

# Numerical Study of Pest Population Size at Various Diffusion Rates

Natalia Petrovskaya, Nina Embleton, Sergei Petrovskii

**Abstract** Estimation of the population size from spatially discrete sampling data is a routing task of ecological monitoring. This task may however become quite challenging in case the spatial data are sparse. The latter often happens in nationwide pest monitoring programs where the number of samples per field or area can be reduced, due to resource limitation and other reasons, to just a few. In this rather typical situation, the standard approaches become unreliable. Here we develop an alternative approach to obtain an estimate of the population size from sparse spatial data by considering numerical integration of the population density over a coarse grid. We first show that the species diffusivity is a controlling parameter that directly affects the complexity of the density distribution. We then obtain the conditions on the grid step size (i.e. the distance between two neighboring samples) allowing for the integration with a given accuracy at different diffusion rates. We consider how the accuracy of the population size estimate may change if the sampling positions are spaced non-uniformly. Finally, we discuss the implications of our findings for pest monitoring and control.

---

Natalia Petrovskaya

School of Mathematics, University of Birmingham, Birmingham, B15 2TT, U.K.,  
e-mail: n.b.petrovskaya@bham.ac.uk

Nina Embleton

School of Mathematics, University of Birmingham, Birmingham, B15 2TT, U.K.,  
e-mail: embleton@for.mat.bham.ac.uk

Sergei Petrovskii

Department of Mathematics, University of Leicester, Leicester, LE1 7RH, U.K.,  
e-mail: sp237@le.ac.uk

## 1 Introduction

Theoretical ecologists routinely operate with quantities like average population densities and/or population sizes, apparently assuming that they can be measured in the field with sufficient accuracy. Indeed, there is a variety of approaches to estimate the population size depending on the species taxonomy and biological traits [37]. However, it is almost never measured directly, e.g. by counting all the animals in a given field or forest. Much more typically, an estimate is obtained through collecting samples and their subsequent analysis, e.g. by using statistical methods [32]. The accuracy of the estimate then depends significantly on the number of samples. This has long been a focus of applied statistical analysis, yet there are some issues that remain rather poorly understood. The matter is that the focus of statistical methods have been more on calculating the variance in the sampling data (and on the relation between the variance and the mean [38]) rather than on the mean density itself.

The essence of the problem can be readily seen from the following example. Let  $u_0, \dots, u_{N-1}$  are the values of the population density of a given species obtained at the location of the samples  $\mathbf{r}_0, \dots, \mathbf{r}_{N-1}$ , respectively, where  $N$  is thus the number of samples. In order to obtain the average population density  $\bar{u}$  and/or the (total) population size  $I$  in an area  $A$ , this information must somehow be ‘integrated’ over the area. A commonly used statistical approach to estimate the population size is based on the arithmetic average [35]:

$$I \approx \tilde{I} = A\hat{u}, \quad \text{where} \quad \hat{u} = \frac{1}{N} \sum_{n=0}^{N-1} u_n \approx \bar{u}. \quad (1)$$

This approach works well when  $N$  is sufficiently large because the theory predicts that  $\hat{u}$  converges to  $\bar{u}$  when  $N$  tends to infinity. However, if  $N$  is not large, the application of Eq. (1) become questionable, especially when the density distribution is not spatially homogeneous but exhibits some form of aggregation. We want to mention it here that mathematically rigorous criteria assessing the minimum number of samples required to obtain a robust estimate of the population size are largely missing, and the decision about the optimum number of sampling locations is often made based on the intuition [3].

The crucial question for approach based on (1) is if the number of samples can always be made large enough to ensure that Eq. (1) is valid. Technically, in a particular scientific study, the number of samples can indeed be large (e.g. a few hundreds per an agricultural field), which seems to provide reliable information about the population size [1, 10]. However, the first observation we make here is that the idea ‘the more the better’ does not always work in case of sampling, e.g. because excessively large number of samples in a given area may have a disruptive effect on the behavior of the monitored animals, thus resulting in biased counts.

The second observation is that the situation becomes much worse when the information is required about the abundance of a pest species. The matter is that pest species are usually subject of nationwide or regional monitoring programs. This implies that the information is collected (i.e. samples are taken) simultaneously across

a large region. Due to resource limitation, it means that the number of samples then may become as small as just one or a few per unit area or per field [15]. Moreover, even under an idealized assumption of unlimited resources, a large number of samples in an agricultural field would hardly be possible anyway. Apart from the unpredictable potential effect on the monitored population mentioned above, sampling introduces a disturbance to agricultural procedures. The pest monitoring specialists would not be allowed to make this disturbance large as it can damage the agricultural product (e.g. crops) significantly, hence making the protective measures rather senseless.

Therefore, a challenge of pest monitoring is to obtain reliable information about the pest abundance from sparse spatial data, i.e. from a small number  $N$  of samples. In other words, we need to be able to obtain a robust estimate of the pest population size in the range of  $N$  where the application of Eq. (1) is likely to become unreliable.

In our study, we develop an alternative approach to address this challenge, basing on ideas different from that of the statistical analysis. Here we restrict our consideration to a hypothetical 1D case. (We mention it here that arranging the sample locations along a line sometimes appear to be more effective than arranging them on a 2D grid; e.g. see [1].) We start with the case when the sampling positions  $x_n$  ( $n = 0, \dots, N-1$ ) are equidistant, i.e.  $x_{n+1} = x_n + h$  where  $h > 0$  is constant. Equation (1) can then be written as

$$\tilde{I} = \sum_{n=0}^{N-1} u_n h \approx \int_a^b u(x) dx = I, \quad (2)$$

where  $h = L/(N-1)$ ,  $x_0 = a$ ,  $x_{N-1} = b$  and  $L = (b-a)$  is the size of the domain. It is readily seen that Eq. (2) coincides with the simplest method of numerical integration. This coincidence is not just by chance: a closer look at the problem shows that estimation of the population size based on the values of population density at discrete space (i.e. the position of the sampling points; see Fig. 1) is exactly the same as the general problem of numerical integration [19]. We therefore can make use of a vast variety of tools and methods of numerical integration accumulated in the field of numerical mathematics, e.g. see [7]. In section 1.1 below, we briefly revisit (to the extent required by the goals of this paper) the main ideas of numerical integration and reveal the problems that arise when we apply these ideas to the population size estimation from sparse spatial data.

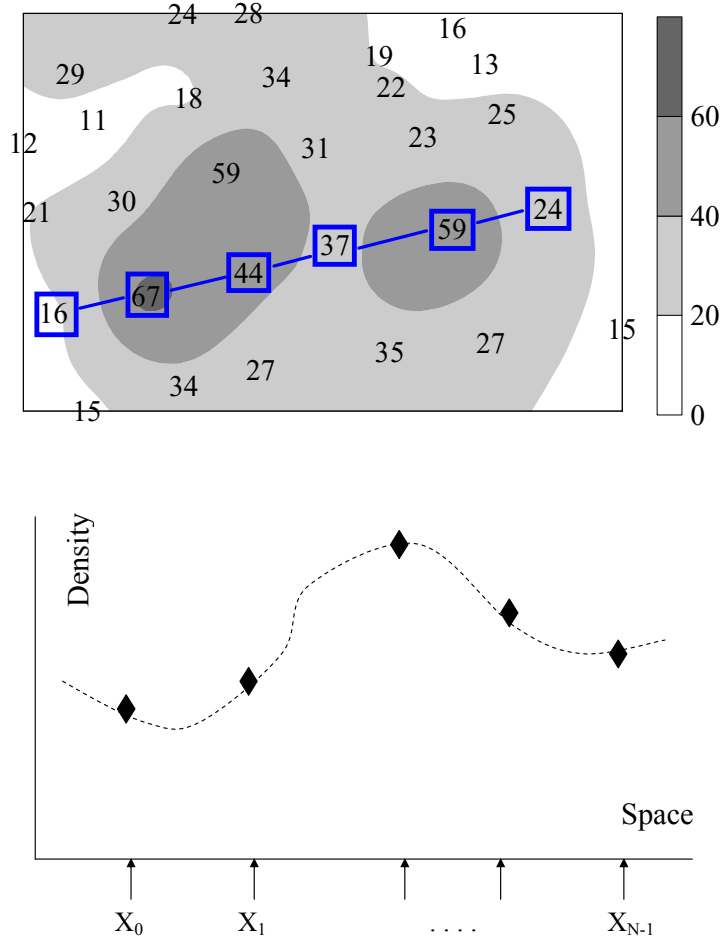
### 1.1 Numerical integration on coarse grids: The problem outline

A standard problem of numerical integration is to approximate the integral  $I$  by a sum  $\tilde{I}$ :

$$I = \int_a^b u(x) dx \approx \tilde{I}, \quad (3)$$

where the particular expression for the sum  $\tilde{I}$  depends on the choice of the integration rule; one option is given by Eq. (2), some more advanced options will be considered in section 2.1. In its turn, the change of integration to summation implies that, instead of the integrand  $u(x)$  defined on a continuous domain  $[a, b]$ , we are provided with a discrete set of values  $u(x_0), u(x_2), \dots, u(x_{N-1})$  (Fig. 1, bottom).

For any method of numerical integration, an essential requirement is that the approximation  $\tilde{I} \approx I$  should be accurate enough to meet the condition  $e \leq \varepsilon$ , where



**Fig. 1** (top) An example of field data collected in a field study on an insect pest [10], the numbers show the number of insect caught at the corresponding location in space, the boxed numbers show the samples along a transect; (bottom) a sketch of the numerical integration problem, the diamonds show the population density at the position  $X_0, \dots, X_{N-1}$  of the samples while the actual continuous density distribution (shown by the dashed curve) remains unknown.

$\varepsilon$  is the given tolerance and the integration error  $e$  is defined as

$$e = \frac{|I - \tilde{I}|}{|I|}. \quad (4)$$

In the below, we refer to the set of points  $x_n$ ,  $n = 0, 1, \dots, N-1$  in the domain  $[a, b]$  as a computational grid  $G$ . The location of the grid nodes is generally defined as  $x_{n+1} = x_n + h_n$ , where  $h_n > 0$  is the grid step size. The grid is called uniform if the grid step size is constant,  $h_n \equiv h = (b - a)/(N - 1)$ , and non-uniform otherwise. In ecological applications, the integrand function  $u(x)$  has the meaning of the density of the pest population, while the grid nodes  $x_n$  are the points where the samples are taken. Hence the density  $u(x)$  becomes a discrete function available at points  $x_n$  only (see Fig. 1).

The accuracy of ecological data is usually not very high and hence the error tolerance  $\varepsilon \sim 0.25 - 0.3$  is regarded as acceptable [17, 34]. However, even this relatively undemanding level of required accuracy cannot always be provided when the function  $\{u_n \equiv u(x_n), n = 0, \dots, N-1\}$  is integrated on a *coarse grid*, i.e. where the number  $N$  of nodes is small. The lack of information about the integrand function  $u(x)$  may lead to an inaccurate evaluation of the integral (3) and the numerical integration of sparse data may result in a large integration error.

Meanwhile, it has been shown in [19, 20] that an integral estimate  $\tilde{I}$  computed on coarse grids does not necessarily lie beyond the range of accuracy required in real-life ecological problems. The results obtained in [19, 20] show that the accuracy of integration on coarse grids is defined by the spatial heterogeneity of the integrand function. For instance, the examples considered in [19, 21] demonstrate that, when coarse grids are considered, numerical integration of a monotone function gives a considerably better accuracy than the integration of a function that has several ‘humps’ or oscillates rapidly. In its turn, the spatial structure of the population density  $u(x)$  (i.e. the integrand) is determined by several physical/biological parameters, in particular, by diffusion. In the next section, we consider the effect of diffusion in more detail.

## 1.2 Spatial heterogeneity and the effect of diffusion

We begin with a simple yet illuminating example when the population density distribution is described by the scalar diffusion equation, thus neglecting for the moment the impact of population multiplication and the interspecific interactions:

$$\frac{\partial u(x, t)}{\partial t} = D \frac{\partial^2 u}{\partial x^2}, \quad (5)$$

where  $D$  is the diffusion coefficient due to the self-movement of individuals [16].

For the purposes of this section, we consider a population in the unbounded domain,  $-\infty < x < \infty$ . The population density distribution over space, i.e. the solution

of Eq. (5), depends on the initial conditions. In case of a point-source release at a position  $x_0$ , it is well known that

$$u(x, t) = \frac{I}{\sqrt{4\pi Dt}} \exp\left(-\frac{(x-x_0)^2}{4Dt}\right). \quad (6)$$

It is readily seen that the characteristic width of the distribution (6), i.e. the characteristic length of the spatial heterogeneity, is given as

$$\Delta \sim \sqrt{Dt}, \quad (7)$$

where the sign  $\sim$  means ‘up to a constant coefficient’. The solutions of the diffusion equation obtained for some other ecologically sensible initial conditions, e.g. for a release over a finite domain, possess similar properties (see [23], section 9.3), i.e. the characteristic size of the arising spatial heterogeneity is given by (7). A more general approach based on the analysis of dimensions shows that this is, in fact, a generic property of the diffusion equation. Briefly, the matter is that the diffusion equation contains a single parameter, the diffusion coefficient  $D$ , and its dimension is  $\text{distance}^2 \cdot \text{time}^{-1}$ . Therefore, for any given time  $t$ , the only quantity with the dimension of length is  $\sqrt{Dt}$ ; see [2] for more details.

The next level of complexity is a single-species model with multiplication, i.e. a diffusion-reaction equation. Consider a particular case when reproduction is described by the logistic function:

$$\frac{\partial u(x, t)}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + \alpha u \left(1 - \frac{u}{K}\right), \quad (8)$$

where  $\alpha$  is the per capita growth rate and  $K$  is the carrying capacity. The dimension of  $\alpha$  is  $\text{time}^{-1}$  and hence the only way to create a quantity with the dimension of length from the parameters of Eq. (8) is

$$\Delta_{fr} \sim \sqrt{D/\alpha}. \quad (9)$$

For a wide class of initial conditions, in the large-time limit Eq. (8) describes a travelling front [14] and then  $\Delta_{fr}$  gives the characteristic length of the system’s spatial heterogeneity, i.e. the width of the front.

In case of multi-species systems, e.g. as described by a system of diffusion-reaction equations, application of the dimensions analysis is less instructive as such systems contain more than one parameter with the dimension of time or inverse time, and more than one diffusion coefficient. However, there are some alternative approaches. Let us assume that all diffusion coefficients have the same value  $D$ . Consider the case when the corresponding non-spatial system has a unique positive state and this state is as an unstable focus. In this case, the system is known to develop complex, chaotic spatiotemporal pattern sometimes referred to as the “biological turbulence.” [13]. The characteristic length  $\Delta_g$  of the emerging multi-hump spatiotemporal pattern, i.e. the width of a single hump, is then given as [24]

$$\Delta_g = 2\pi c^* \left( \frac{D}{\max \operatorname{Re}(\lambda)} \right)^{1/2}, \quad (10)$$

where  $\max \operatorname{Re}(\lambda)$  is the maximum real part of the eigenvalues of the linearized system and  $c^*$  is a numerical coefficient of the order of unity. Note that, since  $\max \operatorname{Re}(\lambda)$  has the dimension of time, Eq. (10) is in a good agreement with the dimensions analysis; in fact, it can be regarded as a generalization of Eq. (9).

An observation important for our analysis is that, in all three cases (7), (9) and (10) the characteristic length of the spatial heterogeneity is proportional to  $\sqrt{D}$ , i.e.

$$\Delta_g = \omega \sqrt{D}, \quad (11)$$

where  $\omega$  is a factor that can depend on the parameters of the intra- and interspecific interactions, but not on the diffusion coefficient.

### 1.3 Goals and the road map

The main goal of this study is to evaluate the effectiveness of the methods of numerical integration as a possible tool to obtain a reliable estimate of the population size from spatially discrete sampling data. We are especially interested in the case when the values of the population density are only available on a coarse grid, i.e. when the samples are taken only at a few spatial locations. This is a typical situation in pest monitoring programs.

A straightforward approach to increase the integration accuracy is to make the number of nodes in the computational grid sufficiently large in order to resolve the spatial heterogeneity. However, simulations show that, in practice, an acceptable accuracy can be obtained with a much smaller number of nodes, even in case of an ‘extreme aggregation’ when a peak of the population density may fall almost completely in between of two subsequent nodes. In this paper, we make a more quantitative insight into this problem. In particular, we relate this issue to the species diffusivity. Since the theory predicts that the hump width depends on the diffusivity, we can evaluate the grid step size required to accurately integrate a complex spatial pattern in terms of the diffusion coefficient, assuming that the latter is known as a biological trait of the given pest species.

The second issue is the impact of the spatial structure of the grid itself. The previous analysis was done under condition that the grid is uniform. However, this situation is hardly realistic as a certain variation in the samples location is inevitable, even in a very carefully designed field study. An important question therefore is how the accuracy of integration may be affected by the effects of non-uniformness.

The paper is organized as follows. In the next section, we introduce a population dynamics model that we use to generate ecologically meaningful population distributions for various diffusion rates. We then describe the numerical integration method designed to evaluate the population size and apply it to spatial population

distributions of different complexity. In section 3, we perform a detailed mathematical analysis of the impact of the diffusion rates on the accuracy of numerical integration on a coarse grid. In section 4, we investigate the effect of the grid's non-uniformness on the population size estimation. Finally, in section 5 we summarize our findings and discuss their implications for the pest monitoring practices.

## 2 The insight from the ecological model

In order to assess the effectiveness of our approach to integrate discrete sampling data, we now need data. Note that, to make a sensible assessment, we need to know not only the values of the population density arranged along a line (e.g. see Fig. 1, top) but also the actual population size to compare our estimate to. However, field data satisfying this requirement are rarely available. Moreover, to study the effect of the grid step size (i.e. the effect of different sample spacing) on the accuracy of the estimate, we need to compare the results obtained on different grids, which is almost impossible to obtain in the field.

For the above reasons, instead of field data, here we use the population density distribution generated by an ecological model. Specifically, we use the spatially explicit Rosenzweig–MacArthur model which, in dimensionless variables, has the following form [14]:

$$\frac{\partial u(x,t)}{\partial t} = d \frac{\partial^2 u}{\partial x^2} + u(1-u) - \frac{uv}{u+h}, \quad (12)$$

$$\frac{\partial v(x,t)}{\partial t} = d \frac{\partial^2 v}{\partial x^2} + k \frac{uv}{u+h} - mv. \quad (13)$$

Here  $u$  and  $v$  are the dimensionless densities of prey and predator, respectively, at time  $t$  and position  $x$  where  $t > 0$  and  $0 < x < 1$ . The distances are therefore measured in fractions of the original domain length  $L$ . (See [13, 19] for more details with regard to the choice of the dimensionless variables and parameters.) The dimensionless diffusion coefficient  $d$  quantifies the species diffusivity due to the movement of the individuals. For the sake of simplicity, we assume it to be the same for both species.

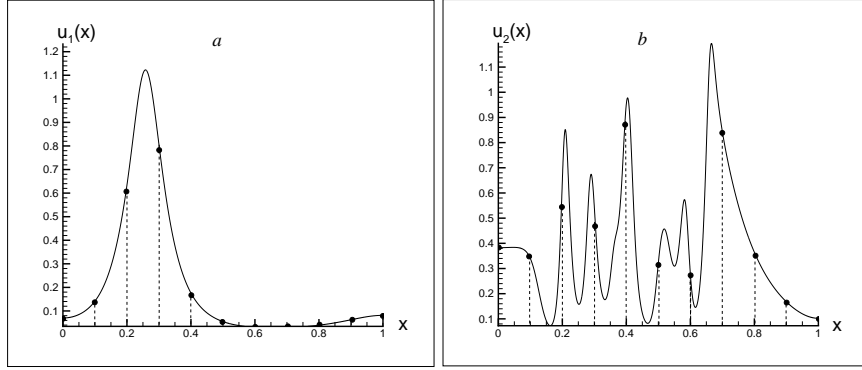
It is readily seen that the relation between the diffusion coefficient and the characteristic length of the system's spatial heterogeneity remains exactly the same as it was in the dimensional units, i.e.

$$\delta_g = \frac{\Delta_g}{L} = \omega \sqrt{d}. \quad (14)$$

Here the coefficient  $\omega$  depends on the system's parameters, cf. Eq. (10). However, an extensive numerical study performed in [24, 25] revealed that, in the predator-prey system (12–13), its value is relatively robust to changes in the parameter values, typically being about 25.



An important feature of the Eqs. (12–13) is that interaction between reaction and diffusion is known [13] to result in pattern formation, e.g. see Fig. 2, where the properties of the pattern<sup>1</sup> depend on the value of dimensionless diffusivity  $d$ . In particular, for  $d$  being on the order of 1 or larger, the solution  $u(x, t)$  will be a monotone function of  $x$ , which means that the local population oscillations are almost synchronized over the entire domain. However, oscillations at different positions can become de-synchronized for  $d \ll 1$  (see [24]). In the latter case the initial conditions  $u(x, 0)$ ,  $v(x, 0)$  evolve to an ensemble of irregular humps and hollows. For an intermediate value of  $d$ , the pattern can consist of just one or a few peaks only (see Fig. 2a), while the number of humps increases for smaller values of  $d$  resulting in oscillations shown in Fig. 2b. From an ecological perspective, it means that in a domain of a given length a slowly diffusing population is more likely to form a spatial pattern than a fast one.



**Fig. 2** Ecological test cases. Typical spatial distribution of the pest population density in the model (12–13) for the values of the dimensionless diffusivity  $d = 10^{-4}$  (a) and  $d = 10^{-5}$  (b). The continuous functions  $u_1(x)$  and  $u_2(x)$  are presented by solid lines, while the function values available for integration on a coarse grid are shown as black filled circles.

From a numerical viewpoint, the above conclusion means that one may expect lower accuracy of integration when the size of slowly diffusing pest population is evaluated. On the contrary, the complex spatial structure of the population density of a fast diffusing pest may be not well resolved on coarse grids because we have to deal with an oscillating integrand function  $u(x)$ . Thus our next step is to compute the integration error for the functions shown in Fig. 2 to establish the link between spatial heterogeneity of the integrand function and the accuracy of numerical integration when the number of grid nodes is small.

<sup>1</sup> At least, for any  $t$  not too small, in order to avoid the effect of the initial conditions.

## 2.1 The method of numerical integration

In this subsection we briefly discuss a method that we use for numerical integration of a discrete function  $u_n$ ,  $n = 0, \dots, N-1$ . The numerical integration technique we employ in order to compute the integral (3) requires a user to replace the integrand function  $u(x)$  at each grid subinterval  $c_n = [x_n, x_{n+1}]$ ,  $n = 0, \dots, N-2$ , by a local polynomial of degree  $K$ ,

$$p_K^n(x) = \sum_{k=0}^K a_{kn} x^k,$$

where the expansion coefficients  $a_{kn}$  are reconstructed independently at each subinterval  $c_n$  (as it is indicated by the subscript  $n$  in their notation). The integral (3) is then evaluated as

$$I = \int_a^b u(x) dx \approx \sum_{n=0}^{N-2} I_n, \quad (15)$$

where the integral  $I_n$  is readily computed over the grid cell  $c_n$  as  $I_n = \int_{x_n}^{x_{n+1}} p_K^n(x) dx$ .

The details of the implementation of the composite integration rule (15) can be found in [19, 21]. Let us note here that the numerical technique we use in the problem is the same as the Newton-Cotes family of methods of numerical integration [7] if uniform grids ( $h_n \equiv h = (b-a)/(N-1)$ ) are considered. In particular, the polynomial degrees  $K=0$ ,  $K=1$  and  $K=2$  correspond to the well-known methods of numerical integration such as the midpoint rule, the trapezoidal rule and the Simpson rule, respectively. These are the first three methods from the Newton-Cotes family. However, our approach is more flexible as it allows one to deal with non-uniform grids where the grid step size  $h_n \neq \text{const}$ .

One important observation about the integral evaluation is that asymptotic error estimates for the approximation (15) will depend on the polynomial degree  $K$ . It has been shown in [19, 21] that the integration error (4) can be evaluated on uniform grids as

$$e = Ch^{K+1}, \quad (16)$$

where  $C = \text{const}$ . The estimate (16), however, only holds on fine grids where the grid step size  $h$  is very small (ultimately, tends to zero). Meanwhile, we have demonstrated in our previous work [19, 20] that the asymptotic error estimates do not hold on coarse grids where, generally speaking, we cannot reduce the integration error by using higher order polynomials. Hence, other ways of controlling the accuracy of integration have to be established when one has to deal with coarse grids, where  $N$  is small, and in the following subsection we discuss this problem.

## 2.2 Estimating the pest population size when the number of traps is small: numerical test cases

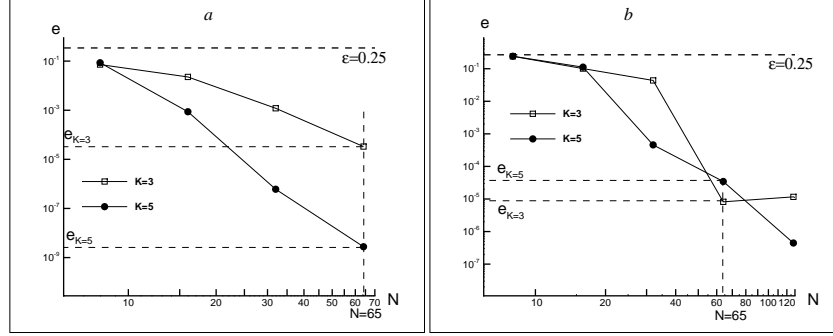
The discussion of the accuracy of numerical integration requires us to understand how to actually compute the integration error for the ecological distributions of Fig. 2. The problem is that the computation of the integration error (4) is based on the knowledge of the exact answer. However, the analytical solution of the system (12–13) is not known and we cannot compute the integral  $I$  as required by the definition (4). Therefore, we have to define the ‘exact’ value of the integral when the pest population density represented by the function  $u(x)$  is integrated. For this purpose we compute a numerical solution to the system (12–13) on a very fine uniform grid  $G_f$  of  $N_f = 2^{15} + 1 \equiv 32769$  nodes, and we consider the result as the ‘exact’ solution to the problem. The corresponding value of the solution integral  $I$  is then considered as the exact integral that is compared with the value of the integral on a given coarse grid  $G_c$ .

For the purpose of our study, we are going to compute the integration error as a function of the number of grid nodes,  $e = e(N)$ , as we want to understand what happens to the approximation  $\tilde{I}$  when we increase or decrease the number of grid nodes. A usual technique to generate a finer uniform grid from a coarser one is to halve each grid subinterval by inserting a new node at the subinterval midpoint. Let us denote the number of grid subintervals as  $\hat{N}$ , where we have  $\hat{N} = N - 1$ . We generate a sequence of uniform grids, where the number of subintervals on each grid is defined as  $\hat{N} = s\hat{N}_0$ . The number  $\hat{N}_0$  of grid subintervals on the initial grid is taken  $\hat{N}_0 = 8$  and the scaling coefficient  $s$  varies as  $s = 2^m, m = 0, 1, 2, \dots, 12$ . The integrand function  $u(x)$  is then readily available at nodes of each grid generated as above, as we simply project it from the fine grid  $G_f$  where it has originally been computed. Hence the integration error (4) can be easily defined on any grid in the sequence to obtain the convergence rate  $e(N)$  of our numerical method.

Let us refer to the density distributions shown in Fig. 2a and Fig. 2b as  $u_1(x)$  and  $u_2(x)$  respectively. We integrate  $u_1(x)$  and  $u_2(x)$  and compute the error (4) on each uniform grid generated as above. The integration error as a function of the number  $N$  of grid nodes for the integrand function  $u_1(x)$  is shown in Fig. 3a, while the error for the function  $u_2(x)$  is displayed in Fig. 3b. The integration error is shown on a logarithmic scale. In both cases the error is computed for approximation by polynomials of degree  $K = 3$  and  $K = 5$ .

It is readily seen from the figure that the behavior of the error curve depends on the integrand function. For the function  $u_1(x)$  the convergence results are in a good agreement with the error estimate (16). Namely, the polynomial approximation with  $K = 5$  provides better accuracy if we compare it with the  $K = 3$  approximation on each grid in the sequence; see Fig. 3a. The integration error is always within the required range  $e < 0.25$ , as we already have  $e \approx 0.1$  on the initial grid of  $N_0 = 9$  nodes. Hence the coarse initial grid has a sufficient number of nodes to provide an accurate estimate of the integral in case that the distribution  $u_1(x)$  is considered.

Meanwhile, for the function  $u_2(x)$ , the use of higher order polynomials to approximate the integrand does not always result in a more accurate approximation on coarse grids. It can be seen from Fig. 3b that the integration error of a higher order polynomial approximation ( $K = 5$ ) remains about the same as the error of  $K = 3$  approximation. Moreover, the error  $e_{K=5}$  can even be greater than  $e_{K=3}$  as it is shown in Fig. 3b for a grid of  $N = 65$  nodes. As we have already discussed, the complex multi-peak spatial pattern  $u_2(x)$  may require a finer grid to resolve function's spatial oscillations. Also, despite the initial grid of  $N_0 = 9$  nodes still providing the accuracy acceptable for ecological applications, the error  $e \approx 0.25$  is essentially bigger in comparison with the integration error obtained on the same grid for the integrand function  $u_1(x)$ .



**Fig. 3** The integration error (4) as a function of the number of grid nodes for ecological test cases shown in Fig. 2. (a) The density distribution  $u_1(x)$ . The error  $e_{K=3}$  of the approximation by polynomials of degree  $K = 3$  remains always bigger than the error  $e_{K=5}$  computed for the approximation by polynomials of degree  $K = 5$ . (b) The density distribution  $u_2(x)$ . The approximation by high order polynomials ( $K = 5$ ) cannot always provide better accuracy.

The above examples demonstrate that, while it is sufficient to have a grid of several nodes in order to provide accurate integration results for a simple spatial distribution, the same number of grid nodes may give the accuracy beyond the acceptable range if a more complex spatial pattern is considered. The error behavior, when the convergence rate does not follow its asymptotic value (16), is called “a coarse grid problem” [19, 21]. Since the error cannot be controlled based on the estimate (16), it becomes extremely important to understand what factors determine the error on coarse grids where the spatial structure of the integrand function is not well resolved. That will be done in the next section.

### 3 The impact of the diffusion rates on the accuracy of numerical integration

In this section we derive the functional relationship between the diffusion coefficient and the grid step size required to provide a good integration accuracy. Our previous discussion revealed that using higher order polynomials to approximate the integrand function does not necessarily result in a more accurate estimate of the integral on coarse grids. Hence we now reduce our attention to a technically simple yet illuminating case when the integrand function is approximated by linear polynomials ( $K = 1$ ).

Let a nonnegative function  $u(x)$  have a ‘hump’ (i.e., a local maximum) at the interval  $[0, 1]$ . The first assumption we make for our analysis is that the hump can be handled as a quadratic function. Namely, let us introduce the subinterval  $[x_0, x_2]$  of the length  $2h$  in the vicinity of the hump<sup>2</sup> (see Fig. 4). We then assume that in the vicinity of the hump the integrand  $u(x)$  can be considered as

$$u(x) \approx g(x) = B - A(x - x_1)^2, \quad x \in [x_0, x_2],$$

where  $A > 0$ ,  $B > 0$  and the function  $g(x)$  has the maximum at the interval midpoint  $x_1 = x_0 + h$ . We also require  $g(x)$  to be a nonnegative function over the interval  $[x_0, x_2]$ , that is  $g(x_0) = g(x_2) = B - Ah^2 > 0$ . That gives us the following condition relating  $A$ ,  $B$  and  $h$ :

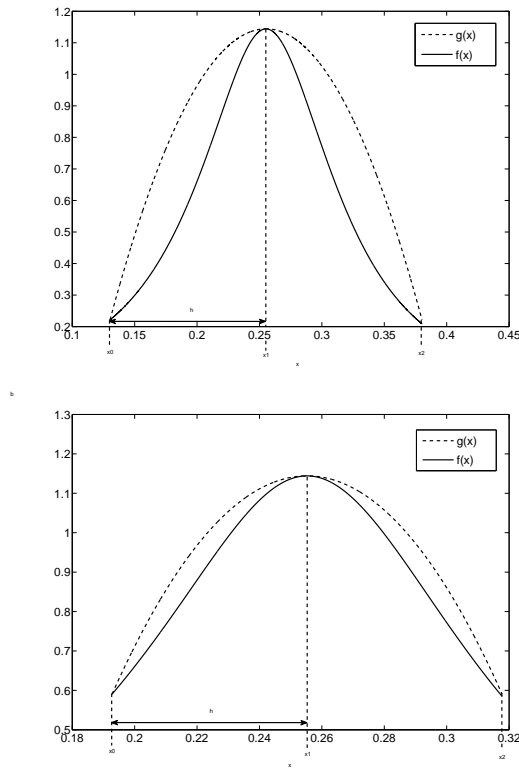
$$h^2 < \frac{B}{A}. \quad (17)$$

The examples of the approximation of a hump by a quadratic function for the pest population density  $u_1(x)$  are shown in Fig. 4, where various choice of points  $x_0$ ,  $x_1$  and  $x_2$  in the vicinity of the hump is illustrated. The details of such approximation can be found in the Appendix.

It is obvious that we introduce an additional error to the integration problem when we tackle a hump as a quadratic function (see the discussion in Appendix). However, as we will see below, such approximation enable us to make correct conclusions about a grid step size that should be recommended for accurate integration of the function  $u(x)$ . Thus our next step is to investigate what happens when we replace the quadratic function  $g(x)$  (and, therefore, the original function  $u(x)$ ) with two linear polynomials in the vicinity of the hump, as our method of numerical integration requires us to do. We first consider a uniform grid where one of grid nodes is located at the maximum point. We then study the case of an arbitrary location of a maximum point on a coarse uniform grid. Finally, we discuss non-uniform grids to understand what impact the grid distortion will make on the integration error.

---

<sup>2</sup> Note that the notation  $x_0, x_1, x_2$  we use to discuss the hump approximation is not the same as the numeration of grid nodes we introduced in the previous section. In other words, the ‘endpoints’  $x_0$  and  $x_2$  are arbitrarily located interior points of the interval  $[0, 1]$ .



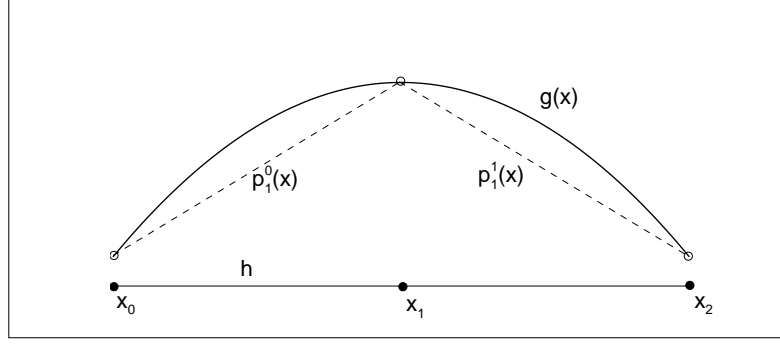
**Fig. 4** Hump approximation for the population density distribution  $u_1(x)$ . (a) Quadratic approximation with  $h = 0.125$ . (b) Quadratic approximation with  $h = 0.0625$ .

### 3.1 Uniform grid

Let the quadratic function  $g(x)$  be integrated at the interval  $[x_0, x_2]$  where we consider a local grid of two subintervals  $c_0 = [x_0, x_1]$  and  $c_1 = [x_1, x_2]$ , the node location being  $x_1 = x_0 + h$  and  $x_2 = x_1 + h$ . We use linear polynomials  $p_1^n(x) = \sum_{k=0}^1 a_{kn}x^k$  at each grid cell  $c_n$ ,  $n = 0, 1$ , where we reconstruct polynomial coefficients from the condition  $p_1^n(x_k) = g(x_k)$ ,  $k = 0, 1, 2$ , as  $p_1^0(x) = B + Ah(x_1 - x)$  and  $p_1^1(x) = B - Ah(x - x_1)$ . The approximation of a quadratic function by linear polynomials over a grid of the two subintervals is illustrated in Fig. 5.

We then compute the approximate integral  $\tilde{I}$  as

$$\tilde{I} = \int_{x_0}^{x_1} p_1^0(x) dx + \int_{x_1}^{x_2} p_1^1(x) dx = 2Bh - Ah^3,$$



**Fig. 5** Approximation of a quadratic function by linear polynomials over a uniform grid of 3 nodes.

while the exact integral is

$$I = \int_{x_0}^{x_2} g(x) dx = 2Bh - \frac{2Ah^3}{3}. \quad (18)$$

Consider the error of integration (4) and let us require that  $e < 0.25$ . Correspondingly, we have  $|I - \tilde{I}| = \frac{Ah^3}{3}$ . Therefore, we obtain:

$$\frac{Ah^3}{3} < \frac{1}{4} \left| 2Bh - \frac{2Ah^3}{3} \right|. \quad (19)$$

Solving (17) and (19) together and taking also into account that  $I > 0$  (as  $B > 0$  and  $g(x) \geq 0$  for any  $x \in [x_0, x_2]$ ), we obtain  $h < h_0 = \sqrt{B/A}$ .

In order to reveal the impact of diffusion, we now define the ‘hump width’  $\delta_g$  of the quadratic function  $g(x)$  as the distance between its roots, so that  $\delta_g = 2\sqrt{B/A}$ . Correspondingly, we obtain that the required accuracy  $e < 0.25$  is ensured for  $h < h_0 = \delta_g/2$ . Finally, recalling that the characteristic length of the spatial heterogeneity is given by Eq. (14) and substituting it into the expression above, we arrive at

$$h < h_0 = \frac{\omega\sqrt{d}}{2}. \quad (20)$$

Whatever is the value of the diffusion coefficient  $d$ , condition (20) is sufficient to integrate the ‘hump’ with the desirable accuracy  $e < 0.25$ . The limiting value  $e = 0.25$  is reached for  $h_0$ .

### 3.2 The analysis of the grid step size for ecological distributions

In this subsection we validate our findings – in particular, condition (20) – by considering the density distributions shown in Fig. 2. We first study the function  $u_1(x)$  that has a single hump. The aim of our numerical test is to find the number  $N^*$  of grid nodes sufficient for accurate integration of the pest population density  $u_1(x)$ . In other words, we integrate the function  $u_1(x)$  over the domain  $[0, 1]$  on a sequence of uniform grids and compute the corresponding integration error (4). We then look for the grid step size  $h^*$  whose value provides us with the integration error  $e < 0.25$ . That should give us the number  $N^* \approx 1/h^*$  of grid nodes (or the number of samples in the pest monitoring problem) required to resolve the spatial heterogeneity. The number  $N^*$  (or the grid step size  $h^*$ ) obtained in this straightforward integration procedure is then compared with the estimate (20).

Let us note it again that the hump itself is not, of course, a quadratic function and the integration error obtained for the integrand  $u_1(x)$  is not the same as the integration error derived for the quadratic function. However, since a single hump can be approximated by a quadratic function with good accuracy (see Appendix), we expect that the results of our numerical experiment will be in reasonable agreement with the estimate (20) that can be obtained from the information about the diffusion coefficient only.

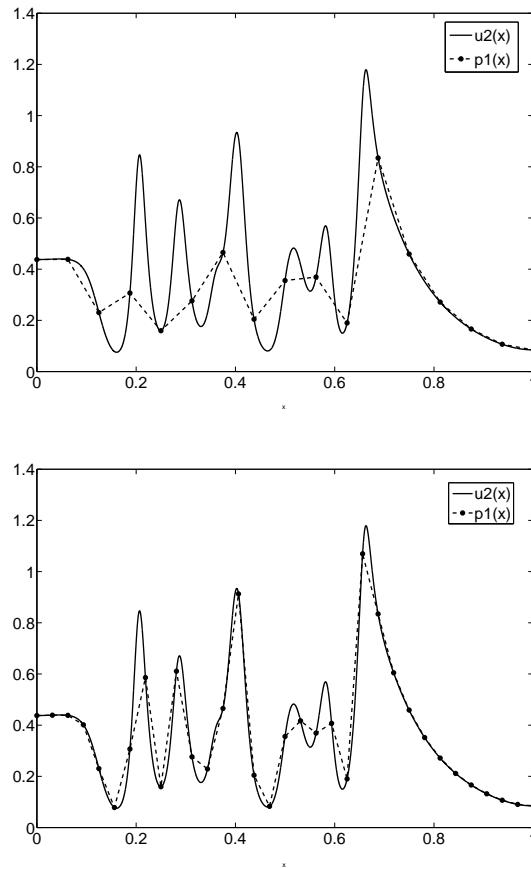
The estimate (20) gives us the value  $h^* \sim 0.12$  for the diffusion coefficient  $d = 10^{-4}$  used to generate the density distribution  $u_1(x)$ . The integration error when the function  $u_1(x)$  is approximated by linear polynomials over a uniform grid of  $N$  nodes is shown in Table 2. We compute the error (4) on a very coarse grid of 2 subintervals, we then refine the grid by halving each grid subinterval, compute the integral error again and repeat the refinement procedure until the error is smaller than the threshold value  $e = 0.25$ . It can be seen from the table that the results of numerical integration are in good agreement with our estimate (20). While a very coarse grid does not provide the accuracy  $e \leq 0.25$ , the grid of  $N = 9$  nodes ( $h = 0.125$ ) gives the integration error much smaller than the required limit  $e = 0.25$ . Hence the number of grid nodes can be evaluated as  $N^* \approx 9$ .

N	3	5	9	17
h	0.5	0.25	0.125	0.0625
e	0.6948	0.5459	0.0823	0.0036

**Table 1** The integration error (4) for the density distribution  $u_1(x)$  on a sequence of uniformly refined grids with grid step size  $h$ . The integrand function  $u_1(x)$  is approximated by piecewise linear polynomials ( $K = 1$ ) on each grid in the sequence.

Consider now the density distribution  $u_2(x)$  shown in Fig. 2b. The diffusion coefficient used to generate the distribution  $u_2(x)$  is  $d = 10^{-5}$ . Hence the estimate (20) of the grid step size is  $h \sim 0.03$ . In other words, a uniform grid should contain about 30 nodes in order to guarantee the integration error  $e < 0.25$ .





**Fig. 6** Piecewise linear approximation for the population density distribution  $u_2(x)$ . (a) Approximation using  $N = 17$  grid nodes (b) Approximation using  $N = 33$  nodes.

The values of the integration error (4) are shown in Table 3. A substantial jump in accuracy is evident when the grid is refined from  $N = 17$  to  $N = 33$  nodes. This is further illustrated by Fig. 6 where it can be seen that for 16 subintervals, the majority of the humps in  $u_2(x)$  are approximated by a single polynomial and the spatial heterogeneity is not well resolved. When the grid is refined to 32 subintervals, all but the two of the humps are approximated by two or more linear polynomials. Approximating a hump with a single linear polynomial is equivalent to the approximation of a quadratic by linear polynomials over a local grid of two nodes instead of considering three nodes for the approximation. That extreme case will be discussed in more detail in the next section.

At the same time it is worth noting here that for the density distribution  $u_2(x)$  the integration error is not entirely the same as expected from our analysis, as the

integration error actually remains within the required range  $e < 0.25$  on any grid that we use in our computations. We believe that this may happen because of the ‘cancelation effect’ that may arise when underestimated contribution of the humps is balanced by overestimated contribution of the hollows. However, we would like to emphasize that the error value cannot be predicted on coarse grids. In other words, while the estimate (20) guarantees the error  $e < 0.25$  on a grid of  $N = 33$  nodes, it cannot be said a priori what the error is on coarse grids with  $N < 33$ .

N	3	5	9	17	33
h	0.5	0.25	0.125	0.0625	0.03125
e	0.1579	0.1567	0.2193	0.1304	0.0001

**Table 2** The integration error for the density distribution  $u_2(x)$  on a sequence of uniformly refined grids. The integrand function  $u_2(x)$  is approximated by piecewise linear polynomials, see Fig.6.

### 3.3 Arbitrary location of the peak on a uniform coarse grid

In the previous subsection we assumed that there are three grid nodes in the region of the hump and the position of the central node coincides with the position of the maximum. Especially the last assumption is not entirely realistic because in applications to pest monitoring the position of the population density would usually be unknown. Hence, two practically important questions that arise from our analysis above are (i) how the integration error changes when the maximum is not at the position of the node (see Fig. 7) and (ii) whether we can make the grid even coarser, e.g. what will be the integration accuracy if just one grid node is used in the subdomain where the hump is located. In other words, we are now interested in the situation given by  $N = 3$  and  $N = 5$  in Table 3.2 when the entire hump is located in between two grid nodes. The error shown in Table 3.2 is quite large, but can we possibly make it any smaller with the same number of nodes?

Consider a regular grid consisting of three nodes,  $x_0, x_1 = x_0 + h$  and  $x_2 = x_0 + 2h$ . Let a population density distribution have a hump within the interval  $[x_0, x_2]$ . We approximate the hump by a quadratic function. Let us define the approximation  $g(x)$  of the hump as

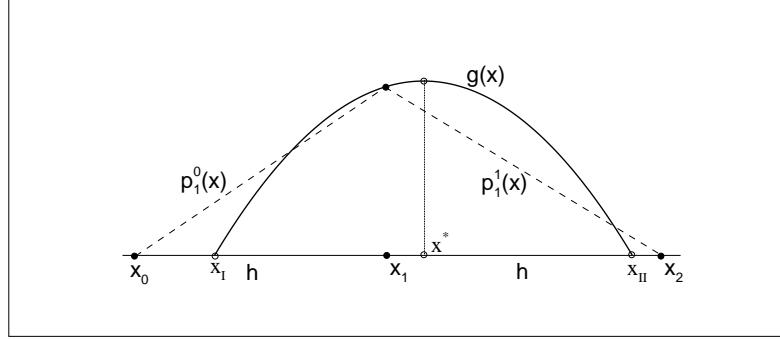
$$g(x) = \begin{cases} B - A(x - x^*)^2, & \text{if } x \in [x_I, x_{II}], \\ 0, & \text{otherwise.} \end{cases}$$

In the approximation above  $x^*$  is the location of the maximum point, which is now different from the node  $x_1$ , and the values  $x_I$  and  $x_{II}$  are the roots of  $g(x)$ . We can express  $x^*$  in terms of the grid nodes as

$$x^* = x_I + \gamma h = x_0 + h(\gamma + 1),$$

where  $\gamma \in [0, 1]$ . The roots  $x_I$  and  $x_{II}$  are then given by

$$x_I = x_0 + h(\gamma + 1) - \sqrt{B/A} \quad \text{and} \quad x_{II} = x_0 + h(\gamma + 1) + \sqrt{B/A}.$$



**Fig. 7** Piecewise linear approximation of  $g(x)$ .

The exact integral of  $g(x)$  in the vicinity of the hump is thus

$$I = \int_{x_0}^{x_2} g(x) dx \equiv \int_{x_I}^{x_{II}} g(x) dx = \frac{2}{3} B \delta_g, \quad (21)$$

where  $\delta_g$  is the hump width as above. We now approximate  $g(x)$  by two piecewise linear polynomials as follows (see Fig. 7)

$$g(x) \approx \begin{cases} p_1^0(x), & \text{if } x \in [x_0, x_1], \\ p_1^1(x), & \text{if } x \in [x_1, x_2]. \end{cases}$$

An approximated value  $\tilde{I}$  of the integral (22) is then obtained by integrating the piecewise linear approximation of the function  $g(x)$ :

$$\tilde{I} = \sum_{k=0}^{k=1} \left( \int_{x_k}^{x^{k+1}} p_1^k dx \right) = h (B - A \gamma^2 h^2). \quad (22)$$

We now require the integration error (4) to be  $e \leq 0.25$ , which means that

$$0.75I \leq \tilde{I} \leq 1.25I, \quad (23)$$

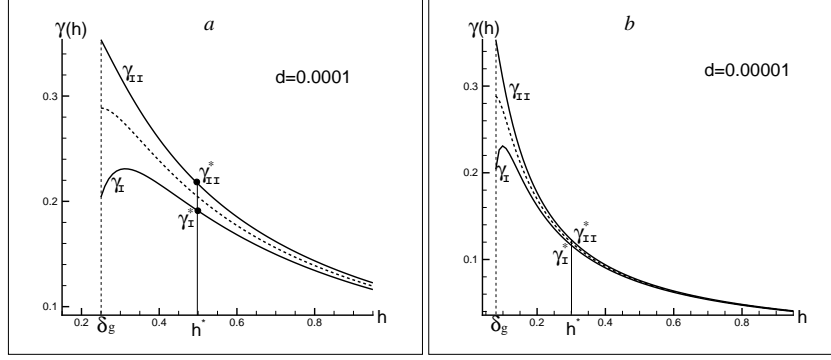
where  $I > 0$ . Consider the lower bound of the inequality (23) and find the values  $\gamma_{II}$  of parameter  $\gamma$  for which the equation  $\tilde{I} = 0.75I$  holds. Substituting  $I$  and  $\tilde{I}$  in the above we obtain

$$Bh - A\gamma^2 h^3 = \frac{1}{2}B\delta_g.$$

Hence

$$\gamma_{II}(h, \delta_g) = \frac{\delta_g}{2h} \sqrt{\frac{2h - \delta_g}{2h}}, \quad (24)$$

where we should require the grid step size  $h > \delta_g/2$  to get  $\gamma_{II}$  as a real number for any fixed  $\delta_g$ . That also makes our analysis consistent with our previous assumption that the grid is very coarse; see item (ii) at the beginning of this section.



**Fig. 8** The function  $\gamma(h)$  for various values of the dimensionless diffusivity  $d$ . The part of the  $(h, \gamma)$  plane between the two solid curves gives the parameter range where the integration is done with the required accuracy  $e \leq 0.25$ .

We then consider the upper bound of (23) and find the values  $\gamma_I$  that satisfy the equation  $\tilde{I} = 1.25I$ . The parameter  $\gamma_I$  as a function of the grid step size  $h$  and the hump width  $\delta_g$  is given by

$$\gamma_I(h, \delta_g) = \frac{\delta_g}{2h} \sqrt{\frac{6h - 5\delta_g}{6h}}, \quad h > \frac{5\delta_g}{6}. \quad (25)$$

The hump width  $\delta_g$  is defined by the diffusion coefficient  $d$ , so that  $\gamma$  in expressions (24) and (25) becomes a function of  $h$  only for a given value of  $d$ . The curves  $\gamma_I(h)$  and  $\gamma_{II}(h)$  are shown in Fig. 8a and Fig. 8b for the dimensionless diffusivity  $d = 10^{-4}$  and  $d = 10^{-5}$  respectively. The range of  $h$  is chosen in both cases as  $h \in [\delta_g, 1]$ , where  $\delta_g$  is calculated from the estimate (14).

For any given value of  $d$ , the conditions (24) and (25) define the parameter range where integral is computed with the required accuracy. Indeed, let us fix the grid step size at a certain hypothetical  $h = h^*$  (see Fig. 8) and compute  $\gamma_I^* = \gamma_I(h^*)$  and  $\gamma_{II}^* = \gamma_{II}(h^*)$ . It then follows from the inequality (23) that for any  $\gamma_I^* \leq \gamma \leq \gamma_{II}^*$  the error is  $e \leq 0.25$ . Also, let us mention that, for any fixed  $h$ , there exists the value of  $\gamma$  for which  $\tilde{I} = I$ ; its value is readily obtained from (21) and (22):

$$\gamma(h) = \frac{\delta_g}{2h} \sqrt{\frac{3h - 2\delta_g}{3h}}, \quad h > \frac{2\delta_g}{3}. \quad (26)$$

One straightforward yet important observation that can be made from Fig. 8 is that the domain where the error is  $e \leq 0.25$  is getting smaller when we decrease the diffusivity  $d$ . In other words, a narrow hump ( $\delta_g \rightarrow 0$ ) is getting ‘lost’ on a very coarse grid with the grid step size  $h \gg \delta_g$ . Another interesting observation is that installing a grid node at the location of the maximum point (which corresponds to  $\gamma = 0$ ) does not at all result in the smallest possible integration error as Eq. (26) clearly gives the value  $\gamma(h) > 0$  (see the dashed curve in Fig. 8 where  $\tilde{I} = I$ ).

#### 4 Nonuniform grid

Our next task is to evaluate the integration error on a non-uniform grid, where we want to find the condition on the grid step size  $h$  that ensures the required accuracy  $e < 0.25$  for a given hump width  $\delta_g$ .

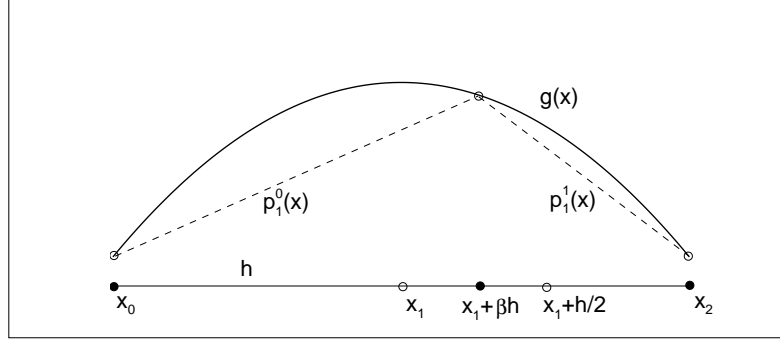
In order to make an insight into this issue, we use the same approach as in Section 3.1. We consider a single-hump distribution which we approximate with the quadratic function  $q(x)$ . However, the function  $g(x)$  is now integrated on a grid of three nodes  $\{x_0, \tilde{x}_1, x_2\}$ , where the central node  $x_1$  is now moved to the position  $\tilde{x}$  while the maximum of the integrand remains at the midpoint  $x_1$  of the domain  $[x_0, x_2]$ ; see Fig. 9. In other words, the new grid is obtained from a uniform grid  $\{x_0, x_1, x_2\}$  of Section 3.1 by the following mapping:

$$x_1 \rightarrow \tilde{x}_1 = x_1 + \beta h, \quad (27)$$

where  $\beta$  is a parameter quantifying the degree of the non-uniformness,  $0 < \beta < 1/2$ . The lower limit  $\beta = 0$  thus corresponds to the original uniform grid. The upper limit  $\beta = 1/2$  corresponds to the case when  $\tilde{x}_1$  is the midpoint of the subinterval  $[x_1, x_2]$  (see Fig. 9).

From the ecological viewpoint the transformation (27) with  $0 < \beta < 1/2$  means that for some practical reason one cannot provide equidistant location of samples in the area where the measurements are made. In other words, we cannot provide sampling at the midpoint  $x_1$  of the interval  $[x_0, x_2]$  (for example, because of a natural obstacle, such as a tree) and have to install a sample somewhere in the neighborhood but still close to the point  $x_1$ . It is important to note that a hump in the density distribution still remains well resolved, as we are still allowed to use three grid points to integrate it.

We now apply the technique described previously in Section 3.1 on the non-uniform grid. The exact integral  $I$  is still given by (18). The approximate value  $\tilde{I}$  of the integral is computed as



**Fig. 9** Approximation of a quadratic function by linear polynomials over a non-uniform grid of 3 nodes.

$$\tilde{I} = \int_{x_0}^{x_1 + \beta h} p_1^0(x) dx + \int_{x_1 + \beta h}^{x_2} p_1^1(x) dx.$$

The linear polynomials are now given by  $p_1^0(x) = B - Ah(x_1 - \beta x_0) - Ah(\beta - 1)x$  and  $p_1^1(x) = B - Ah(x_1 + \beta x_2) - Ah(1 + \beta)x$ . Substituting  $p_1^0(x)$  and  $p_1^1(x)$  in the integrals above, we obtain

$$\tilde{I} = 2Bh - Ah^3 - A\beta^2 h^3.$$

Again we require that the error (4) should be  $e < 0.25$ . Substituting the expressions for  $I$  and  $\tilde{I}$  in the condition  $|I - \tilde{I}| < 0.25I$  and taking into account the condition (17), we arrive at

$$h^2 < \frac{B}{A(1 + 2\beta^2)}.$$

Recall that  $B/A = \delta_g^2/4$  where  $\delta_g$  is the hump width. Making use of Eq. (14) that relates the hump width to the diffusion rate, we obtain:

$$h < \frac{\omega\sqrt{d}}{2\sqrt{1 + 2\beta^2}}. \quad (28)$$

Therefore, the upper bound for  $h$  is a monotonously decreasing function of  $\beta$ . For the extreme value  $\beta = 1/2$  we obtain that  $h = \sqrt{2/3}h_0$ , where  $h_0$  is the restriction (20) on the grid step size obtained for the uniform grid where  $\beta = 0$ . Substituting (20) into Eq. (28), we arrive at

$$h < \omega\sqrt{d/6}. \quad (29)$$

Condition (29) gives us the information on how to choose the grid step size if we want to have the relative error  $e < 0.25$  on a non-uniform grid.

We now want to reveal how the integration error depends on the degree of the grid distortion in case the restriction (29) is ignored. Let us set  $h = h_0$ . For the fixed value  $h_0$ , e.g. as defined by the condition (20), the error becomes a function of  $\beta$ ,

$$e_q(h_0, \beta) = \frac{|I(h_0) - \tilde{I}(h_0, \beta)|}{|I(h_0)|} = \frac{1 + 3\beta^2}{4}. \quad (30)$$

It is readily seen from the expression above that for  $\beta = 0$  the integration error is  $e_q = 0.25$ , as required. Since the error  $e_q(h_0, \beta)$  is a monotone function of  $\beta$ , it reaches its maximum  $e_q = 7/16 \approx 0.44$  at  $\beta = 1/2$ . Hence, moving a node away from the maximum point on a grid with the fixed grid step size  $h = h_0$  can increase the error of integration almost twice.

In conclusion, let us consider the extreme case when  $\beta \rightarrow 1$  in the transformation (27), i.e. when  $\tilde{x}_1$  closely approaches  $x_2$ . From the integration viewpoint, the singular value  $\beta = 1$  means the transition to a coarser grid where we are now allowed to use only two grid nodes instead of three. Correspondingly, we have a single linear polynomial in the vicinity of the hump instead of having two of them as considered in Section 3.1. It then readily follows from the restriction (28) that we should set

$$h = \frac{h_0}{\sqrt{3}}$$

in order to obtain a sufficiently accurate estimate of the integral.

## 5 Discussion and conclusions

Estimation of pest abundance is a key topic in many ecological monitoring and control programs. Their ultimate goal is to provide robust and timely recommendations on the application of pesticides, e.g. once the pest abundance exceeds a certain threshold [36].

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 practice, the information about species presence is usually obtained through collecting samples. The population size, which is an integral of the population density over the area, has to be evaluated based on the values of the population density that are known only at the position of samples. This is a conventional problem of numerical integration. Indeed, integration of sampled data frequently arises in experimental work as well as in computational applications [8, 12, 40]. However, the situation with pest monitoring is essentially different from a standard problem of numerical integration. The matter is that the number of samples collected over an agricultural field usually cannot be made large. Evaluating pest population size becomes a problem of numerical integration of a discrete function obtained on a very coarse grid. Thus the issue of integration accuracy becomes a crucial one, as we only have sparse data to deal

with. Following the approach developed in our recent work [19, 21], in this paper we discuss this issue in much detail.

The emphasis of this paper is on identifying the factors that can affect the accuracy of integration on coarse grids. We showed that the diffusion determines the spatial heterogeneity of the integrand function and that, in turn, is a crucial factor for accurate numerical integration. We demonstrated how the knowledge of the diffusion rate in the problem can be used to obtain an accurate estimate of the pest population size. Alternatively, this knowledge can be used to define the minimum number  $N$  of samples sufficient for accurate evaluation of the pest population size. It should be mentioned here that optimization of the number of samples required to provide robust estimates is an important issue for pest monitoring programs [3, 4, 18].

The main results of our study are itemized below:

- We showed that the problem of obtaining a robust estimate of the population size from sparse spatial data can, in principle, be solved by applying the methods and ideas of numerical integration;
- We obtained condition (20) for the grid step size (i.e. the distance between the sampling locations) to ensure that the estimate of the population size is obtained with a required accuracy (the error being less than 25%) for given diffusion rates. The analytical prediction (20) is in excellent agreement with simulation results, see Tables 1 and 2.

We mention it here that the analysis of the simulation results shows that, in case the population density has a complex multi-hump spatial structure, an accurate estimate of the population size can sometimes be obtained on an very coarse grid consisting of just 3 nodes; see the second column in Table 2 and the last paragraph of Section 3.2.

Note that the coefficient  $\omega$  determining the characteristic length of the spatial pattern (see Eqs. (11) and (14)) may vary depending on the parameters of intra- and interspecific interactions. Once these parameters are known, its value can be estimated theoretically, cf. Eqs. (10) and (11). In ecological practice, the value of  $\omega$  can be extracted from available field data (e.g. from previous studies on the given species) by fitting Eq. (11) to the characteristics of the observed spatial pattern.

- We obtained the accuracy estimates (23–24) in case the population is aggregated inside a single narrow hump and the grid is very coarse, so that the hump is ‘resolved’ by just one node. Even in this rather extreme case, there is a parameter range where the numerical integration evaluates the population size with a required accuracy.

A closer look at the integration of a narrow peak on a coarse grid suggests that it may lead to a paradigm shift [20] when the integration results should be interpreted probabilistically rather than deterministically. The matter is as follows. There is a range of the peak’s positions with regard to the grid nodes where the peak can be integrated with a sufficient accuracy, outside of this range the accuracy becomes unacceptably low. The problem is that, especially in the routine monitoring, the position of the peak would not be known in advance. Integration of the sampling data



would then provide a result that could be accurate in some cases but inaccurate in other cases. This is a typical problem with uncertainty, and a standard approach to deal with it is to quantify different possible outcomes with probability. The conditions (23–24) can then be used to estimate the probability of accurate integration. Indeed, taking into account that  $0 < \gamma < 1$ , the probability of accurate integration with a given value of  $h$  is determined by the distance between the curves  $\gamma_I(h)$  and  $\gamma_{II}(h)$  (see Fig. 8) along the vertical line  $h = \text{const}$ . For instance, for  $d = 0.0001$  it is about 0.18 if  $h = 0.25$  but less than 0.05 if  $h = 0.5$ .

- We considered the effect of the grid non-uniformness, i.e. when a grid node is moved from its ‘regular’ position, on the accuracy of our approach. This is a practically important issue because the grid of sampling positions can hardly be made precisely uniform either as a result of the human factor or because of peculiarities of the landscape structure. We showed that the accuracy of integration is robust with respect to a small variation in the node’s position. For the case of a larger variation, we obtained conditions (28–30) describing what should be the average grid step size to maintain the required accuracy and/or what the accuracy is going to be should the step size be chosen irrelevantly.

Our study suggests a few directions for future work. First, an extension of our approach onto a 2D case should be made. The results obtained here are in a good qualitative agreement with the results of the numerical study made in [20] for the 2D case. However, a modification of the analytical methods that we used in this paper will require considerable work before they can be applied to a 2D grid.

Second, in this paper we validated our approach using the numerical data obtained from an ecological model. Application of the methods of numerical integration to data on invertebrate sampling made in [20] led to an encouraging result. However, a further validation is necessary by applying our method to field data obtained in different environments, for different species and on different spatial grids.

In conclusion, a more general comment should be made. In order to reveal the effect of species diffusivity on the accuracy of the population size estimation, we used the diffusion-based theoretical framework. Correspondingly, the dynamics of the population density is described by diffusion or diffusion-reaction equations and the diffusivity is quantified by the diffusion coefficient  $D$  with the dimension as  $\text{distance}^2 \cdot \text{time}^{-1}$ . This description implies that the individual animal movement is the Brownian motion when the mean squared displacement  $\langle r^2(t) \rangle$  grows with time linearly:

$$\langle r^2(t) \rangle \sim Dt. \quad (31)$$

The corresponding dispersal kernel is then given by a normal distribution; see Eq. (6).

That may rise a question about the generality of our results. Indeed, there have been a growing amount of evidence that some animal species perform a faster dispersal (often referred to as the anomalous diffusion or “superdiffusion,” or Lévy flight) when the mean squared displacement shows growth faster than linear:

$$\langle r^2(t) \rangle \sim \mathcal{D}t^\nu, \quad (32)$$

where  $\nu > 1$  and  $\mathcal{D}$  is a coefficient similar to the diffusion coefficient in its meaning but having a different dimension, i.e.  $\text{distance}^2 \cdot \text{time}^{-\nu}$ . The dispersal kernel in this case has a fatter tail, e.g. showing either exponential or power law rate of decay at large distances. However, the main result of the dimensions analysis still holds, i.e. there is the only quantity with the dimension of length, although its expression becomes slightly different:

$$\Delta_a \sim \sqrt{\mathcal{D}t^\nu}, \quad (33)$$

cf. Eq. (7).

There have been several studies concerned with the relation between the spatial heterogeneity and the ‘diffusivity’ in a broader sense. For instance, it has been shown in [22] that the characteristic length is a power-law function of the coefficient  $\mathcal{D}$  (with the exponent larger than  $\frac{1}{2}$ ) in the case of a clearly non-Brownian motion in a turbulent environment. The dependence of the rate of decay in the population density on the combination  $x/(\mathcal{D}t^\nu)$  rather than on  $x$  alone was proved in [9]. These results point out that the diffusivity rate, considered in a somewhat broader sense, still is a controlling factor that determines the characteristics of the spatial heterogeneity. Therefore, our results and conclusions about its impact on the accuracy of the population size estimation are not restricted to the case of the standard Fickian diffusion and the corresponding Brownian motion of individuals, but should remain valid in a more general case.

**Acknowledgements.** This study was partially supported by The Leverhulme Trust through grant F/00-568/X.

## Appendix: Approximation of a hump by a quadratic function

Let  $u(x)$  be an integrand function that has a local maximum (a ‘hump’) at point  $x_1$ , where  $x_1 \in [0, 1]$ . Consider points  $x_0 = x_1 - h$  and  $x_2 = x_1 + h$ , where  $h > 0$  is an arbitrary parameter defining the ‘hump width’. For instance, the value  $h$  can be defined from the condition that  $u(x_2) = 0.1u(x_1)$ . The examples of the choice of  $h$  will be given further in the text for a particular problem under consideration.

Once we know the function values  $u_m \equiv u(x_m)$ ,  $m = 0, 1, 2$ , we can approximate  $u(x)$  by a quadratic polynomial. That is a well-known interpolation problem (e.g., see [6]) and below we give a brief description of this technique.

To find the equation of the quadratic  $g(x)$ , the function values are generally needed at three points, so that the coefficients of the function  $g(x)$  can be reconstructed using the conditions  $g(x_m) = u(x_m)$ ,  $m = 0, 1, 2$ . However, in our case it is more convenient to write a quadratic function in the form

$$g(x) = B - A(x - x_1)^2,$$

because we require  $g(x)$  to have the same maximum as the hump that it replaces. The coefficients  $A$  and  $B$  are then obtained by using just the two collocation conditions  $g(x_m) = u(x_m)$ ,  $m = 0, 1$ . Thus, the hump is replaced by a quadratic which is symmetric about the location of the maximum  $x = x_1$ .

We now introduce the interpolation error  $e_{int}(x)$  in order to evaluate what we miss when we replace a hump  $u(x)$  with the function  $g(x)$ . The function  $e_{int}(x)$  is defined at any point  $x$  of the interval  $[x_0, x_2]$  as

$$e_{int}(x) = |u(x) - g(x)|.$$

We then consider the maximum distance between the functions  $u(x)$  and  $g(x)$ ,

$$e_{max} = \max_{x \in [x_0, x_2]} e_{int}(x).$$

The maximum interpolation error  $e_{max}$  depends on  $h$ , as it is demonstrated by the following example. Consider the approximation of a hump by a quadratic function for the pest population density  $u_1(x)$ . The coefficients  $A$  and  $B$  for the quadratic function  $g(x)$  are defined from the collocation conditions as

$$A = \frac{u_1(x_0) - u_1(x_1)}{h^2}, \quad B = u_1(x_1).$$

The interpolation error  $e_{max}$  incurred by replacing the hump in the population distribution  $u_1(x)$  by a quadratic function is shown in Table 3. As  $h$  decreases, so does the size of the interpolation error. This is further illustrated by the quadratic approximations shown in Fig. 4.

h	0.125	0.0625	0.0312	0.0156
$e_{max}$	0.3325	0.0922	0.0123	0.0011

**Table 3** The interpolation error for the quadratic functions approximating the hump of  $u_1(x)$  for various values of  $h$ .

Once the integrand function  $u(x)$  has been replaced by a quadratic function in the vicinity of the hump, we can integrate the function  $g(x)$  by a chosen numerical method. Let us apply the method outlined in Section 2.1 to both functions  $u_1(x)$  and  $g(x)$  to integrate them in the vicinity of the hump. Consider the integration error local to the hump, i.e. on the interval  $[x_0, x_2]$ , where the function  $u(x)$  and its corresponding quadratic replacement  $g(x)$  are approximated by piecewise linear polynomials. The integration error (4) computed for the function  $u_1(x)$  and for the quadratic function  $q(x)$  is denoted in Table 4 as  $e_u$  and  $e_q$ , respectively.

It can be seen from Table 4, that  $e_q$  provides a sufficiently reliable estimate for the integration error (4). This conclusion is further confirmed by the results of Table

h	0.125	0.0625	0.0312	0.0156
$e_u$	0.0641	0.0464	0.0279	0.0091
$e_q$	0.1839	0.0961	0.0341	0.0096

**Table 4** The integration error (4) when the integral is computed in the vicinity of the hump. The integration errors are computed for the density distribution  $u_1(x)$  (the row  $e_u$ ), and its quadratic approximation  $g(x)$  (the row  $e_q$ ). The functions are approximated by piecewise linear polynomials.

5 where we integrate both  $u_2(x)$  and  $g(x)$  in the vicinity of the first hump in the multi-peak distribution  $u_2(x)$  (see Fig. 2b). Thus our assumption that the density distribution  $u(x)$  can be approximated by a quadratic function in the vicinity of a hump is justified by computation of the interpolation error and the integration error and such approximation can be used for further theoretical and numerical analysis.

h	0.0312	0.0156	0.0078	0.0039
$e_u$	0.0545	0.0532	0.0267	0.0080
$e_q$	0.1986	0.1024	0.0325	0.0086

**Table 5** The integration error for the first hump in density distribution  $u_2(x)$ , and its quadratic approximation  $g(x)$ . The functions are approximated by piecewise linear polynomials.

## References

1. 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.
2. Barenblatt, G.I. 1996 *Scaling, Self-Similarity, and Intermediate Asymptotics*. Cambridge: Cambridge University Press.
3. Boag, B., Mackenzie, K., McNicol, J.W. & Neilson, R. 2010 Sampling for the New Zealand flatworm. *Proceedings Crop Protection in Northern Britain 2010*, 45-50.
4. Buntin, G.D. 1994 Developing a primary sampling program. In *Handbook of Sampling Methods for Arthropods in Agriculture* (L.P. Pedigo & G.D. Buntin, Eds.), pp. 99-115. Boca Raton: CRC Press.
5. Burn, A.J. 1987 *Integrated Pest Management*. New York: Academic Press.
6. Davis, P.J. 1975 *Interpolation and Approximation*. Dover Pub.
7. Davis, P.J. & Rabinowitz, P. 1975 *Methods of Numerical Integration*. New York: Academic Press.
8. Dunn, S.M., Constantinides, A. & Moghe, P.V. 2006 Numerical methods in biomedical engineering. New York: Elsevier.
9. Giona, M. & Roman, H.E. 1992 Fractional diffusion equation for transport phenomena in random media. *Physica A* **185**, 87-97.
10. Holland, J.M., Perry, J.N. & Winder, L. 1999 The within-field spatial and temporal distribution of arthropods in winter wheat. *Bull. Ent. Res.* **89**, 499-513.
11. Lewis, M.A. & Kareiva, P. 1993 Allee dynamics and the spread of invading organisms. *Theor. Popul. Biol.* **43**, 141-158.

12. Liao, S.X. & Pawlak, M. 1996 On image analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(3), 254-266.
13. Malchow, H., Petrovskii, S.V. & Venturino, E. 2008 *Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulations*. London, UK: Chapman & Hall / CRC Press.
14. Murray, J.D. 1989 *Mathematical Biology*. Berlin, Germany: Springer.
15. Northing, P. 2009. Extensive field based aphid monitoring as an information tool for the UK seed potato industry. *Aspects of Applied Biology* **94**, 31-34.
16. Okubo, A. & Levin, S. 2001 *Diffusion and Ecological Problems: Modern Perspectives*. Berlin: Springer.
17. Pascual, M.A. & Kareiva, P. 1996 Predicting the outcome of competition using experimental data: Maximum likelihood and Bayesian approaches. *Ecology* **77**, 337-349.
18. Perry, J.N. 1996 Simulating spatial patterns of counts in agriculture and ecology. *Comput. Electr. Agri.* **15**, 93-109.
19. Petrovskaya, N.B. & Petrovskii, S. 2010 The coarse-grid problem in ecological monitoring. *Proc. R. Soc. A* **466**, 2933-2953.
20. Petrovskaya, N.B., Petrovskiy, S.V., Murchie, A.K. 2011 Challenges of ecological monitoring: estimating population abundance from sparse trap counts. (*Journal of Royal Society Interface*, under revision).
21. Petrovskaya, N.B. & Venturino, E. 2011 Numerical integration of sparsely sampled data. *Simulat. Modell. Pract. Theory*, **19**(9), 1860-1872.
22. Petrovskii, S.V. 1999 On the plankton front waves accelerated by marine turbulence. *J. Mar. Sys.* **21**, 179-188.
23. Petrovskii, S.V. & Li, B.-L. 2006 Exactly solvable models of biological invasion. Chapman & Hall / CRC, Boca Raton.
24. Petrovskii, S.V., Li, B.-L. & Malchow, H. 2003 Quantification of the spatial aspect of chaotic dynamics in biological and chemical systems. *Bull. Math. Biol.* **65**, 425-446.
25. Petrovskii, S.V. & Malchow, H. 2001 Spatio-temporal chaos in an ecological community as a response to unfavorable environmental changes. *Advances in Complex Systems* **4**, 227-250.
26. Petrovskii, S.V., & McKay, K. 2010 Biological invasion and biological control: A case study of the gypsy moth spread. *Aspects of Applied Biology* **104**, 37-48.
27. Petrovskii, S.V., Morozov, A.Y. & Venturino, E. 2002 Allee effect makes possible patchy invasion in a predator-prey system. *Ecology Letters* **5**, 345-352.
28. 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. *Biological Invasions* **7**, 771-793.
29. Raworth, D.A. & Choi, W.-J. 2001 Determining numbers of active carabid beetles per unit area from pitfall-trap data. *Ent. Exp. Appl.* **98**, 95-108.
30. Roblin, E. 1988 Chemical control of Japanese knotweed (*Reynoutria japonica*) on river banks in South Wales. *Aspects of applied biology* **16**, 201-206.
31. Scott, R. & Marrs, R.H. 1984 Impact of Japanese knotweed and methods of control. *Aspects of applied biology* **11**, 291-296.
32. Seber, G.A. 1982 *The Estimation of Animal Abundance and Related Parameters*. London, UK: Charles Griffin.
33. 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.
34. 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.
35. Snedecor, G.W. & Cochran, W.G. 1980 *Statistical Methods*. Ames: The Iowa State University Press.
36. Stern, V.M. 1973 Economic thresholds. *Ann. Rev. Entomol.* **18**, 259-280.
37. Sutherland, W.J. (Ed.) 1996 *Ecological Census Techniques: A Handbook*. Cambridge, UK: Cambridge University Press.
38. Taylor, L.R., Woiwod, I.P., & Perry, J.N. (1979) The negative binomial as a dynamic ecological model for aggregation, and the density dependence of k. *Journal of Animal Ecology* **48**, 289-304.

39. Turchin, P. 2003 *Complex Population Dynamics: a Theoretical / Empirical Synthesis*. Princeton, USA: Princeton University Press.
40. Yaroslavsky, L., Moreno, A. & Campos, J. 2005 Frequency responses and resolving power of numerical integration of sampled data. *Optics Express* **13(8)**, 2892-2905.