Testing models of bee foraging behavior through the analysis of pollen loads and floral density data

Philippe Marchand[a,]*, Alexandra N. Harmon-Threatt[b] and Ignacio Chapela[a]

[a] Department of Environmental Science, Policy and Management, 130 Mulford Hall, University of California, Berkeley, CA 94720-3114 USA.

[b] Department of Entomology, University of Illinois at Urbana-Champaign, 505 South Goodwin Ave., Urbana, IL 61801 USA.

* Corresponding author at: National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, MD 21401 USA. Tel.: 1-443-743-2230. E-mail address: pmarchand@sesync.org (P. Marchand)

Abstract

The composition of social bees' corbicular pollen loads contains information about both the bees' foraging behavior and the surrounding floral landscape. There have been, however, few attempts to integrate pollen composition and floral landscape to test hypotheses about foraging behavior. Here, we present an individual-based model that generates the species composition of pollen loads given a foraging model and a spatial distribution of floral resources. We apply this model to an existing dataset of inflorescence counts and bumble bee pollen loads sampled at different field sites in California. For two out of three sites, a foraging model consisting in correlated random walks with constant preferences for each plant species provides a plausible fit for the observed distribution of pollen load content. Pollen load compositions at the third site could be explained by an extension of the model, where different preferences apply to the choice of an initial foraging patch and subsequent foraging steps. Since this model describes the expected level of pollen load differentiation due solely to the spatial clustering of conspecific plants, it provides a null hypothesis against which more complex descriptions of behavior (e.g. flower constancy) can be tested.

Keywords

approximate Bayesian computation, bumble bee, individual-based model, pollination, random walk

Published in *Ecological Modelling* 313, 41-49. doi: 10.1016/j.ecolmodel.2015.06.019 .

© 2015. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

1. Introduction

Along with nectar, pollen constitutes a major food source for social bee colonies. For example, a typical colony of honey bees (*Apis mellifera*) consumes 40 kg of pollen a year, requiring millions of foraging trips (Stanley and Linskens 1974). When foraging, both honey bees and bumble bees (*Bombus* spp.) accumulate pollen in the corbicula (or pollen baskets) located on their hind legs. By determining the composition of these pollen loads, which can be sampled non-destructively by capturing bees in the field or installing traps at the entrance of the hive or nest, pollination ecologists can obtain a record of the foragers' visits to different floral types.

Bee pollen analysis has been used to estimate the foraging range of a colony (Beil et al. 2008), compare vegetation profiles between widely separated locations (Diaz-Losada et al. 1998), study changes in resource use by one or more colonies over time (Aronne et al. 2012, Boff et al. 2011) and predict the amount of pollen flow between transgenic and conventional crop varieties (Ramsay et al. 2003). Most published studies report aggregated results – i.e. the average proportion of pollen sources at the hive level – and until recently, there were few attempts to model the variability in composition between pollen loads (de Valpine and Harmon-Threatt 2013).

Although it may simplify the determination of colony-level components of foraging behavior, the aggregation of pollen loads obscures other aspects of foraging – such as the area covered in a single bout or the level of flower constancy shown by pollinators – that could be investigated by analyzing pollen samples from individual bees. Extracting this information from a given case study requires, in addition to the compositional pollen data, an estimate of the spatial distribution of visited plants as well as a modelling framework that links plant distribution, foraging behavior and pollen load composition.

In this study, we establish such a link – between floral landscape, foraging behavior and pollen load composition – through an individual-based modelling approach. Using estimates of the spatial distribution of each pollen source in the field, our model simulates foraging paths by following stochastic rules describing the bees' movement and floral preferences. Its output is a distribution of pollen load compositions corresponding to the specific field configuration and foraging parameters.

We present a simple parametrization of the model, where foraging paths consist of correlated random walks and floral preferences are summarized in fixed vectors describing the relative attractiveness of each species. We apply the resulting model to a data set that includes pollen counts from bumble bees (*Bombus vosnesenskii*) foraging at different sites in Northern California as well as inflorescence counts sampled on a quadrat grid at each site. We demonstrate through a sensitivity analysis that the effective parameter space of the movement model can be reduced as many parameters have redundant effects on the model output. Finally, we use approximate Bayesian computation (ABC) to estimate the model's parameters based on a comparison of simulated and observed summary statistics – specifically, the average prevalence of each species' pollen and the compositional differentiation between individual pollen loads.

2. Materials and methods

2.1. Model specification

The model description in this section follows the ODD (Overview, Design concepts, Details) protocol (Grimm et al. 2006, 2010).

2.1.1. Purpose

Our model aims to predict the composition of bee pollen loads given the spatial distribution of pollen sources and parameters of bee foraging behavior (shape of foraging paths and preferences for certain floral species); conversely, the model would serve to infer parameters of foraging behavior from the observed composition of pollen loads sampled from a known floral landscape.

2.1.2. Entities, state variables and scales

The spatial distribution of floral resources is described as a density field **d**, where $d_j(x,y)$ is the density (inflorescences m^{-2}) of species j at the field coordinates (x,y). Inflorescences (clusters of flowers arranged on a stem) are used as the basic floral unit since their abundance was found to be a better predictor of resource use for the different sites studied here (Harmon-Threatt, unpublished data). From empirical estimates of **d** on a rectangular grid of n_x by n_y grid points with spacings of Δ_x and Δ_y , the model interpolates the density field at any point within the grid (see interpolation submodel below).

Each run of the model simulates the activity of n_b foraging bees based on two sets of parameters, which are the same for all bees in a given run. The first set of parameters is used in modelling foraging paths as correlated random walks (CRW) and includes: the number of inflorescences visited on the path (n_s); the root mean square (RMS) value of the step length (l_s); the mean cosine of the turning angle distribution (ρ); the frequency (f_j) and RMS length (l_j) of occasional larger steps or "jumps" in the path. Details of the CRW model are presented in the submodels section below. The second set of parameters describes the relative attractiveness of each species to foragers and includes two vectors, $\mathbf{a}_{init}$ and $\mathbf{a}_{succ}$: the former affects bee preferences for the starting point of the foraging bout while the latter affects bee preferences for successive flower visits. This distinction is motivated by the observation that bees may be attracted to a patch by a specific floral type, but will also visit less preferred flowers located in the same patch (Seifan et al. 2014). Therefore, we expected $\mathbf{a}_{succ}$ to show either the same or more even (less discriminative) preferences than $\mathbf{a}_{init}$.

During the simulation, the state of a bee is represented by its current position (x,y), its current direction θ (based on a line from the last to the current position), the plant species visited at the current position as well as the total number of steps taken to this point. The output of a model run is the composition matrix $\boldsymbol{P}$ where P_{ij} represents the proportion of the j^{th} floral species in the i^{th} bee's pollen load. A summary of the parameters and outputs of the model is presented in Table 1.

We note that while all positions and distances in the field are expressed in meters, the foraging parameters l_s and l_j are dimensionless. As we explain in the CRW submodel description, the actual step size is adjusted dynamically based on the local inflorescence density.

2.1.3. Process overview and scheduling

The model simulates each forager's activity as an independent realization of the following stochastic algorithm:

- The bee selects one of the grid points as the first position on its foraging path. Its initial direction (θ, in radians) is selected uniformly over ($-\pi$, π). The probability of starting from a given point is proportional to a weighted sum of the densities of all species present at that point, with weights given by $\mathbf{a}_{init}$. That is, using j as an index over species and k as an index over grid points:

$$Pr(k) = \frac{\sum a_{init(j)} d_j(x_k, y_k)}{\sum \sum a_{init(j)} d_j(x_k, y_k)} \quad . \quad (1)$$

Foraging paths are restricted to start on one of the grid points because the interpolation method does not produce a simple analytical form of the density field. Thus sampling the initial position in continuous space would be impractical.

- Starting from that point, the bee's path is simulated using the CRW submodel.
- The inflorescence density at each point in the path, which serves both to scale the CRW step size and to determine the probability of visiting each species, is calculated by the interpolation submodel.
- The inflorescence type visited at each point in the path is selected with a probability proportional to the local density of each species, this time weighted by $\mathbf{a}_{succ}$:

$$Pr(j, x, y) = \frac{a_{(j)} d_j(x, y)}{\sum a_{(j)} d_j(x, y)} \quad . \quad (2)$$

- The proportion of each species in the bee's pollen load is taken to be equal to the proportion of inflorescences of that species among those visited.

2.1.4. Design concepts

Here we highlight some of the basic principles or key assumptions that are implicit in the model overview above. First, we treat inflorescence density as a continuous quantity. This is primarily done for computational efficiency, since at the densities and field sizes typical of our data, the number of individual inflorescences by field is of the order of 10^4–10^5. Although the foraging path and floral choices are based on a stochastic model, we assume the parameters of that model are fixed for all foragers in a given simulation. Finally, our model does not take into account the variation in pollen rewards between inflorescences of the same species or across different species.

Notably absent from this model are the collective aspects of social bee foraging behavior – the interactions between foragers, colony-level information gathering processes about the floral landscape, etc. Without explicitly modelling these processes, we can consider that the attractiveness vectors $\mathbf{a}_{init}$ and $\mathbf{a}_{succ}$ reflect foraging objectives or priorities learned at the colony level, and that conditional on these parameters, the paths of individual foragers are independent. Our algorithm for selecting an initial foraging point implicitly assumes that the colony has some knowledge of the whole field and apportions foraging effort between different areas in proportion to the density and attractiveness of species present.

The model also ignores how an individual forager's behavior may adapt to information gathered within a bout, such as the diminution of pollen rewards obtained from a preferred species. This would be an important factor to include in future iterations of this model, for cases where data on the variation of pollen rewards between inflorescences is available.

As stated above, the main observable output of our model is the a matrix $\boldsymbol{P}$ of pollen load compositions, from which we calculate the summary statistics used to evaluate the model's agreement with field data. In addition to the average proportion of each species across pollen loads, denoted as the vector $\mathbf{p}$, we are interested in the level of differentiation between pollen loads. We quantify the latter using an analog of F_{ST}, the fixation index of population genetics:

$$F_{ST} = \frac{\sum_j \sigma_{p_j}^2}{1 - \sum_j p_j^2}, \quad (3)$$

where p_j and σ_{pj}^2 are respectively the mean and variance of the proportion of species j across pollen loads. In this context, F_{ST} can be interpreted as the portion of the total species diversity (more specifically, the Gini-Simpson index) that is due to variation between, as opposed to within, pollen loads. In particular, $F_{ST} = 1$ if only a single species is represented in each pollen load; in our model, this would suggest that monospecific floral clusters are large compared to the typical bee foraging area.

2.1.5. Initialization

Since we use a Bayesian framework to infer foraging parameters, as described in section 2.3, their values for each model run are selected from a prior distribution.

2.1.6. Input data

All aspects of the field description, including the number of plant species, the sampling grid size and spacings, and the density estimates for each grid point are supplied as fixed input for the model.

2.1.7. Submodels

Density interpolation: We use inverse distance weighing (IDW) to interpolate the inflorescence density between grid points. The density of each species at the bee's current position is estimated as the weighted average of its density at the four nearest grid points, with weights proportional to some inverse power of the distance to those grid points. When comparing different techniques for the interpolation of weed densities in a field, Dille et al. (2003) found that IDW with an exponent of 2 or 4 performed as well or better than more computationally-intensive methods such as kriging. We used an exponent of 2 for our main results, while verifying the effect of changing this exponent as part of the sensitivity analysis.

Correlated random walk: Studies of foraging behavior often model foraging paths as variants of the simple random walk (Codling et al. 2008). Here, we simulate movement between inflorescences as correlated random walks (CRW) with n_s steps. The turning angle (change in θ)

between steps follows a wrapped Cauchy distribution and is generated through this formula from Bartumeus et al. (2005):

$$\Delta\theta = 2\arctan\left[\left(\frac{1-\rho}{1+\rho}\right)\tan\left[\pi\left(u-\frac{1}{2}\right)\right]\right], \quad (4)$$

where u is a uniform random variable over (0,1). The parameter ρ corresponds the mean cosine of the turning angle. When $\rho = 0$, the turning angle is uniformly distributed and the path is an uncorrelated random walk; in the limit $\rho = 1$, the path is a straight line.

To account for variation in inflorescence density across the field, we scale the length of each step of the random walk by a factor d_{nn}, an estimate of the "nearest-neighbor distance" between inflorescences, equal to the inverse square root of the total density at the current position of the bee:

$$d_{nn}(x,y) = \left(\sum_j d_j(x,y)\right)^{-1/2}. \quad (5)$$

This choice is motivated by the results of Levin and Kerster (1969), who found that the length of foraging steps is proportional to the spacing between inflorescences for multiple bee and plant taxa. In our model, the length L of a step follows a Rayleigh distribution with a root mean square (RMS) value equal to the product $d_{nn}\, l_s$. It is generated from the following formula:

$$L = d_{nn}\, l_s \sqrt{-\ln v}, \quad (6)$$

where v is a uniform random variable over (0,1). Since this distribution may underestimate the variability in step size along the foraging path, we consider an alternative parametrization where the bee may, with a probability f_j, take larger steps or "jumps" of RMS length l_j, which replaces l_s in Eq. (6).

A few restrictions were added to implement this basic CRW model. We set an absolute upper bound l_{max} (in meters) to prevent the step length from becoming unrealistically large in low-density regions of the field. Furthermore, if a step would result in the bee exiting the boundaries of the field grid, this step was rejected and a new step was generated, with the turning angle chosen from a uniform distribution (i.e. forcing $\rho = 0$) to prevent the bee getting stuck in a corner.

The individual-based model was implemented in Fortran 95 in order to rapidly perform the large number of simulations required for the ABC analysis. The model code is included in the electronic supplementary materials.

2.2. Bee pollen and plant data

We applied our model to a pollen data set collected by A.N. Harmon-Threatt (unpublished study). Harmon-Threatt sampled pollen loads from individual foragers of the species *Bombus vosnesenskii* at five different 1-ha sites at Mt. Diablo State Park and Briones Regional Park in Contra Costa County, California in 2009. Each site was separated by at least 1 km to limit overlap in foraging range of bee species. Sampling was conducted every two weeks between late May and early August resulting in 5 sampling rounds. Sites were only sampled when bees were present resulting in a total of 15 site-date combinations. Bees were hand netted on flowers within the 1-ha site and one pollen load was removed for later identification of plant species visited and to prevent recapturing the same bee. Between 4 and 21 bees were captured at each site-date with

a mean of 15.4 and a total of 231 bees. Pollen loads were homogenized and stained with fuchsin to allow identification of 300 randomly selected pollen grains under the microscope. Pollen samples from each blooming species were collected throughout the sampling to provide a library for comparison. Plant diversity was sampled in 100 one-m^2 quadrats arranged in a regular grid in the field: 10 quadrats spaced by 10 m in one direction, 5 quadrats spaced by 20 m each in the other, for a total area of 90 m x 80 m inside the one hectare site.

For the present work, we only considered plant species that composed at least 10% of one or more pollen loads. Trace pollen amounts from other species are less likely to be the result of active pollen foraging: they could be incidentally collected when foraging for nectar and later groomed into pollen baskets. In our analysis, we also ignored pollen loads where over 10% of the pollen was either unidentified or from species not found in any of the quadrats sampled at that site. This left us with three sites (BRD, MD4 and MD5) and two sampling dates per site where multiple species were present in significant amounts and enough (>10) pollen loads were available for analysis. California poppy (*Eschscholzia californica*) was represented in pollen loads from every site, while yellow star-thistle (*Centaurea solstitalis*), wide-bannered lupine (*Lupinus microcarpus*), fewflower clover (*Trifolium oliganthum*) and hairy vetch (*Vicia villosa*) constituted a major pollen source for at least one of the sites. For each pollen load, we calculated the proportion of each pollen type as if only those major species were present; these proportions formed the dataset which we compared to our simulation results.

By using the pollen proportions rather than the raw counts as our data, we ignore the additional variance caused by (multinomial) sampling of pollen from each bee. Preliminary simulation results incorporating this multinomial sampling confirmed that its contribution to F_{ST} ($\sim 1/N$, where $N = 300$ pollen grains per bee) was small enough to be neglected for the sake of reducing computation time, especially considering that the small number of bees sampled introduces much greater variability in F_{ST}.

2.3. Model fitting and checking

We fit our model to the observed bee pollen statistics by approximate Bayesian computing (ABC), a technique developed to infer parameters of stochastic simulation models for which the exact likelihood is intractable. ABC originated in the field of population genetics, where the high-dimensionality of genetic datasets and the large number of possible system histories (genetic trees) often prevent the efficient computation of full data likelihoods, even with Markov chain Monte Carlo (MCMC) methods (Tavaré et al. 1997, Beaumont et al. 2002). In the last five years, a increasing number of studies have applied ABC to various ecological models, including stage-structured population dynamics (Scranton et al. 2014) and community assembly (Jabot and Chave 2011).

In its most basic form, ABC consists in performing multiple simulations of the model, with parameters drawn from a prior distribution, then accepting those parameter values that provide results close enough to the data (based on a chosen distance measure and set of summary statistics) as an estimate of the posterior distribution. It provides a better approximation of the true likelihood if the chosen statistics are sufficient with respect to the parameters of interest and if the tolerance range (the distance above which a simulation is rejected) is narrow. The approximation of the posterior distribution of each parameter can be further improved by

performing a linear regression of this parameter as a function of the summary statistics, then using this regression to correct each parameter value towards the observed values of those statistics (Beaumont et al. 2002).

We used the 'abc' package (Csilléry et al. 2012) in R (R Development Core Team 2008) to estimate model parameters from the output of our simulation program. The 'abc' package uses the Euclidean distance to compare simulated and observed vectors of summary statistics, scaling each statistic by an estimate of its standard deviation to prevent any component from dominating the distance measure. It produces a posterior distribution by accepting parameter values for which the simulated summary statistics lie within a specified tolerance range, then corrects these parameter values based on a linear regression, as described above.

The 'abc' package also includes a cross-validation function that successively picks different individual outputs from the simulated set, treats that particular output as the "data" and attempts to predict the parameters that generated it by applying the ABC algorithm with the remaining simulations in the set. We used cross-validation to choose a tolerance rate for ABC that would maximize the precision of posterior estimates.

We fit the model separately at each of the three sites, using a vector of summary statistics that combined the mean proportions **p** and the level of differentiation F_{ST} for each of the two dates at that site. We left out the last component of **p** for each site-date combination, since it is fixed by the constraint that proportions sum to 1. We first attempted to fit a model with a single set of attractiveness coefficients ($\mathbf{a}_{init} = \mathbf{a}_{succ}$); if that model could not have plausibly produced the observed summary statistics at a given site, we considered a different model where the values $\mathbf{a}_{succ}$ may be more uniform than $\mathbf{a}_{init}$ (for reasons stated in section 2.1.2 above). We used posterior predictive checks (Gelman et al. 1996) to evaluate model goodness of fit. Specifically, we performed multiple simulations of the data using the median parameter values from the posterior distribution and obtained a predicted distribution of summary statistics. These include **p**, F_{ST} as well as an additional statistic corresponding to the number of mixed pollen loads – defined arbitrarily as those where no more than 90% of the pollen comes from a single species.

3. Results

3.1. Pollen and inflorescence data summary

For each site and date, Table 2 shows the number of pollen loads used in our analysis as well as the summary statistics calculated from the composition data. Pollen loads sampled at site MD5 were notably less differentiated – as indicated by the low F_{ST} and high number of mixed loads – than those from the other two sites. Table 3 indicates the mean inflorescence density per site and date for the five main pollen sources, calculated as the average of the inflorescence counts over all 1-m^2 quadrats. From these two tables, we note that each of the five plant species was a significant pollen source (forming at least 10% of at least one pollen load) for each site and date where it was present in quadrats.

3.2. Sensitivity analysis

For a given field configuration (distribution of inflorescences) and fixed values of the

attractiveness coefficients $\mathbf{a}_{init}$ and $\mathbf{a}_{succ}$, we found that the various parameters of the random walk model only influenced the simulated summary statistics based on their effect on the RMS distance (D) between the first and last inflorescences visited (Fig. 1). That is, any parameter variation that increased D by a certain amount – either by adding steps (n_s), making each step longer (l_s), adding occasional large steps (l_j) or increasing directional correlation in the path (ρ) – had approximately the same effect on the output statistics. (The specific parameter sets used to produce the results in Fig.1 are included in the electronic supplementary materials as Table S1.) This finding is consistent with the results obtained by Marchand (2013) using a simpler version of the model that did not account for variations in inflorescence density or differences in attractiveness between species. It allows us to reduce the dimensionality of our parameter space in the following ABC analysis by only varying one of the random walk parameters.

In most cases, increasing D led primarily to a reduction in F_{ST}, which is to be expected as bees have more probability of leaving monospecific plant clusters. Since our model's random walks are not directionally biased, longer walks also increase the chance of bees encountering less attractive species that are present at high densities, although our simulations show this effect is small (Fig. 1a). A special case occurs when the bees always start foraging on the same species (i.e. $a_{init} = 1$ for that species), but are less discriminative on successive steps (Fig. 1b). As we will see from the ABC results, this behavior best explains the observed statistics at site MD5. In that case, the primary effect of longer walks is to increase the representation of the less attractive species, while F_{ST} varies little with D.

As previously noted, we interpolated the inflorescence densities from quadrat counts using the IDW method with a exponent of 2, i.e. the weight assigned to each grid point is proportional to its inverse square distance from the point where the density is to be interpolated. The choice of IDW exponent is expected to affect the predicted pollen load differentiation, since lower exponents result in smoother density fields and less demarcated species clusters. We found that using an exponent of 1 reduces F_{ST} by 10% to 20% from the values shown in Fig. 1, while using an exponent of 4 increases F_{ST} by 5% to 15%. Based on the relationship between F_{ST} and D (e.g. in Fig. 1a), a change in the interpolation exponent would thus affect the inferred value of the random walk parameters for a given empirical value of F_{ST}. In the absence of plant density at a smaller scale, it is difficult to determine the optimal exponent value for this dataset; however, it should be kept in mind that this choice introduces an additional source of uncertainty above that reported by the ABC analysis below.

3.3. Parameter estimation by ABC

For each site, we generated 100,000 sets of parameters. Each parameter set was used to simulate 100 pollen loads for both inflorescence distributions (i.e. different sampling dates) at that site. We picked $\mathbf{a}_{init}$ from a uniform distribution over k-simplexes (vectors that sum to 1), where k is the number of species present at that site. When fitting the model with $\mathbf{a}_{succ}$ different from $\mathbf{a}_{init}$, we picked each component of $\mathbf{a}_{succ}$ uniformly between the corresponding component of $\mathbf{a}_{init}$ and the value making all species equally preferred, e.g. if there are two species and $a_{init} = 0.75$ for species 1, then a_{succ} for that species is picked from the interval (0.5, 0.75). The last component was automatically set by the unit sum condition. Since our sensitivity analysis showed that only one random walk parameter had to be adjusted, we chose to vary l_s and picked it from a uniform

distribution over (0.5, 10). The other random walk parameters were fixed as $n_s = 100$, $\rho = 0$ and $f_j = 0$ (i.e. no "jumps"). For all results reported in this paper, we set the maximum step length (l_{max}) to 20 m.

Based on cross-validation results, we chose a tolerance rate of 1% for the ABC analysis (i.e. keep 1000 parameter sets to estimate the posterior distribution), which in most cases resulted in the lowest cross-validation error. However, we note that changing the tolerance rate between 0.5% and 10% had only a small effect on either the cross-validation error or the posterior estimates. Cross-validation plots are included as Fig. S1 in the electronic supplementary materials. We also found that a correction based on local linear regression was more effective in reducing cross-validation error than one based on ridge regression, and therefore used the former method to produce the estimates below.

We chose the median of the posterior distribution as a point estimate of each parameter and the 95% Bayesian credible interval (or BCI, i.e. the 95% central range of the posterior) to indicate the precision of this estimate (Table 4). The estimates of $l_s < 1$ at BRD and MD4 do not necessarily mean that the RMS step length is less than the local inter-inflorescence distance; a more plausible scenario is that $n_s < 100$, producing an equivalent decrease in the typical foraging distance D. The attractiveness coefficients, which express relative preferences, are difficult to compare between sites: the only two species shared by more than one site are poppy and vetch, and the latter is only marginally present at MD5.

We did not retain the model with $\mathbf{a}_{init} = \mathbf{a}_{succ}$ at site MD5, since it failed to qualitatively reproduce a key characteristic of that site: a high prevalence of poppy pollen combined with a low pollen load differentiation. The alternative model's best fit suggests that bees always prefer to start foraging on poppy over thistle, but are less discriminative for successive steps. Under this specific scenario, our sensitivity analysis shows that varying the random walk parameters has a much smaller effect on pollen load differentiation (Fig. 1b). This may explain why the posterior range of l_s is comparable to its prior distribution at MD5, indicating that our data provides little information on that parameter (Table 4).

Setting each parameter to the median of its posterior distribution, we performed 1000 replicate simulations with the same number of pollen loads by site and date as our data. Most observed values of $\mathbf{p}$, F_{ST} and the number of mixed pollen loads lie within the 95% central range of simulated statistics, although due to the small sample sizes – between 16 and 21 pollen loads by site and date – those ranges are relatively large (Fig. 2). The exceptions occur at site BRD, where the prevalence of the minority species (poppy) on June 5th is above the simulated range, and the pollen load differentiation on June 14th is slightly below the 95% central range (<= 2.1% of simulations). We note that the very low bound of the prediction interval for F_{ST} on June 5th is due to simulations where the minority species is nearly absent from pollen loads and F_{ST} thus approaches zero. F_{ST} is undefined when only one species is present in the whole sample.

We could improve the fit of our base model (with $\mathbf{a}_{init} = \mathbf{a}_{succ}$) for sites BRD and MD4 – unlike MD5 – by allowing the parameters to vary between the two sampling dates. In particular, temporal changes in the attractiveness coefficients may be expected based on each species' peak pollination time. However, this more flexible model produced equal or greater cross-validation

errors for all parameters, which indicates overfitting (results not shown here).

4. Discussion

In this study, we present an individual-based approach that relates the compositional statistics of bee pollen loads to the spatial distribution of floral resources through the simulation of individual foraging paths. A major strength of this approach is that it can accommodate complex descriptions of both the floral landscape and foraging behavior. We can use approximate Bayesian computation to estimate parameters from this type of models without the need to explicitly calculate the likelihood function. One application of this modelling framework is to infer foraging parameters from pollen load data and surveys of the floral resources, as illustrated in this study. Alternatively, if foraging parameters can be estimated by independent observations, such as radio-tracking of bees (Osborne et al. 1999), the model could serve to infer properties of the spatial distribution of floral resources (such as their level of clustering) based on the compositional statistics of pollen loads.

Under a relatively simple parametrization of the model, where foraging paths follow correlated random walks and bee preferences are described by an attractiveness coefficient associated to each species, we could reproduce key compositional statistics of pollen loads sampled from *B. vosnesenskii* foragers across multiple field sites, including one statistic (the number of mixed pollen loads) that was not directly used in fitting the model. However, for the sample sizes of this study (~ 20 pollen loads by site and date), our model also predicts a high level of variability in those summary statistics. Increasing the number of pollen loads sampled – while identifying the same number of grains per pollen load – would be the simplest way to reduce the portion of the variance that is due to the sampling process and thus increase the power of this type of analysis. The ongoing development of robust genetic methods for identifying single pollen grains (Isagi and Suyama 2011, Bektaş and Chapela 2014) offers a higher-throughput alternative to morphological identification methods, and could allow the analysis of larger pollen samples as needed to effectively discriminate between foraging models.

Even with a large sampling variance, our analysis could identify foraging behavior that was qualitatively different at one of the sites in our dataset (MD5). At that site, bees exhibited a very strong preference for California poppy (*E. californica*) pollen over yellow star-thistle (*C. solistitalis*), and the two plants were rarely found in the same quadrats, yet bees had more mixed pollen loads and less unifloral poppy pollen loads than at the other sites. Those results fit the model only by assuming that bees always start on poppy but are less discriminative as they continue foraging. The cause of this particular behavior remains unclear, although we can hypothesize that the strong preference for poppy pollen at MD5 combined with its low abundance could lead to a rapid depletion of this resource, forcing bees to seek less preferred pollen. Testing this hypothesis would require an extension of the model to incorporate the variation in pollen rewards between flowers as well as pollen depletion.

In a different analysis of this pollen load content and plant abundance dataset, Harmon-Threatt and Kremen (2015) found that the overall mix of pollen sources visited by the bees was not random, but tended to produce a balanced nutrient intake (e.g. amino acid content) at each site. This result does not necessarily imply that individual foraging bouts should include a diversity of pollen types, which would be unoptimal when plants are spatially clustered by species. However,

it shows that the bees' relative preferences for one species over another depend not only on factors intrinsic to the two species (e.g. quantity of pollen by flower), but also on the overall composition of the floral landscape as it determines which nutrients would be limiting.

By modelling foraging bouts within the confines of a 90 m x 80 m field, and by excluding pollen loads predominantly composed of species not found within the sampling quadrats, our analysis ignores foraging behavior at a larger scale. Even if bumble bees can travel large distances to exploit floral patches, both optimal foraging theory and empirical observations suggest that they follow compact paths while foraging, visiting neighboring inflorescences and only leaving a patch when rewards fall below a critical threshold (Zimmerman 1982, Lefebvre et al. 2007). The appropriateness of our field scale for the analysis of pollen load composition is also supported by the observation than less than 10% of pollen loads from the sites studied contained above trace levels of pollen from external or unidentified sources.

From a theoretical point of view, one interesting result from our simulations is that for foraging paths represented as correlated random walks, the pollen load composition statistics (for a given field configuration) only depend on the RMS distance between the ends of the path, rather than on the specific random walk parametrization. This result holds even if the plant density varies across the field (if the foraging steps scale with plant spacing) and if the bees prefer certain species (as long as those preferences are fixed). Alternative stochastic models represent bee foraging paths as biased rather than correlated random walks, where the bias may be in the direction of patches with preferred species or higher plant density (e.g. Hanoteaux and al. 2013). In further development of this model, we could investigate to which extent the results reported here generalize to biased random walks.

For the simplest version of our model (with $\mathbf{a}_{init} = \mathbf{a}_{succ}$), the attractiveness coefficients determine the average prevalence of each species in pollen loads, whereas the typical scale of foraging paths (indicated by D) determines the level of differentiation (species sorting) between pollen loads. This model thus formalizes a “null hypothesis” according to which pollen load differentiation is due solely to spatial factors, i.e. monospecific clusters in the field and the limited distance covered in a foraging bout. Departures from the model's predictions would provide evidence of more complex foraging behavior: an excess of unifloral pollen loads would suggest flower constancy, i.e. the tendency of individual foragers to stick to one species per bout (Chittka et al. 1999) , whereas an excess of mixed pollen loads could suggest that bees are actively seeking multiple resources or – as in our alternate model for site MD5 – that they are relaxing their preferences in the course of a foraging bout.

Acknowledgements
We thank the associate editor and two anonymous reviewers for their helpful feedback leading to improvements of the paper. P.M. acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and the Fonds de recherche du Québec – Nature et technologies.

References

Aronne G, Giovanetti M, Guarracino MR, de Micco V (2012) Foraging rules of flower selection applied by colonies of Apis mellifera: ranking and associations of floral sources. Funct Ecol 26:1186–1196. doi: 10.1111/j.1365-2435.2012.02017.x

Bartumeus F, da Luz MGE, Viswanathan GM, Catalan J (2005) Animal search strategies: a quantitative random-walk analysis. Ecology 86:3078–3087. doi: 10.1890/04-1806

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Beil M, Horn H, Schwabe A (2008) Analysis of pollen loads in a wild bee community (Hymenoptera: Apidae) — a method for elucidating habitat use and foraging distances. Apidologie 39:456–467. doi: 10.1051/apido:2008021

Bektaş A, Chapela I (2014) Loop-mediated isothermal amplification of single pollen grains. J Integr Plant Biol (in press). doi: 10.1111/jipb.12191

Boff S, Luz CFP, Araujo AC, Pott A (2011) Pollen analysis reveals plants foraged by africanized honeybees in the Southern Pantanal, Brazil. Neotrop Entomol 40:47–54. doi: 10.1590/S1519-566X2011000100007

Chittka L, Thomson JD, Waser NM (1999) Flower constancy, insect psychology, and plant evolution. Naturwissenschaften 86:361–377. doi: 10.1007/s001140050636

Codling EA, Plank MJ, Benhamou S (2008) Random walk models in biology. J R Soc Interface 5:813–834. doi: 10.1098/rsif.2008.0014

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). Methods Ecol Evol 3:475–479. doi: 10.1111/j.2041-210X.2011.00179.x

De Valpine P, Harmon-Threatt AN (2013) General models for resource use or other compositional count data using the Dirichlet-multinomial distribution. Ecology 94:2678–2687. doi: 10.1890/12-0416.1

Diaz-Losada E, Ricciardelli-D'Albore G, Saa-Otero MP (1998) The possible use of honeybee pollen loads in characterising vegetation. Grana 37:155–163. doi: 10.1080/00173139809362660

Dille JA, Milner M, Groeteke JJ, et al. (2003) How good is your weed map? A comparison of spatial interpolators. Weed Sci 51:44–55. doi: 10.1614/0043-1745(2002)051[0044:HGIYWM]2.0.CO;2

Gelman A, Meng X-L, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. Stat sinica 6:733–760.

Grimm V, Berger U, et al. (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198:115–126. doi: 10.1016/j.ecolmodel.2006.04.023

Grimm V, Berger U, DeAngelis DL, et al. (2010) The ODD protocol: A review and first update. Ecol Model 221:2760–2768. doi: 10.1016/j.ecolmodel.2010.08.019

Hanoteaux S, Tielbörger K, Seifan M (2013) Effects of spatial patterns on the pollination success of a less attractive species. Oikos 122:867–880. doi: 10.1111/j.1600-0706.2012.20801.x

Harmon-Threatt AN, Kremen C (2015) Bumble bees selectively use native and exotic species to maintain nutritional intake across highly variable and invaded local floral resource pools. Ecol Entomol (in press).

Isagi Y, Suyama Y (2010) Single-pollen genotyping. Springer, Tokyo

Jabot F, Chave J (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. The American Naturalist 178:E37–E47. doi: 10.1086/660829

Lefebvre D, Pierre J, Outreman Y, Pierre J-S (2007) Patch departure rules in Bumblebees:

evidence of a decremental motivational mechanism. Behav Ecol Sociobiol 61:1707–1715. doi: 10.1007/s00265-007-0402-6
Levin DA, Kerster HW (1969) The Dependence of bee-mediated pollen and gene dispersal upon plant density. Evolution 23:560. doi: 10.2307/2406853
Marchand P (2013) Statistical methods for the detection and space-time monitoring of DNA markers in the pollen cloud. PhD dissertation, Department of Environmental Science, Policy and Management, University of California, Berkeley, USA.
Osborne JL, Clark SJ, Morris RJ, et al. (1999) A landscape-scale study of bumble bee foraging range and constancy, using harmonic radar. J Appl Ecol 36:519–533. doi: 10.1046/j.1365-2664.1999.00428.x
R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Ramsay G, Thompson C, Squire G (2003) Quantifying landscape-scale gene flow in oilseed rape. Final report of DEFRA Project RG0216, Department for Environment, Food and Rural Affairs, UK
Scranton K, Knape J, de Valpine P (2014) An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. Ecology 95:1418–1428. doi: 10.1890/13-1065.1
Seifan M, Hoch E-M, Hanoteaux S, Tielbörger K (2014) The outcome of shared pollination services is affected by the density and spatial pattern of an attractive neighbour. J Ecol 102:953–962. doi: 10.1111/1365-2745.12256
Stanley RG, Linskens HF (1974) Pollen: biology, biochemistry, management. Springer, Berlin
Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. Genetics 145:505–518.
Zimmerman M (1982) Optimal foraging: Random movement by pollen collecting bumblebees. Oecologia 53:394–398.

Table 1 Main variables and parameters of the bee foraging model

Model inputs	
$\mathbf{a}_{init}$	Vector representing the relative attractiveness of each plant species when selecting the initial foraging position
$\mathbf{a}_{succ}$	Vector representing the relative attractiveness of each plants species for successive foraging steps
$\mathbf{d}(x,y)$	Vector of densities (inflorescences m^{-2}) for each of the plant species at position (x,y) in the field
n_s	Number of steps (inflorescences visited) in a foraging bout
l_s	Root mean square (RMS) length for a step (dimensionless [a])
f_j	Frequency of jumps (occasional large steps)
l_j	RMS length for a jump (dimensionless [a])
ρ	Mean cosine of the wrapped Cauchy distribution; represents the directional correlation between successive steps
Output and summary statistics	
$\boldsymbol{P}$	Pollen load composition matrix: p_{ij} is the proportion of species j in pollen load i
$\mathbf{p}$	Vector representing the average proportion of each species (over all pollen loads)
F_{ST}	Differentiation in composition between pollen loads (see definition in text)

[a] When determining the RMS step or jump length, l_s and l_j are multiplied by the inverse square root of the local inflorescence density (see text).

Table 2 Summary statistics of the bee pollen data

Site	Date	Number of pollen loads	Species 1 *p*	Species 2 *p*	Species 3 *p*	F_{ST}	Number of mixed loads [a]
BRD	05-Jun-2009	16	Vetch 0.73	Poppy 0.27		0.82	3
BRD	19-Jun-2009	17	Vetch 0.73	Poppy 0.27		0.61	5
MD4	14-May-2009	17	Poppy 0.46	Lupine 0.45	Clover 0.09	0.73	6
MD4	03-Jun-2009	16	Poppy 0.77	Lupine 0.23		0.91	1
MD5	17-Jun-2009	19	Poppy 0.79	Thistle 0.20	Vetch 0.01	0.24	12
MD5	01-Jul-2009	21	Poppy 0.81	Thistle 0.19		0.26	13

[a] Pollen loads containing no more than 90% of any single species.

Table 3 Inflorescence density of pollen sources by site and date

Site	Date	Mean density (inflorescences m^{-2}) [a]				
		Clover	Lupine	Poppy	Thistle	Vetch
BRD	05-Jun-2009			0.14		2.12
BRD	19-Jun-2009			1.48		4.76
MD4	14-May-2009	1.42	0.70	2.68		
MD4	03-Jun-2009		0.26	0.60		
MD5	17-Jun-2009			1.04	1.72	0.10
MD5	01-Jul-2009			0.76	15.30	

[a] Average count from fifty 1 m^2 quadrats at each site/date pair. An empty cell means no inflorescences of the plant were observed in any quadrat.

Table 4 Parameter estimates by approximate Bayesian computation

Site	$\mathbf{a}_{\mathbf{init}}$ Species: Median (95% BCI[a])	$\mathbf{a}_{\mathbf{succ}}$ Species: Median (95% BCI) [b]	l_s Median (95% BCI)
BRD	Vetch: 0.44 (0.39–0.49)	same as $\mathbf{a}_{\mathbf{init}}$	0.5 (0.2–1.3)
MD4	Poppy: 0.35 (0.30–0.39) Lupine: 0.56 (0.48–0.64)	same as $\mathbf{a}_{\mathbf{init}}$	0.8 (0.6–1.0)
MD5	Poppy: 0.86 (0.61–1.00) Thistle: 0.00 (0.00–0.00)	Poppy: 0.61 (0.51–0.78) Thistle: 0.24 (0.02–0.39)	4.7 (2.1–12.8)[c]

[a] BCI: Bayesian credible interval.

[b] Only estimated for site MD5 where the model with $\mathbf{a}_{\mathbf{init}} = \mathbf{a}_{\mathbf{succ}}$ failed to qualitatively reproduce summary statistics.

[c] Cases where the 95% BCI has a range comparable to the prior (0.5–10 for l_s).

Figure captions

Fig. 1 Pollen load statistics (**p**, F_{ST}) predicted by model simulations as a function of the RMS distance between the first and last foraging step (D), for two field configurations: (a) MD4 on 14-May-2009 and (b) MD5 on 01-Jul-2009. For each field, the attractiveness coefficients were fixed as (a) $\mathbf{a}_{init} = \mathbf{a}_{succ}$ = (0.20 [poppy], 0.73 [lupine], 0.07 [clover]) and (b) $\mathbf{a}_{init}$ = (1 [poppy], 0 [thistle]), $\mathbf{a}_{succ}$ = (0.5, 0.5), while the random walk parameters varied over the following ranges: n_s (100–800), l_s (1–4), ρ (0–0.92), l_j (1–12.3) with f_j = 0.1. Grey lines represent linear regression fits. Ten thousand pollen loads were simulated for each set of parameters, resulting in a standard deviation of under 0.005 for each statistic shown here.

Fig. 2 Values of (a) **p**, F_{ST} and (b) the number of mixed pollen loads in 1000 simulations of the data, using the median parameters estimated by ABC. Filled circles and error bars represent the median and 95% central range of the simulated statistics, while open circles show the observed statistics. Note that p_3 only applies to site-date pairs (MD4-14-May and MD5-17-Jun) where three pollen species were found above trace levels.

Figure 1

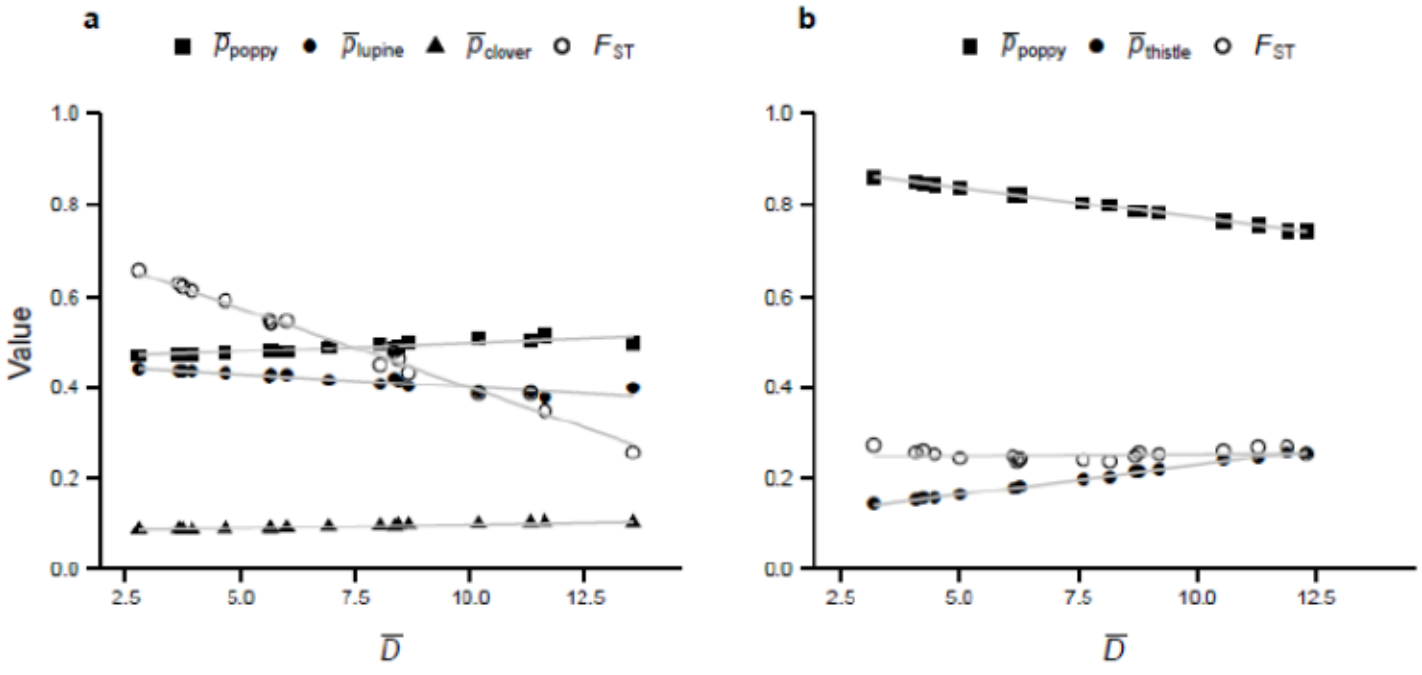
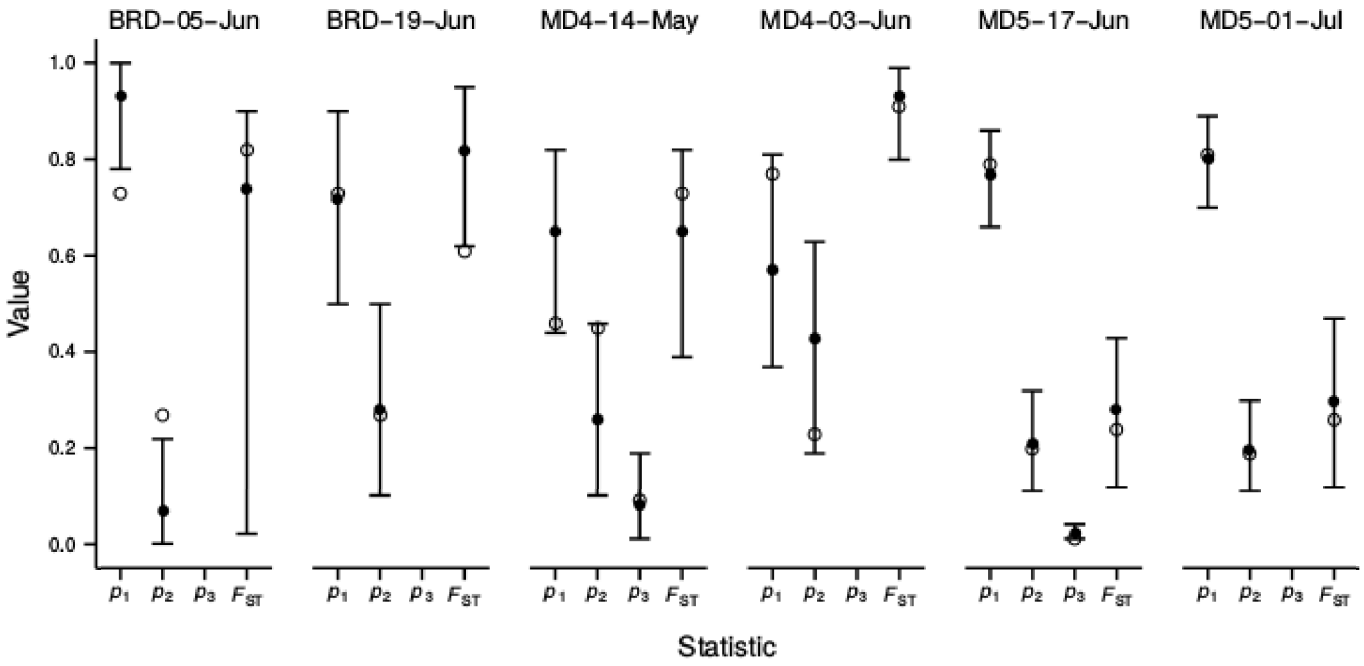
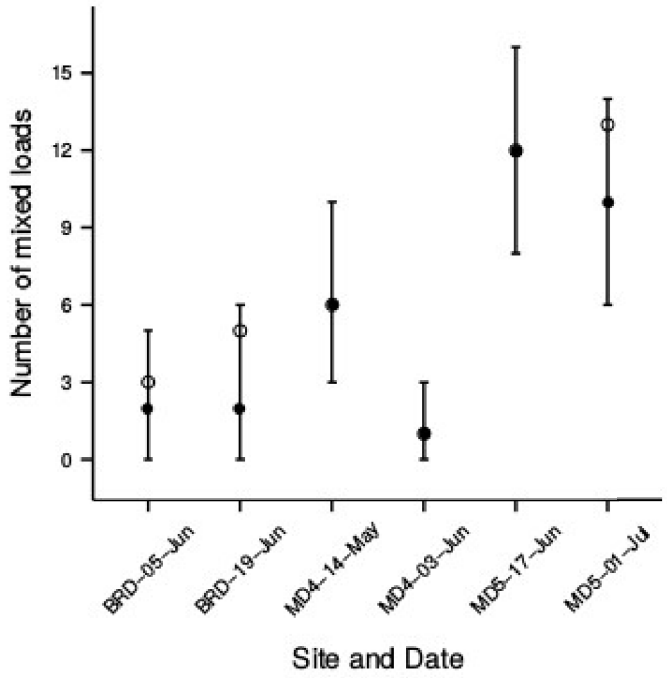


Figure 2

a


BRD−05−Jun
BRD−19−Jun
MD4−14−May
MD4−03−Jun
MD5−17−Jun
MD5−01−Jul
Value
1.0
0.8
0.6
0.4
0.2
0.0
p_1 p_2 p_3 F_{ST}
Statistic

b


Number of mixed loads
15
12
9
6
3
0
BRD−05−Jun
BRD−19−Jun
MD4−14−May
MD4−03−Jun
MD5−17−Jun
MD5−01−Jul
Site and Date

Supplementary electronic material for :

Testing models of bee foraging behavior through the analysis of pollen loads and floral density data

Fortran 95 code used to simulate bee foraging and produce summary statistics for approximate Bayesian computation (ABC) analysis.

```
PROGRAM bee_sim2f
    ! THIS PROGRAM SIMULATES BEES IN TWO FIELDS WITH SAME FORAGING PARAMETERS
    ! The program performs NREP simulations of n_bee bees foraging in a field
    ! and outputs species composition of pollen loads for use in ABC analysis

    IMPLICIT NONE
    INTEGER, PARAMETER :: FILENUM = 1
    REAL, PARAMETER :: PI = 3.141593, TWOPI = 6.283185, EPS = 1.0E-6 ! EPS: tolerance from 0
    INTEGER, PARAMETER :: NREP = 100000 ! number of simulations to run

    ! Grid dimensions (nx * ny), field size and grid steps (in meters)
    INTEGER, PARAMETER :: NX = 10, NY = 5
    REAL, PARAMETER :: XF = 90.0, YF = 80.0, XGS = 10.0, YGS = 20.0

    ! Input data
    INTEGER :: n_bee, n_spc, n_stp ! # bees by sim., # of plant species, # of steps by bee
    INTEGER :: i_sim, i_rep, i_spc, i_fld ! counter variables for above
    REAL :: l_stp, l_max, l_jmp, f_jmp, rho ! random walk parameters
    REAL, ALLOCATABLE :: a_init(:), a_succ(:) ! attractiveness of species to bees (initial
                                              ! step, successive steps)
    REAL :: evenp ! fraction to hold "even preference" (a_succ = 1/n_spc for all species)
    REAL, ALLOCATABLE :: quad_counts(:,:,:)   ! quad_counts(X,Y,SPC) # inflorescences of
                                              ! SPC for quadrat at (X,Y) (current field)

    REAL, ALLOCATABLE :: quad_counts2(:,:,:,:) ! quad_counts for both fields

    ! Output: p by species by bee, average p and variance by species, fst
    REAL, ALLOCATABLE :: p_sim(:,:), pm(:), var(:)
    REAL :: fst

    ! Initialize RNG
    INTEGER :: i_seed, ssize, clock
    INTEGER, ALLOCATABLE :: seed(:)
    call random_seed(size=ssize)
    allocate(seed(ssize))
    call system_clock(count=clock)
    seed = clock + 37 * (/ (i_seed - 1, i_seed = 1, ssize) /)
    call random_seed(put = seed)


    call load_data

    n_bee = 100
    l_max = 20.0
    allocate(a_init(n_spc), a_succ(n_spc))
    allocate(p_sim(n_spc,n_bee), pm(n_spc), var(n_spc))
    open(unit=FILENUM, file="beesim_out.txt", action='write')

    do i_rep = 1, NREP
       ! Generate a random simplex for a_init
        do
            do i_spc = 1, n_spc-1
                call random_number(a_init(i_spc))
            end do
```

```
        if (sum(a_init(1:n_spc-1)) < 1) exit
    end do
    a_init(n_spc) = 1 - sum(a_init(1:n_spc-1))

    ! Generate a_succ w/ condition that it's "more uniform" than a_init
    evenp = 1.0 / real(n_spc)
    do
        do i_spc = 1, n_spc-1
            call random_number(a_succ(i_spc))
            a_succ(i_spc) = evenp + a_succ(i_spc)*(a_init(i_spc) - evenp)
        end do
        if (sum(a_succ(1:n_spc-1)) < 1) exit
    end do
    a_succ(n_spc) = 1 - sum(a_succ(1:n_spc-1))

    ! Pick l_step between (0.5,10) (dimensionless, as explained in article)
    call random_number(l_stp)
    l_stp = l_stp*9.5 + 0.5

    ! Set other random walk parameters
    n_stp = 100
    rho = 0.0
    f_jmp = 0.0
    l_jmp = 0.0

    ! Perform simulations (get_psim subroutine) and calculate statistics
    ! Repeat for both fields
    do i_fld = 1,2
        quad_counts = quad_counts2(:,:,:,i_fld)
        do i_sim = 1, n_bee
            p_sim(:,i_sim) = get_psim()
        end do
        pm = sum(p_sim,2)/n_bee
        var = sum(p_sim**2,2)/n_bee - pm**2
        fst = sum(var) / (1 - sum(pm**2))
        write(FILENUM,*) a_init(1:n_spc-1), a_succ(1:n_spc-1), l_stp, pm, fst
    end do
end do

close(FILENUM)


CONTAINS

    SUBROUTINE load_data
        ! Input plant distr. data (inflorescences per 1m^2 quadrats for each species)
        open(unit=FILENUM, file="quad_counts.txt", action='read')
        read(FILENUM,*) n_spc
        allocate(quad_counts(NX,NY,n_spc))
        allocate(quad_counts2(NX,NY,n_spc,2))
        read(FILENUM,*) quad_counts2
        close(FILENUM)
    END SUBROUTINE load_data

    FUNCTION dens(s,p)
        REAL :: dens
        INTEGER, INTENT(IN) :: s
        REAL, INTENT(IN) :: p(2)
        INTEGER :: ix, iy
        REAL :: dx, dy, w00, w01, w10, w11

        ! Interpolates the density of species s at point p(x,y)
        ! using inverse square distance weights (w00..w11)
        ! (ix,iy) is "bottom-left" closest data point; (dx,dy) distance to that point
        ix = int(p(1)/XGS) + 1
        iy = int(p(2)/YGS) + 1
```

```
    dx = p(1) - (ix-1)*XGS
    dy = p(2) - (iy-1)*YGS
    if (dx < EPS .and. dy < EPS) then
        dens = quad_counts(ix,iy,s)
    else
        w00 = 1.0/(dx**2 + dy**2)
        w01 = 1.0/(dx**2 + (YGS-dy)**2)
        w10 = 1.0/((XGS-dx)**2 + dy**2)
        w11 = 1.0/((XGS-dx)**2 + (YGS-dy)**2)
        dens = (w00*quad_counts(ix,iy,s) + w01*quad_counts(ix,iy+1,s) &
                + w10*quad_counts(ix+1,iy,s) + w11*quad_counts(ix+1,iy+1,s)) &
               / (w00 + w01 + w10 + w11)
    end if
END FUNCTION dens

FUNCTION spc_id(p)
    INTEGER :: spc_id
    REAL, INTENT(IN) :: p(2)
    INTEGER :: s
    REAL :: wdens(n_spc)
    REAL :: run_sum, tot_sum, rnd

    ! Choose species at point p (prop. to density*attractiveness)
    do s = 1, n_spc
        wdens(s) = dens(s,p) * a_succ(s)
    end do
    tot_sum = sum(wdens)
    run_sum = 0.0
    call random_number(rnd)
    rnd = rnd * tot_sum
    do s = 1, n_spc
        run_sum = run_sum + wdens(s)
        if (rnd <= run_sum) then
            spc_id = s
            exit
        end if
    end do
END FUNCTION spc_id

FUNCTION p_init()
    REAL :: p_init(2)
    INTEGER :: s, ix, iy
    REAL :: tot_mat(NX,NY)
    REAL :: tot_sum, run_sum, rnd

    ! Pick initial foraging point (among points with data)
    ! with prob. prop. to preference-weighted total density (tot_mat)
    tot_mat = 0.0
    do s = 1, n_spc
        tot_mat = tot_mat + a_init(s)*quad_counts(:,:,s)
    end do
    tot_sum = sum(tot_mat)
    run_sum = 0.0
    call random_number(rnd)
    rnd = rnd * tot_sum
    outer: do ix = 1, NX
        do iy = 1, NY
            run_sum = run_sum + tot_mat(ix,iy)
            if (rnd <= run_sum) then
                p_init(1) = (ix-1)*XGS
                p_init(2) = (iy-1)*YGS
                exit outer
            end if
        end do
    end do outer
END FUNCTION p_init
```

```
    FUNCTION get_psim()
        REAL :: get_psim(n_spc)
        INTEGER :: i, s
        REAL :: u, v, rnd
        REAL :: tot_dens, dnn, dl, theta, dth, theta_new
        INTEGER :: s_count(n_spc)
        REAL :: b_path(2,n_stp)

        ! Simulate a single bee foraging trip using a
        ! correlated random walk (coords. in b_path)
        ! Returns the proportion of each species
        b_path = 0.0
        b_path(:,1) = p_init()
        call random_number(theta)
        theta = (theta-0.5)*TWOPI
        tot_dens = 0.0
        do s = 1, n_spc
            tot_dens = tot_dens + dens(s,b_path(:,1))
        end do
        dnn = 1.0 / sqrt(tot_dens)  ! Expected local nearest-neigbour dist.

        do i = 2, n_stp
            do ! until suitable step is found
                call random_number(u)
                call random_number(v)
                call random_number(rnd)
                if (rnd <= f_jmp) then
                    dl = min(dnn*l_jmp*sqrt(-log(u)), l_max)
                else
                    dl = min(dnn*l_stp*sqrt(-log(u)), l_max)
                end if
                dth = 2.0*atan((1.0-rho)/(1.0+rho) * tan(PI*(v-0.5)))
                theta_new = theta + dth
                if (abs(theta_new) > PI) then
                    ! Bring theta_new within (-PI,PI) if needed
                    theta_new = theta_new - sign(TWOPI, theta_new)
                end if
                b_path(1,i) = b_path(1,i-1) + dl*cos(theta_new)
                b_path(2,i) = b_path(2,i-1) + dl*sin(theta_new)
                if (b_path(1,i) > 0 .and. b_path(1,i) < XF &
                     .and. b_path(2,i) > 0 .and. b_path(2,i) < YF) then
                    tot_dens = 0.0
                    do s = 1, n_spc
                        tot_dens = tot_dens + dens(s,b_path(:,i))
                    end do
                    if (tot_dens > EPS) then
                        dnn = 1.0 / sqrt(tot_dens)
                        theta = theta_new
                        exit
                    end if
                end if
            end do
        end do

        s_count = 0
        do i = 1, n_stp
            s = spc_id(b_path(:,i))
            s_count(s) = s_count(s) + 1
        end do
        get_psim = real(s_count) / real(n_stp)
    END FUNCTION

END PROGRAM bee_sim2f
```

Table S1 Random walk parameters and simulation outputs for the sensitivity analysis

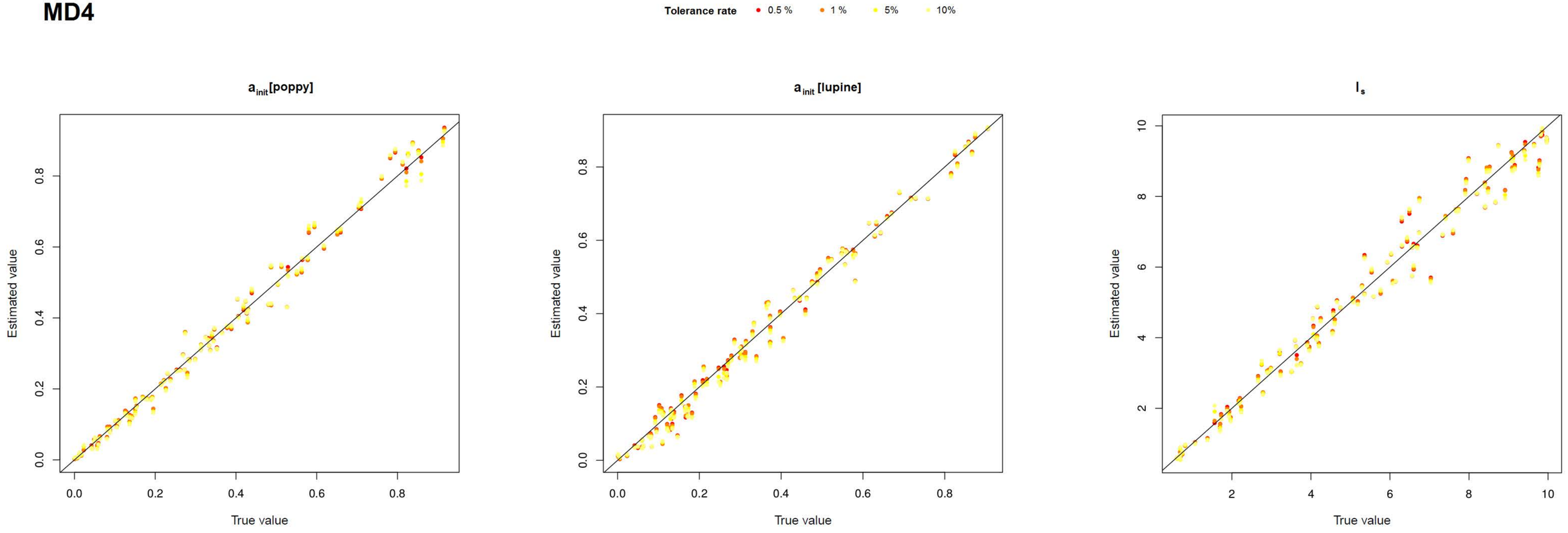
Parameters [a]				**Output for MD4 on 14-May-2009** [b]					**Output for MD5 on 01-Jul-2009** [c]			
n_s	l_s	ρ	l_j	D	p_{poppy}	p_{lupine}	p_{clover}	F_{ST}	D	p_{poppy}	$p_{thistle}$	F_{ST}
100	1	0	n/a	2.78	0.471	0.441	0.088	0.657	3.21	0.858	0.142	0.274
200	1	0	n/a	3.95	0.473	0.438	0.088	0.611	4.49	0.842	0.158	0.253
400	1	0	n/a	5.65	0.482	0.428	0.090	0.539	6.28	0.820	0.180	0.240
600	1	0	n/a	6.92	0.489	0.417	0.094	0.489	7.60	0.801	0.199	0.242
800	1	0	n/a	8.03	0.496	0.408	0.096	0.449	8.71	0.785	0.215	0.248
100	2	0	n/a	5.68	0.483	0.427	0.090	0.543	6.29	0.819	0.181	0.246
100	3	0	n/a	8.66	0.500	0.402	0.098	0.434	9.19	0.781	0.219	0.253
100	3.5	0	n/a	10.19	0.510	0.390	0.100	0.388	10.56	0.761	0.239	0.263
100	4	0	n/a	11.63	0.516	0.378	0.105	0.349	11.9	0.743	0.257	0.271
100	1	0.3	n/a	3.65	0.474	0.437	0.089	0.628	4.08	0.848	0.152	0.258
100	1	0.5	n/a	4.66	0.477	0.434	0.089	0.593	5.01	0.836	0.164	0.246
100	1	0.66	n/a	6.01	0.479	0.429	0.092	0.547	6.22	0.822	0.178	0.236
100	1	0.8	n/a	8.37	0.487	0.418	0.094	0.48	8.16	0.797	0.203	0.234
100	1	0.92	n/a	13.56	0.498	0.399	0.103	0.258	12.3	0.743	0.257	0.254
100	1	0	3	3.73	0.475	0.437	0.088	0.622	4.24	0.844	0.156	0.261
100	1	0	5.6	5.62	0.483	0.426	0.091	0.551	6.13	0.821	0.179	0.248
100	1	0	9	8.44	0.491	0.411	0.098	0.463	8.79	0.786	0.214	0.255
100	1	0	12.3	11.31	0.505	0.392	0.103	0.385	11.3	0.754	0.246	0.269

[a] $f_j = 0.1$ for all cases when "jumps" are used in the path (i.e. l_j is not listed as n/a).
[b] $\mathbf{a}_{init} = \mathbf{a}_{succ}$ = (0.20 [poppy], 0.73 [lupine], 0.07 [clover])
[c] $\mathbf{a}_{init}$ = (1 [poppy], 0 [thistle]), $\mathbf{a}_{succ}$ = (0.5, 0.5)

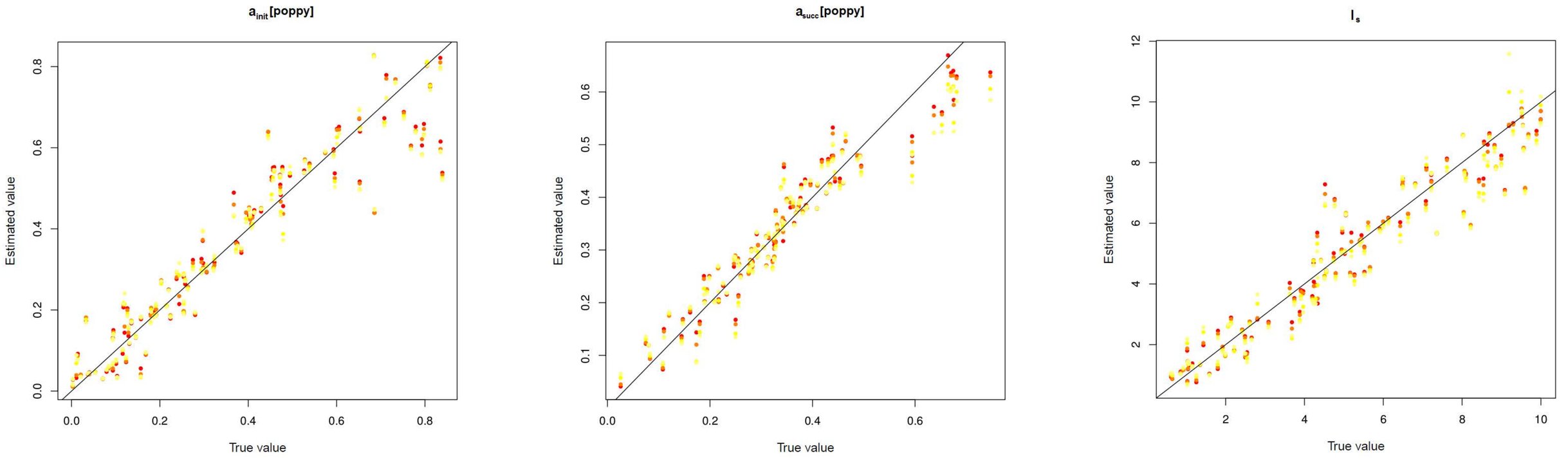
Supplementary figure captions

Fig. S1 Cross-validation plots produced by the 'abc' package in R, for sites MD4 and MD5. For each of the 100 repeats of the cross-validation procedure, the output of one of the simulations is chosen as the observed statistics, and the parameter values for this simulation ("True value") are inferred by performing the ABC analysis based on all the other simulations ("Estimated value"). The process was repeated using four different values of the tolerance rate. A line representing exact inference (Estimated value = True value) is shown on each plot.

MD4


Tolerance rate
0.5 %
1 %
5%
10%
a init [poppy]
a init [lupine]
l s
Estimated value
True value

MD5


a init [poppy]
a succ [poppy]
l s
Estimated value
True value