are needed for characterising stratified dispersal require long computation run-time, which may mean that full characterisation of the Bayesian posterior distribution is not possible. However even in such cases, Bayesian methods are valuable because they calibrate model parameters in a way that is both rigorous and transparent.

Our study builds upon previous work that used a simple model to identify wind advection as a necessary process to explain the dispersal of an introduced *E. eretmocerus* population [25]. Three shortcomings of this earlier study were that: (1) the diffusion processes occurring during wind-borne flight were not modelled explicitly, but rather an approximation was made by assuming that spread occurred over the entire spatial grid cell; (2) model fitting was performed based upon a match between presence/absence in the model and the data, rather than between inferred and predicted densities; and (3) active flight behaviour was not considered in a probabilistic sense, but operated as a global switch mechanism based on changing environmental conditions.

In this study, diffusion is modelled explicitly during both local and wind-borne dispersal modes, and active flight behaviour is modelled nonlinearly based on a probabilistic interpretation of flight decisions given environmental variables. Simplicity and mathematical transparency were valued over mechanistic detail in order to assess the performance of our hypotheses. Additionally, we utilise a Bayesian framework to infer observed parasitisations from modelled population densities and thus estimate the parameter values of the parasitoid dispersal mode by comparing to data. Parameters from a related whitefly parasitoids [26] were used to construct the priors for the Bayesian framework.

2. Methods

Data were taken from a first-time biocontrol release of a small parasitoid wasp, *E. hayati*, near the town of Kalbar in eastern Australia, described in detail in earlier studies [25] and [7]. In the earlier study [25], a model for the parasitoid dispersal was fitted using the Kalbar data, and then tested on a later and separate first-time release near the town of Carnarvon in Western Australia. This model, upon which our study builds, assumed that dispersal was via simple wind-advection and was fitted by matching emergence of parasitoids in the field from the F_1 generation to the predicted presence of females from the F_0 or release generation. More specifically, the earlier model fitted three parameters: (1) a maximum wind-speed at which females would undertake a wind-borne flight, (2) a diurnal time-window during which the wind-borne flights could occur, and (3) a dispersal distance factor f that

scaled flight time and wind speed to displacement distance. Importantly, in this earlier model, no diffusion processes were explicitly modelled at the landscape scale; rather, a grid cell size of 500×500 m was used, which in effect spread females evenly over the area of each cell after every wind-borne flight. In contrast, in this study, we model diffusion explicitly.

We outline the additional specific assumptions of the previous model below for easy comparison with our new model. It was assumed that females would undertake at most one wind-borne flight per day, with equal probability of occurring during any 30 min time interval within the time-window provided that the average wind-speed during that interval was lower than the maximum permitted. Another simplifying assumption was that females would travel at the same speed and in the same direction as the averaged wind, and that the flight time was the same for all individuals. Flight distance was found by multiplying wind speed and flight time by the factor f. Oviposition occurred for the females' whole lifetime, such that presence of a female at a location translated directly to detection of emerging F_1. It was assumed that variance in development time was low such that the predicted presence of females in the model could be fitted to the emergence data with a single time-shift.

As a starting point for the new model analysed in this paper, we will maintain the following assumptions from [25] about *E. hayati* flight. We refer the reader to the earlier paper for detailed discussion of these assumptions [25].

- Each wasp will undertake at most one wind-borne flight per day. This is a simplifying assumption, however we extend the previous work by allowing the probability to undertake the daily flight to be fitted by the model.
- The mean direction of this flight will be in the same direction as the average wind velocity during the flight, and to a distance proportional to the wind's speed. The proportionality factor describing wind speed to drift is fitted by the model.
- Wind velocity is assumed spatially constant in the study area, though it varies temporally according to collected data. It should be noted that during the original release experiment, wind measurements were recorded at 1.8 m off the ground at the release point only, and wind vectors were averaged over 30 minute intervals. Therefore wind variability across the landscape, vertically (e.g. boundary layer effects, eddies), and on shorter timescales (e.g gusting, calms) are neglected.
- Wind-borne flights have a fixed duration, initially set at 30 min. This duration is comparable to flight durations measured in flight-chamber experiments for another *Eretmocerus* species [27].
- Wasps are less likely to initiate wind-borne flight in higher wind velocities. In many insect species, individuals will delay flight when wind speeds make flight difficult or dangerous (e.g. [28]). In the previous study [25], a parameter specifying the maximum wind speed at which flights occurred was found to be a key parameter determining fit between the model and data.

Daily movement for each wasp can be thought of as consisting of two cases: (1) the wasp takes one wind-borne flight at some point during the day, which we model as drift diffusion, or (2) the wasp only moves around locally, which we model as diffusion with possibly a slight drift in the day's (weighted) average wind direction (see Results). For the purposes of this model, both modes of dispersal will be

considered passive and independent of the movement of other wasps; we leave it to future studies to determine the relative effect of factors such as host distribution, landscape heterogeneity, and wasp aggregation due to mating.

Departing from [25], we will explicitly model both diffusive processes by assuming that the probability distribution for each wasp's dispersal on a given day is determined via a special case of the Fokker-Planck equation

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i}[\mu_i(t)p(\mathbf{x},t)] + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial^2}{\partial x_i \partial x_j}[D_{ij}p(\mathbf{x},t)] \tag{1}$$

with wind drift vector $\boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t))^T$ and diffusion tensor

$$D_{ij} = \sum_{k=1}^{2} \sigma_{ik}\sigma_{jk}.$$

In this equation, the first spatial derivative models drift in the direction of $\boldsymbol{\mu}(t)$ while the second derivative models Brownian motion with two dimensional diffusivity D. Since the initial condition for Eqn. 1 is given to be a point mass at the origin (each day's dispersal is considered separately), solutions are given analytically as a bivariate normal distribution function with wind-based mean $\boldsymbol{\mu}(t)$ and covariance matrix D which we will assume is non-singular so that it can be specified in terms of correlation ρ and standard deviations in the x and y directions, σ_x and σ_y, respectively.

For wasps that only move around locally, $\boldsymbol{\mu}$ is taken to be a small constant (based on average wind direction) or zero so that the dynamics are primarily diffusion. Additionally, the diffusion tensor is assumed to be smaller than that for wind-based flight. For wasps taking a wind-based flight, the timing of the flight is considered to be active behaviour by the wasp based on the time of day and past and current wind conditions. The effective $\boldsymbol{\mu}$ for a given wasp could be calculated as the integral of the wasp's wind drift while it was in flight, and this is aggregated over all possible flight take-off times based on takeoff time probabilities achieve the time-dependent drift vector $\mu(t)$ seen in Eqn. 1. Total wasp dispersal for the day is then a realisation of a mixture distribution obtained from the two cases of wind-borne and local movement, based on the probability of taking a wind-borne flight during the day given environmental conditions. Finally, we take each day's dispersal to be independent of all previous days so that we can run this process on a discrete daily time step for as long as desired, aggregating to obtain the total wasp dispersal at the end.

Active behaviour: initiating wind-borne flight. During any given day d, we assume that the probability of a wasp initiating a wind-borne flight at time t (measured in hours) is described by a density function $h(t, \mathbf{w}_d(t); \theta_h)$ where $\mathbf{w}_d(t)$ is a function giving the day's wind velocity at time t and θ_h is a set of shape parameters required to define the components described below (see Appendix A for a complete, explicit mathematical description of this function). Integrating $h(t, \mathbf{w}_d(t); \theta_h)$ from $t = 0$ to $t = 24$ hours yields a value between zero and one which represents the probability that a wasp takes a wind-borne flight during the day. This function completely describes the active-flight component of our model, and in our formulation, h can be understood as having three main components: (1) a probability density function $f(t)$ based on daylight availability which specifies the hours in

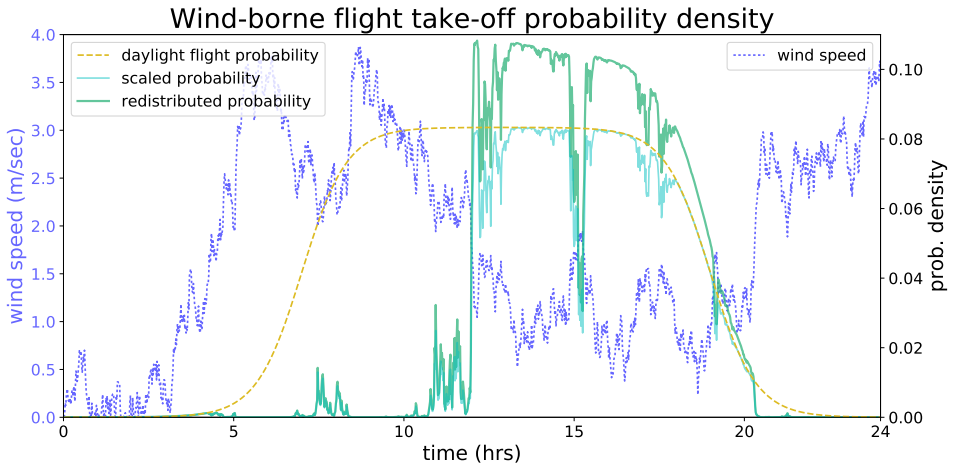

FIGURE 1. The probability at any time of starting a wind-borne flight is based on nonlinear daylight response and wind-speed preference, plus memory of earlier conditions.

which wasps are most likely to fly, (2) a scaling function $g(||\mathbf{w}_d(t)||)$ based upon wind velocity, and (3) a redistribution function that raises the probability of taking a wind-borne flight later in the day if conditions were previously unfavourable.

Our implementation of $f(t)$, the daylight probability density function, is the difference of two logistic functions scaled by the integral of $f(t)$ from $t = 0$ to $t = 24$ hours (see Fig. (1), yellow dashed line). It requires four parameters, a_1, a_2, b_1, and b_2, the first two of which locate the centre of the first and second logistic respectively, and the second two modify the steepness of the incline and decline of the functions. Similarly, $g(||\mathbf{w}_d(t)||)$ is implemented as a single, decreasing logistic function which takes on values between zero and one depending on how probable it is that a wasp would fly at that wind speed. Its two parameters, a_g and b_g, locate and scale the function respectively (see Appendix A for explicit equations for f, g, and all other functions).

A naive approach to implementing active flight take-off behaviour would be to simply multiply these two functions together, obtaining a density for the time that wind-borne flight is initiated. This approach, however, assumes that each moment's probability is independent of wind conditions earlier in the day. The behaviour it models could be imagined as a process where, before the day starts, each wasp randomly chooses a moment t_0 to initiate wind-borne flight and then actually takes the flight with probability $g(||\mathbf{w}_d(t_0)||)$. If conditions are poor, the wasp only moves in a more local manner.

To allow previous conditions to increase the likelihood of flight later in the day, we combine f and g in a way that considers the average decrease of probability in all previous moments weighted by the cumulative distribution function of f (lost flight opportunity later in the day has less effect than if the poor conditions occurred early). The result we call $h(t, \mathbf{w}_d(t); \theta_h)$. An example of this redistribution can be seen in Fig. (1), and for specifics of the function, we direct the reader to Appendix A.

Probability aggregation. Having determined the probability that a wasp initiates wind-borne flight at any point during the day, the total spatial probability distribution for each wasp's movement at the end of a given day is a weighted average of local diffusive movement and drift-diffusion based on wind conditions during flight. The expected spatial population density of wasps is given by multiplying the number of wasps by this probability distribution. Assuming that movement on each day is determined independent of all previous days, we convolute the distribution of each successive day to the current wasp distribution in order to determine each new, aggregate spatial density. This process yields the final model simulations seen in Fig. (3).

Bayesian modelling framework. In order to compare model results to field data and thus assess the model's relative performance under differing parameter choices, it is necessary to introduce a framework that can estimate the likelihood of observing the collected data under a given model scenario. The details of this framework are made more complex by the fact that on the landscape scale, wasps were not directly observed in the field but rather leaves were collected and observed for emerging offspring over a number of days. Each collection location (sentinel field) varied in size, and collection was not standardised in any rigorous way - collectors were instructed to search out leaves holding the host species, ideally displaying signs of having been parasitised. The time duration of this collection was unspecified. In the release field, leaves were collected in the same collection grid as had been previously used for direct observation of wasps [7]. The location of each sentinel field and the release field can be seen in Fig. (2).

To produce the likelihood of the model given emergence data from the sentinel fields, we assumed that the effective area canvassed by the collector in each field was roughly equivalent between fields and that parasitised hosts do not leave or enter the field prior to collection. Each sentinel field was then assigned an observation probability describing the likelihood that each given parasitised host present in the field during collection would be collected and a wasp later observed to emerge. To account for the differing size of the sentinel fields, these probabilities were given a Beta prior with the mean chosen to be the fraction of the field effectively canvassed by the collector, thus conveying a priori information that a parasitised host in a larger field is less likely to be collected. A more detailed description of the priors is given in Appendix B. Fecundity was assumed constant in time for the duration of the study, based on [29].

To model the number of parasitised hosts present in each field at collection time, we assumed that time from oviposition to emergence was 19 to 25 days, distributed approximately according to a truncated normal distribution with variance of 2 days. Although literature on laboratory observed times often suggests a shorter period [30, 31], the field data observed by Kristensen et al. necessitated the use of this longer time distribution [7, 25] based on the number of emergences observed at longer times from collection. The mean number of collected parasitised hosts could then be calculated based on the population density history and the probability of collection in each field. Actual per-day emergence observations were then assumed to be a Poisson distribution parameterised by this mean, resulting in a likelihood of the emergence data given the model. Calculations for the release field were done in a similar manner, but restricted to the collection grid.

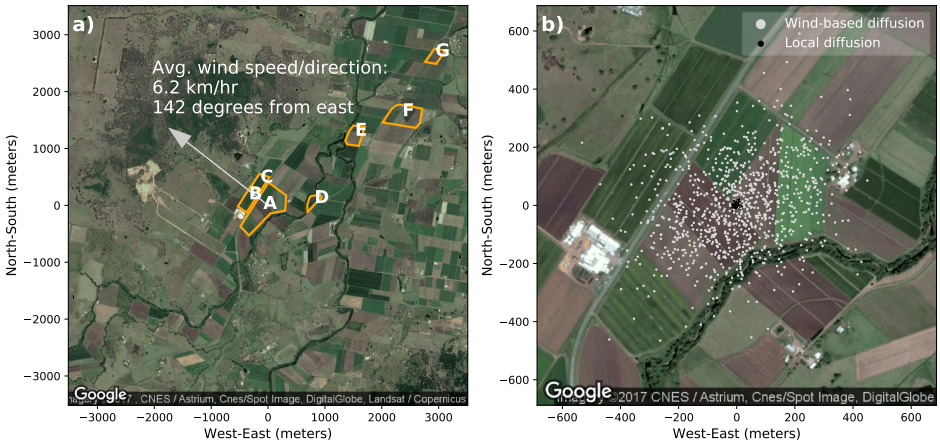

FIGURE 2. (a) Location of release field 'A' and data collection fields 'B'-'G', with arrow indicating average wind direction. (b) Visualisation of wind-based (white dots) and local (black dots) diffusion from the release point after one day based on fitted parameters.

Due to the 25×25 m cell resolution of the model, adult count data collected in each cardinal direction out to a maximum distance of 22 m was not included in this study. Direct observation of adults in the release field grid was included (see Fig. 4) and used to fit model parameters alongside emergence data. The likelihood of the data given the model was once again assumed to be Poisson. To calculate the mean of the distribution, we assumed the probability of observing a given parasitoid present in a grid cell was equivalent for each cell scaled by the number of leaves turned over, which was either 90 or 270 depending on the cell. Parasitoid movement during the data collection period was not considered.

Implementation details. In order to implement the wasp dispersal model in a simulation, certain decisions must be made concerning the time and space discretisation. In discretising time, we approximated all described processes using a left Riemann sum: more precisely, we converted all probability densities that were a function of time (e.g. $h(t, \mathbf{w}_d(t); \theta_h)$ and its subfunctions) into probability mass functions by assuming that wind velocity is constant on each time interval, which we chose to be 1 minute in length. Since wind data was only provided in 30 minute intervals, we used a linear interpolation to specify velocity on a per-minute basis.

The grid resolution for spatial discretisation was chosen as a balance between computational cost when calculating maximum a posteriori parameter estimates and the need to maintain a fine enough resolution to place each point in the release field collection grid into different discretisation cells. Since all wasp movement represents a solution to Eqn. 1 starting from a point mass, and thus is a normal distribution, spatial discretisation requires evaluating the bivariate normal cumulative distribution function to acquire total density in each cell. Though we implement this calculation using a Fortran library [32] through SciPy, this process typically represents the primary computational bottleneck since it must be done at each

point in the time discretisation for as many cells as constitutes the non-negligible area of the probability distribution, defined in our code as the smallest square block of cells centered at the mean which integrates to at least 0.999.

Our implementation of the model is written in Python 3.5 and has been specifically designed for general reuse and transparency, including basic documentation and pervasive comments within the code. Daily dispersal is computed in parallel utilising the Python multiprocessing library, and the convolution of each resulting distribution if via FFT may be performed on a GPU if CUDA and the necessary Python libraries are installed. Field data is organised using the Pandas library, which has some similarities to R. The Bayesian modelling framework is built and implemented using the PyMC library. While the time required to run the model for a given parameter set is dependent on domain size, resolution, and dispersal variance, typical run times for a domain of 64 km^2 are around one minute for 19 simulated days. The source code for this implementation available and maintained at https://github.com/mountaindust/Parasitoids [33].

3. Results

Results for the maximum a posteriori estimate of the model parameters (Table A2) reveal local and wind-based diffusion coefficients that are largely in line with findings in the literature [7]. Local movement was found to be uncorrelated, with a standard deviation of several meters in either direction. The data further suggested that there was no local drift based on average wind direction. For wind-based diffusion, results suggested standard deviations around 0.15 km with an east-west bias. This bias is likely caused by the species' phototactic response to the rising and setting of the sun [34, 35, 36]. In contrast to local movement, this diffusion was found to be positively correlated. The resulting slope of the best linear unbiased prediction of y given x for this distribution is 0.84 which, in its north-eastern direction, corresponds to 40 degrees north from east. This is roughly the direction of the sentinel fields E and G (Fig. 2), so it is likely that this is an effort on the part of the maximum a posteriori algorithm to either spread wasps further in these fields' direction or to make up for a fixed flight duration, as this is approximately the average wind heading for winds that wasps actually fly in (the average heading for wind speeds under 1 m/sec is 57.4 degrees north from east - see Fig. A1 in the appendix).

Wind-based flight time was held constant at 30 min when estimating parameters, with wasps able to fly a distance proportional to the per-minute wind speed. The constant of proportionality relating wind speed to parasitoid drift, μ_r, was fit to data as 1.18. This value is similar to what was found by [7] to provide the best fit for data collected from the Kalbar site and reinforces a roughly 30 minute flying period. However, maximum a posteriori parameter estimates fit the model so that wasps fly during a more extended portion of the day than might be expected, roughly 7:18 am to midnight. At the Kalbar site, at the time of year the field study was conducted, sunrise and sunset occurred at 6:00 am and 8:00 pm, respectively. We consider it likely that these extended flight times reflect lack-of-fit to emergence data further away from the release site as described below. Flight probability as a function of wind speed was fit to be 0.5 at 1.26 m sec^{-1}. With the fitted scaling parameter $b_w = 3.91$, the probability of wind-advected flight with wind speed of 0 m sec^{-1} at takeoff is roughly 0.99, and at 2 m sec^{-1}, it is about 0.05.

Fig. 3a-d shows the model prediction for wasp density (number / $625m^2$) 3, 6, 9, and 19 days post release, respectively. Predictions of 3, 6, and 9 days were chosen to reflect the times at which data for direct observation counts of adult parasitoids were recorded for the release field (see Fig. 4). 19 days post release, leaves were collected in each of 7 fields (red outlines in Fig. 3a-d, labelled A-G in Fig. 3e-f) and subsequently observed for emergences (Fig. 3e) in order to assess long-distance dispersal. Fig. 3e-f shows emergence observations (number of emergences) (e) and the model's projected emergences (number / $100m^2$) in the field (f) from hosts in a parasitised state 19 days post release using the parameters in Table A2 in the appendix.

As can be seen in Fig. 3f, our model broadly captures emergence trends in fields close to the release field A, but largely underestimates the number of parasitoids in fields further away. Also of particular note is the dual spatial-scale nature of the wasp distribution in Fig. 3a-c. Over the entire landscape, the distribution may be characterised as heavy-tailed with the tail skewed roughly in the average direction of the wind (northwest). However, on a more local spatial scale where the wasp density is higher (yellow and green), the distribution is skewed instead to the northeast in a direction more closely following the average low-velocity wind heading. This dual-spatial scale dynamic continues up until 13 days post release, when local dynamics begin to merge with the average full velocity profile of the wind. At 16 days post release, the wasp density is similar in character to that seen in Fig. 3d, where it should be noted that the centre of the distribution actually lies to the north of the release field rather than the northwest, even though the distribution is skewed to the northwest with the average wind. Unfortunately, there were no sentinel fields in the northwest direction for direct comparison with data.

Fig. 4 shows the model's projected density of parasitoids (individuals / $100m^2$) over the release field as a surface plot along with the observed parasitoid densities normalised by collection methodology (shown as bars). The largest peak in this local projected parasitoid density falls northwest and northeast of the release site, which correspond to the average wind direction and average low-velocity wind direction respectively. This is in general agreement with observed data and the northeast direction is further represented in the model outside the release field by a pocket of relatively high population density that predominately aligns to the northeast through 9 days post release (Fig. 3a-c).

These observations are also evident when examining the maximum a posteriori estimates for parameters in the Bayesian modelling framework (Table A3 in the appendix). Specifically, the data suggests a probability of 0.037 to observe any single given emerging parasitoid in field G that was incubating on the collection day. A priori, this number seems rather high given that the parasitised host is around a millimetre in size, possibly out of view on foliage, and could be anywhere in the sentinel field. Probabilities fitted for fields closer to the release point were anywhere from 0.00031 to 0.0034, with the exception of field D which was 0.033 and field E which was 0.0123. We note that the probability for observing a given emerging wasp in the release field was 0.064, but in this case, the area was heavily restricted to a sampling grid, and the probability only applies to incubating wasps located within this grid.

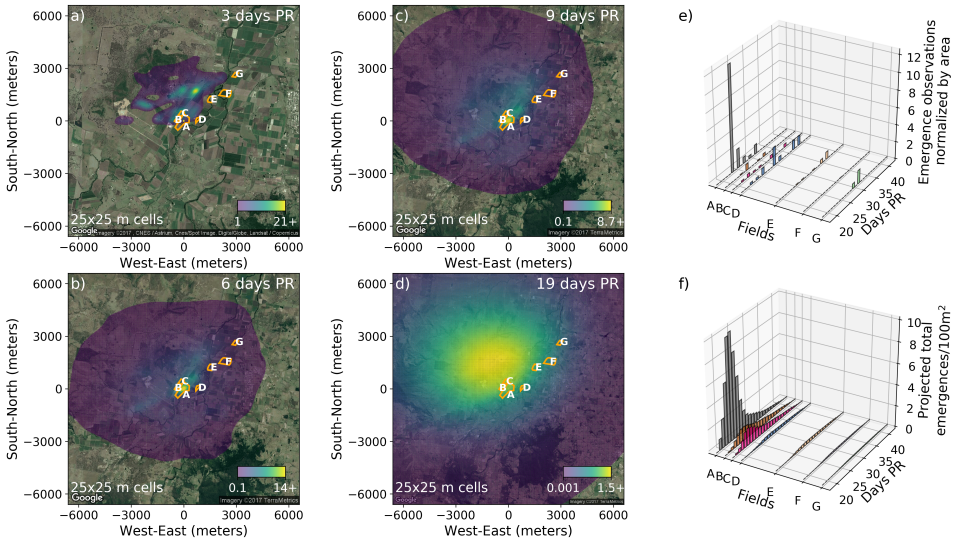

FIGURE 3. Results from the spatial model of parasitoids showing the expected mean number of individuals per 25x25 m cell (a) 3 days, (b) 6 days, (c) 9 days, and (d) 19 days post release, corresponding to data collection days. (e) Parasitoid emergence observation data per 100 m^2 (collection within rigorous sampling grid in the release field) and 10000 m^2 (sentinel fields).

(f) model projected in-field emergences per 100 m^2.

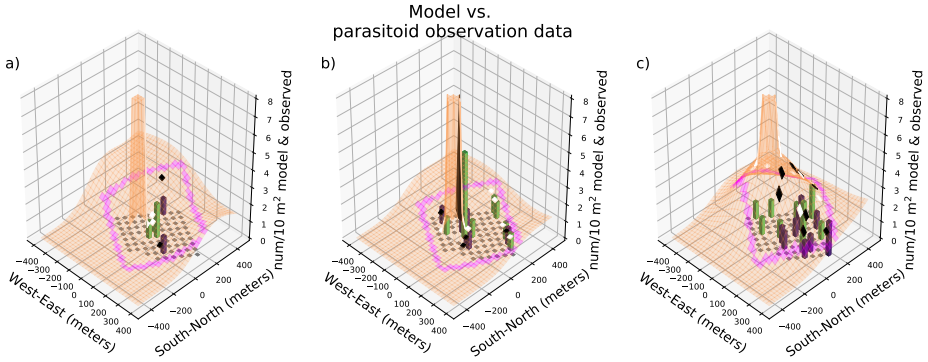

FIGURE 4. Spatial model of parasitoid numbers per 10 m^2 (surface plot) together with counts of observed wasps (bars) in the release field collection grid (a) 3 days, (b) 6 days, and (c) 9 days post release. Green bars (white boxes) denote 270 leaves sampled within 9 m^2; blue bars (black boxes) denote 90 leaves sampled within 1 m^2. The magenta box is roughly the grid boundary on the surface.

In addition to maximum a posteriori results, we fed the model into an adaptive Metropolis-Hastings Monte Carlo Markov Chain (MCMC) algorithm in order to analyse the shape of the posterior distribution for each of the model's parameters.

This algorithm features block-updating utilising a multivariate normal jump distribution whose covariance is tuned during sampling [37, 38]. Given that our model is spatially explicit, with a discrete daily time-step, and a Bayesian modelling layer connecting proxy data to predicted wasp density, analytical derivatives are not available and therefore Langevin and Hamiltonian MCMC methods (such as those seen in [23]) cannot be used. Additionally, posterior distributions could not be obtained. It is important to note that, even after coding for efficiency, realisations of a fully spatial, multi-scale model take far longer to obtain than is common in statistical Bayesian posterior analysis. One million model realisations represents a typical number for fully formed and converged posteriors. In our case, 10,000 model realisations requires approximately one week to produce. Therefore, after 220,000 model samples representing 3 months of computation time, the algorithm has yet to converge to a posterior distribution.

4. Discussion

On a spatial scale of less than 1 km, the model reproduces results seen in the field, suggesting that passive drift-diffusion with active behaviour at take-off may often be sufficient to capture the spread of small insects from point release locally. Despite the coarse nature of our field data, the Bayesian framework was clearly able to separate out two modes of diffusion, local and wind-based, when given similar priors (see Table A1 in the appendix). The east-west phototactic bias noted in the literature was also reproduced on the landscape scale. A strong fitted correlation coefficient for wind-based flight suggests that the model could be improved by allowing flight duration to vary, which would roughly stretch the resulting distribution along the direction of the wind. Finally, our model indicates that within the release field and the fields immediately surrounding it, the predominant local dispersal direction may be forecast for up to two weeks by averaging over wind velocities with magnitude less than 1 m/sec.

On spatial scales over 1 km, our results strongly suggest that drift-diffusion based solely on spatially homogeneous and temporally averaged wind velocities are insufficient to capture the long-distance dispersal of parasitoid wasps in the field. This fact is evident both in the inflated observation probabilities returned by the Bayesian framework for the far sentinel field (G) and a midrange sentinel field (D), and by the increasingly poor fit of the model to emergence data as the collection field gets further away from the release point (Figure 3). A positively correlated wind-based diffusion pattern and extended flight hours also point to this conclusion.

The layout of the fields was roughly perpendicular to the average wind direction, yet parasitoids reached high densities in further fields. As noted in the previous work [7, 25], wind speed during the study was related to wind direction such that high wind speeds tended to point perpendicular to the further fields and lower wind speed tended to flow parallel. Therefore a model that constrains parasitoids to only fly when wind speeds are relatively low allows parasitoids to remain in the cultivated area and disperse to far fields. However even with this behavioural mechanism included, the parasitoid densities predicted by the model are low compared to the data. We suggest two possible non-mutually exclusive explanations for the higher-than-predicted concentrations of parasitoids in far fields: (1) variability of wind-flow, both spatially (across the landscape and as a function of height) and temporally, or (2) parasitoids' ability to direct landing towards cultivated land.

First, on spatial scales on the order of a kilometre or more, it is likely important to resolve the spatial and smaller scale temporal variability of the wind. In order to obtain an analytical solution, our model assumed uniform wind speed and direction in space at any given point in time. Wind speed does vary as a function of height, and parasitoids that are advected upward on, for example, thermal updrafts may be carried significantly farther than those flying near the ground [39]. Previous work suggests that other tiny insects may suspend active flight once airborne, potentially extending flight times and distances travelled [40]. Furthermore, heterogeneous landscapes and vegetation will create spatially varying winds [41]. Complex structures such as rolling eddies, convection cells, and other instabilities generated by density stratifications may also be present and greatly alter flight trajectories [42]. Note that characterisation of these complex structures requires additional spatial and temporal resolution of wind, and we were not able to capture these effects in this study given the limited wind data. Gusts may also play a role in longer than predicted dispersal distances. Given that the parasitoids strongly prefer to fly on calm days, we do not think that gusts alone can explain the differences we see between actual and predicted long distance dispersal. Overall, our results suggest that it is important to resolve the spatially and temporally varying air flow above the landscape in order to correctly predict the dispersal of passively moving, wind-borne organisms.

Second, while very small insects in the wind-stream are generally regarded as 'passive' dispersers, and cannot control their direction of flight, they can potentially control entry to and exit from the wind-stream There is some evidence that visual flight-arrest triggers may be used by small insects to land in habitat that is more likely to provide them with needed food and reproduction resources, thus reducing the risk that wind-borne dispersal holds of transporting individuals to resource-poor areas. In this study, the agricultural area formed a thin strip (approximately 3 km wide) along a creek, and the wider landscape was predominantly livestock pasture. If the parasitoids were triggered to land by the greener irrigated crop fields compared to the browner surrounding landscape, and particularly if the parasitoids are capable of detecting the oncoming colour difference that signalled the edge of the agricultural area, then this may have reduced the proportion of parasitoids leaving the agricultural area. Most phytophagous insects are responsive to plant-related light wavelengths as triggers to land [43, 44]. They show an attraction to plant-related colours, and are influenced by colour contrasts that occur at field edges or between crop types [45, 46]. In related *Eretmocerus eremicus*, females in particular have been demonstrated to respond to a plant cue in a wind tunnel experiment [36], and UV-absorbing plastic sheeting impairs the ability of *Eretmocerus mundus* to locate plants containing hosts [47]. This raises the possibility that parasitoids in the wind-stream may respond to visual stimuli in order to reduce the risk of being carried past potentially resource-rich areas [48], which would explain the higher concentrations of parasitoids indicated by the data compared to that predicted by the model.

5. Conclusions

Matches and mismatches between data and model predictions allow us to refine our hypotheses about the mechanisms of movement. In this study, a Bayesian framework was used to rigorously connect a mechanistic, mathematical simulation

model to sparse proxy data in order to verify a stratified dispersal mechanism. It was capable of identifying two separate diffusion kernels associated with local and wind-borne dispersal modes respectively, and the exogenous condition (wind speed) that likely triggers the switch between dispersal modes. The model also indicated that, while conditional wind-borne dispersal provides some of the explanation for how the parasitoid successfully dispersed to other host-containing fields on the landscape scale, it could not explain how the parasitoid did so in such high numbers. Therefore two possible explanations – variability of the windflow across the landscape scale, or the use of visual host-habitat cues as triggers for landing – are proposed. The Bayesian framework provided here can be used to test either of these hypotheses once the appropriate data is collected.

Data accessibility

All software and data needed to reproduce the results in this study is maintained at https://github.com/mountaindust/Parasitoids and provided open source and free of charge under the terms of the GNU GPLv3 license agreement. For other licensing arrangements, please contact the lead author. A snapshot of the software and data at time of acceptance is available with DOI (10.5281/zenodo.556462) as release version 1.0.0 [33].

Competing interests

The authors have no competing interests to report.

Authors' contributions

Christopher Strickland designed the model, authored all code, interpreted the results, and drafted the manuscript and figures. Nadiah Kristensen provided the data, ecological motivation for the model, and drafted the manuscript. Laura Miller conceived of the study, coordinated the study, and edited the manuscript.

Acknowledgements

The authors wish to acknowledge Richard Smith for his extremely helpful guidance in formulating the Bayesian framework used to fit model parameters to data. We would also like to thank our four anonymous reviewers for their insightful and detailed comments that greatly helped improve our manuscript.

Funding

This material was based upon work partially supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute and CBET-1511427 and DMS-1151478 to L.A.M. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Hastings A, Cuddington K, Davies KF, Dugaw CJ, Elmendorf S, Freestone A, et al. The spatial spread of invasions: new developments in theory and evidence. Ecology Letters. 2005;8(1):91–101.

[2] Cronin JT, Reeve JD. Host–parasitoid spatial ecology: a plea for a landscape-level synthesis. Proceedings of the Royal Society of London B: Biological Sciences. 2005;272(1578):2225–2235.

[3] Zera AJ, Denno RF. Physiology and ecology of dispersal polymorphism in insects. Annual Review of Entomology. 1997;42(1):207–230.

[4] Blakeslee A, McKenzie C, Darling J, Byers J, Pringle J, Roman J. A hitchhiker's guide to the Maritimes: anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland. Diversity and Distributions. 2010;16(6):879–891.

[5] Suarez AV, Holway DA, Case TJ. Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. Proceedings of the National Academy of Sciences. 2001;98:1095–1100.

[6] Schellhorn NA, Parry HR, Macfadyen S, Wang Y, Zalucki MP. Connecting scales: Achieving in-field pest control from areawide and landscape ecology studies. Insect science. 2015;22(1):35–51.

[7] Kristensen NP, De Barro PJ, Schellhorn NA. The initial dispersal and spread of an intentional invader at three spatial scales. PLOS ONE. 2013;8(5):e62407. Available from: `http://dx.doi.org/10.1371/journal.pone.0062407`.

[8] Colombari F, Battisti A. Spread of the introduced biocontrol agent Torymus sinensis in north-eastern Italy: dispersal through active flight or assisted by wind? BioControl. 2016;61(2):127–139.

[9] Van Dyck H, Baguette M. Dispersal behaviour in fragmented landscapes: routine or special movements? Basic and Applied Ecology. 2005;6(6):535–545.

[10] Wang Z, Wu J, Shang H, Cheng J. Landscape connectivity shapes the spread pattern of the rice water weevil: a case study from Zhejiang, China. Environmental management. 2011;47(2):254–262.

[11] Evans E, Bolshakova V, Carlile N. Parasitoid dispersal and colonization lag in disturbed habitats: biological control of cereal leaf beetle metapopulations. Journal of Applied Entomology. 2015;139(7):529–538.

[12] Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, et al. A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences. 2008;105(49):19052–19059.

[13] Poelman EH, Oduor AM, Broekgaarden C, Hordijk CA, Jansen JJ, Van Loon JJ, et al. Field parasitism rates of caterpillars on Brassica oleracea plants are reliably predicted by differential attraction of Cotesia parasitoids. Functional Ecology. 2009;23(5):951–962.

[14] Williams IH, Frearson DJ, Barari H, McCartney A. First field evidence that parasitoids use upwind anemotaxis for host-habitat location. Entomologia experimentalis et applicata. 2007;123(3):299–307.

[15] Weyman GS. A review of the possible causative factors and significance of ballooning in spiders. Ethology, Ecology, and Evolution. 1993;5(3):279–291.

[16] Weyman GS. Laboratory studies of the factors stimulating ballooning behavior by linyphiid spiders (Araneae, Linyphiidae). Journal of Arachnology. 1995;25:75–84.

[17] Cronin TW, Forward RB. Vertical migration cycles of crab larvae and their role in larval dispersal. Bulletin of Marine Science. 1986;39(2):192–201.

[18] Reynolds AM, Reynolds DR. Aphid aerial density profiles are consistent with turbulent advection amplifying flight behaviours: abandoning the epithet 'passive'. Proceedings of the Royal Society B: Biological Sciences. 2009;276(1654):137–143.

[19] Ellison AM. Bayesian inference in ecology. Ecology Letters. 2004;7:7 509–520.

[20] Ovaskainen O, Rekola H, Meyke E, Arjas E. Bayesian methods for analyzing movements in heterogeneous landscapes from mark–recapture data. Ecology. 2008;89(2):542–554.

[21] Forcada J. Can population surveys show if the Mediterranean monk seal colony at Cap Blanc is declining in abundance? J Appl Ecol. 2000;37:171–181.

[22] Barrowman NJ, Myers RA, Hilborn R, Kehler DG, Field CA. The variability among populations of coho salmon in the maximum reproductive rate and depensation. Ecol Appl. 2003;13:784–793.

[23] House T, Ford A, Lan S, Bilson S, Buckingham-Jeffery E, Girolami M. Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. Journal of The Royal Society Interface. 2016;13(121):20160279.
[24] Cummings CL, Alexander HM, Snow AA, Rieseberg LH, Kim MJ, Culley TM. Fecundity selection in a sunflower crop-wild study: can ecological data predict crop allele changes? Ecol Appl. 2002;12:1661–1671.
[25] Kristensen NP, Schellhorn NA, Hulthen AD, Howie LJ, Barro PJD. Wind-borne dispersal of a parasitoid: The process, the model, and its validation. Environ Entomol. 2013 dec;42(6):1137–1148. Available from: http://dx.doi.org/10.1603/en12243.
[26] Liu TX, Stansly PA, Gerling D. Whitefly parasitoids: distribution, life history, bionomics, and utilization. Annual review of entomology. 2015;60:273–292.
[27] Bellamy DE, Byrne DN. Effects of gender and mating status on self-directed dispersal by the whitefly parasitoid Eretmocerus eremicus. Ecological Entomology. 2001;26(6):571–577.
[28] Walters KFA, Dixon AFG. The effect of temperature and wind on the flight activity of cereal aphids. Ann Appl Biol. 1984;104:17–26.
[29] Villanueva-Jimenez JA, Schellhorn NA, De Barro PJ. Comparison between two species of *Eretmocerus* (Hymenoptera: Aphelinidae): Reproductive performance is one explanation for more effeffect control in the field. Biological Control. 2012;63:333–338.
[30] Yang NW, Wan FH. Host suitability of different iinstar of *Bemisia tabaci* biotype B for the parasitoid *Eretmocerus hayati*. Biological Control. 2011;59:313–317.
[31] Zhang YB, Yang NW, Sun LY, Wan FH. Host instar suitability in two invasive whiteflies for the nanatural occurring parasitoid *Eretmocerus hayati* in China. Journal of Pest Science. 2015;88(2):225–234.
[32] Genz A. Numerial computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics. 1992;1(2):141–149.
[33] Strickland C. mountaindust/Parasitoids: Code release corresponding to 2017 Royal Society publication. Zenodo; 2017.
[34] Blackmer J, Byrne D. Environmental and physiological factors influencing phototactic flight of *Bemisia tabaci*. Physiological Entomology. 1993;18:336–342.
[35] Blackmer J, Byrne D. Flight behaviour of *Bemisia tabaci* in a vertical flight chamber: effect of time of day, sex, age and host quality. Physiological Entomology. 1993;18:223–232.
[36] Blackmer JL, Cross D. Response of *Eretmocerus eremicus* to skylight and plant cues in a vertical flight chamber. Entomologia Experimentalis et Applicata. 2001;100:295–300.
[37] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. Bernoulli. 2001;7(2):223–242.
[38] Patil A, Huard D, Fonnesbeck C. PyMC : Bayesian Stochastic Modelling in Python. Journal of Statistical Software. 2010;35(4). Available from: http://dx.doi.org/10.18637/jss.v035.i04.
[39] Chapman JW, Reynolds DR, Smith AD, Smith ET, Woiwod IP. An aerial netting study of insects migrating at high altitude over England. B Entomol Res. 2004;94(02):123–136. Available from: http://dx.doi.org/10.1079/ber2004287.
[40] Santhanakrishnan A, Robinson AK, Jones S, Low AA, Gadi S, Hedrick TL, et al. Clap and fling mechanism with interacting porous wings in tiny insect flight. J Exp Biol. 2014;217:3898–909. Available from: http://dx.doi.org/10.1242/jeb.084897.
[41] Nathan R, Sapir N, Trakhtenbrot A, Katul GG, Bohrer G, Otte M, et al. Long-distance biological transport processes through the air: Can nature's complexity be unfolded in silico? Divers Distrib. 2005;11:131–137. Available from: http://dx.doi.org/10.1111/j.1366-9516.2005.00146.x.
[42] Suter RB. An aerial lottery: The physics of ballooning in a chaotic atmosphere. Journal of Arachnology. 1999;27:173–189. Available from: http://www.jstor.org/stable/3705999.
[43] Finch S, Collier R. Host-plant selection by insects–a theory based on 'appropriate/inappropriate landings' by pest insects of cruciferous plants. Entomologia experimentalis et Applicata. 2000;96(2):91–102.
[44] Parry HR. Cereal aphid movement: general principles and simulation modelling. Movement ecology. 2013;1(1):1.
[45] Kostal V, Finch S. Influence of background on host-plant selection and subsequent oviposition by the cabbage root fly (Delia radicum). Entomologia experimentalis et applicata. 1994;70(2):153–163.

[46] Ciss M, Parisey N, Dedryver CA, Pierre JS. Understanding flying insect dispersion: Multiscale analyses of fragmented landscapes. Ecological informatics. 2013;14:59–63.
[47] Chiel E, Messika Y, Steinberg S, Antignus Y. The effect of UV-absorbing plastic sheet on the attraction and host location ability of three parasitoids: Aphidius colemani, Diglyphus isaea and Eretmocerus mundus. BioControl. 2006;51(1):65–78.
[48] Compton SG. Sailing with the wind: dispersal by small flying insects. In: Bullock JM, Kenward RE, Hails RS, editors. Dispersal Ecology: 42nd Symposium of the British Ecological Society. Blackwell Publishing, Oxford, UK; 2002. p. 113–133.

(1) University of North Carolina-Chapel Hill, Department of Mathematics, Phillips Hall, CB #3250, Chapel Hill, NC 27599, U.S.A.

(2) Statistical and Applied Mathematical Sciences Institute (SAMSI), 19 T.W. Alexander Drive, P.O. Box 14006, Research Triangle Park, NC 27709, U.S.A.
E-mail address: wcstrick@email.unc.edu
URL: http://www.christopherstrickland.info

(3) National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077
E-mail address: nadiah@nadiah.org

(4) University of North Carolina-Chapel Hill, Department of Biology, Coker Hall, CB #3280, Chapel Hill, NC 27599, U.S.A.
E-mail address: lam9@email.unc.edu