Modelling mobile agent-based ecosystem services using kernel-weighted predictors

Paul W. Goedhart1 | Marjolein E. Lof2 | Felix J. J. A. Bianchi3 | Hans (J.) M. Baveco4 | Wopke van der Werf2

1Biometris, Wageningen University and Research, Wageningen, The Netherlands
2Centre for Crop Systems Analysis, Wageningen University, Wageningen, The Netherlands
3Farming Systems Ecology, Wageningen University, Wageningen, The Netherlands
4Wageningen Environmental Research (Alterra), Wageningen University and Research, Wageningen, The Netherlands

Correspondence
Wopke van der Werf
Email: wopke.vanderwerf@wur.nl

RESEARCH ARTICLE

Abstract
1. Agriculture benefits from ecosystem services provided by mobile agents, such as biological pest control by natural enemies and pollination by bees. However, methods that can generate spatially explicit predictions and maps of these ecosystem services based on empirical data are still scarce.

2. Here we propose a generic statistical model to derive kernel functions to characterize the spatial distribution of ecosystem services provided by mobile agents. The model is similar in spirit to a generalized linear model, and uses data of landscape composition and ecosystem services assessed at target sites to estimate parameters of the kernel. The approach is tested in a simulation study and illustrated by an empirical case study on parasitism rates of the diamondback moth Plutella xylostella.

3. The simulation study shows that the scale parameter of the exponential power kernel can be estimated with limited bias, whereas estimation of the shape parameter is difficult. For the case study the model provides biologically relevant estimates for the kernel associated with parasitism of P. xylostella. These estimates can be used to generate ecosystem service maps for existing or planned landscapes. The case study reveals that predictions can be sensitive to the parameter values for the width and shape of the kernel, and to the link function used in the statistical model.

4. In the last two decades numerous empirical studies assessed ecosystem services at target sites and related these to the surrounding landscape. Our method can take advantage of these data by estimating underlying kernels that can be used to map the spatial distribution of ecosystem services. However, empirical data that can discriminate between alternative kernel shapes remain critical.

KEYWORDS
biocontrol, conservation planning, natural pest control, pollination, source–sink dynamics, spatial distribution, spatial extent, statistical model
1 | INTRODUCTION

Ecosystems provide a range of goods and services that are essential for human well-being (Daily, 1997). Some ecosystem services, such as carbon sequestration and soil conservation, are closely associated with habitats and show little spatial dynamics or interlinkages at the landscape level. In contrast, ecosystem services that are based on mobile organisms, such as pollinators, natural enemies of pests and seed dispersers, are highly dynamic in space and time, and the strength of the ecosystem service depends on factors operating at the landscape level (Bianchi, Booij, & Tscharntke, 2006; Kremen et al., 2007). With growing concerns about the deterioration of ecosystem services (Millennium Ecosystem Assessment, 2005), there is a need for methods that can quantify and map ecosystem services at relevant spatial scales to support land-use management and planning for ecosystem service conservation.

In the last two decades, many studies have assessed the intensity of mobile agent-based ecosystem services, typically by placing sentinels in a range of landscapes, and regressing the measured ecosystem service intensity against landscape variables observed in concentric rings or circles of varying width around the points of measurement (Thies & Tscharntke, 1999). This regression approach has been widely used to identify source habitats for service-providing organisms, and to assess the spatial extent of the associated agent-based ecosystem services (e.g., Bianchi, Goedhart, & Baveco, 2008; Haenke, Scheid, Schaefer, Tscharntke, & Thies, 2009; Holzschuh, Dormann, Tscharntke, & Steffan-Dewenter, 2013; Schmidt & Tscharntke, 2005; Thies, Steffan-Dewenter, & Tscharntke, 2003). We will refer to this method as spatial indexed regression (SIR).

However, SIR has important shortcomings (Miguet, Fahrig, & Lavigne, 2017). First, the rings are discrete but the agent movement process is continuous, which represents a fundamental conceptual mismatch. Second, the chosen radii of concentric rings are arbitrary and small perturbations of the radii can affect significance of regression coefficients and therefore the conclusion of a study. Third, the method is inevitably very parameter rich: for every habitat variable of interest a parameter needs to be estimated for each ring or circle around the target site. Fourth, regressions for subsequent rings are assumed independent which they are not. Fifth, an analysis per ring has reduced power as compared to an overall analysis of landscape effects on ecosystem service provisioning. Moreover, as a result of limitations 1, 2 and 4, the SIR approach is not well suited for making spatial maps of services.

Here we propose a generalization of the SIR approach which overcomes these limitations while using the same empirical sentinel data. The method assumes that service-providing mobile agents disperse from source habitats to a target location. Furthermore, it is assumed that there is a dose–response relationship between agent abundance at the target and the ecosystem service provided. The dispersal process of mobile agents is modelled using a two-dimensional distribution, or kernel for short, such as the bivariate normal distribution. Such kernels have few parameters and have been widely used to characterize spatial movement processes in ecology (Bullock et al., 2017; Clark, Silman, Kern, Macklin, & HilleRisLambers, 1999; Snäll, O’Hara, & Arjas, 2007).

The approach outlined here is based on the following premises: (1) ecosystem service providing agents are colonizing the surrounding landscape from source habitats such as forests; (2) service-providing agents are radiating out from the source habitats following a kernel which is a decreasing function of the distance between the source and the target (cf. Miguet et al., 2017); (3) the abundance of agents is assumed to be proportional to the area of source habitat; and (4) there is a continuous dose–response relationship between the abundance of agents at a target site and the ecosystem service they provide at the target. Note that premises (1), (3) and (4) are also implicitly made in the SIR approach.

It follows from (1) and (2) that the abundance of the ecosystem service providing agents at a given target site is a weighted sum of contributions from different sources in the surrounding landscape where the weights are defined by the kernel and the distance between source and target. In this way the spatial structure of service-providing agents is captured into a single predictor, i.e. the kernel-weighted sum of contributions. In most cases the abundance of mobile agents at sources remains unknown, whereas the land use is known, and therefore premise (3) is required. In the sequel, the weighted sum of contributions will either be called abundance of agents or, whenever more appropriate, amount of source habitat. Conceptually, there is a proportionality factor between this distance-weighted amount of source habitat at the target and the abundance of the agents at the target. This proportionality factor represents the density of agents provided per unit area of source habitat. In practice, this proportionality factor is not identifiable if the abundance of agents at the target is not measured, and in the sequel, we set it therefore to one. In contrast to the SIR approach where separate regressions are conducted for each ring or circle, premise (4) states that there is only a single dose–response regression model, which links the abundance of agents at the target site to the ecosystem service at the target site. Fitting this model to ecosystem service data yields estimates of the kernel (which describes the way agents are radiating out from a source), and the parameters of the dose–response model (which translates the abundance of agents at a target site to a measured ecosystem service). The kernel model and the dose–response model are fitted simultaneously.

The purpose of this methodology is twofold. First, to identify source habitats in the landscape that are associated with ecosystem services provided by mobile agents, and to assess the spatial scale at which these agents influence ecosystem services. This objective is congruent with the SIR methodology. Second, to map the estimated spatial distribution of the resulting ecosystem services across the landscape by projecting the kernel and dose–response model around source habitats. While the mapping procedure reflects the methods proposed by Lonsdorf et al. (2009) for pollination services and Jonsson et al. (2014) for biological control, our methodology requires a relatively low number of parameters that can directly be estimated from empirical data. The purpose of the paper was to outline the modelling methodology, analyse its performance using a simulation study, and provide a proof of concept by applying it to a case study.
2 | MATERIALS AND METHODS

We first introduce kernel functions for the spatial distribution of ecosystem service providing agents around patches of a single type of source habitat. We then introduce the dose–response model that links the abundance of agents at the target to the delivery of ecosystem services at the target. Then we explain how the kernel and dose–response model are simultaneously fitted to data, and finally we outline how the resulting model can be used to map ecosystem services. Two applications are provided for illustration: a simulation study that explores whether the model parameters are identifiable from simulated data, and an empirical study based on the ecosystem service of parasitization of caterpillars by parasitoid wasps.

2.1 | Abundance of ecosystem service providing agents arriving in a target site

We make the simplifying assumption that the mobile agents associated with the ecosystem service move randomly without a preference for direction and independently of the structure of the landscape. This implies that the spatial distribution of mobile agents around a point source can be described with a kernel \( f(r) \) that is only a function of the distance \( r \) to the source. The kernel \( f(r) \) thus represents in a single dimension the "proportion" of agents active at a distance \( r \) from the source (details in Appendix S1).

Mobile agents at the target site may originate from multiple source habitat patches at different distances from the target site. To obtain a mathematical expression for the total abundance of agents arriving at a target consider a target site which is placed at the origin \( (r = 0) \). Define \( a(r, \theta) \) as the abundance of the ecosystem service providing agents at a location \( (r, \theta) \), i.e. at a distance \( r \) and angle \( \theta \) from the target location. Note that \( a(r, \theta) \) is in fact a density, i.e. the number of agents per unit area, which we conveniently refer to as abundance. The abundance of the agents from this point source which arrives at the target site is then given by the product \( a(r, \theta) f(r) \). The total abundance of agents \( T \) arriving at a target site is then the integral over all contributions from source habitats:

\[
T = \int_0^\infty \int_0^{2\pi} a(r, \theta) f(r) \, r \, dr \, d\theta
\]  

(1)

The extra term \( r \) in the integral is due to the transformation from Cartesian coordinates to polar coordinates (Nathan, Klein, Robledo-Arnuncio, & Revilla, 2012) (Appendix S1).

In practical applications, it is usually not feasible to fully specify the agent abundance function \( a(r, \theta) \). However, the area of source habitat (e.g. forest) in a series of non-overlapping rings can usually be obtained from geographical information systems and this area is used as a proxy for the number of agents provided. We approximate the total agent abundance \( T \) in Equation 1 in the following way. Denote \( r_0 = 0, r_1, r_2, \ldots, r_n \) as a set of ordered radii which define \( n \) rings around a target, and define \( A_j \) as the total amount of source habitat in ring \( j \). Further define the mean function value \( F_j \) of the kernel \( f(r) \) in ring \( j \) as

\[
F_j = \frac{1}{t_j-t_{j-1}} \int_{t_{j-1}}^{t_j} f(r) \, dr
\]  

(2)

The total abundance of agents \( T \) at the target site (Equation 1) can then be approximated by the summation over all rings of the product of \( A_j \) and \( F_j \) (details in Appendix S1):

\[
T = \sum_j A_j F_j
\]  

(3)

Obviously, the approximation becomes worse when the rings become wider and their number decreases. Moreover, when the rings do not cover the complete range where \( f(r) \) and \( a(r, \theta) \) are both positive, the total amount \( T \) will be underestimated. Here, we elaborate the model using one type of source habitat, but extensions with multiple types of source habitats are possible.

2.2 | The exponential power kernel

The shape of the kernel \( f(r) \) is important because kernels with fat tails provide higher agent abundances at locations farther from source habitats than thin-tailed kernels. It is therefore convenient to have a family of kernels with an extra shape parameter. Such a family is given by the exponential power kernel, see, e.g. Clark et al. (1999) and Snäll et al. (2007):

\[
f(r) = \frac{c}{2\pi a^2 \Gamma(2/c)} \exp\left(-\frac{r^c}{a}\right)
\]  

(4)

in which \( \Gamma(\cdot) \) is the gamma function, \( a \) is a scale parameter and \( c \) is a shape parameter. This flexible kernel includes some familiar ones as special cases. For \( c = 2 \) a rotationally symmetric bivariate normal distribution with variance \( a^2/2 \) in both directions is obtained. For \( c = 1 \) the kernel is exponentially shaped and has fatter tails than for \( c = 2 \). For \( c = 0.5 \) a square root exponential kernel is defined which has fat tails (Kot, Lewis, & van den Driessche, 1996). Online Appendix S2 provides the mean value \( F_j \) in Equation 2 for the exponential power kernel. An alternative parameterization of this kernel employs, instead of \( a \), the mean displacement \( \mu \) of the associated marginal one dimensional probability density (online Appendix S2). This parameterization is used in the simulation study. Online Appendix S3 shows, for some special cases of the exponential kernel and using artificial landscapes, that the approximation of the total \( T \) in Equation 3 is quite satisfactory when 20 or more rings are used.

2.3 | Linking the abundance of mobile agents to the ecosystem service

A dose–response model is required to describe the link between the abundance \( T \) of the providers and the measured ecosystem service at the target site. This dose–response model will usually take the form of a generalized linear regression model (McCullagh & Nelder, 1989; Zuur, Ieno, Walker, Saveliev, & Smith, 2009). Suppose the ecosystem service measurement follows some statistical probability distribution with mean \( \lambda \), and there is interest in the relation between \( \lambda \) and the abundance \( T \). In general the dose–response model has the form
where \( g(\cdot) \) is a link function. For models of proportion (e.g., biological control effect or efficacy of pollination compared to “full pollination”) a logit link would be appropriate, while a log link or identity link may apply for other services (e.g., pollinator visits, pollen load or fruit weight per plant).

For a set of ecosystem service measurements, the parameters \( \alpha, c, \beta_0, \) and \( \beta_1 \) in Equation 5 can be estimated by means of maximum likelihood. This is an iterative fitting process because estimates for the beta’s depend on the kernel parameters and vice versa. For fixed values of the kernel parameters \( \alpha \) and \( c \) the predictor \( T(\alpha, c) \) can be calculated for every target, and the regression parameters \( \beta_0 \) and \( \beta_1 \) can be estimated with standard software for generalized linear models. In case the values of \( \alpha \) and/or \( c \) are unknown, which will usually be the case, a general maximization routine, or a grid search for \( \alpha \) and \( c \), can be used to optimize the likelihood.

Goedhart, Lof, Bianchi, Baveco, and van der Werf (2018) provides \( r \) code to fit the logistic regression model for the case study, for a fixed value of \( c \).

### 2.4 | Prediction of the spatial distribution of the ecosystem service for new landscapes

The spatial distribution of the ecosystem service, more precisely the mean \( \lambda \) in Equation 5, can be predicted and mapped when the spatial locations of source habitat patches, the parameter values \( \alpha \) and \( c \) for the exponential power kernel \( f(r) \) and parameters \( \beta_0 \) and \( \beta_1 \) of the dose–response model are known. Consider a landscape map with a superimposed regular square grid. Suppose that the amount of source habitat in grid cell \( k \) equals \( A_k \) and denote the centre of the cell by \( (x_k, y_k) \). The abundance of agents \( T(x, y) \) in a target point \( (x, y) \), resulting from all grid cells, can then be approximated by the following sum, which is a slightly modified form of Equation 3:

\[
T(x, y) \approx \sum_k A_k f\left( \sqrt{(x-x_k)^2 + (y-y_k)^2} \right)
\]  

This is then translated to \( \lambda \), employing Equation 5, to give the spatial distribution of the ecosystem service at every point \( (x, y) \) which can be used for mapping. Equation 6 is an approximation because the amount \( A_k \) is assumed to be located in the centre of the grid cell, while in reality it may be scattered across the cell, and because the kernel \( f(r) \) is assumed to be constant in the grid cell. The approximation will thus become better when the size of the grid cells becomes smaller. The cells \( (x, y) \), for which \( T(x, y) \) is required, should be embedded in a larger two-dimensional space such that the extended space fully covers the kernel. Direct calculation of Equation 6 can be time consuming for large grids. Two-dimensional convolution based on the Fast Fourier Transform (FFT) of \( A_k \) and \( f(r) \) offers a computationally efficient alternative (Bracewell, 2000). Goedhart et al. (2018) provides \( r \) code to map the parasitism rate for the case study, using the logistic and exponential link function.

### 2.5 | Testing model performance in a simulation study

The performance of the logistic regression model for parasitism rates using the exponential power kernel was evaluated by means of a simulation study in which we determined bias and precision of the estimated parameters. The simulation study employs 30 landscapes and the amount of source habitat at each landscape is defined by a Beta distribution. The exponential power kernel was used with shape parameter \( c = 0.5, 1 \) and \( 2 \), and mean displacement parameter \( \mu = 1 \). True parasitism rates ranged between 13% and 95%. Full details of the setup of the simulation study are given in Appendix S4.

### 2.6 | Analysis of an empirical case study

Bianchi et al. (2008) investigated how landscape composition affects parasitism rates of larvae of the diamond black moth, *Plutella xylostella*, in 22 organic Brussels sprout fields across the Netherlands. When the caterpillars are parasitized, one or more parasitoids will emerge from the pupa, and the host is killed. Hence, parasitization is an example of the ecosystem service of biological control. Observed parasitism rates ranged between 4% and 94% and were related to landscape variables measured in circles with radius 0.15, 0.5, 1 and 5 km. Univariate analysis, using a generalized linear mixed model with the binomial distribution and a logistic link, indicated that parasitism rates were positively related to area of forest, forest edge and road verges at different radii. The dataset is re-analysed here, assuming that the number of parasitized *P. xylostella* per field follows a quasi-binomial distribution with parasitism probability \( x \), and the total number of recovered larvae per field was taken as binomial totals. The quasi-binomial distribution was used because the data are heavily overdispersed with respect to the binomial distribution, and therefore we used maximum quasi-likelihood rather than maximum likelihood (McCullagh & Nelder, 1989). In addition to the logit link, an exponential link, given by \( x = 1 - \exp(-\beta, T(\alpha, c)) \), which is also known as the one-hit model (Turner, 1975), was used and the results were compared. Note that the exponential model lacks the intercept parameter \( \beta_0 \) of Equation 5. Landscape variables were re-assessed in 20 rings, rather than the four rings as in Bianchi et al. (2008), with radii 0.1, 0.2,..., 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0, 3.5, 4.0 and 5.0 km using land use map LGNS (Hazeu, 2005) and topographical map VIRIS, both with a grid cell-size of 25 m. Area of forest and of forest edge in the 20 rings, as well as parasitism rates in the 22 fields, are provided in Goedhart et al. (2018). The logistic regression model with kernel-weighted predictor \( T(\alpha, c) \) was fitted using a golden section search for \( \alpha \), for a fine grid of 65 fixed values of \( c \) between 0.2 and 20. For the exponential link a general optimization routine in GenStat (VSN International, 2015) was used for \( c = 0.5, 1 \) and 2 only. The estimated parameters for predictor area of forest edge, for fixed values \( c = 0.5 \) and \( c = 2 \) combined with the logistic and the exponential link, were used to map the predicted rate of parasitism at the national scale of the Netherlands employing a grid with cell size of 25 m. Here we present the predicted ecosystem service for an area of 3.5 × 3.5 km with area of forest edge given in Figure 1.
FIGURE 1  Spatial distribution of forest edge, expressed as a percentage of the area of forest edge in 25 × 25 m grid cells, for a 10 × 10 km area from the Dutch national map. The predicted parasitism rates resulting from the kernel model for the case study are shown in Figure 2 for the central 7 × 7 km area in this figure. A wider area is displayed because the kernel model also takes sources outside the central area into account. Arable land is not indicated on this map.

3 | RESULTS

3.1 | Simulation study

Simultaneous estimation of the shape parameter $c$ and the mean displacement $\mu$, or equivalently the scale parameter $\alpha$, of the exponential power kernel was difficult (Appendix S4). This was particularly evidenced by wide confidence intervals for $c$. Therefore, in a second simulation study the shape parameter $c$ was fixed. Simulated data for $c = 0.5, 1$ and 2 were analysed for models with fixed values $c = 0.5, 1$ and 2, respectively, to investigate the effect of misspecification of $c$ on the estimates of the kernel displacement parameter $\mu$, and the regression parameters $\beta_0$ and $\beta_1$. When the simulated data were analysed with the true value of $c$, the mean estimates were close to their true values with limited bias (Appendix S4). Misspecification of $c$, however, gave relatively large bias for the true value $c = 0.5$, and small bias for the true value $c = 2$.

3.2 | Empirical case study

Re-analysis of the dataset of parasitism rates of larvae of $P. xylostella$ in 22 organic Brussels sprout fields with univariate logistic regression for each of the twenty individual rings revealed that area of forest was a significant predictor for parasitism rates in 12 out of 20 rings, while area of forest edge was significant in four rings only, all within 0.8 km of the target site (results not shown). These findings indicate that parasitism rate was associated with both forest-based landscape variables, and that the effect of forest ranges further than that of forest edge.

Fitting the kernel model with a logistic link revealed that, for area of forest, the mean deviance of the fitted model did not depend on the shape parameter $c$, with deviance values ranging between 7.79 and 7.99 for values of $c$ between 0.2 and 20 (data not shown). For area of forest edge the mean deviance ranged from 7.76 to 9.46, and the minimum value was attained for $c = 20$. For both forest predictors the quasi likelihood ratio interval for $c$ covered the whole interval from 0.2 to 20 indicating that, similar to the simulation study, the shape parameter $c$ of the kernel is not identifiable for the data at hand. Therefore, we fitted the kernel model for fixed values $c = 0.5, 1$ and 2 (Table 1). The estimates for $\beta_0$ and $\beta_1$ were relatively independent of the value of $c$, and so was the mean displacement parameter $\mu$ for area of forest. However, when using area of forest edge as predictor, the estimate of $\mu$ for $c = 0.5$ was twice as large as for $c = 2$. For area of forest the estimate of the mean displacement $\mu$ was such that the kernel extended well beyond the 5 km limit. This was evidenced by cumulative probabilities of the marginal kernel $H(5)$ (see equation S2.4 in Appendix S2) at $r = 5$ km that were below 1: 0.90 for $c = 2$ and 0.77 for both $c = 1$ and 0.5. This suggests that parasitism rates at the target site can be influenced by forest areas at distances beyond 5 km. In contrast, for area of forest edge, the cumulative probability at a distance of 5 km, $H(5)$, equalled 1 for all three values of $c$, indicating that the chosen domain of 5 km was large enough to estimate the effect of forest edge. For area of forest the correlation between the estimates of $\alpha$ and $\beta_1$ was around 0.95 and the correlations between $\alpha$ and $\beta_0$, and between $\beta_0$ and $\beta_1$ were in absolute value smaller than 0.5. For area of forest edge the largest correlation was between $\alpha$ and $\beta_0$ with value −0.84, and the other correlations were smaller than 0.67. This indicates that, similar to the simulation study, the scale parameter $\alpha$ of the kernel cannot be estimated independently from the regression parameters $\beta_0$ and $\beta_1$. 
When employing the exponential link function instead of the logit link function, somewhat larger mean deviances were obtained (results not shown) indicating a poorer fit. This is not surprising since this model has one parameter less, i.e. \( \beta_0 \), than the model employing the logit link and is thus less flexible.

The projections of the estimated effect of forest edge on parasitism rates for the logistic and the exponential link for \( c = 0.5 \) and \( c = 2 \) indicate that the choice of the kernel shape parameter \( c \) and link function can have major consequences for the predicted spatial distribution of the ecosystem service (Figure 2). The model with the logistic link and \( c = 2 \) had the lowest mean deviance providing the most plausible map.

### Table 1

Mean deviance and parameter estimates for the logistic kernel model applied to the case study on parasitism rates of *Plutella xylostella* (Bianchi et al., 2008), for fixed values of the kernel shape parameter \( c = 0.5, c = 1 \) and \( c = 2 \). Parameters \( \alpha \) and \( \mu \) relate to the mean of the kernel and \( \beta_0 \) and \( \beta_1 \) are regression parameters. The model was fitted for two choices of predictors: area of forest and area of forest edge

| \( c \) | Mean deviance | \( \alpha \) | \( \mu \) | \( \beta_0 \) | \( \beta_1 \) |
|-------|---------------|-------------|-------------|-------------|-------------|
| Area of forest |
| 0.5  | 7.92          | 0.182       | 3.649       | -1.844      | 0.080       |
| 1.0  | 7.98          | 1.781       | 3.561       | -1.839      | 0.080       |
| 2.0  | 7.99          | 3.279       | 2.906       | -1.834      | 0.068       |
| Area of forest edge |
| 0.5  | 8.73          | 0.046       | 0.914       | -1.930      | 0.238       |
| 1.0  | 8.43          | 0.276       | 0.552       | -1.940      | 0.233       |
| 2.0  | 8.14          | 0.520       | 0.461       | -1.967      | 0.235 |

### Figure 2

Predicted spatial distribution, for the central area depicted in Figure 1, of parasitism rates of larvae of *Plutella xylostella* expressed as percentages for the case study, with area of forest edge as predictor, and the logistic and exponential link with fixed shape parameters \( c = 0.5 \) and \( c = 2 \).
scale, suggesting that functional spatial scales may extend beyond the maximum scale in the analysis (O’Rourke et al., 2011; Schmidt, Thies, Nentwig, & Tscharntke, 2008). Re-analysis of the case study dataset of P. xylostella parasitism using the kernel approach indicated that the spatial extent of biocontrol services associated with area of forest may well exceed 5 km, while in the original analysis, using four landscape sectors up 2 km from the target, the highest significance was found for forest area at a scale of 500 m radius (Bianchi et al., 2008). This 10-fold difference in spatial scale highlights the sensitivity of the analysis method used, and sets an agenda for further work on the analyses of the functional spatial scale of ecosystem service providers.

Our approach uses measurements of ecosystem services in landscape replicates, differing in the amount and spatial distribution of source habitats, to estimate the spatial distribution of mobile agent-based ecosystem services around source habitats. In the conceptual framework, an estimate of the total abundance of agents $T$ at the target is calculated from the kernel. However, this is only an intermediate result that is not further analysed because, in landscape scale studies which focus on the quantification of ecosystem services, the abundance of agents at target sites is rarely recorded (e.g. Bianchi et al., 2008). However, when measurements of the mobile agents are available, these can be directly modelled as a function of the distance to sources. Examples of this approach are pollen grain and fungal spore dispersal (Shaw, Harwood, Wilkinson, & Elliott, 2006), pollinating bee visitation rates (Olsson, Bolin, Smith, & Lonsdorf, 2015) and abundance of wild bees (Kennedy et al., 2013).

Recently, Miguet et al. (2017) used a similar approach to the one proposed here, combining kernel modelling with generalized linear regression, to estimate the abundance of mobile agents in landscapes. They use the nonlinear link function primarily to add additional flexibility to the kernel modelling, whereas in our approach the link function reflects a mechanistic description of the relationship between the abundance of agents and the service provided. While the conceptual underpinnings differ, the mathematical elaboration is similar. Further work is needed to compare these approaches in more depth, and to determine the best framework for the analysis of sentinel data.

Our method assumes that the kernel is spatial invariant and that intervening habitats do not interfere with the movement of the agents. These assumption are also made in the conventional SIR approach and in the framework of Miguet et al. (2017). The assumption of spatial invariance, i.e. an identical kernel for every target location, is widely used in geostatistics and is the starting point of any spatial analysis. Local effects, such as crop type or management practices at the target site (Kennedy et al., 2013), or heterogeneity of the landscape around the target (Kennedy et al., 2013; Olsson et al., 2015) can be explored by adding these as extra covariates or factors to the regression model. Moreover additional, possibly intervening, habitats, e.g. tree rows blocking movement or attractive habitats arresting the agents (Wratten et al., 2003), can also be added to the regression model. For instance, a logistic model with two source habitats $x_1$ and $x_2$, then becomes $\logit(x) = \beta_0 + \beta_1 T(x_1; \alpha_1, c_1) + \beta_2 T(x_2; \alpha_2, c_2)$ in which every source habitat type can have its own kernel parameters $\alpha$ and $c$. This may result in negative regression coefficients representing disservices. In our experience, however, a typical dataset does not contain enough landscape replicates (usually around 20) to estimate so many parameters. Hence, the idea of adding further complexity will require a major boost in data collection effort, beyond what is currently standard. Moreover, adding additional habitats as explanatory variables would still not account for the interaction between habitats as a result of their spatial positioning with respect to the target. While the assumption of a spatially invariant kernel simplifies reality, the assessment of kernels currently seems the most promising way to integrate the “action at a distance” of resource habitats as pioneered in the SIR approach. Our model can be fitted in turn for single habitat types as predictors. Model selection using AIC or BIC can then be used to select the habitat type with the greatest explanatory power (Bolker, 2008).

The methodology presented here has similarities to regression with a functional predictor (Ramsay & Silverman, 2005) which, in our case, would amount to using $\sum \delta_i X$ as a predictor instead of Equation 3. In functional regression the $\delta_i$ parameters are estimated using a penalty on the smoothness of the $\delta$’s resulting in a series of smooth estimates. Since we are assuming that mobile agents are radiating out of sources, giving less agents further away, it is more elegant to replace the $\delta$’s by the kernel based decreasing values $F_i$ of Equation 2, which are based on the mechanistic dispersal kernel, and to estimate the few parameters of the smooth kernel instead. Heaton and Gelfand (2011) used kernel-averaged predictors, assuming a spatial process for both the predictor and the response. In our case a spatial process for the predictor is not necessary because the amount of source habitat in every ring is known. Moreover, a spatial process for the response is not relevant for our data because the parasitism rate was observed at locations which were far apart, as is typically the case in the SIR approach.

Application of the kernel method resulted in some practical guidelines. First, the precise radii of the landscape sector rings are not critical, although the approximation in Equation 3 becomes worse when the number of rings decreases, and care needs to be taken that the largest ring exceeds the spatial scale of the ecosystem service of interest. The use of twenty rings or more with sufficient radii near the target site seemed to suffice. Rather than using rings, the model can also be fitted using a regular square grid thereby replacing Equation 3 by Equation 6. Second, the simultaneous fitting of the kernel parameters $\alpha$ and $c$ was cumbersome in both the simulation (30 target sites) and the case study (22 target sites). The simulation with truly binomially distributed response showed that confidence intervals for the shape parameter $c$ were very wide, and this will worsen when there is overdispersion relative to the binomial distribution. Fitting exponential power kernels with shape parameter $c = 0.5, 1$ and $2$ to mark-recapture data of the parasitoid Diadegma semiclausum also revealed that models with different c values received similar support from the data (Bianchi, Schellhorn, & van der Werf, 2009), confirming that $c$ is hard to estimate. Therefore, it will often be necessary to choose a fixed value for $c$ in advance. Clark et al. (1999) also assumed a fixed value for $c$ and then estimated $\alpha$. For our purpose, the exponential power kernels with $c = 0.5, 1$ and $2$ provided enough variation in shape. A list of alternative kernels is provided by Bullock et al. (2017) who found...
that the exponential power kernel was one of two best performing kernels when fitted to plant dispersal data. Third, in this paper we used the example of parasitism rates with a binomial or quasi-binomial distribution. However, our method is flexible, and any linear or generalized linear model can be used in combination with kernel-averaged landscape variables. Fourth, the kernel approach allows addition of extra predictors and/or a second kernel-weighted habitat predictor. The latter was tested for the case study by fitting a logistic regression model with the two habitats area of forest and area of forest edge, again for fixed values of $c = 0.5$, 1 and 2. However, the mean deviance of this double kernel model was only slightly smaller than those for the single kernel models, indicating no relevant improvement of fit for the case study.

The ecosystem service kernels can be projected on any map that contains data on source habitat. On the one hand this is convenient as the empirically derived kernels can be employed at areas outside the original study area, assisting in management and planning of areas of interest where no data are available. On the other hand, this also holds the risk that kernels are applied to regions for which they are not representative. For instance, key ecosystem service providing agents may show different responses in regions that differ in climate, management or vegetation. The comparison of ecosystem service kernels derived across regions could provide useful information on the variation in the spatial extent of ecosystem services and highlight to what extent kernels are region-specific. As a general rule, the use of kernels in the same ecophysical region as where the experimental data originate seems reasonable.

ACKNOWLEDGEMENTS

Work reported in this paper was financially supported by the Dutch Ministry of Economic Affairs (BO Functional Agrobiodiversity) and by the European Union’s Seventh Framework Programme for research and technological development, through grant no. 311879 (QuESSA: Quantification of Ecosystem Services for Sustainable Agriculture). We thank Henk Meeuwsen and Harold Kuipers† (Alterra) for conducting GIS analyses.

AUTHORS’ CONTRIBUTIONS

W.W. conceived the idea. P.W.G. developed the methodology, performed the simulation study and wrote the R code. F.J.J.A.B and J.M.B collected the empirical case study data. P.W.G analysed the empirical data with the help of co-authors. All authors jointly wrote the manuscript, contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Data and R code for the analysis of the empirical case study on parasitism rates of larvae of the diamond black moth are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.3vS90 (Goedhart et al., 2018).

ORCID

Paul W. Goedhart http://orcid.org/0000-0002-9586-9862

REFERENCES

Bianchi, F. J. J. A., Booij, C. J. H., & Tscharntke, T. (2006). Sustainable pest regulation in agricultural landscapes: A review on landscape composition, biodiversity and natural pest control. Proceedings of the Royal Society, Series B Biological Sciences, 273, 1715–1727. https://doi.org/10.1098/rspb.2006.3530

Bianchi, F. J. J. A., Goedhart, P. W., & Baveco, J. M. (2008). Enhanced pest control in cabbage crops near forest in The Netherlands. Landscape Ecology, 23, 595–602. https://doi.org/10.1007/s10109-008-9219-6

Bianchi, F. J. J. A., Schellhorn, N. A., & van der Werf, W. (2009). Predicting the time to colonization of the parasitoid Diadegma semiclausum: The importance of the shape of spatial dispersal kernels for biological control. Biological Control, 50, 267–274. https://doi.org/10.1016/j.biocontrol.2009.04.014

Bolker, B. M. (2008). Ecological models and data in R. Princeton, NY: Princeton University Press.

Bracewell, R. N. (2000). The Frontier transform and its applications (3rd ed.). New York, NY: McGraw-Hill.

Bullock, J. M., González, L. M., Tamme, R., Götzenberger, L., White, S. M., Pärtel, M., & Hooftman, A. A. P. (2017). A synthesis of empirical plant dispersal kernels. Journal of Ecology, 105, 6–19. https://doi.org/10.1111/1365-2745.12666

Clark, J. S., Silman, M., Kern, R., Macklin, E., & HilleRisLambers, J. (1999). Seed dispersal near and far: Patterns across temperature and tropical forests. Ecology, 80, 1475–1494. https://doi.org/10.1890/0012-9658(1999)080[1475:SDNAPF]2.0.CO;2

Daily, G. C. (1997). Nature’s services: Societal dependence on natural ecosystems. Washington, DC: Island Press.

Goedhart, P. W., Lof, M. E., Bianchi, F. J. J. A., Baveco, J. M., & van der Werf, W. (2018). Data and R code from: Modelling mobile agent-based ecosystem services using kernel-weighted predictors. Dryad Digital Repository. https://doi.org/10.5061/dryad.3vS90

Haenke, S., Scheid, B., Schaefer, M., Tscharntke, T., & Thies, C. (2009). Increasing syrphid fly diversity and density in sown flower strips within simple vs. complex landscapes. Journal of Applied Ecology, 46, 1106–1114. https://doi.org/10.1111/j.1365-2664.2009.01685.x

Hazeu, G. W. (2005). Landelijk Grondgebruikstatistisch Nederland (LGN5): Vervoudiging, nauwkeurigheid en gebruik. Alterra-rapport 1213, Alterra, Wageningen, The Netherlands. http://library wur.nl/WebQuery/wurpubs/fulltext/17654

Heaton, M. J., & Gelfand, A. E. (2011). Spatial regression using kernel averaged predictors. Journal of Agricultural, Biological, and Environmental Statistics, 16, 233–252. https://doi.org/10.1007/s13253-010-0050-6

Holzschuh, A., Dormann, C. F., Tscharntke, T., & Steffan-Dewenter, I. (2013). Mass-flowering crops enhance wild bee abundance. Oecologia, 172, 477–484. https://doi.org/10.1007/s00442-012-2515-5

Jonsson, M., Bommarco, R., Ekobm, B., Smith, G. H., Bengtsson, J., Caballero-Lopez, B., ... Freckleton, R. (2014). Ecological production functions for biological control services in agricultural landscapes. Methods in Ecology and Evolution, 5, 243–252. https://doi.org/10.1111/2041-210X.12149

Kennedy, C. M., Lonsdorf, E., Neel, M. C., Williams, N. M., Ricketts, T. H., Winfree, R., ... Kremen, C. (2013). A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. Ecology Letters, 16, 584–599. https://doi.org/10.1111/ele.12082

Kot, M., Lewis, M. A., & van den Driessche, P. (1996). Dispersal data and the spread of invading organisms. Ecology, 77, 2027–2042. https://doi.org/10.2307/2265698

Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., ... Ricketts, T. H. (2007). Pollination and other ecosystem
services produced by mobile organisms: A conceptual framework for the effects of land-use change. Ecology Letters, 10, 299–314. https://doi.org/10.1111/j.1461-0248.2007.01018.x

Lonsdorf, E., Kremen, C., Ricketts, T., Winfree, R., Williams, N., & Greenleaf, S. (2009). Modelling pollination services across agricultural landscapes. Annals of Botany, 103, 1589–1600. https://doi.org/10.1093/aob/mcp069

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models (2nd ed.). London, UK: Chapman and Hall.

Miguet, P., Fahrig, L., & Lavigne, C. (2017). How to quantify a distance-dependent landscape effect on a biological response. Methods in Ecology and Evolution, 8, 1717–1724. https://doi.org/10.1111/2041-210X.12830

Millennium Ecosystem Assessment (2005). Ecosystems and human well-being: Synthesis. Washington, DC: Island Press.

Nathan, R., Klein, E., Robledo-Arnuncio, J. J., & Revilla, E. (2012). Dispersal kernels: Review. In J. Clobert, M. Baguette, T. G. Benton, & J. M. Bullock (Eds.), Dispersal ecology and evolution (pp. 187–210). Oxford, UK: Oxford University Press.

Olsson, O., Bolin, A., Smith, H. G., & Lonsdorf, E. V. (2015). Modeling pollinating bee visitation rates in heterogeneous landscapes from foraging theory. Ecological Modelling, 316, 133–143. https://doi.org/10.1016/j.ecolmodel.2015.08.009

O’Rourke, M. E., Rienzo-Stack, K., & Power, A. G. (2011). A multi-scale, landscape approach to predicting insect populations in agro-ecosystems. Ecological Applications, 21, 1782–1791. https://doi.org/10.1890/10-0241.1

Ramsay, J. O., & Silverman, B. W. (2005). Functional data analysis (2nd ed.). New York, NY: Springer.

Schmidt, M. H., Thies, C., Nentwig, W., & Tscharntke, T. (2005). Landscape context of sheetweb spider (Araneae: Linyphiidae) abundance in cereal fields. Journal of Biogeography, 32, 467–473. https://doi.org/10.1111/j.1365-2699.2004.01244.x

Shaw, M. W., Harwood, T. D., Wilkinson, M. J., & Elliott, L. (2006). Assembling spatially explicit landscape models of pollen and spore dispersal by wind for risk assessment. Proceedings of the Royal Society B, 273, 1705–1713. https://doi.org/10.1098/rspb.2006.3491

Snäll, T., O’Hara, R. B., & Arjas, E. (2007). A mathematical and statistical framework for modelling dispersal. Oikos, 116, 1037–1050. https://doi.org/10.1111/j.0030-1299.2007.15604.x

Thies, C., Steffan-Dewenter, I., & Tscharntke, T. (2003). Effects of landscape context on herbivory and parasitism at different spatial scales. Oikos, 101, 18–25. https://doi.org/10.1034/j.1600-0706.2003.12567.x

Thies, C., & Tscharntke, T. (1999). Landscape structure and biological control in agroecosystems. Science, 285, 893–895. https://doi.org/10.1126/science.285.5429.893

Turner, M. E. (1975). Some classes of hit-theory models. Mathematical Biosciences, 23, 219–235. https://doi.org/10.1016/0025-5564(75)90037-1

VSN International (2015). Genstat for windows (18th ed.). Hemel Hempstead, UK: VSN International. http://www.vsni.co.uk/

Wratten, S. D., Bowie, M. H., Hickman, J. M., Evans, A. M., Sedcole, J. R., & Tylianakis, J. M. (2003). Field boundaries as barriers to movement of hover flies (Diptera: Syrphidae) in cultivated land. Oecologia, 134, 605–611. https://doi.org/10.1007/s00442-002-1128-9

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. New York, NY: Springer. https://doi.org/10.1007/978-0-387-87458-6

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.