

---

## Challenges of ecological monitoring: estimating population abundance from sparse trap counts

Natalia Petrovskaya, Sergei Petrovskii and Archie K. Murchie

*J. R. Soc. Interface* published online 10 August 2011  
doi: 10.1098/rsif.2011.0386

---

### References

**This article cites 37 articles, 4 of which can be accessed free**

<http://rsif.royalsocietypublishing.org/content/early/2011/08/04/rsif.2011.0386.full.html#ref-list-1>

### P<P

Published online 10 August 2011 in advance of the print journal.

### Subject collections

Articles on similar topics can be found in the following collections

[biomathematics](#) (270 articles)

[environmental science](#) (690 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *J. R. Soc. Interface* go to: <http://rsif.royalsocietypublishing.org/subscriptions>

---

# Challenges of ecological monitoring: estimating population abundance from sparse trap counts

Natalia Petrovskaya<sup>1,\*</sup>, Sergei Petrovskii<sup>2</sup> and Archie K. Murchie<sup>3</sup>

<sup>1</sup>*School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK*

<sup>2</sup>*Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK*

<sup>3</sup>*Applied Plant Science and Biometrics Division, Agri-Food and Biosciences Institute, Belfast BT9 5PX, UK*

Ecological monitoring aims to provide estimates of pest species abundance—this information being then used for making decisions about means of control. For invertebrate species, population size estimates are often based on trap counts which provide the value of the population density at the traps' location. However, the use of traps in large numbers is problematic as it is costly and may also be disruptive to agricultural procedures. Therefore, the challenge is to obtain a reliable population size estimate from sparse spatial data. The approach we develop in this paper is based on the ideas of numerical integration on a coarse grid. We investigate several methods of numerical integration in order to understand how badly the lack of spatial data can affect the accuracy of results. We first test our approach on simulation data mimicking spatial population distributions of different complexity. We show that, rather counterintuitively, a robust estimate of the population size can be obtained from just a few traps, even when the population distribution has a highly complicated spatial structure. We obtain an estimate of the minimum number of traps required to calculate the population size with good accuracy. We then apply our approach to field data to confirm that the number of trap/sampling locations can be much fewer than has been used in many monitoring programmes. We also show that the accuracy of our approach is greater than that of the statistical method commonly used in field studies. Finally, we discuss the implications of our findings for ecological monitoring practice and show that the use of trap numbers 'smaller than minimum' may still be possible but it would result in a paradigm shift: the population size estimates should be treated probabilistically and the arising uncertainty may introduce additional risk in decision-making.

**Keywords:** ecological monitoring; pest control; sparse data; coarse grid; numerical integration

## 1. INTRODUCTION

Ecosystems are under pressure owing to anthropogenic impact, which has increased considerably over the last few decades [1]. This pressure can significantly affect the structure of ecological communities, often enhancing population outbreaks of harmful species (e.g. as a result of a collapsed top-down control, cf. [2]). In particular, biological invasions have been identified as one of the main reasons of ecosystem degradation and biodiversity loss around the world [3]. The situation is exacerbated by the ongoing environmental/climatic changes that may affect areals as well as abundances. A comprehensive ecological monitoring is therefore necessary in order to provide sufficiently detailed and timely information about species that can potentially cause problems, such as, for instance, alien species that may become dangerous pests. Well-known examples of the latter are given by

invasion of the gypsy moth in the USA [4], invasion of Japanese knotweed in the UK [5,6] and invasion of *Mnemiopsis leidyi* in the Black and Caspian Seas [7,8].

In terrestrial ecosystems, especially in agri-ecosystems, ecological monitoring is usually a part of the integrated pest management programme [9,10], which recommends application of pesticides once the pest abundance exceeds a certain threshold [11]. Use of chemical pesticides has its obvious drawbacks in additional costs, including wheeling damage to the crop and operator costs (sometimes substantial) added to the agricultural product, and in potential damage caused to the environment [12]. Although it does not seem possible to fully avoid the use of pesticides at the current state of agricultural science and technology, it is certainly possible to further optimize their use. One way to achieve this goal is to provide robust information about the abundance of a given target species. In a somewhat different ecological context, accurate information about population abundance is required to trace the spread of harmful invading

\*Author for correspondence (n.b.petrovskaya@bham.ac.uk).



Figure 1. A typical study area with the traps (seen as grey spots in the grass) installed at the nodes of a rectangular  $12 \times 12$  grid with 2 m spacing. This experimental layout has been used to collect data on flatworm abundance (see §5). (Online version in colour.)

species: control measures are much more effective when applied at the right time and in the right place.

There are several methods to collect information about species abundance depending on their biological traits. In this paper, we mostly focus on the problems associated with invertebrates. For invertebrates such as insects or planarians, a common method to collect field data regarding the species' abundance is trapping. A number of traps are installed across the monitored area (figure 1), e.g. in a field or grassland, they are emptied on a regular basis (e.g. daily or weekly), their content is analysed, different species identified and counted. The trap counts are then used to estimate the population density of the problematic species at the position of traps, e.g. by dividing the trap counts by the effective catchment area (cf. [13,14]).

Exhaustive information about species' presence in a given area is given by its population size, i.e. by the total number of its individuals. In order to get reliable estimates of the population size, the information about population abundance obtained at some particular locations must somehow be 'extended' over the whole monitored area. This is routinely performed by statistical analysis of the samples [15]. Indeed, application of statistical methods in ecology has a long and successful history [16–18]. Estimation of the population size is usually based on the arithmetic mean: let  $u_1, u_2, \dots, u_N$  be the population densities at  $N$  different positions in space (i.e. position of the traps), then the average density  $\bar{u}$  is estimated as [16]

$$\bar{u} \approx \frac{1}{N} \sum_{i=1}^N u_i, \quad \text{so that} \quad M \approx A\bar{u}, \quad (1.1)$$

where  $M$  is the population size and  $A$  is the total monitored area.

There are, however, a few problems in implementing of equation (1.1). The probability theory predicts that equation (1.1) gives the exact value of the population size when  $N$  becomes infinitely large. In practical applications,  $N$  is of course finite, which means that the theoretical results obtained in the large- $N$  limit may not be valid. While in a particular scientific study, the number  $N$  of traps per given area can be quite large, e.g. of the order of hundreds, in routine pest monitoring programmes  $N$  rarely exceeds 20 (cf. [19]) and, in some cases, it can be as small as one or a few traps per field [20]. There are several reasons why the number of traps cannot be larger, for instance:

- handling a large number of traps is labour- and resource-consuming and hence is often not possible, especially in nationwide or regional ecological monitoring programmes where traps may be operated simultaneously in many fields;
- traps introduce a disturbance into the field (cf. figure 1). Installing a large number of them can damage the corresponding agricultural product considerably, thus making the subsequent protective measures rather senseless;
- large number of traps may also have a disruptive effect on the behaviour of the monitored animals and hence result in a biased estimate of the population size.

Application of equation (1.1) in the case of small  $N$  increases the level of error of the population estimate and hence remains disputable. In particular, the question arises as to what can be the minimum number of

traps required to provide a robust estimate of the population size. While equation (1.1) may work well for a small  $N$  when the population spatial distribution is random or homogeneous, its application is problematic when the population distribution exhibits a considerable degree of aggregation. However, the latter is a much more common case in ecology than the former [21], e.g. either owing to the heterogeneity of environment [22] or owing to the self-organized animal grouping and pattern formation [23,24]. Several methods have therefore been developed to estimate the required number of traps for a pest sampling programme where the trap catches display aggregated spatial distribution. Most often these are based on the negative binomial distribution or Taylor's power law, which provide indexes of aggregation [25]. However, these techniques still require an initial intensive sampling phase to determine the expected parameters of the frequency distribution.

The second and even more important problem is that equation (1.1) and the existing statistical methods for estimating the required number of traps do not take space into account. That is, they do not use the spatial information available to its fullest extent (cf. [26,27]). In particular, they do not contain any information about the distance between the traps. Meanwhile, the area associated with each trap as well as the way how this area is accounted for (e.g. by ascribing a larger 'weight' to the trap counts collected from a bigger area) are likely to be of crucial importance, especially with attractive traps such as colour traps, light traps, pheromone-baited traps, etc., which are most commonly used in pest monitoring programmes as they provide a selective catch.

In practice, the conclusion about the optimum number of traps is often compromised by cost and agonomic considerations, and can be based solely on the intuition of the grower or crop adviser. Although the importance of intuition is widely recognized, so that 'one should not underestimate the processing power of the human brain' [28], yet it can hardly be accepted as an adequate substitute for a rigorous quantitative analysis. One of the continuing problems of pest management is getting scientific principles of ecological monitoring implemented at commercial production. Correspondingly, there is a strong need for conceptually simple and yet scientifically sound and effective methods to extract higher quality information from sparse spatial data.

In this paper, we develop a new method that is based on ideas different from statistical analysis and hence may provide an alternative to the existing approaches. Moreover, we will show that, in many cases, it actually appears to be significantly more accurate and/or more effective than the commonly used statistical methods, in particular than those based on equation (1.1).

It has been recently shown that the situation when the population density is only known at certain locations can be interpreted as a numerical integration problem [29,30]. The conventional problem of numerical integration is well known and well studied, and many efficient methods have been developed to compute the integral of a given function with a required accuracy (e.g. [31]).

However, while we can benefit from using the numerical integration techniques, their application in ecological monitoring is not at all straightforward. Contrary to the standard problem of numerical integration where a computational grid can be refined to ensure required accuracy of the integration, the grid of traps in a routine monitoring procedure cannot be refined for the reasons listed above. Calculation of the population size therefore requires numerical integration of the population density defined on a very coarse grid. Integration on a coarse grid is a challenging mathematical problem that has been poorly studied in the literature. Most of the results of the numerical integration theory, such as, for instance, estimates of the convergency rate which make it possible to compare the accuracy of different methods, simply do not apply to the coarse grid integration. Since the grid cannot be refined, the whole concept of convergency appears to be irrelevant.

In this paper, we consider the problem of integration of sparse data in the context of calculation of the population size as required by the goals of ecological monitoring. In §2, we revisit a few of the most commonly used methods of numerical integration. We then apply these methods to different types of spatial population distribution either obtained from a population dynamics model or generated straightforwardly and show that a reasonably accurate population size estimate (with integration error consistently less than 25%) can be obtained on a coarse grid of 25 nodes (e.g.  $5 \times 5$  traps in a square domain). Moreover, we show that a reasonable estimate of the population size can often be obtained even on a coarser grid of just nine nodes (i.e.  $3 \times 3$  traps). We then apply the methods of numerical integration to some available field data and show that, apart from some cases of extreme spatial aggregation, an estimate of the population size with accuracy less or about 20–30% can indeed be obtained on a grid of nine traps. Finally, we discuss the implications of our findings for ecological monitoring practices and show that they may lead to a paradigm shift and a new concept of ecological monitoring.

## 2. METHODS OF NUMERICAL INTEGRATION

### 2.1. The problem statement

Let a function  $f(x,y)$  be the density of a pest population in a given two-dimensional domain  $\tilde{D}$ . We assume that the information about the pest population density is collected through trapping where traps are placed at the nodes of a rectangular grid. From a computational viewpoint, the trapping procedure means that the function  $f(x,y)$  is not available at an arbitrary point  $P = (x,y) \in D$ . Instead, we have to deal with a discrete function  $f_{ij} \equiv f(x_i, y_j)$ ,  $i = 1, \dots, N_x$ ,  $j = 1, \dots, N_y$  defined on a Cartesian grid consisting of the total number  $N^2 = N_x \times N_y$  nodes.

For the sake of simplicity, in this paper, we restrict our study by considering a rectangular domain  $\tilde{D}$ . For convenience, we transform (using a linear mapping) domain  $\tilde{D}$  into the unit square  $D = [0,1] \times [0,1]$ , where

a Cartesian grid can easily be generated. Namely, let us consider a set of points  $x_i$ ,  $i = 1, \dots, N_x$  at the interval  $[0,1]$ , where we require that  $x_1 = 0$ ,  $x_{i+1} = x_i + h_x$ ,  $i = 1, \dots, N_x - 1$  and the grid step size  $h_x$  is defined as  $h_x = 1/(N_x - 1)$ . Consider now a set of points  $y_j$ ,  $j = 1, \dots, N_y$  in the domain  $[0,1]$  and generate a one-dimensional grid in the  $y$ -direction as  $y_1 = 0$ ,  $y_{j+1} = y_j + h_y$ ,  $j = 1, \dots, N_y - 1$ ,  $h_y = 1/(N_y - 1)$ . The grid node position is then given as  $(x_i, y_j)$ .

Once a computational grid has been generated and the data are available at grid nodes, there are several methods of numerical integration that can be applied in an ecological monitoring problem. We briefly revisit them below.

## 2.2. Composite rules of numerical integration

A conventional approach to integration over a Cartesian grid is to consider a composite rule of integration where the integral over the unit square  $D$  is replaced by the sum of integrals evaluated at each grid cell (e.g. [31]). Consider the grid cell  $c_{ij}$  defined as  $c_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ , where  $i = 1, \dots, N_x - 1$ ,  $j = 1, \dots, N_y - 1$ . A composite rule of integration therefore reads as

$$I = \int_0^1 \int_0^1 f(x, y) dx dy = \sum_{i,j} I_{ij}, \quad (2.1)$$

where

$$I_{ij} = \iint_{c_{ij}} f(x, y) dx dy. \quad (2.2)$$

Hence, the integration problem is reduced to the integral evaluation at each rectangular subdomain  $c_{ij}$ . Different integration rules use different ‘local’ approximation of the integrand  $f(x, y)$  at each grid cell  $c_{ij}$ .

**2.2.1. The midpoint rule.** The simplest evaluation of the integral (2.2) can be performed under the assumption that the function  $f(x, y)$  is approximated by a constant at each grid cell. Such approximation results in the following formula:

$$I_{ij} \approx A_{ij} f(x_i, y_j), \quad (2.3)$$

where  $A_{ij} = h_x h_y$  for the interior nodes, i.e. for  $i = 2, \dots, N_x - 1$ ,  $j = 2, \dots, N_y - 1$ . At boundary points, where one of the conditions  $i = 1$ ,  $i = N_x$ ,  $j = 1$  or  $j = N_y$  holds, we have  $A_{ij} = (1/2)h_x h_y$  and  $A_{ij} = (1/4)h_x h_y$  at the four corners of the domain.

**2.2.2. The trapezoidal rule.** The trapezoidal rule of integration implies the approximation of  $f(x, y)$  by a linear function in each grid cell  $c_{ij}$ . Correspondingly, the integral of  $I_{ij}$  is evaluated as

$$I_{ij} \approx \frac{h_x h_y}{4} [f(x_i, y_j) + f(x_{i+1}, y_j) + f(x_i, y_{j+1}) + f(x_{i+1}, y_{j+1})]. \quad (2.4)$$

**2.2.3. The Simpson rule.** The Simpson rule of integration is another ‘local’ integration rule, where the

integrand  $f(x, y)$  is approximated by a quadratic polynomial at each subdomain. The application of this rule in the cell  $c_{ij}$  requires that the data  $f(x, y)$  are available at points  $(x_{i+q}, y_{j+r})$ , where  $q = 0, 1/2, 1$  and  $r = 0, 1/2, 1$ . The notations  $q = 1/2$  and  $r = 1/2$  are used for the midpoints in the  $x$ - and  $y$ -directions, that is  $x_{i+1/2} = 0.5(x_i + x_{i+1})$  and  $y_{i+1/2} = 0.5(y_i + y_{i+1})$ . The function  $f(x, y)$  is then integrated in the cell  $c_{ij}$  by the Simpson rule as

$$I_{ij} \approx \frac{h_x h_y}{36} [f(x_i, y_j) + f(x_i, y_{j+1}) + f(x_{i+1}, y_j) + f(x_{i+1}, y_{j+1}) + 4(f(x_i, y_{j+1/2}) + f(x_{i+1/2}, y_j) + f(x_{i+1}, y_{j+1/2}) + f(x_{i+1/2}, y_{j+1})) + 16f(x_{i+1/2}, y_{j+1/2})]. \quad (2.5)$$

Note that the straightforward application of the Simpson rule given by equation (2.5) corresponds to the case when the grid can be refined by introducing the new ‘half-integer’ nodes. In the case we consider in this paper, the number of nodes cannot be increased. Therefore, in order to apply the Simpson rule, we have to increase the size of integration cells by considering  $h_x \rightarrow 2h_x$ ,  $h_y \rightarrow 2h_y$ ; we can then use the even-numbered nodes instead of the half-integer ones. As an immediate consequence, it means that the number of grid nodes in each direction  $x, y$  must be odd.

## 2.3. Least-squares polynomial approximation

One alternative to the composite rules discussed above is to consider a ‘global’ polynomial approximation, where the function  $f(x, y)$  is replaced by a single polynomial over the entire domain of integration. Consider a polynomial  $p_K(x, y)$  of degree  $K$ ,

$$p_K(x, y) = \sum_{m=1}^M a_m \phi_m(x, y), \quad (2.6)$$

where  $\phi_m(x, y) = (x - x_0)^{m_1} (y - y_0)^{m_2}$  are polynomial basis functions,  $m_1 + m_2 = 0, 1, \dots, K$ . We then require that

$$f(x, y) \approx p_K(x, y), \quad (2.7)$$

where the basis functions  $\phi_m(x, y)$ ,  $m = 1, \dots, M$  are now linked to the central point of the unit square,  $x_0 = 1/2$ ,  $y_0 = 1/2$ . Hence, integral (2.1) can be computed as

$$I = \int_0^1 \int_0^1 f(x, y) dx dy \approx \int_0^1 \int_0^1 \sum_{m=1}^M a_m \phi_m(x, y) dx dy = \sum_{m=1}^M a_m I_m, \quad (2.8)$$

where the partial integrals

$$I_m = \int_0^1 \int_0^1 \phi_m(x, y) dx dy, \quad m = 1, \dots, M,$$

are easily available in closed form for polynomial functions  $\phi_m(x, y)$ .

The polynomial coefficients in expansion (2.6) can be defined by interpolation or by least-squares approximation. In our work, we use a least-squares method to define the coefficients  $a_m$ . The least-squares procedure is explained in appendix A.

#### 2.4. A coarse grid problem

A conventional approach to the choice of a numerical integration method is to evaluate the error that appears when the exact integral is approximated by a numerical rule. Provided the grid step size  $h = \max(h_x, h_y)$  is sufficiently small, the integration error  $e$  usually depends on  $h$  as  $e = Ch^k$ , where  $C$  is a constant and  $k \geq 1$  [31]. If two methods of numerical integration are compared, then the most suitable candidate for the problem will be the method that has a smaller integration error. In other words, if the integration method  $A$  has the error  $e_A = O(h^{k_1})$  and method  $B$  has the error  $e_B = O(h^{k_2})$ , where  $k_1 > k_2$ , then the first method should give a more accurate result on a grid with a fixed number of grid nodes. Thus, an integration method with faster convergence rate (i.e. a method that has a greater value of  $k$  in the error estimate) is thought of as a more accurate method that should be recommended for numerical integration.

Meanwhile, a crucial feature of the numerical integration problem we deal with is that the above conclusion is not necessarily true for coarse grids, i.e. for grids with a small number  $N^2$  of nodes. It has been shown in Petrovskaya & Venturino [30] that the asymptotic error estimates in the form  $e = Ch^k$  cannot be implemented when sparse data are integrated, because we cannot provide an accurate estimate of the constant  $C$  on coarse grids. The estimate of  $C$  depends on the integrand function  $f(x, y)$ , while discrete data  $f_{ij}$  available on coarse grids may not have the same properties as their continuous counterpart  $f(x, y)$ . In other words, the method that has slower convergence rate can nevertheless have a smaller integration error on coarse grids, where the integrand function is not well resolved. Hence, our next task is to compare the integration error for all the methods above and to conclude if any of them can provide a user with acceptable accuracy on coarse grids.

### 3. NUMERICAL TEST CASES I: ECOLOGICAL MODEL

The ultimate goal of our study is to apply the methods of numerical integration to trap count data of pest monitoring with the purpose to estimate, with a reasonable accuracy, the population size in a given area. We aim to understand the relative efficiency of different integration methods and also show how the number of traps can be minimized without any essential loss of information about pest abundance.

However, the currently available field data appropriate for our purposes are rare, in particular, because the standard methods of integration imply certain restrictions on the traps number and position. Indeed, the methods discussed above can be applied only when the traps are installed at the nodes of a rectangular

grid but not to the case of their arbitrary position. Besides, the Simpson method requires that the number of traps in each direction must be odd.

Therefore, in order to evaluate the effectiveness and accuracy of numerical integration on a coarse grid, we first consider simulation data obtained from an ecologically sound mathematical model of population dynamics. Specifically, we consider the following spatially explicit predator–prey model with the Allee effect (e.g. [28,32]):

$$\begin{aligned} \frac{\partial U(X, Y, T)}{\partial T} = & D_1 \left( \frac{\partial^2 U}{\partial X^2} + \frac{\partial^2 U}{\partial Y^2} \right) + \left( \frac{4\nu}{(K - U_0)^2} \right) \\ & \times U(U - U_0)(K - U) - \frac{AUV}{U + B} \end{aligned} \quad (3.1)$$

and

$$\frac{\partial V(X, Y, T)}{\partial T} = D_2 \left( \frac{\partial^2 V}{\partial X^2} + \frac{\partial^2 V}{\partial Y^2} \right) + \kappa \frac{AUV}{U + B} - MV. \quad (3.2)$$

Here  $U$  and  $V$  in equations (3.1) and (3.2) are the densities of prey and predator, respectively, at time  $T$  ( $T > 0$ ) and position  $(X, Y)$ . Note that, depending on a particular application, either of the two species can be the pest. We consider population dynamics in a square-shaped domain so that  $0 < X < L$  and  $0 < Y < L$ , where  $L$  can be interpreted as the typical size of a given agricultural field. At the domain boundaries, the zero-flux conditions are used. Coefficients  $D_1$  and  $D_2$  describe species diffusivity owing to the movement of the individuals, parameter  $K$  is the prey carrying capacity,  $U_0$  is the Allee threshold density ( $0 < U_0 < K$ ),  $\nu$  is the maximum prey *per capita* growth rate (cf. [33]),  $A$  is the predator attack rate,  $B$  is the half-saturation prey density,  $\kappa$  is the food assimilation efficiency coefficient and  $M$  is the predator mortality. For the sake of simplicity, below we assume that  $D_1 = D_2 = D$ . Note that this assumption is not ecologically unrealistic: available estimates of the range of diffusivity values often overlap between different species, even for species from different taxonomic groups ([34], §3.7).

Here we want to mention that the equations (3.1) and (3.2) are, in an appropriate parameter range, in a good qualitative and sometimes quantitative agreement with experimental data ([24], §11.3) and therefore can be used to simulate ecologically realistic spatial density distributions.

In order to obtain the population distributions, equations (3.1) and (3.2) are solved numerically. For convenience, we first introduce dimensionless variables. The domain length  $L$  makes a convenient scale for the coordinates,  $x = X/L$  and  $y = Y/L$ , where the new variables  $x$  and  $y$  are dimensionless,  $0 < x < 1$ , and  $0 < y < 1$ . In a similar way, the carrying capacity  $K$  makes a convenient scale for the prey population density,  $u = U/K$ . By introducing dimensionless predator density as  $v = V/(\kappa K)$  and dimensionless time as  $t = aT$ , where  $a = A \kappa K/B$ , equations (3.1) and

(3.2) take the following form:

$$\frac{\partial u(x, y, t)}{\partial t} = d_1 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + \beta u(u - b)(1 - u) - \frac{uv}{1 + \Lambda u} \quad (3.3)$$

and

$$\frac{\partial v(x, y, t)}{\partial t} = d_2 \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) + \frac{uv}{1 + \Lambda u} - mv, \quad (3.4)$$

where  $\Lambda = K/B$ ,  $b = U_0/K$ ,  $\beta = 4vBK/(A\kappa(K - U_0)^2)$ ,  $m = M/a$ ,  $d_1 = D_1/(aL^2)$  and  $d_2 = D_2/(aL^2)$  are dimensionless parameters.

For the parameter range where the local prey–predator dynamics is oscillatory, equations (3.3) and (3.4) are known to generate a rich variety of spatio-temporal patterns [24]. The left column in figure 2 shows three different cases (the corresponding parameter values, the initial conditions and other technical details are given in appendix B). Figure 2*a* shows a smooth distribution of the prey density over the area, it reaches its maximum value (shown in red) near the  $x$ -axis and drops to zero in the right-hand side of the domain. Figure 2*b* shows the almost homogeneous prey distribution inside the round area (shown in yellow) separated from the rest of the domain, where the population is absent, by a steep gradient in the population density. This case may correspond to the species spread from the place of its initial introduction in the centre of the domain. Figure 2*c* also shows the species spread but following a different dynamical regime when the population distribution in the wake of the population front forms a distinct patchy structure (cf. [35]).

Now, we are going to check whether the method of numerical integration on a coarse grid can be applied to population distributions shown in figure 2. Note that, in order to estimate the integration error, we need the exact value of the integral (i.e. the population size), which is not known because equations (3.3) and (3.4) cannot be solved analytically. To overcome this difficulty, we first solve equations (3.3) and (3.4) numerically on a very fine grid  $G_f$  that has  $N_x = N_y = N_f = 2^{10} + 1$  grid nodes at each direction, and consider the result as the exact solution to the problem. Consequently, the corresponding value  $I_{N_f}$  of the solution integral is considered as the exact integral that will be compared with the value  $I_N$  of the integral on a coarse grid. Coarse grids  $G_c$  are generated as grids where the number of nodes at each direction is  $N_x = N_y = N = 2^s + 1$ ,  $1 \leq s \leq 7$ , and a projection of the ‘exact’ solution obtained on  $G_f$  onto each coarse grid is considered. Once the function  $u(x, y)$  is available at nodes of a coarse grid, the numerical integration is performed and the integration error is computed as

$$e = \frac{|I_{N_f} - I_N|}{I_{N_f}}. \quad (3.5)$$

An advantage of using simulation data instead of real data to test the integration methods on a coarse

grid is that we can easily make the grid as refined as we want by adding more grid nodes, which is difficult in empirical field studies. Correspondingly, we can consider how the integration accuracy changes depending on the number of grid nodes. Error (3.5) as a function of the number  $N$  of grid nodes at each direction, starting from the minimum number of  $N = 3$ , is shown in the right column of figure 2. The dashed red horizontal line shows the hypothetical value of desirable error of 25 per cent. It is readily seen from figure 2 that, for the coarse grid with  $N = 5$ , the Simpson rule and the least-squares method provide good accuracy of less than 25 per cent consistently over all three different types of population distribution. Even on a very coarse grid with  $N = 3$  (so that the total number of grid nodes is just  $N^2 = 9$ ), methods of numerical integration provide a very good accuracy (less than 10%) for the density distribution shown in figure 2*a* and a reasonable accuracy<sup>1</sup> less or about 50 per cent in the case of figure 2*c*.

The population distributions shown in figure 2 correspond to three qualitatively different cases but they, of course, do not exhaust all the possibilities. An interesting and ecologically important pattern is given by the ‘patchy invasion’ [38–40] when, contrary to what is shown in figure 2*e*, the strongly heterogeneous patchy spatial distribution of a spreading population is not preceded by the propagation of any continuous front. The left column of figure 3 shows the snapshots of the prey population density at three different stages of the patchy invasion, i.e. (a) early, (b) intermediate, and (c) late (the parameters and other details are given in appendix B). The right column shows the corresponding accuracy of numerical integration. Rather counterintuitively, except for the early stage of the patchy spread, even a very coarse grid provides a reasonable accuracy which is consistently less than 50 per cent for  $N = 3$  and less than 30 per cent for  $N = 5$ .

#### 4. NUMERICAL TEST CASES II: EXTREME AGGREGATION

In §3, we showed that numerical integration on coarse grids provides good accuracy in the case where the population is distributed more or less over the whole area, no matter whether this distribution is simple (e.g. like the one shown in figure 2*a*) or may have a complex structure (figure 3*e*). However, the accuracy appears to be much lower when the population is aggregated within one or a few small areas which may fall in between the sampling grid nodes (cf. figure 3*a*).

It should be noted that, even on a very coarse grid, such a narrow peak in the spatial population distribution is not necessarily undetectable. It depends on the peak location with respect to the position of the traps (i.e. position of the grid nodes) if its contribution to the population size is properly accounted. The actual problem is that the location of the peak is not known *a priori*. Therefore, it is impossible to distinguish in advance between the cases when the numerical

<sup>1</sup>Note that, in many ecological studies, the relative error of 100 per cent (i.e.  $e \sim 1$ ) is still regarded as acceptable [36,37].

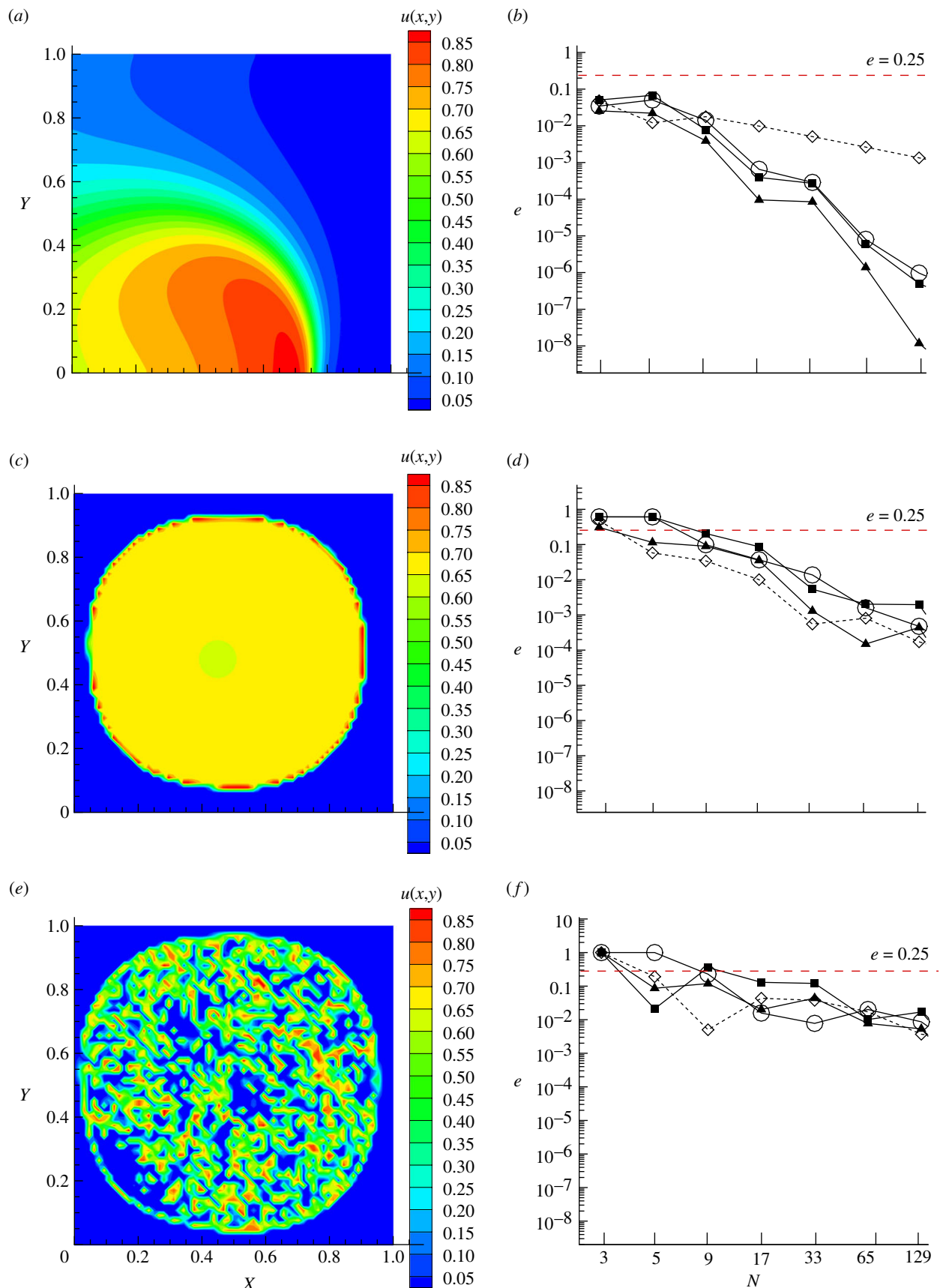


Figure 2. (a,c,e) Different population distributions as predicted by the population dynamics model (3.3) and (3.4) for different parameter values (see appendix B). (b,d,f) The error  $e$  obtained by numerical integration rules on a sequence of uniformly refined grids. Solid line with filled square, midpoint rule (MR); solid line with open circle, trapezoidal rule (TR); solid line with filled triangle, Simpson's rule (SR); dashed line with open diamond, least-squares method (LS).  $N$  is the number of grid nodes in each direction on a uniform Cartesian grid. The red dashed line labeled as  $e = 0.25$  shows the hypothetical value of desirable error of 25%.



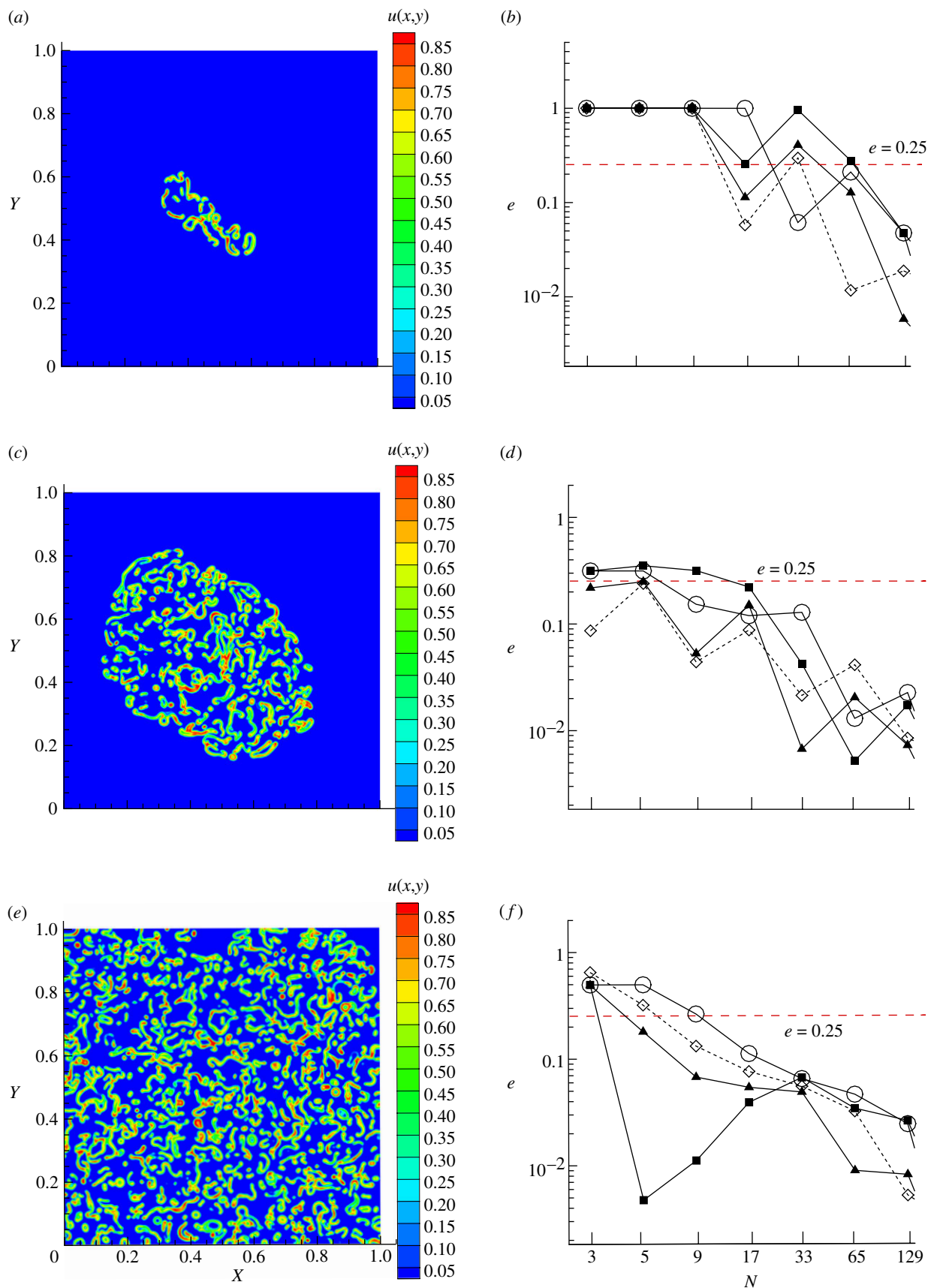


Figure 3. (a,c,e) Snapshots of the population density at different stages of the patchy invasion; the parameters and technical details are given in appendix B. (b,d,f) The error  $e$  obtained by numerical integration rules on a sequence of uniformly refined grids. Solid line with filled square, midpoint rule (MR); solid line with open circle, trapezoidal rule (TR); solid line with filled triangle, Simpson's rule (SR); dashed line with open diamond, least-squares method (LS).  $N$  is the number of grid nodes in each direction on a uniform Cartesian grid. The red dashed line labeled as  $e = 0.25$  shows the hypothetical value of desirable error of 25%.

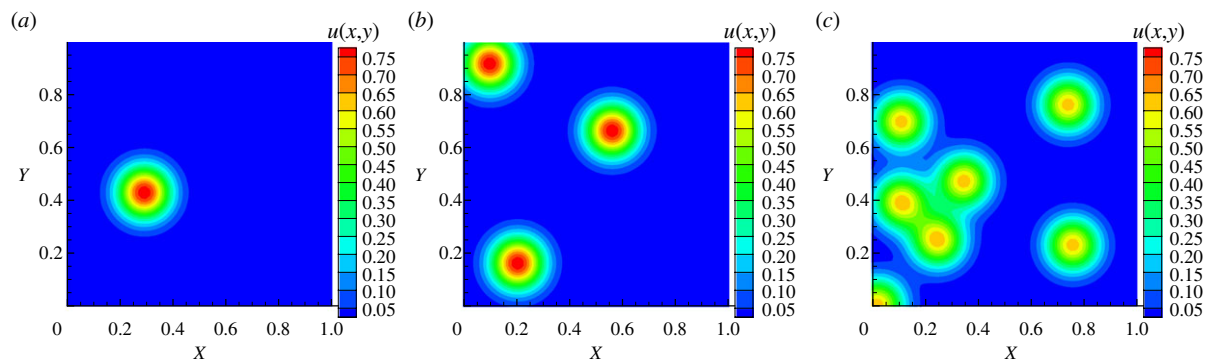


Figure 4. Snapshots of the population density in the case of (a) a single peak distribution, (b) a three peak distribution and (c) a seven peak distribution. The standard deviation of each peak is  $\sigma = 0.05$  and the position is chosen randomly. (Online version in colour.)

Table 1. The integration error (3.5) on a coarse grid of  $N^2 = 9$  nodes for the single peak distributions, for different integration rules:  $e_{\text{MR}}$  for the midpoint rule,  $e_{\text{TR}}$  for the trapezoidal rule,  $e_{\text{SR}}$  for the Simpson rule,  $e_{\text{LS}}$  for the least-square method and  $e_{\text{stat}}$  for the arithmetic mean-based statistical approach (see equation (1.1)).

| dataset | $e_{\text{MR}}$ | $e_{\text{TR}}$ | $e_{\text{SR}}$ | $e_{\text{LS}}$ | $e_{\text{stat}}$ |
|---------|-----------------|-----------------|-----------------|-----------------|-------------------|
| 1       | <b>0.063</b>    | 0.532           | <b>0.167</b>    | 1.000           | <b>0.167</b>      |
| 2       | 0.897           | 0.788           | 0.954           | 0.587           | 0.816             |
| 3       | 4.328           | 4.328           | 3.736           | 1.000           | 3.736             |
| 4       | 0.935           | 0.935           | 0.885           | 1.000           | 0.971             |
| 5       | 0.720           | 0.415           | 0.876           | <b>0.120</b>    | 0.502             |
| 6       | 0.968           | 0.968           | 0.972           | 1.000           | 0.972             |
| 7       | 0.750           | 0.750           | 0.556           | 1.000           | 0.889             |
| 8       | <b>0.216</b>    | <b>0.216</b>    | 0.393           | 1.000           | 0.652             |
| 9       | 4.404           | 1.702           | 3.804           | 1.000           | 3.804             |
| 10      | 0.982           | 0.982           | 0.982           | 1.000           | 0.985             |

integration on a coarse grid may still provide a reasonable estimate and when it fails. However, it is possible to assess the frequency of different outcomes if we fix the grid and treat the location of the peak as random.

In order to make a quantitative insight into this matter, we now reduce the grid size to a minimum sensible case, e.g.  $N^2 = 9$  or  $N^2 = 25$  nodes, and consider another series of numerical tests. We do not need any population dynamics model now as we can generate the desirable spatial distributions directly. We start with the case when the population consists of a single peak distribution (figure 4a). We describe the shape of the peak by a normal distribution:

$$u(x, y) = \frac{U_0}{4\pi\sigma^2} \exp\left(-\frac{(x - \hat{x}_1)^2 + (y - \hat{y}_1)^2}{4\sigma^2}\right), \quad (4.1)$$

where  $U_0$  is a scaling factor, the variance  $\sigma^2$  is fixed at a small hypothetical value and  $\hat{x}_1$  and  $\hat{y}_1$  are random numbers distributed uniformly between 0 and 1, so that the position  $(\hat{x}_1, \hat{y}_1)$  of the peak is chosen randomly. We consider 10 different realizations (numbered sequentially in the first column of table 1) and calculate the population size and the corresponding error for each of them. We use the methods of numerical integration described in §2 as well as the statistical method given

by equation (1.1). The results are given in tables 1 and 2. Expectedly, we observe that the typical accuracy of the integration on a very coarse grid with  $N^2 = 9$  nodes is low. In fact, only in one or two cases (shown in bold in table 1) out of 10, the accuracy is within the desirable 25 per cent; see also table 2a. However, we point out that the number of cases, where the accuracy is within 90 per cent (which may still be acceptable for some applications, see footnote 1), is five out of 10 for four integration rules (the least-squares method appears to be somewhat less effective).

In order to further reveal the tendency, we repeat this procedure in the case where the population distribution consists of a few peaks:

$$u(x, y) = \frac{U_0}{4\pi\sigma^2} \sum_{j=1}^J \exp\left(-\frac{(x - \hat{x}_j)^2 + (y - \hat{y}_j)^2}{4\sigma^2}\right), \quad (4.2)$$

where  $\hat{x}_j$  and  $\hat{y}_j$  ( $j = 1, \dots, J$ ) are random numbers and  $J$  is the total number of peaks. Table 2b,c shows respectively, the results obtained for  $J = 3$  and  $J = 7$  (figure 4b,c). Interestingly, the frequency of the ‘good’ cases does not change much when the number of peaks increases from one to three. However, it increases significantly when the number of peaks increases from three to seven. Note that even in the latter case, the population density is negligibly small in more than 80 per cent of the area, yet the midpoint and trapezoidal rules ensure the acceptable accuracy within 90 per cent in all the cases considered.

We also mention here that, in all three cases considered above, i.e. for  $J = 1$ ,  $J = 3$  and  $J = 7$ , there is no significant difference in accuracy between the statistical method (1.1) and the numerical integration rules.

The situation, however, changes remarkably if the integration is considered on a slightly less coarse grid, i.e. when the number of nodes increases from  $N^2 = 9$  to  $N^2 = 25$  (table 2d). Now, even in the extreme situation of the single peak distribution ( $J = 1$ ), the Simpson rule ensures the desirable accuracy (within 25%) in 10 out of 10 cases. Almost the same good accuracy is provided by the least-squares method. The efficiency of other integration rules is comparable with the Simpson rule in the low accuracy range of

Table 2. The integration accuracy of different methods for the distributions consisting of  $J$  peaks. The first column gives the number  $N^2$  of the nodes in the grid, columns 4–8 give the number of cases with the accuracy in the prescribed range.

|          | $N^2$ | peaks ( $J$ ) | accuracy range (%)    | midpoint | trapezoidal | Simpson | least squares | equation (1.1) |
|----------|-------|---------------|-----------------------|----------|-------------|---------|---------------|----------------|
| <i>a</i> | 9     | 1             | $\varepsilon \leq 25$ | 2        | 1           | 1       | 1             | 1              |
|          | 9     | 1             | $\varepsilon \leq 50$ | 2        | 2           | 2       | 1             | 1              |
|          | 9     | 1             | $\varepsilon \leq 90$ | 5        | 5           | 5       | 2             | 5              |
| <i>b</i> | 9     | 3             | $\varepsilon \leq 25$ | 0        | 0           | 1       | 0             | 3              |
|          | 9     | 3             | $\varepsilon \leq 50$ | 3        | 3           | 3       | 1             | 4              |
|          | 9     | 3             | $\varepsilon \leq 90$ | 7        | 7           | 8       | 3             | 8              |
| <i>c</i> | 9     | 7             | $\varepsilon \leq 25$ | 3        | 4           | 3       | 2             | 4              |
|          | 9     | 7             | $\varepsilon \leq 50$ | 6        | 7           | 4       | 2             | 6              |
|          | 9     | 7             | $\varepsilon \leq 90$ | 10       | 10          | 9       | 3             | 10             |
| <i>d</i> | 25    | 1             | $\varepsilon \leq 25$ | 3        | 2           | 10      | 8             | 1              |
|          | 25    | 1             | $\varepsilon \leq 50$ | 7        | 6           | 10      | 10            | 3              |
|          | 25    | 1             | $\varepsilon \leq 90$ | 9        | 10          | 10      | 10            | 8              |

$\varepsilon \leq 90\%$  but appears to be much lower in the high accuracy range of  $\varepsilon \leq 25\%$ .

## 5. APPLICATION TO FIELD DATA

In the previous sections, we demonstrated that numerical integration on a grid as coarse as  $N^2 = 9$  nodes can provide a reliable estimate of the population size even when the population spatial distribution has a highly complicated structure. Now, we are going to further validate our approach by applying it to field data of ecological monitoring. The experimental data we use have been collected for a New Zealand flatworm population (*Arthurdendyus triangulatus*) [41]. This flatworm is an alien species originally brought from New Zealand and eventually having established itself over a large part of the British Isles [42] as well as having potential to colonize Europe [43]. The New Zealand flatworm is regarded as a potentially dangerous pest because of its predatory impact on earthworms [44], which may have a severe disruptive effect on the corresponding ecosystem, in particular, causing a decline in the soil fauna and strongly affecting earthworm predators like moles and badgers [45,46].

For this reason, flatworms have been a subject of controversy, intensive study and monitoring, particularly, in Northern Ireland [42,47]. Here, we use the results obtained in a field study performed in January–March of 2002. The goal of the study was to reveal the features of flatworms' spatial distribution in grassland, including a population assessment. An area of a flatworm-infested grassland field at Newforge Lane, Belfast, Northern Ireland, was used for the investigation (World Geodetic System  $54^\circ 33' 24''$  N,  $5^\circ 56' 47''$  W, Irish Grid J330698). The data on flatworms abundance at different locations were collected by means of trapping (figure 1). Shelter traps were heavy duty black polythene sacks, filled with 5 kg of gravel/sand mixed in a ratio of 4 : 1 and measuring  $50 \times 30$  cm. *Arthurdendyus triangulatus* typically shelters on the soil surface under stones, wood and discarded plastic, so shelter traps such as these provide a relatively cheap and convenient method of sampling

and have been widely used [48–50]. The traps were positioned in the nodes of  $12 \times 12$  uniform grid with 2 m spacing, they were examined every week and the numbers of flatworms caught were counted. The features of the experimental design (in particular, trap spacing) are consistent with other similar studies on flatworms (e.g. [49]) and hence can be regarded as typical for this species.

The six distributions of trap counts obtained in the field study are shown in figure 5. Over the sampling period, the  $12 \times 12$  grid traps caught 465–748 flatworms per week. The initial two adjacent sampling weeks, gave counts of 624 and 544 flatworms, equating to a total of 1168 individual flatworms within the grid area [41]. For our purposes, from the trap data collected in the field, we extract the sub-grid with  $N_x = N_y \equiv N = 11$  nodes at each direction. That has been carried out because the Simpson rule requires an odd number  $N$  of grid nodes. An example of the corresponding trap data is given in table 3.

Dividing the trap counts at each location by the area of the grid cell (i.e. by  $4 \text{ m}^2$ ), the trap counts are linked to the local population density [13,14]. Having then integrated the population density over the fine grid of  $N^2 = 121$  nodes, we reproduce the total number of collected flatworms (shown in the first row of table 4 for each of distributions from figure 5*a–f*). This number is considered as the exact value of the population size.

We then recalculate the population size using a very coarse grid of  $N^2 = 9$  nodes by the midpoint rule (the row marked as  $I_{\text{MR}}$  in table 4), by the trapezoidal rule (the row  $I_{\text{TR}}$ ), by the Simpson rule (the row  $I_{\text{SR}}$ ), by the least-squares approximation (the row  $I_{\text{LS}}$ ) and by the baseline statistical approach given by equation (1.1) (the row  $I_{\text{stat}}$ ). For each of these rules, we compute the integration error (3.5) (the rows marked as  $e_{\text{MR}}$ ,  $e_{\text{TR}}$ ,  $e_{\text{SR}}$ ,  $e_{\text{LS}}$  and  $e_{\text{stat}}$ , table 4).

It is readily seen from table 4 that, apart from case (a), integration on the grid of  $N^2 = 9$  nodes gives an estimate of the total population with error consistently less or about 30 per cent. Note that, even in the 'unfortunate' case (a) where the information is largely lost because the patch of a high flatworm population

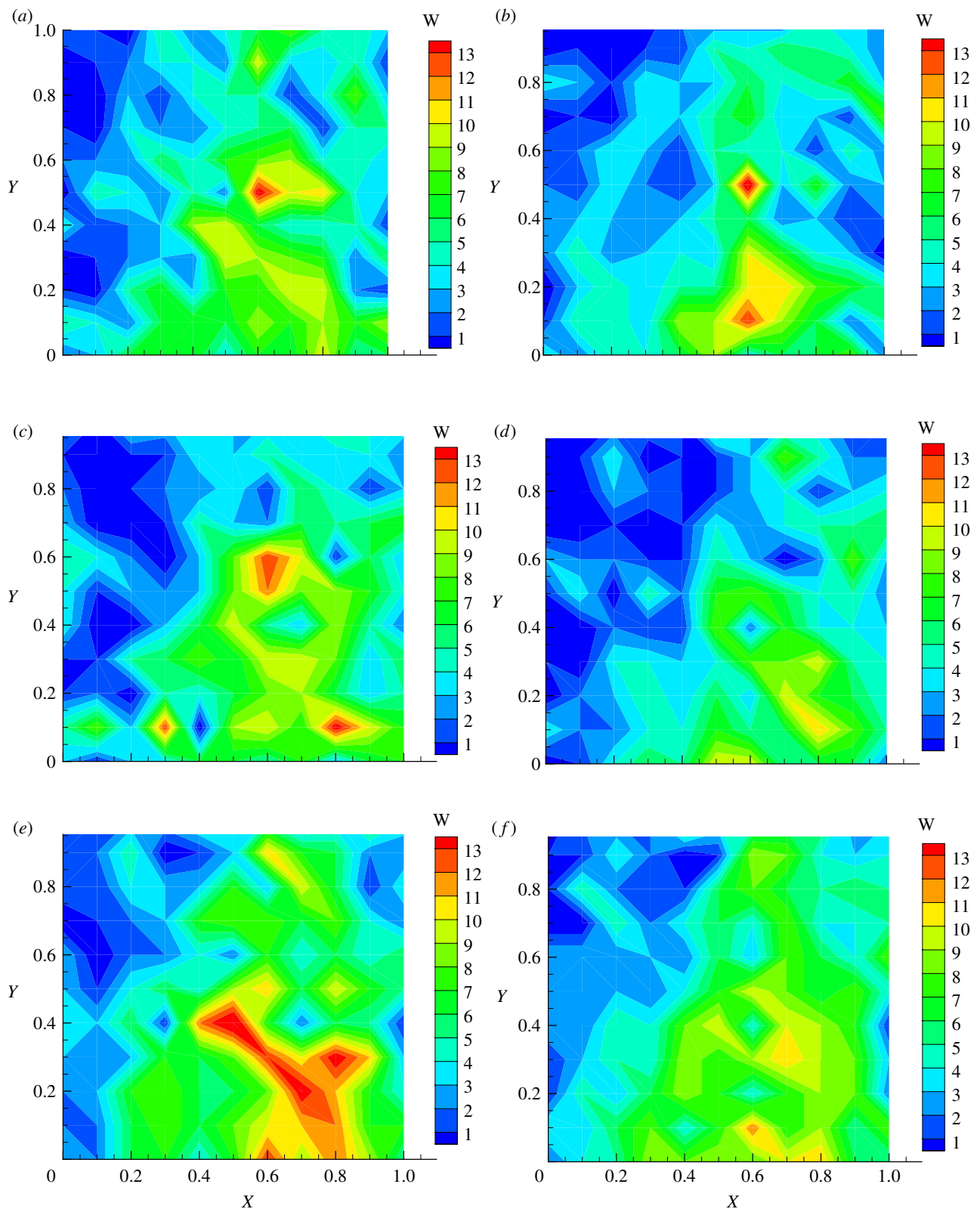


Figure 5. The field data on flatworm spatial distribution over the study area obtained at six different times (see details in the text).

density falls in between the grid nodes, the error of the population size estimate still remains within 55 per cent. Such accuracy can still be considered as acceptable for large-scale monitoring programmes (cf. [20]).

The value of the error averaged over the six cases appears to be just 22 per cent for all four numerical integration rules and 30 per cent for the statistical rule

(1.1). It shows that a robust information about the population size of flatworm population can be obtained using much less traps per unit area (or much larger inter-trap distance) than it is usually used. It also demonstrates a higher accuracy of the numerical integration rules compared with the standard statistical approach.

Table 3. An example of the flatworm trap count data (cf. figure 5a). The square brackets show the traps that form the coarse  $3 \times 3$  grid, the corresponding counts (shown in bold) are used to calculate the coarse-grid estimate of the flatworm population size (table 4).

|            |   |   |   |   |            |    |    |    |   |            |
|------------|---|---|---|---|------------|----|----|----|---|------------|
| <b>[2]</b> | 1 | 0 | 5 | 2 | <b>[4]</b> | 6  | 8  | 5  | 5 | <b>[3]</b> |
| 0          | 1 | 2 | 4 | 4 | 2          | 10 | 3  | 3  | 5 | 1          |
| 1          | 0 | 3 | 1 | 4 | 5          | 5  | 1  | 4  | 8 | 3          |
| 1          | 0 | 4 | 2 | 2 | 4          | 5  | 6  | 1  | 5 | 5          |
| 2          | 2 | 2 | 6 | 4 | 7          | 8  | 9  | 4  | 5 | 3          |
| <b>[0]</b> | 5 | 4 | 2 | 5 | <b>[2]</b> | 14 | 10 | 11 | 4 | <b>[3]</b> |
| 4          | 1 | 2 | 3 | 9 | 10         | 6  | 4  | 4  | 5 | 1          |
| 0          | 1 | 2 | 3 | 2 | 10         | 9  | 8  | 7  | 2 | 6          |
| 1          | 0 | 5 | 7 | 3 | 7          | 7  | 10 | 10 | 2 | 1          |
| 5          | 4 | 2 | 6 | 7 | 6          | 9  | 7  | 9  | 7 | 9          |
| <b>[2]</b> | 3 | 6 | 7 | 7 | <b>[4]</b> | 7  | 6  | 10 | 5 | <b>[6]</b> |

## 6. DISCUSSION

Ecological monitoring is an essential part of the integrated pest management and control. Its main goal is to provide information about the abundance of problem species, which is performed through collecting information in the field. This information is then used as a basis for decision-making, for instance, about application of pesticides. The reliability of the information on pest abundance is therefore an issue of high importance. On the other hand, collecting field data on a large scale, as required by national or regional pest surveys, is costly. A challenging problem is therefore to obtain sufficiently accurate estimates of pest abundance keeping the resources spent on collecting field data collection at a minimum possible value.

For invertebrate species, the information about population abundance is usually obtained from trap counts. The trap counts are then re-calculated into the population density at the position of the traps (e.g. [13,14]). However, as the number of traps cannot be made large, it gives the value of population density only at several locations. The statistical methods to estimate the total population size based on this sparse information remain difficult, in particular, owing to the majority of pests displaying some form of non-random aggregated distribution [21]. The commonly used statistical approaches [25,18] are not fully satisfactory as they are based on restrictive assumptions about the properties of the population spatial distribution or neglecting the spatial aspect at all. The determination of the optimum number of traps within a sampling programme often relies on properties of the frequency distribution of counts rather than the actual pattern [27].

In our previous work [29], we considered a hypothetical one-dimensional case and showed that this problem can, in principle, be solved by applying methods of numerical integration. The population size is obtained by integrating the population density on a coarse grid, where the grid nodes are given by the traps' position. In this paper, we applied this approach to a realistic two-dimensional case using both numerical data (in particular, simulation data from a mathematical model of population dynamics) and field data on some invertebrate species (New Zealand flatworms).

We used several different methods of numerical integration along with a baseline statistical method and showed that a reliable estimate of the population size (with a good accuracy within 25%) can be obtained on a grid consisting of just  $5 \times 5 = 25$  nodes, even in the cases when the population density exhibits a highly complex spatial structure (e.g. figure 3c) or shows extreme spatial aggregation (e.g. figure 4a). In fact, for some of the numerical tests, as well as for the field data, even a smaller grid of  $3 \times 3 = 9$  nodes is sufficient to provide a good accuracy, although the results become more sensitive to the details of the density distribution. This is in good agreement with the results of other field studies. In particular, Boag *et al.* [50] analysed the results of trap counts on a spatially aggregated flatworm population and suggested that a minimum trap number of 15 was necessary to achieve a good population estimate, although they cautioned that this is a compromise between statistical accuracy and fieldwork practicalities.

We also mention here that in almost all cases, the statistical approach based on the arithmetic mean, see equation (1.1), appears to be less accurate than (some of) the integration rules; in general, the Simpson rule seems to be the most accurate one.

### 6.1. Explicit space

The simulation results shown in figures 2–4 are obtained in the dimensionless space. It evokes questions as to what spatial scale the shown population distributions may correspond in reality and how we can relate their scale to the number of grid nodes required for an accurate numerical integration. Recalling that, in simulations, the distances are measured as fractions of the domain size (see the definition of the dimensionless variables above equations (3.3) and (3.4)), for a hump with a characteristic dimensionless size  $\delta_{\text{hump}}$ , we obtain that its real size  $\Delta_{\text{hump}}$  is calculated as

$$\Delta_{\text{hump}} = \delta_{\text{hump}} L. \quad (6.1)$$

Obviously, for a given size  $L$  of the domain, a smaller  $\delta_{\text{hump}}$  corresponds to a smaller  $\Delta_{\text{hump}}$ . However, let us note that  $L$  is a parameter which, generally speaking, can be made different in different simulations, e.g. by considering fields of different size. Therefore, we can address this situation by assuming that the size  $\Delta_{\text{hump}}$  of the hump is the same but the size of the domain is different. Indeed, there is significant theoretical and empirical evidence (cf. [23,51–53]) that, for a given species, the typical size of animals' spatial aggregates can be regarded as a species' trait as it remains approximately the same for different habitats.

We therefore assume that the distribution shown in figure 2a corresponds to the 'microscale' of a single hump in a small field, whereas the multi-hump distribution shown in figure 3e corresponds to a 'macroscopic' spatial scale of the same population but in a much larger field. Having estimated the number of humps at each direction, we conclude that the spatial scale of the distribution shown in figure 3e is about 20 times larger than that of figure 2a. Interestingly, the accuracy of the numerical integration on the same

Table 4. The population size  $I$  and the integration error  $e$  for different integration rules for the flatworm field data on the fine ( $N^2 = 121$ ) and the coarse ( $N^2 = 9$ ) grids. Notations are as in table 1, labels (a–f) correspond to the spatial distribution shown in figure 5.

| field data          | (a)   | (b)   | (c)   | (d)   | (e)   | (f)   |
|---------------------|-------|-------|-------|-------|-------|-------|
| $I, N^2 = 121$      | 544   | 459   | 543   | 419   | 651   | 611   |
| $I_{MR}, N^2 = 9$   | 269   | 319   | 512   | 450   | 531   | 488   |
| $e_{MR}$            | 0.506 | 0.305 | 0.056 | 0.074 | 0.184 | 0.202 |
| $I_{TR}, N^2 = 9$   | 269   | 319   | 512   | 450   | 531   | 488   |
| $e_{TR}$            | 0.506 | 0.305 | 0.056 | 0.074 | 0.184 | 0.202 |
| $I_{SR}, N^2 = 9$   | 247   | 325   | 589   | 544   | 636   | 561   |
| $e_{SR}$            | 0.545 | 0.291 | 0.085 | 0.299 | 0.023 | 0.082 |
| $I_{LS}, N^2 = 9$   | 250   | 350   | 550   | 567   | 600   | 567   |
| $e_{LS}$            | 0.540 | 0.237 | 0.013 | 0.352 | 0.078 | 0.073 |
| $I_{stat}, N^2 = 9$ | 289   | 300   | 456   | 344   | 444   | 411   |
| $e_{stat}$          | 0.469 | 0.346 | 0.161 | 0.178 | 0.317 | 0.327 |

coarse grid does not at all drop down 20 times; in fact, using the midpoint integration rule, the accuracy of the population size estimate on the grid of  $5 \times 5$  nodes appears to be even higher for figure 3e than for figure 2a. Therefore, contrary to what has been observed before in the one-dimensional case (cf. [29]), in order to achieve reasonable accuracy, the numerical grid must not necessarily resolve all the solution's heterogeneities. (A heuristic explanation for this can be that, owing to the high irregularity of the complex 'multipatch' spatial pattern, the overshoots and undershoots at different grid nodes may balance each other.) Having in mind the practical applications of our approach, it leads to a highly counterintuitive conclusion that the number of traps required to obtain a reliable estimate of the population size does not directly depend on the size of the monitored area, and the same minimum number of traps (e.g. nine, as required by the Simpson rule) may be sufficient for a small agricultural field as well as for a large one.

In order to relate the simulation results to the real scale of metres or kilometres, we consider the characteristic scale of the pattern, e.g. an average linear size of a single hump. Its value is known to be species-specific. For instance, it is estimated to be between 30 and 50 m for many common agricultural insect pests [51,52] but it can be much smaller for soil invertebrates; e.g. the results shown in figure 5 suggest the value of the order of several metres. Therefore, the pattern shown in figure 3e may correspond to an insect population distribution over a field of roughly  $1 \times 1$  km. Thus, we conclude that a grid of  $5 \times 5$  traps should be sufficient to provide robust information about the population size over the area of up to 100 ha for insect species and over the area of about 4 ha for flatworms. A conceptual possibility of using smaller grids of traps is discussed below.

## 6.2. Ultra-coarse grids: a paradigm shift?

A traditional approach to numerical integration assumes that the integral can be calculated with a given accuracy. Indeed, in computational mathematics, any prescribed accuracy can be reached by refining the numerical grid (i.e. by increasing the number of nodes

per unit area) as the theory predicts that the calculated value converges to the exact value when the grid spacing tends to zero. In the case that the ideas of numerical integration are applied to ecological monitoring, the situation becomes qualitatively different: we have to deal with sparse data (e.g. population density) obtained at certain given locations (traps position). The numerical grid is therefore fixed and cannot be refined.

Convergency of the calculated integral to the exact value implies that, as a general tendency, the integration error should be the smaller the more refined the grid is, i.e. the larger the number of nodes is. That, in principle, makes it possible to give a robust estimate of the population size with a reasonable accuracy even on a fixed grid if the number of nodes is not too small. Ideally, the number of nodes should be sufficiently large in order to resolve the inhomogeneities in the population spatial distribution [29]. This requires some *a priori* information about the distribution features, which is often unknown. Moreover, careful resolution of small size heterogeneities such as narrow peaks would require a number of nodes per unit area much larger than it is normally available in ecological monitoring programmes [29]. In practical terms, however, we have observed that, using the Simpson integration rule, a good accuracy of less than 25 per cent can be reached<sup>2</sup> on a grid consisting of just 25 nodes, even that this grid apparently does not resolve the single peak distributions well enough (figure 4a).

A grid of 25 traps is still larger than that used in practice (cf. [20]) so there is a need for further optimization. A question is whether any useful information can be obtained on a smaller grid. Obviously, for an 'ultra-coarse' grid, the above approach is not valid any more: indeed, a smaller grid of nine nodes largely fails when applied to highly aggregated population distributions consisting of one or a few narrow peaks (see §4). The problem is that the position of the peak(s) is not known in advance. Hence, it is not possible to adjust the position of the grid nodes (traps) to the position of the peaks. A narrow peak can still be 'integrated' properly but only if it is located close to one of the

<sup>2</sup>Except for some degenerate cases (cf. figure 3a).

grid nodes. This is a typical example of a problem with uncertainty, which can have more than one outcome, e.g. depending on relative position of the peaks and traps. A standard way to deal with this type of problem (e.g. [54]) is by introducing probabilities of different outcomes (table 2). That leads to a paradigm shift: instead of considering the integration result *per se*, we now have to interpret it probabilistically. For instance, whatever is the estimate obtained, there is a probability  $p_1$  that the integration error is within 25 per cent, a probability  $p_2$  that it is within 50 per cent, etc. The results of the monitoring should be interpreted accordingly by considering the risks associated with each of the probability ranges. Although this approach is unlikely to be informative if the monitoring is limited to one particular field or area, it may be effective for nationwide programmes where the corresponding probabilities can be obtained from comparison of the results between different fields, with the results of case studies on the given species, and with simulations.

### 6.3. Concluding remarks

The main conclusions from our study are as follows:

- numerical integration rules (in particular, Simpson rule) can provide a significantly more accurate estimate of the population size from sparse trap data than the standard statistical approach. Application of integration rules is not based on any assumption about the population distribution and remains effective even when the distribution exhibits a complex spatial structure;
- the minimum number of traps required to provide a robust estimate of the population size is found to be 25 (on a rectangular grid). Integration on a grid of  $5 \times 5$  traps keeps the error consistently within 25 per cent. This result shows only slight dependence on the spatial scale and hence is applicable to large areas as well as to small ones; and
- the use of a smaller grid of traps is possible but it may result in a paradigm shift. Integration on an ultra-coarse grid cannot provide an estimate of the population size with any prescribed accuracy because of the insufficient information (uncertainty) about the population distribution properties. Instead, the results of the integration can be treated probabilistically by considering the integration error as a random variable. Results of a large-scale ecological monitoring should be interpreted accordingly by considering additional risks associated with different accuracy ranges.

Our study leaves a number of open questions. On the theoretical side of our approach, for the methods considered in this paper, it is required that the values of the population density are known at the nodes of a rectangular grid, so that the traps' position must be chosen accordingly. It remains unclear how the accuracy of the integration may change if the traps are installed at arbitrary locations. Obviously, the power of our approach would increase considerably should we be able to place the

traps flexibly. However, the latter requires application of more advanced rules of numerical integration. A similar observation can be made with regards to the domain's geometry. Application of the methods considered here is restricted to a rectangular domain. In order to ensure a broad practical application of our approach, the integration rules should be generalized onto the case of an arbitrary field shape.

A related question is to what extent the integration accuracy may be affected if the traps are not laid exactly at the grid points but at some neighbouring position, e.g. as a result of environmental heterogeneity. Our preliminary analysis indicates that relatively small deviations (up to approx. 25% of the inter-trap distance) from a uniform grid do not seriously affect the integration error. A more detailed consideration of this problem will become a focus of a separate study.

On a more practical side, there can be some possible limitations that we do not take into account in this study but those can affect the application of our approach. In particular, spatial and temporal variations in trapping results may reflect not only variations in densities, but also errors connected to the trapping methods used (e.g. how efficient are the traps). The effect of such factors may call for a larger trapping effort than our theoretical study indicates. Future research should develop a more quantitative approach to address these issues.

The authors thankfully acknowledge instructive criticism from four anonymous reviewers. This study was partially supported by The Leverhulme Trust through grant F/00-568/X (to S.P.). S.P. also acknowledges support from the U.K. East Midlands Development Agency through Innovation Fellowship 'QuantiFly'.

## APPENDIX A. THE LEAST-SQUARES APPROXIMATION

Consider points  $P_i = (x_i, y_i)$ ,  $i = 1, \dots, N$  and let  $f_1, f_2, \dots, f_N$  be the values of the function  $f(x, y)$  at points  $P_1, P_2, \dots, P_N$ , respectively. We have to fit the data  $f_i$ ,  $i = 1, \dots, N$  to the function

$$u(x, y) = \sum_{m=1}^M u_m \phi_m(x, y), \quad M \leq N, \quad (\text{A } 1)$$

where  $\phi_m(x, y)$ ,  $m = 1, \dots, M$ , are basis functions, and the expansion coefficients  $(u_1, u_2, \dots, u_M)$  are considered as fitting parameters. The polynomial basis functions  $\phi_m(x, y)$  are given by

$$\phi_m(x, y) = (x - x_0)^{k_1} (y - y_0)^{k_2}, \quad (\text{A } 2)$$

where  $k_1 + k_2 = 0, 1, \dots, K$ , and the  $K$  defines the highest polynomial degree of the expansion (A 1). For any given  $K$ , the number of basis functions is determined as  $M = (K + 1)(K + 2)/2$ . The point  $P_0 = (x_0, y_0)$  is a fixed point in the domain  $D$ . We take  $P_0 = (0.5, 0.5)$  when the function  $f(x, y)$  is reconstructed in the unit square  $D = [0, 1] \times [0, 1]$ .

The least-squares approach considers the vector  $\mathbf{u} = (u_1, u_2, \dots, u_M)$  as the best fit to a given dataset,

if  $\mathbf{u}$  minimizes the following merit function  $F^2$ :

$$F^2 = \sum_{i=1}^N \left[ \frac{\left( f_i - \sum_{m=1}^M u_m \phi_m(P_i) \right)}{\sigma_i} \right]^2, \quad (\text{A } 3)$$

where  $\sigma_i$  are the weights of the method. Thus, the parameters  $u_m$  can be found from the  $M$  conditions

$$\frac{\partial F^2}{\partial u_m} = 0, \quad m = 1, \dots, M,$$

which are called the normal equations of the method. Taking into account the definition (A 3), we obtain the normal equations in the following form

$$\sum_{i=1}^N \frac{1}{\sigma_i^2} \left[ f_i - \sum_{j=1}^M u_j \phi_j(P_i) \right] \phi_m(P_i) = 0, \quad (\text{A } 4)$$

$$m = 1, \dots, M.$$

Introducing the weighted data  $\mathbf{b}$  and the design matrix  $[D]$  as

$$b_i = \frac{U_i}{\sigma_i} \quad \text{and} \quad D_{ij} = \frac{\phi_j(P_i)}{\sigma_i}, \quad i = 1, \dots, N,$$

$$j = 1, \dots, M,$$

the normal equations can be written as  $[A] \mathbf{u} = \mathbf{r}$ , where the matrix  $[A] = [D]^T [D]$ , and the right-hand side  $\mathbf{r} = [D]^T \mathbf{b}$ . They are to be solved for the vector of parameters  $\mathbf{u} = (u_1, \dots, u_M)$ ,

$$\mathbf{u} = [A]^{-1} \mathbf{r}. \quad (\text{A } 5)$$

## APPENDIX B. SIMULATION PARAMETERS AND THE INITIAL CONDITIONS

The population density snapshot shown in figure 2a is obtained at  $t = 50$  for parameters  $\beta = 3$ ,  $b = 0.28$ ,  $\Lambda = 0.5$ ,  $m = 0.48$ ,  $D = 10^{-6}$  and the initial conditions when both populations are distributed over the domain:

$$u(x, y, 0) = u_* + \varepsilon_{ux}x + \varepsilon_{uy}y \quad (\text{B } 1)$$

and

$$v(x, y, 0) = v_* + \varepsilon_{vx}x + \varepsilon_{vy}y, \quad (\text{B } 2)$$

where  $\varepsilon_{ux} = 0.007$ ,  $\varepsilon_{uy} = 0.008$ ,  $\varepsilon_{vx} = 0.008$  and  $\varepsilon_{vy} = -0.007$ .

Figures 2c,e and 3 are obtained for the initial conditions when both populations are concentrated around the place of the original introduction, e.g. around the domain centre:

$$u(x, y, 0) = u_0 \quad \text{if } x_{11} < x < x_{12} \quad \text{and} \\ y_{11} < y < y_{12}, \quad \text{otherwise } u(x, y, 0) = 0 \quad (\text{B } 3)$$

and

$$v(x, y, 0) = v_0 \quad \text{if } x_{21} < x < x_{22} \quad \text{and} \\ y_{21} < y < y_{22}, \quad \text{otherwise } v(x, y, 0) = 0, \quad (\text{B } 4)$$

where  $u_0$  and  $v_0$  are the initial prey and predator densities, respectively, and  $x_{ij}$ ,  $y_{ij}$  (where  $i = 1, 2$  and  $j = 1, 2$ ) are parameters with obvious meaning.

The snapshot shown in figure 2c is obtained at  $t = 800$  for  $m = 0.5$  (other parameters are the same as in figure 2a) and the initial conditions (B 3 and B 4) with  $x_{11} = 0.42$ ,  $x_{12} = 0.53$ ,  $y_{11} = 0.45$ ,  $y_{12} = 0.55$ ,  $x_{21} = 0.42$ ,  $x_{22} = 0.48$ ,  $y_{21} = 0.45$  and  $y_{22} = 0.51$ .

The snapshot shown in figure 2e is obtained at  $t = 1200$  for  $m = 0.45$  (other parameters are the same as in figure 2a,c) and the same initial conditions as in figure 2c.

Figure 3 is obtained for  $m = 0.414$  (other parameters are the same as in figure 2) and the same initial conditions as figure 2c,e. The snapshots are shown for (a)  $t = 450$ , (c)  $t = 1500$  and (e)  $t = 3500$ .

## REFERENCES

- Levin, S. A. (ed.) 2000 *Encyclopedia of biodiversity*. New York, NY: Elsevier.
- Cain, M. L., Bowman, W. D. & Hacker, S. D. 2011 *Ecology*, 2nd edn. Sunderland, MA: Sinauer.
- Pimentel, D. 2002 *Biological invasions: economic and environmental costs of alien plant, animal and microbe species*. New York, NY: CRC Press. (doi:10.1201/9781420041668)
- Liebholt, A. M., Elkinton, J. & Williams, D. 2000 What causes outbreaks of the gypsy moth in North America. *Popul. Ecol.* **42**, 257–266. (doi:10.1007/PL00012004)
- Scott, R. & Marrs, R. H. 1984 Impact of Japanese knotweed and methods of control. *Asp. Appl. Biol.* **11**, 291–296.
- Roblin, E. 1988 Chemical control of Japanese knotweed (*Reynoutria japonica*) on river banks in South Wales. *Asp. Appl. Biol.* **16**, 201–206.
- Ivanov, V. P., Kamakin, A. M., Ushivtzev, V. B., Shiganova, T. A., Zhukova, O., Aladin, N., Wilson, S. I., Harbison, R. & Dumont, H. J. 2000 Invasion of the Caspian Sea by the comb jellyfish *Mnemiopsis leidyi* (Ctenophora). *Biol. Invas.* **2**, 255–258. (doi:10.1023/A:1010098624728)
- Kideys, A. 2002 Fall and rise of the Black Sea ecosystem. *Science* **297**, 1482–1484. (doi:10.1126/science.1073002)
- Metcalfe, R. L. 1982 *Introduction to insect pest management*. London, UK: Wiley.
- Burn, A. J. 1987 *Integrated pest management*. London, UK: Academic Press.
- Stern, V. M. 1973 Economic thresholds. *Ann. Rev. Entomol.* **18**, 259–280. (doi:10.1146/annurev.en.18.010173.001355)
- Jepson, P. C. & Thacker, J. R. M. 1990 Analysis of the spatial component of pesticide side-effects on non-target invertebrate populations and its relevance to hazard analysis. *Funct. Ecol.* **4**, 349–355. (doi:10.2307/2389596)
- Byers, J. A., Anderbrant, O. & Lofqvist, J. 1989 Effective attraction radius: a method for comparing species attractants and determining densities of flying insects. *J. Chem. Ecol.* **15**, 749–765. (doi:10.1007/BF01014716)
- Raworth, D. A. & Choi, W.-J. 2001 Determining numbers of active carabid beetles per unit area from pitfall-trap data. *Entomol. Exp. Appl.* **98**, 95–108. (doi:10.1046/j.1570-7458.2001.00761.x)
- Davis, P. M. 1994 Statistics for describing populations. In *Handbook of sampling methods for arthropods in agriculture* (eds L. P. Pedigo & G. D. Buntin), pp. 33–54. Boca Raton, FL: CRC Press.



- 16 Snedecor, G. W. & Cochran, W. G. 1980 *Statistical methods*. Ames, IA: The Iowa State University Press.
- 17 Seber, G. A. 1982 *The estimation of animal abundance and related parameters*. London, UK: Charles Griffin.
- 18 Sutherland, W. J. (ed.) 1996 *Ecological census techniques: a handbook*. Cambridge, UK: Cambridge University Press.
- 19 Mayor, J. G. & Davies, M. H. 1976 A survey of leather-jacket populations in south-west England, 1963–1974. *Plant Pathol.* **25**, 121–128. (doi:10.1111/j.1365-3059.1976.tb01939.x)
- 20 Northing, P. 2009 Extensive field based aphid monitoring as an information tool for the UK seed potato industry. *Asp. Appl. Biol.* **94**, 31–34.
- 21 Taylor, L. R., Woiwod, I. P. & Perry, J. N. 1978 The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* **47**, 383–406. (doi:10.2307/3790)
- 22 Comins, H. N., Hassell, M. P. & May, R. M. 1992 The spatial dynamics of host-parasitoid systems. *J. Anim. Ecol.* **61**, 735–748. (doi:10.2307/5627)
- 23 Okubo, A. 1986 Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Adv. Biophys.* **22**, 1–94. (doi:10.1016/0065-227X(86)90003-1)
- 24 Malchow, H., Petrovskii, S. V. & Venturino, E. 2008 *Spatiotemporal patterns in ecology and epidemiology: theory, models, and simulations*. London, UK: Chapman and Hall/CRC Press.
- 25 Buntin, G. D. 1994 Developing a primary sampling program. In *Handbook of sampling methods for Arthropods in agriculture* (eds L. P. Pedigo & G. D. Buntin), pp. 99–115. Boca Raton, FL: CRC Press.
- 26 Perry, J. N. & Hewitt, M. 1991 A new index of aggregation for animal counts. *Biometrics* **47**, 1505–1518. (doi:10.2307/2532402)
- 27 Perry, J. N. 1996 Simulating spatial patterns of counts in agriculture and ecology. *Comput. Electron. Agric.* **15**, 93–109. (doi:10.1016/0168-1699(96)00005-1)
- 28 Turchin, P. 2003 *Complex population dynamics: a theoretical/empirical synthesis*. Princeton, NJ: Princeton University Press.
- 29 Petrovskaya, N. B. & Petrovskii, S. 2010 The coarse-grid problem in ecological monitoring. *Proc. R. Soc. A* **466**, 2933–2953. (doi:10.1098/rspa.2010.0023)
- 30 Petrovskaya, N. B. & Venturino, E. 2011 Numerical integration of sparsely sampled data. *Simulat. Model. Pract. Theory* **19**, 1860–1872. (doi:10.1016/j.simpat.2011.05.003)
- 31 Davis, P. J. & Rabinowitz, P. 1975 *Methods of numerical integration*. New York, NY: Academic Press.
- 32 Murray, J. D. 1989 *Mathematical biology*. Berlin, Germany: Springer.
- 33 Lewis, M. A. & Kareiva, P. 1993 Allee dynamics and the spread of invading organisms. *Theor. Popul. Biol.* **43**, 141–158. (doi:10.1006/tpbi.1993.1007)
- 34 Shigesada, N. & Kawasaki, K. 1997 *Biological invasions: theory and practice*. Oxford, UK: Oxford University Press.
- 35 Sherratt, J. A., Lewis, M. A. & Fowler, A. C. 1995 Ecological chaos in the wake of invasion. *Proc. Natl Acad. Sci. USA* **92**, 2524–2528. (doi:10.1073/pnas.92.7.2524)
- 36 Pascual, M. A. & Kareiva, P. 1996 Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology* **77**, 337–349. (doi:10.2307/2265613)
- 37 Sherratt, J. A. & Smith, M. 2008 Periodic travelling waves in cyclic populations: field studies and reaction diffusion models. *J. R. Soc. Interface* **5**, 483–505. (doi:10.1098/rsif.2007.1327)
- 38 Petrovskii, S. V., Morozov, A. Y. & Venturino, E. 2002 Allee effect makes possible patchy invasion in a predator–prey system. *Ecol. Lett.* **5**, 345–352. (doi:10.1046/j.1461-0248.2002.00324.x)
- 39 Petrovskii, S. V., Malchow, H., Hilker, F. M. & Venturino, E. 2005 Patterns of patchy spread in deterministic and stochastic models of biological invasion and biological control. *Biol. Invas.* **7**, 771–793. (doi:10.1007/s10530-005-5217-7)
- 40 Petrovskii, S. V. & McKay, K. 2010 Biological invasion and biological control: a case study of the gypsy moth spread. *Asp. Appl. Biol.* **104**, 37–48.
- 41 Murchie, A. K. & Harrison, A. J. 2004 Mark-recapture of ‘New Zealand flatworms’ in grassland in Northern Ireland. *Proc. Crop Protection in Northern Britain 2004*, 63–67.
- 42 Cannon, R. J. C., Baker, R. H. A., Taylor, M. C. & Moore, J. P. 1999 A review of the status of the New Zealand flatworm in the UK. *Ann. Appl. Biol.* **135**, 597–614. (doi:10.1111/j.1744-7348.1999.tb00892.x)
- 43 Boag, B. & Yeates, G. W. 2001 The potential impact of the New Zealand flatworm, a predator of earthworms, in Western Europe. *Ecol. Appl.* **11**, 1276–1286. (doi:10.1890/1051-0761(2001)011[1276:TPIOTN]2.0.CO;2)
- 44 Blackshaw, R. P. & Stewart, V. I. 1992 *Artioposthia triangulata* (Dendy 1894), a predatory terrestrial planarian and its potential impact on lumbricid earthworms. *Agric. Zool. Rev.* **5**, 201–219.
- 45 Boag, B., Neilson, R. & Jones, H. D. 2010 Quantifying the risk to biodiversity by alien terrestrial planarians. *Asp. Appl. Biol.* **104**, 55–61.
- 46 Murchie, A. K. 2010 Between two stools: dealing with the problem of the New Zealand flatworm. *Asp. Appl. Biol.* **104**, 73–78.
- 47 Moore, J. P., Dynes, C. & Murchie, A. K. 1998 Status and public perception of the ‘New Zealand flatworm’, *Artioposthia triangulata* (Dendy), in Northern Ireland. *Pedobiologia* **42**, 563–571.
- 48 Blackshaw, R. P., Moore, J. P. & Alston, R. 1996 Removal trapping to control *Artioposthia triangulata*. *Ann. Appl. Biol.* **129**, 355–360. (doi:10.1111/j.1744-7348.1996.tb05758.x)
- 49 Boag, B., Deeks, L., Orr, A. & Neilson, R. 2005 A spatio-temporal analysis of a New Zealand flatworm (*Arthurden-dyus triangulatus*) population in western Scotland. *Ann. Appl. Biol.* **147**, 81–88. (doi:10.1111/j.1744-7348.2005.00017.x)
- 50 Boag, B., Mackenzie, K., McNicol, J. W. & Neilson, R. 2010 Sampling for the New Zealand flatworm. *Proc. Crop Protection in Northern Britain 2010*, 45–50.
- 51 Holland, J. M., Perry, J. N. & Winder, L. 1999 The within-field spatial and temporal distribution of arthropods in winter wheat. *Bull. Entomol. Res.* **89**, 499–513. (doi:10.1017/S0007485399000656)
- 52 Alexander, C. J., Holland, J. M., Winder, L., Wooley, C. & Perry, J. N. 2005 Performance of sampling strategies in the presence of known spatial patterns. *Ann. Appl. Biol.* **146**, 361–370. (doi:10.1111/j.1744-7348.2005.040129.x)
- 53 Okubo, A. & Levin, S. 2001 *Diffusion and ecological problems: modern perspectives*. Berlin, Germany: Springer.
- 54 Hubbard, D. 2007 *How to measure anything: finding the value of intangibles in business*. New York, NY: John-Wiley Sons.
- 55 Thomas, C. F. G., Parkinson, L. & Marshall, E. J. P. 1998 Isolating the components of activity-density for the carabid beetle *Pterostichus melanarius* in farmland. *Oecologia* **116**, 103–112. (doi:10.1007/s004420050568)