

Evaluation of peak functions on ultra-coarse grids

Natalia Petrovskaya and Nina Embleton

Proc. R. Soc. A 2013 **469**, 20120665, published 28 February 2013

References

This article cites 32 articles, 5 of which can be accessed free
<http://rspa.royalsocietypublishing.org/content/469/2153/20120665.full.html#ref-list-1>

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](#)

Research



CrossMark
click for updates

Cite this article: Petrovskaya N, Embleton N.

2013 Evaluation of peak functions on ultra-coarse grids. *Proc R Soc A* 469: 20120665. <http://dx.doi.org/10.1098/rspa.2012.0665>

Received: 6 November 2012

Accepted: 31 January 2013

Subject Areas:

computer modelling and simulation,
computational physics, applied mathematics

Keywords:

sampled data, sparse data, peak function,
coarse grid, midpoint rule

Author for correspondence:

Natalia Petrovskaya

e-mail: n.b.petrovskaya@bham.ac.uk

Integration of sampled data arises in many practical applications, where the integrand function is available from experimental measurements only. One extensive field of research is the problem of pest monitoring and control where an accurate evaluation of the population size from the spatial density distribution is required for a given pest species. High aggregation population density distributions (peak functions) are an important class of data that often appear in this problem. The main difficulty associated with the integration of such functions is that the function values are usually only available at a few locations; therefore, new techniques are required to evaluate the accuracy of integration as the standard approach based on convergence analysis does not work when the data are sparse. Thus, in this paper, we introduce the new concept of ultra-coarse grids for high aggregation density distributions. Integration of the density function on ultra-coarse grids cannot provide the prescribed accuracy because of insufficient information (uncertainty) about the integrand function. Instead, the results of the integration should be treated probabilistically by considering the integration error as a random variable, and we show how the corresponding probabilities can be calculated. Handling the integration error as a random variable allows us to evaluate the accuracy of integration on very coarse grids where asymptotic error estimates cannot be applied.

1. Introduction

Integration of sampled data arises in a wide class of problems, as in many practical applications, the integrand function is available from experimental measurements only and is therefore given by a discrete set of function values. Numerous examples include acoustics and signal processing [1,2], image reconstruction [3], microbiology [4], ecological applications [5],

etc. A general problem of numerical integration has a long and successful history, and many accurate and efficient methods for integration of sampled data have been designed and documented in the literature [6–9]. Meanwhile the increasing complexity of practical problems has resulted in the recent need to revise existing algorithms as new problems have emerged. One such problem is that of insect pest monitoring, where an accurate evaluation of the population size from the spatial density distribution is required for a given biological species [10].

Let us consider a particular problem of pest insect monitoring in a single agricultural field as an instructive example. An accurate estimate of the total number of pest insects is crucial for making a reliable decision about the use of pesticides in the area where the crop is grown [11]. The data for evaluation of the pest population size is usually collected by trapping. Insect traps are installed at the nodes of a uniform Cartesian grid in the agricultural field, they are exposed for a certain time, and then the traps are emptied and the insects caught are counted [12]. Under the assumption that the number of insects in each trap gives us the true value of the population density obtained at the location of the trap [13], the methods of numerical integration on uniform grids can be employed to estimate the total number of pest insects from the discrete density distribution [14]. However, as we show below, the application of well-known methods, such as the Newton–Cotes integration rules on a uniform grid, is not straightforward at all.

The main difficulty associated with the estimation of the total number of pest insects from trap counts is that the number N of traps cannot be made large enough to ensure that the integral estimate is accurate. Installment of many traps per unit agricultural area would, in itself, bring considerable damage to the agricultural product. Also, trapping is costly and labour consuming, and it introduces a disturbance to agricultural procedures. Hence, the problem of numerical integration has to be solved for a small number N of traps.

From a computational viewpoint, the evaluation of the population size when the number N of traps is small presents the problem of numerical integration of the integrand function on a very coarse uniform Cartesian grid. Moreover, grid adaptation is not possible in the problem as the grid should be generated only once, and N cannot be further increased. This restriction appears because a repeated trapping with an increased number of traps is not available in ecological applications due to the impossibility of reproducing the initial conditions.

Under the restrictions outlined above, the two following questions arise.

- What is the minimum number N_t of traps required to achieve desirable accuracy?
- What can be an alternative measure of accuracy on a coarse grid of traps where $N < N_t$?

Although the processing of sparse data has intensively been studied in various problems from physics and engineering [15–18], to the best of our knowledge, numerical integration of sparse data has not been discussed in the literature. An attempt to address the questions above has been made in our recent work [14,19,20]. In particular, it has been demonstrated in [14] that the answers to those questions depend strongly on the integrand function under consideration. Namely, the accuracy of integration remains acceptable, even on very coarse grids, if the pest population is distributed more or less over the whole area, no matter whether this distribution is close to homogeneous or has a complex heterogeneous structure. On the contrary, the accuracy is very poor when one has to integrate a high aggregation density distribution (i.e. the density function with a single maximum whose support subdomain is small in comparison with the domain area) on a coarse grid.

The high aggregation density distributions are an important class of data that may appear in ecosystems under various conditions. For instance, one common scenario of biological invasion is that the pest species starts spreading from a small localized area and invades the entire domain as time progresses [10]. As the entire pest population is confined to a single subregion within an agricultural field, and the pest population is zero outside that subregion at the beginning of the invasion process, the population density is described as a peak function from a mathematical viewpoint. Obviously the decision about the application of pesticides is best made before the patch of high density spreads over the whole agricultural field. Hence, timely and accurate

evaluation of the total number of pest insects at earlier stages of biological invasion is very beneficial for the cultivation of the agricultural product. At the same time, the application of numerical integration methods is significantly hampered by the fact that the exact location of the peak is not known in the problem. Thus, instead of installing the traps locally (i.e. in the initially infested area) to increase the accuracy of integration, a uniform grid of traps has to be generated over the entire domain. That makes numerical integration of peak functions a very difficult task, as the information about the density function on such a sparse grid may be insufficient for reasonably accurate integration. Namely, it was discussed in [14,20] that the whole peak may be located between nodes of a sparse grid and will therefore be completely missed. This can in turn severely affect the accuracy of the pest population size evaluation. Hence, the questions highlighted above require careful attention when peak functions are considered on coarse grids, and in this paper, we develop a novel probabilistic approach to work with sparse data.

The paper is organized as follows. In §2, we briefly revisit the problem of numerical integration and introduce the concept of ultra-coarse grids for peak functions. An ultra-coarse grid is defined as a grid where the desirable accuracy of integration can only be achieved with probability smaller than 1. It will be shown that integration on ultra-coarse grids cannot guarantee the prescribed accuracy because of the insufficient information (uncertainty) about the integrand function. Instead, the results of the integration should be treated probabilistically by considering the integration error as a random variable with a high magnitude. We calculate the probability of an accurate integral estimate in §3, while in §4 we consider the transition from ultra-coarse grids to coarse grids where the desirable accuracy can be achieved with the probability equal to 1. Numerical examples are considered in §5. A discussion of our results is provided in §6. Our research of ultra-coarse grids will be focused on the one-dimensional case, but the approach we present in the paper can be extended to two-dimensional density distributions.

2. Numerical integration of sparse data

The invasion regime when a spreading pest population forms a strongly heterogeneous spatial distribution, has its one-dimensional counterpart when a peak density function appears somewhere at the unit interval $D = [0, 1]$ [19]. Thus, in the one-dimensional case, a high aggregation density distribution $u(x)$ can be modelled by the following peak function with the support $D_u = [x_I, x_{II}]$ on the unit interval

$$u(x) = \begin{cases} f(x) > 0, & x \in (x_I, x_{II}), \\ 0, & \text{otherwise,} \end{cases} \quad (2.1)$$

where we assume that the function $f(x)$ has a single maximum at point $x^* = 0.5(x_I + x_{II})$. At this stage, we are not interested in a more detailed definition of the function $f(x)$, as particular examples of $u(x)$ will be considered further in the text.

One computational problem related to the study of ecologically meaningful density distributions is that the integral is not available in closed form. Hence, before moving to the discussion of a more general case, we consider several examples where the exact value of the integral is available and the integration error can therefore be computed. A convenient example of a peak function is given by the normal distribution

$$u(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x - x^*)^2}{\sigma^2}\right), \quad (2.2)$$

shown in figure 1a for the peak width $\delta = 6\sigma = 0.25$. Let us numerically integrate (2.2) over the unit interval $D = [0, 1]$. Consider a uniform grid generated in the domain D as $x_{i+1} = x_i + h$, $i = 1, \dots, N - 1$, where the grid step size is $h = 1/(N - 1)$. We employ the midpoint rule of

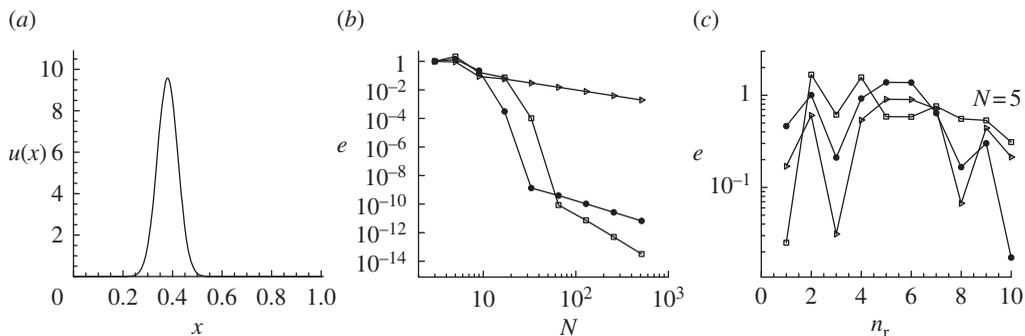


Figure 1. Numerical integration of a peak function. (a) The peak function (2.2), where the peak width is $\delta = 0.25$ and the peak is located at $x^* = 0.38$. (b) The integration error (2.5) computed for the peak function of figure 1a on a sequence of uniformly refined grids. The uniform refinement of the original coarse grid of $N = 3$ nodes does not decrease the integration error, unless the domain of the non-zero density is resolved. (c) The integration error computed when the peak (2.2) is randomly located on the uniform grid of $N = 5$ nodes. The error (2.5) is shown for the 10 realizations n_r of the random variable x^* . (b,c) solid line with filled circle, midpoint rule; solid line with open square, Simpson rule and solid line with open right triangle, statistical rule (2.4).

integration [21] as a baseline integration method in our problem. Once a computational grid has been generated, the integral I is computed by the compound midpoint rule as

$$I = \int_0^1 u(x) dx \approx \tilde{I} = \sum_i w_i u(x_i), \quad (2.3)$$

where $w_i = h$ for the interior nodes $i = 2, \dots, N - 1$ and $w_i = h/2$ at boundary points $i = 1, N$.

In our discussion of the accuracy of numerical integration on coarse grids, we will also compute the integral by the compound Simpson method where the number N is required to be an odd number, $N = 2m + 1$, and the weights in (2.3) are given by $w_i = 4h/3$, $i = 2, 4, \dots, 2m$, $w_i = 2h/3$, $i = 3, 5, \dots, 2m - 1$ and $w_i = h/3$, $i = 1, i = N$. Finally, the third method we employ for numerical integration is a so-called ‘statistical method’ widely used in ecological applications [22]. The method evaluates the integral as

$$I \approx A\bar{u} = \frac{A}{N} \sum_{i=1}^N u(x_i), \quad (2.4)$$

where A is the given area. Since $A = 1$, when we integrate over the unit interval, the method (2.4) is reduced to the evaluation of the mean density of the pest population.

We define the integration error as

$$e = \frac{|I - \tilde{I}|}{|I|}, \quad (2.5)$$

where I is the exact integral and \tilde{I} is the approximate integral computed by the chosen method of numerical integration. The accuracy of integration should be

$$e \leq \tau, \quad (2.6)$$

where τ is a specified tolerance. It is important to note here that in ecological applications, the accuracy requirements on coarse grids are essentially different from those arising in the conventional problem of numerical integration, as the tolerance $0.2 < \tau < 0.5$ is considered as acceptable [23,24]. However, we will see further in the text that even such relatively low accuracy of computations is not always achievable when the number of grid nodes is small.

Let us compute the integration error (2.5) for the density distribution (2.2). Consider first the midpoint rule (2.3) of integration. The error (2.5) of the midpoint rule is shown as a function of the number N of grid nodes in figure 1b. It can be seen from the figure that the integration error of the midpoint rule is not controlled by uniform grid refinement on very coarse grids that are in the focus of ecological research. Namely, the error $e \sim 1$ remains beyond the acceptable range $e \sim 0.2\text{--}0.5$ on grids with $N < \tilde{N}$, where $\tilde{N} \sim 10$ is the upper bound for the realistic number of traps to be considered in ecological applications [25,26].

As the uniform refinement does not decrease the error on coarse grids (see figure 1b), the obvious decision would be to apply a more accurate method of numerical integration (i.e. the Simpson method; see [21]) in order to improve the accuracy of integration on grids with small N . However, it can be seen from figure 1b, where the error of the Simpson method is shown on a sequence of uniformly refined grids, that the Simpson method is not more accurate than the midpoint rule on coarse grids. On the other hand, the method (2.4) is theoretically less accurate than (2.3) [21], but it cannot be said from figure 1b that the method (2.3) is better than (2.4) when coarse grids are considered.

Let us recall that the exact location of the ‘peak subdomain’ D_u is unknown to us. Meanwhile, the value of the approximate integral depends obviously on how many grid nodes are stationed inside the subdomain D_u on a coarse grid. In order to better understand how the integration error (2.5) depends on the location of the peak with respect to the position of the grid nodes, we now consider the peak location x^* in the function (2.2) as a uniformly distributed random variable. Let us fix the number of grid nodes as $N = 5$ and randomly move the peak (2.2) over the domain D 10 times. We integrate the function (2.2) every time that we move the peak, i.e. for each of the 10 realizations n_r of the random variable x^* . The results of numerical integration are shown in figure 1c, where the integration error (2.5) is computed for each of the 10 locations of the peak on the coarse grid of five nodes. For instance, when $n_r = 3$ the peak’s location is $x^* = 0.7013$, while for $n_r = 7$ the same peak is located at $x^* = 0.4188$, etc. It is readily seen from the graph of figure 1c that for the midpoint rule of integration, the error (2.5) depends essentially on the peak location as the error varies as $0.0173 \leq e \leq 1.379$ on the grid with the fixed number of nodes. The same observation with regard to the integration error is true for the Simpson rule and the integration rule (2.4), where it can be seen from the figure that the error’s magnitude varies significantly as we vary the peak location.

It follows from the results of the test case (2.2) that the accuracy of computation cannot be determined when the high aggregation density distributions are considered on grids with a small number of nodes. Since the integration error depends on a random location of the point x^* , the error itself becomes a random variable and a probabilistic approach should be used to evaluate the accuracy. We will refer to the grids, where the accuracy of numerical integration cannot be determined, as *ultra-coarse grids*. Correspondingly, the questions formulated in §1 are reformulated on ultra-coarse grids as follows.

- Given the number N of grid nodes on a uniform grid (the grid step size $h = \text{const.}$), what is the probability of the event $e \leq \tau_0$, where the integration error e is given by (2.5) and τ_0 is the chosen tolerance?
- What is the threshold number N_t of grid nodes when the probability p of achieving an integration error (2.5) smaller than the given tolerance τ_0 becomes $p = 1$?

In the next two sections of our paper, we answer the questions above when peak functions are integrated using the general midpoint rule (2.3).

3. Ultra-coarse grids

Let us expand the density function $u(x)$ given by (2.1) at the maximum point x^* as

$$u(x) = u(x^*) + \frac{1}{2} \frac{d^2 u(x^*)}{dx^2} (x - x^*)^2 + R(x).$$

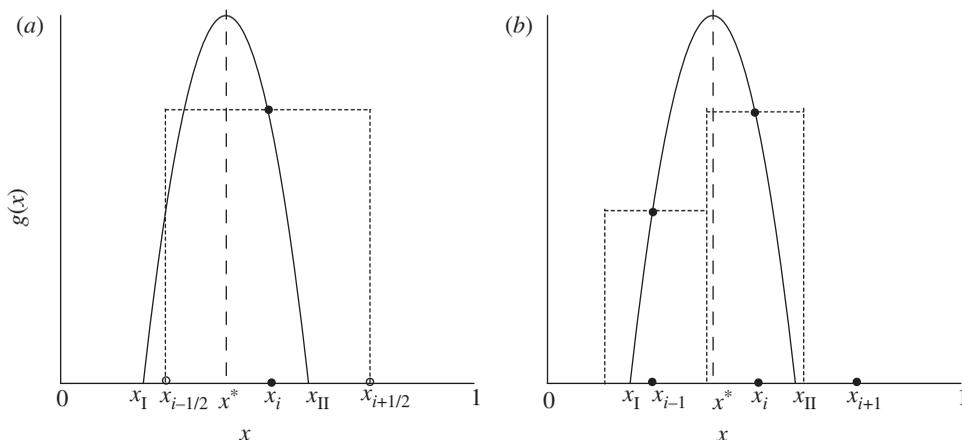


Figure 2. The midpoint integration rule for the peak function. (a) One grid node is located within the subdomain $[x_I, x_{II}]$. (b) Two grid nodes belong to the peak subdomain.

We assume that the remainder $R(x)$ can be neglected in the vicinity of the peak, so that the integrand $u(x)$ becomes

$$u(x) \approx g(x) = \begin{cases} B - A(x - x^*)^2, & x \in [x_I, x_{II}], \\ u(x) = 0, & \text{otherwise,} \end{cases} \quad (3.1)$$

and

where $A = -\frac{1}{2}(d^2u(x^*)/dx^2) > 0$, $B = u(x^*) > 0$. The integral I is then given by

$$I = \int_0^1 u(x) dx \approx \int_{x_I}^{x_{II}} g(x) dx = \frac{2}{3}B\delta, \quad (3.2)$$

where the peak width δ is defined as

$$\delta \equiv x_{II} - x_I = 2\sqrt{\frac{B}{A}}. \quad (3.3)$$

Consider a uniform grid of N nodes generated in the domain $[0, 1]$ as $x_{i+1} = x_i + h$, $i = 1, \dots, N - 1$, where the grid step size is $h = 1/(N - 1)$. We begin our study of ultra-coarse grids with the case when the grid step size $h > \delta$. In other words, we require that the grid is so coarse that either no grid nodes or one grid node fall within the peak subdomain. Since the absence of grid nodes within the peak support is a degenerate case, below we consider one grid node x_i located in the subdomain $[x_I, x_{II}]$ (see figure 2a).

Let the grid step size be

$$h = \alpha\delta, \quad (3.4)$$

where the parameter $\alpha > 1$. Since a quadratic function is symmetric, it is sufficient to consider the subinterval $[x^*, x_{II}]$. The location of the grid node x_i relative to the location of the peak x^* is then given by

$$x_i = x^* + \gamma h, \quad \gamma \in [0, \frac{1}{2}]. \quad (3.5)$$

The midpoint rule approximation of the integral for the geometry $\alpha > 1$ is shown in figure 2a. The integral \tilde{I} computed by the midpoint rule when we use the approximation (3.1) for $u(x)$ is as follows:

$$\tilde{I} = (B - A\gamma^2 h^2)h. \quad (3.6)$$

Hence, the integration error (2.5) is

$$e = \frac{|(2/3)B\delta + A\gamma^2 h^3 - Bh|}{(2/3)B\delta} \leq \tau_0.$$

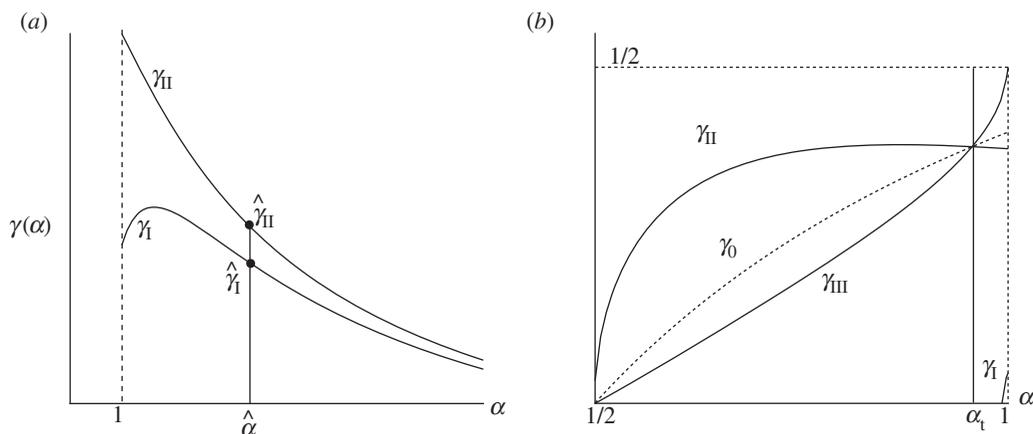


Figure 3. The set of parametric curves defining the admissible range of node locations $\gamma = \gamma(\alpha)$ for which the integration error is $e \leq \tau_0$. (a) The grid step size is $h > \delta$ where δ is the peak width. (b) The grid step size is $\delta/2 \leq h \leq \delta$.

Without loss of generality, let us choose the tolerance as $\tau_0 = \frac{1}{4}$. Taking into account (3.3) and (3.4) and solving the inequalities

$$\frac{3}{4}I \leq \tilde{I} \leq \frac{5}{4}I \quad (3.7)$$

for the node location γ , we obtain the following condition:

$$\gamma_I(\alpha) \leq \gamma(\alpha) \leq \gamma_{II}(\alpha), \quad (3.8)$$

where

$$\gamma_I(\alpha) = \frac{1}{2\alpha} \sqrt{1 - \frac{5}{6\alpha}} \quad \text{and} \quad \gamma_{II}(\alpha) = \frac{1}{2\alpha} \sqrt{1 - \frac{1}{2\alpha}}, \quad (3.9)$$

for the tolerance $\tau_0 = \frac{1}{4}$.

For the sake of the discussion in §4, it is worth noting here that the inequality $\gamma_I(\alpha) \leq \gamma(\alpha)$ implies that $\alpha \geq \alpha_I = \frac{5}{6}$. If $\alpha < \alpha_I$, then $\tilde{I} \leq \frac{5}{4}I$ always holds, and the inequalities (3.8) should be replaced as

$$0 \leq \gamma(\alpha) \leq \gamma_{II}(\alpha). \quad (3.10)$$

Similarly, the inequality $\gamma(\alpha) \leq \gamma_{II}(\alpha)$ requires $\alpha \geq \alpha_{II} = \frac{1}{2}$, the condition that always holds under our previous assumption $\alpha > 1$.

The curves $\gamma_I(\alpha)$ and $\gamma_{II}(\alpha)$ are shown in figure 3a. The conditions (3.9) define the parameter range where the integral is computed with the required accuracy τ_0 . Consider a peak of width δ and let us fix $\alpha = \hat{\alpha} > 1$, so that the grid step size becomes fixed as $h = \hat{h} = \hat{\alpha}\delta$. Compute $\hat{\gamma}_I = \gamma_I(\hat{\alpha})$ and $\hat{\gamma}_{II} = \gamma_{II}(\hat{\alpha})$ on the grid of size \hat{h} . The inequalities (3.8) provide the error $e \leq \tau_0$ for any $\hat{\gamma}_I \leq \gamma \leq \hat{\gamma}_{II}$ (see figure 3a).

Let us assume that the location x^* of the peak maximum can be found at any point of the domain $[0, 1]$ with equal probability, that is the random variable x^* is uniformly distributed. It then readily follows from the above consideration that for the peak width δ , the probability $p(e \leq \tau_0, \alpha)$ of achieving the desired accuracy $e \leq \tau_0$ on a uniform grid with grid step size h is computed as

$$p(e \leq \tau_0, \alpha) = \frac{(\gamma_{II}(\alpha) - \gamma_I(\alpha))}{(\gamma_{\max} - \gamma_{\min})} = 2(\gamma_{II}(\alpha) - \gamma_I(\alpha)),$$

where the entire range of γ is given by $\gamma_{\min} = 0$, $\gamma_{\max} = \frac{1}{2}$. Since the tolerance τ_0 is always fixed as $\tau_0 = \frac{1}{4}$, we further omit it in the definition of the probability function and consider the probability

$p = p(\alpha)$. Straightforward analysis of expressions (3.9) shows that the probability $p(\alpha)$ of achieving the integration error $e \leq \tau_0$ remains $p(\alpha) < 1$ on any ultra-coarse grid where the condition $\alpha > 1$ (i.e. $h > \delta$) holds.

We summarize the findings of this section as follows.

- We have shown that the integration error should be handled as a random variable on ultra-coarse grids where the data available for integration are sparse.
- We have considered a quadratic approximation of the integrand function and found the probability p of an accurate answer when a quadratic polynomial is integrated on an ultra-coarse grid. The relative position of a grid node to the peak function was parametrized by the ratio α of the grid step size h and the width of the peak function δ . It was shown that the probability p of achieving the prescribed accuracy is $p < 1$ for all $\alpha > 1$ (i.e. when either zero or one grid node lie within the support of the peak function).

Meanwhile, the important question that still remains is: if we gradually increase the number of grid nodes N , what is the threshold number N_t for which we have the probability $p = 1$ of the accurate answer (2.6)? The answer to this question will be given in §4.

4. Transition from ultra-coarse grids to coarse grids

In this section, we investigate the transition from grids where the integration error is a random variable to grids where the condition $e \leq \tau_0$ always holds for the given tolerance τ_0 . Let us increase the number N of grid nodes in order to decrease the grid step size h as

$$\frac{\delta}{2} \leq h \leq \delta, \quad (4.1)$$

where δ again is the peak width (3.3). In other words, we now require that $\frac{1}{2} \leq \alpha \leq 1$ when the parametrization (3.4) is used. If the condition (4.1) holds, then either one or two grid nodes belong to the subdomain $[x_I, x_{II}]$ (see figure 2b).

Consider the location (3.5) of grid node x_i . The minimum value γ_0 that provides the location of two grid nodes x_{i-1} and x_i in the subdomain $[x_I, x_{II}]$ is defined from the conditions $x_{i-1} = x^* - \delta/2$ and $x_{i-1} = x_i - h$. We have

$$\gamma_0 = 1 - \frac{\delta}{2h} = 1 - \frac{1}{2\alpha}, \quad (4.2)$$

where the parametrization (3.4) is taken into account.

For $\gamma \in [0, \gamma_0)$, only one grid point belongs to the interval $[x_I, x_{II}]$, and we can use the result (3.9) to compute the probability of accurate integration. Hence, we now focus on the range $\gamma \in [\gamma_0, \frac{1}{2}]$ when two grid points are captured by the peak, as shown in figure 2b.

Let us use again the quadratic approximation (3.1) of the integrand function. Since $g(x_{i-1}) = -A((\gamma - 1)h)^2 + B$ and $g(x_i) = -A(\gamma h)^2 + B$, the integral \tilde{I} is computed by the midpoint rule as follows:

$$\tilde{I} = (g(x_{i-1}) + g(x_i))h = -A((\gamma - 1)^2 + \gamma^2)h^3 + 2Bh.$$

For the sake of simplicity, we require again the tolerance $\tau_0 = \frac{1}{4}$ to arrive at the inequalities (3.7). Consider first the inequality $\tilde{I} \geq \frac{3}{4}I$. Simple algebraic transformation results in

$$\gamma^2(\alpha) - \gamma(\alpha) + C(\alpha) \leq 0,$$

where $C(\alpha) = \frac{1}{2} - (1/16\alpha^3)(4\alpha - 1)$.

Consider the roots $\gamma_{III} = (1 - \sqrt{1 - 4C(\alpha)})/2$ and $\gamma_{IV} = (1 + \sqrt{1 - 4C(\alpha)})/2$ of the equation $\gamma^2 - \gamma + C = 0$. The non-empty range $\gamma \in [\gamma_{III}, \gamma_{IV}]$ exists, if the inequality $4C(\alpha) \leq 1$ holds. Substituting the above expression for $C(\alpha)$ into this inequality, we obtain

$$4\alpha^3 - 4\alpha + 1 \leq 0. \quad (4.3)$$

Numerical solution of equation (4.3) gives us the roots $\alpha_1 \approx -1.1072$, $\alpha_2 \approx 0.2696$ and $\alpha_3 \approx 0.8376$. Hence, if $\alpha \in [\frac{1}{2}, \alpha_3]$, the range $\gamma \in [\gamma_{\text{III}}, \gamma_{\text{IV}}]$ will provide us with the integration error $e \leq \tau_0 = \frac{1}{4}$, where we should also take into account the restriction $\gamma \in [\gamma_0, \frac{1}{2}]$. It readily follows from the above computation that $\gamma_{\text{III}}(\alpha) < \frac{1}{2}$ and $\gamma_{\text{IV}}(\alpha) > \frac{1}{2}$ for any $\alpha \in [\frac{1}{2}, \alpha_3]$.

Consider now the lower boundary γ_0 . The curve $\gamma_{\text{III}}(\alpha)$ intersects the curve $\gamma_0(\alpha)$ at the point α_t (see figure 3*b*). We require $\gamma_0(\alpha_t) = \gamma_{\text{III}}(\alpha_t)$ to obtain the equation

$$8\alpha_t^3 - 8\alpha_t^2 + 1 = 0. \quad (4.4)$$

The roots of (4.4) in the subinterval $[\frac{1}{2}, 1]$ are $\alpha = \frac{1}{2}$ and $\alpha_t \approx 0.8090$. Hence, the curve $\gamma_{\text{III}}(\alpha)$ lies under the curve $\gamma_0(\alpha)$ for $\alpha \in (\frac{1}{2}, \alpha_t)$, and it is above the curve $\gamma_0(\alpha)$ when $\alpha \in (\alpha_t, \alpha_3]$ (see figure 3*b*). Let us also note that $\alpha_t < \alpha_1 < \alpha_3$.

Finally, we consider the inequality $\tilde{I} \leq \frac{5}{4}I$ where after some algebraic transformations we arrive at

$$\gamma^2(\alpha) - \gamma(\alpha) + D(\alpha) \geq 0, \quad (4.5)$$

with $D(\alpha)$ given by $D(\alpha) = \frac{1}{2} - (1/48\alpha^3)(12\alpha - 5)$. The requirement $4D(\alpha) \leq 1$ results in the inequality

$$4\alpha^3 - 4\alpha + \frac{5}{3} \leq 0,$$

which does not have any real roots for $\alpha > 0$. Hence, the inequality (4.5) holds for any value of γ .

Let us now compute the probability $p = p(\alpha)_{\text{theor}}$ of the event that the error is $e \leq \tau_0$ on a grid with fixed grid step size $h = \alpha\delta$, where $\alpha \geq \frac{1}{2}$. The entire domain $\alpha \geq \frac{1}{2}$ is shown in figure 4*a*, where the curves $\gamma(\alpha)$ of figure 3*a, b* are now 'glued' together. In all cases considered below, it is possible that either one node (when $\gamma < \gamma_0$) or two nodes (when $\gamma \geq \gamma_0$) fall within the peak subdomain depending on the value of the parameter γ . The probability p of accurate approximation is then $p = p_1 + p_2$, where p_1 is the probability of an accurate estimate when one node is located in the peak subdomain, and p_2 is the probability of an accurate estimate computed when two nodes belong to the peak subdomain. The whole range of γ is $\gamma_{\text{max}} - \gamma_{\text{min}} = \frac{1}{2}$, and the following cases should be considered for the length of the interval where numerical approximation is accurate.

- $\alpha \in [\frac{1}{2}, \alpha_t]$, where $\alpha_t \approx 0.8090$ has been obtained as a solution to equation (4.4) derived under the condition that the tolerance $\tau_0 = \frac{1}{4}$. Since $\gamma_{\text{III}}(\alpha) \leq \gamma_0(\alpha)$ and $\gamma_{\text{IV}}(\alpha) > \frac{1}{2}$, then for any $\gamma \in [\gamma_0, \frac{1}{2}]$ the conditions (3.7) hold. In other words, if there are two grid points in the subdomain $[x_{\text{I}}, x_{\text{II}}]$, then the integration error is $e \leq \tau_0$, no matter how those grid points are located with respect to the maximum point x^* . The probability of the accurate answer is $p_2(\alpha) = (\frac{1}{2} - \gamma_0(\alpha)) / (\gamma_{\text{max}} - \gamma_{\text{min}}) = 1 - 2\gamma_0(\alpha)$.

Consider now $\gamma \in [0, \gamma_0)$, so that just one grid point is located in the peak subdomain. The admissible range of γ is then given by the inequality (3.8). Let us investigate the position of the curve $\gamma_{\text{II}}(\alpha)$ with respect to the curve $\gamma_0(\alpha)$. Simplifying the equation $\gamma_0(\alpha) = \gamma_{\text{II}}(\alpha)$, we obtain the same cubic equation for α as equation (4.4). Hence, the three curves $\gamma_0(\alpha)$, $\gamma_{\text{II}}(\alpha)$ and $\gamma_{\text{III}}(\alpha)$ intersect in a single point α_t , if we consider the semi-open subinterval $[\frac{1}{2}, 1]$ (see figure 4*b*).

As $\gamma_0(\alpha) \leq \gamma_{\text{II}}(\alpha)$ for $\alpha \in [\frac{1}{2}, \alpha_t]$, the upper bound in the inequality (3.8) should be replaced with $\gamma_{\text{I}}(\alpha) \leq \gamma(\alpha) \leq \gamma_0(\alpha)$. At the same time, the lower bound in (3.8) requires the restriction $\alpha \geq \alpha_1$, which does not hold for $\alpha \in [\frac{1}{2}, \alpha_t]$, and therefore the condition (3.10) should be considered. Hence, the inequality (3.8) is transformed as $0 \leq \gamma(\alpha) \leq \gamma_0(\alpha)$, if there is one grid point in the subdomain $[x_{\text{I}}, x_{\text{II}}]$. This means the integration error is $e \leq \tau_0$ for the grid step size $h < \alpha_t\delta$. As the admissible range becomes $0 \leq \gamma(\alpha) \leq \gamma_0(\alpha)$, the probability is $p_1(\alpha) = 2\gamma_0(\alpha)$. The resulting probability is $p(\alpha)_{\text{theor}} = p_1(\alpha) + p_2(\alpha) = 1$, and the condition $e \leq \tau_0$ holds for any γ if the grid step size is $h = \alpha\delta$, where $\alpha \in [\frac{1}{2}, \alpha_t]$ for the fixed peak width δ .

- $\alpha \in (\alpha_t, \alpha_1]$, where $\alpha_1 = \frac{5}{6} \approx 0.8333$ for $\tau_0 = \frac{1}{4}$ (see §3). From condition (3.10), we now have $p_1(\alpha) = 2\gamma_{\text{II}}(\alpha)$ for $\gamma \in [0, \gamma_0)$ (one node within the peak subdomain) and $p_2(\alpha) = 1 -$

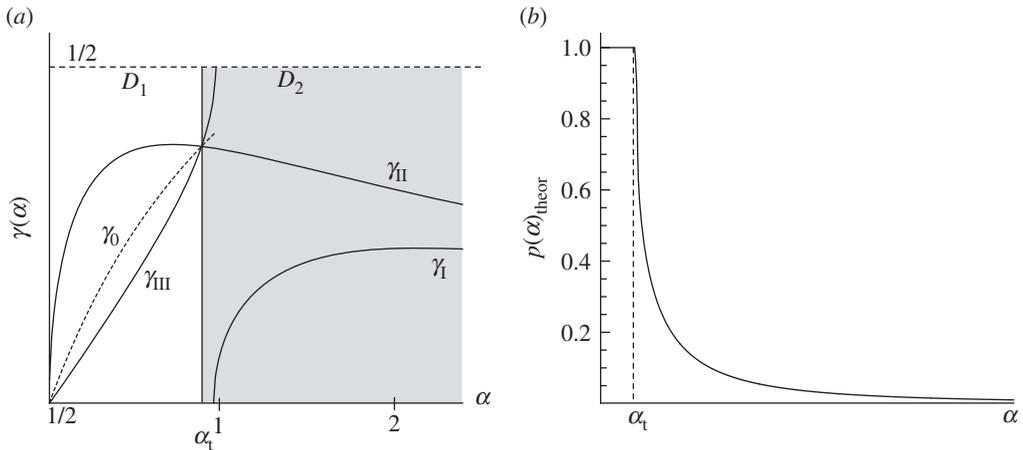


Figure 4. (a) Transition from the probability $p < 1$ to $p = 1$ in the parametric plane (α, γ) . In the domain $D_1 : \alpha \in [\frac{1}{2}, \alpha_t]$, the condition $e \leq \tau_0$ holds for any γ and the probability $p(\alpha) = 1$, while in the domain $D_2 : \alpha > \alpha_t$ (the shaded area in the figure), the integration error becomes a random variable and the probability $p(\alpha) < 1$. (b) The probability $p(\alpha)$ of having the integration error $e \leq \tau_0$ for a peak of the width δ integrated on a grid with the grid step size $h = \alpha \delta$.

$2\gamma_{\text{III}}(\alpha)$ if $\gamma \in [\gamma_0(\alpha), \frac{1}{2}]$ (two nodes within the peak subdomain). The resulting probability is $p(\alpha)_{\text{theor}} = p_1(\alpha) + p_2(\alpha) < 1$.

- $\alpha \in (\alpha_1, \alpha_3]$, where we have $\alpha_3 \approx 0.8376$ from (4.3) for $\tau_0 = \frac{1}{4}$. For this range of α , we have $p_1(\alpha) = 2(\gamma_{\text{II}}(\alpha) - \gamma_1(\alpha))$, as the inequality (3.8) now holds for any $\gamma \in [0, \gamma_0]$ (one node within the peak subdomain). We also have $p_2(\alpha) = 1 - 2\gamma_{\text{III}}(\alpha)$, $\gamma \in [\gamma_0, \frac{1}{2}]$, if two nodes fall within the peak. The probability $p(\alpha)_{\text{theor}} = p_1(\alpha) + p_2(\alpha) < 1$.
- $\alpha > \alpha_3$. We compute the probability p as $p(\alpha)_{\text{theor}} = p_1(\alpha) = 2(\gamma_{\text{II}}(\alpha) - \gamma_1(\alpha))$ (see also §3). The probability $p_2 = 0$ because of the restriction $\alpha \in [\frac{1}{2}, \alpha_3]$ required for the inequality $\tilde{I} \geq \frac{3}{4}I$ when $\gamma \in [\gamma_0, \frac{1}{2}]$.

The function $p(\alpha)_{\text{theor}}$ is shown in figure 4b. For the fixed width δ of the peak, the parameter α_t is the threshold value of the grid step size that provides the transition from ultra-coarse grids to coarse grids. On any grid with $\alpha \leq \alpha_t$ (domain D_1 in figure 4a), the error (2.5) is deterministic in the sense that the condition $e \leq \tau_0$ always holds. On ultra-coarse grids where $\alpha > \alpha_t$ (domain D_2 in figure 4a), the error (2.5) is a random variable as the probability of getting the accurate answer is $p < 1$.

Below, we summarize the main results of §4.

- We have increased the number of nodes on a uniform grid, so that either one or two grid nodes are available for approximating the peak function. For the parametrization introduced in §3, this means the consideration of the case $\alpha \in [\frac{1}{2}, 1]$. We have then found the threshold number N_t of grid nodes such that the integration error e is always $e \leq \tau$, if the number of grid nodes $N \geq N_t$. In other words, the probability of obtaining the desirable accuracy of integration is always $p = 1$ when we approach the threshold N_t .
- It has been shown that the grid step size h_t that corresponds to the threshold number N_t on a uniform grid is a linear function of the peak width δ , that is, $h_t = \alpha_t \delta$, where α_t depends on the chosen tolerance τ only. The coefficient α_t has been computed for the tolerance $\tau_0 = 0.25$ as $\alpha_t \approx 0.8090$.

Let us note that the introduction of ultra-coarse grids extends the existing grid classification. Namely, we consider fine grids as grids where one can rely upon asymptotic error estimates (usually in the form $e = O(h^k)$, e.g. [21]). We can now refer to coarse grids as grids where

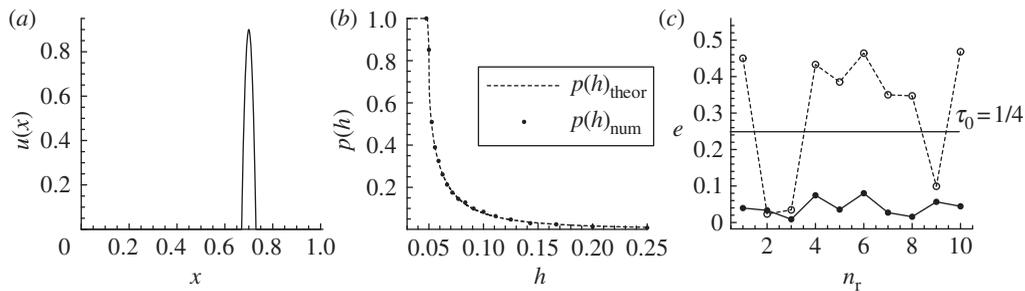


Figure 5. (a) A quadratic function (3.1). The peak width is $\delta = 0.06$ and the parameters are $A = 1000$ and $B = 0.9$. (b) The probability $p(h)_{\text{num}}$ obtained by direct computation agrees with the theoretical results $p(h)_{\text{theor}}$ for the function (3.1) of (a). (c) The integration error for the function (3.1) on an ultra-coarse grid and a coarse grid with a fixed number of nodes. The error (2.5) is shown for the 10 realizations n_r of the random variable x^* , where x^* is uniformly distributed. The probability of achieving an accurate answer (2.6) is $p \approx 0.3$ on an ultra-coarse grid of $N = 18$ nodes (dashed line), while $p = 1$ on a coarse grid of $N = 25$ nodes (solid line).

the asymptotic error estimates may not be applied, but where the probability of obtaining the accuracy goal is $p = 1$. Finally, ultra-coarse grids are grids where the error is a random variable with high magnitude and the probability of obtaining the desirable accuracy is $p < 1$.

The results above have been obtained under the assumption of a quadratic approximation of the integrand function. Hence, numerical investigation of the problem is required when a peak function has a shape different from quadratic in order to verify our findings. This will be carried out in §5.

5. Numerical examples

In this section, we first consider several standard test cases to illustrate our approach to numerical integration on ultra-coarse grids. We then turn our attention to ecologically meaningful density distributions and discuss how the theoretical predictions of §4 work for them.

(a) Standard test cases

We begin our consideration with a quadratic function (3.1), as our first test case is to verify the probability estimate $p(h)_{\text{theor}}$ derived in §4. Let us fix the peak width δ and consider the peak location x^* as a random variable that is uniformly distributed over the interval $[\delta, 1 - \delta]$, as we require that the entire peak is stationed within the unit interval $[0, 1]$. In our test, we provide $n_r = 10^4$ realizations of the random variable x^* on a grid with the fixed number N_l of nodes and compute the integral error (2.5) where we integrate the function (3.1) by the midpoint rule for each realization of x^* . The probability $p(h_l)_{\text{num}}$ of accurate numerical integration is computed as

$$p(h_l)_{\text{num}} = \frac{\hat{n}_r}{n_r}, \quad (5.1)$$

where $h_l = 1/(N_l - 1)$ is the grid step size on a grid of N_l nodes and \hat{n}_r is the number of realizations for which the integration error is $e \leq \tau_0$, $\tau_0 = \frac{1}{4}$. We then increase the number of grid nodes as $N_{l+1} = N_l + 1$ and repeat computation (5.1). We stop when the number N_L of grid nodes results in a grid with the grid step size $h_L \leq \delta/2$.

The quadratic function (3.1) is shown in figure 5a for the peak width $\delta = 0.06$. The probability $p(h)_{\text{num}}$ of accurate numerical integration (2.6) of the function (3.1) is shown in figure 5b. We start from the grid of $N_1 = 5$ nodes and end our computations on the grid of $N_{18} = 22$ nodes where the condition (4.1) still holds. It can be seen from the figure that all values of the probability $p(h_l)$,

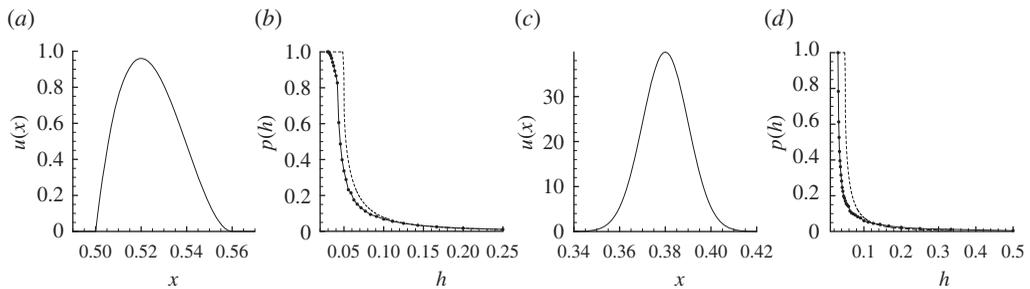


Figure 6. Numerical test cases: (a) the cubic function (5.2) and (c) the normal distribution (2.2). For the functions (a) and (c), the peak width is chosen as $\delta = 0.06$. The probability (5.1) of an accurate answer (2.6) and its comparison with the theoretical curve $p(h)_{\text{theor}}$ obtained for the quadratic function: (b) the probability graph $p(h)_{\text{num}}$ computed for the cubic function (5.2) and (d) the probability $p(h)_{\text{num}}$ for the normal distribution (2.2). (b,d) Dashed line, $p(h)_{\text{theor}}$ and solid line with filled dots, $p(h)_{\text{num}}$.

$l = 1, \dots, 18$, computed by direct evaluation (5.1) lie very close to the theoretical curve $p(h)_{\text{theor}}$, shown as a dashed curve in the figure.

The results of §§3 and 4 are further illustrated in figure 5c where the integration error (2.5) is computed for the function (3.1) on an ultra-coarse grid and a coarse grid. The computation resulting in the graphs shown in figure 5c is similar to the test case discussed in §2 (cf. figure 1c). Namely, we randomly move the peak (3.1) 10 times ($n_r = 10$) on a grid with a fixed number of nodes and compute the integration error every time the peak is moved. It can be seen from the figure the error (2.5) depends on the peak location when we integrate the function (3.1) on an ultra-coarse grid with the number of grid nodes $N = 18$ (i.e. the grid step size is $h = \frac{1}{17}$). The probability of achieving an accurate answer (2.6) is $p \approx 0.3$ (see the graph in figure 5b). Meanwhile, the error is deterministic on a grid of $N = 25$ nodes ($h = \frac{1}{24}$) and the error is $e \leq \tau_0$, no matter where the peak is located.

We now consider several peak functions different from the quadratic function (3.1) in order to understand how the probability estimate obtained for (3.1) will work for them.

A cubic function

$$u(x) = \begin{cases} A \left(x - x^* + \left(\frac{\delta}{3} \right) \right) \left(x - x^* - \left(\frac{2\delta}{3} \right) \right)^2, & x \in \left[x^* - \frac{\delta}{3}, x^* + 2\frac{\delta}{3} \right], \\ 0, & \text{otherwise,} \end{cases} \quad (5.2)$$

presents an interesting test case because the peak is now asymmetric. The function (5.2) is shown in figure 6a, where we still keep the peak width $\delta = 0.06$ for $A = 3 \times 10^4$. The probability graph for the function (5.2) is shown in figure 6b. The graph $p(h)_{\text{num}}$ is in a good agreement with the curve $p(h)_{\text{theor}}$ on very coarse grids ($\alpha > \alpha_3$) where the probability of accurate integration is small. At the same time, the actual probability in the transition layer is smaller than the probability estimate based on the quadratic approximation. One important observation about the graph $p(h)_{\text{num}}$ of figure 6b is that the grid step size h_t^{num} for which the error (2.5) becomes deterministic (i.e. $p(h_t^{\text{num}}) = 1$ and $p(h) < 1$ for any $h > h_t^{\text{num}}$) is smaller than the theoretical estimate

$$h_t = \alpha_t \delta, \quad (5.3)$$

obtained for the quadratic approximation of the integrand function.

The normal distribution (2.2) already discussed in §2 gives us an example of a peak function that is different from zero everywhere in the domain $x \in [0, 1]$ (see figure 6c). However, when we integrate the density function (2.2), 99.7 per cent of the mass will be concentrated within the peak of width $\delta = 6\sigma$, and we can therefore expect a random integration error when we integrate the normal distribution on ultra-coarse grids. The graph $p(h)_{\text{num}}$ computed for the function (2.2) with peak width $\delta = 0.06$ is shown in figure 6d. It can be seen from the figure that the

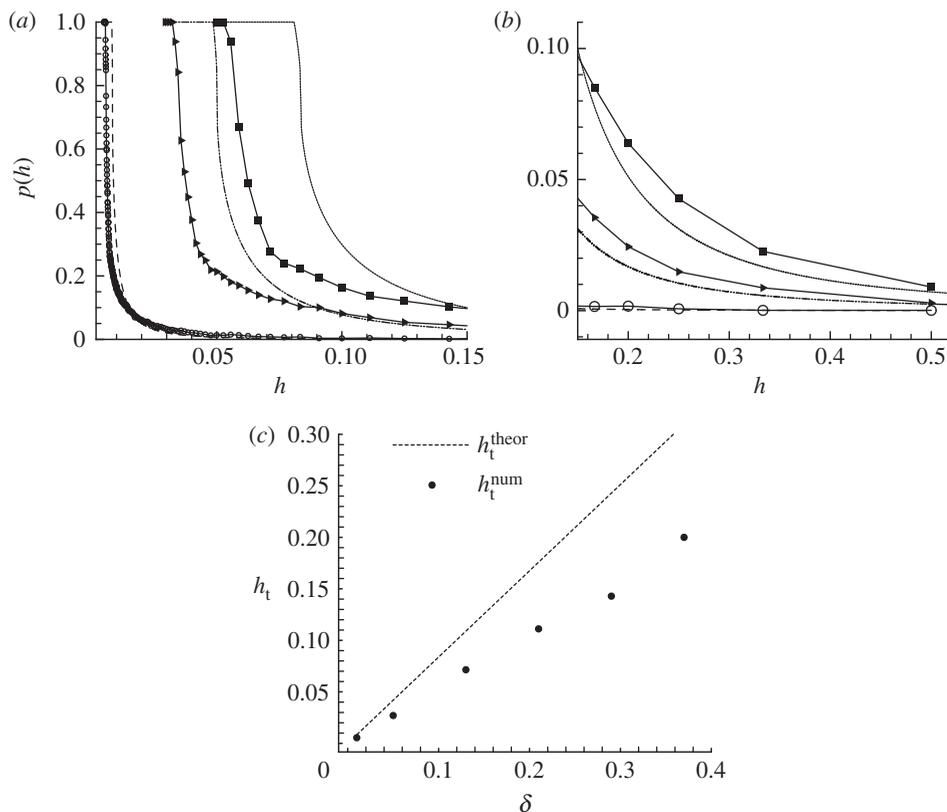


Figure 7. Numerical test cases: (a) the probability graphs $p(h)_{\text{num}}$ computed for the function (5.4) when the peak width δ varies. $\delta = 0.06$: solid line, closed right triangle; $\delta = 0.1$: solid line, closed square; $\delta = 0.01$: solid line, open circle. The computed probability $p(h)_{\text{num}}$ is compared with the theoretical estimate $p(h)_{\text{theor}}$ made for a quadratic function. $\delta = 0.06$: dash-dot-dotted line; $\delta = 0.1$: solid line; $\delta = 0.01$: dashed line. (b) The ‘tail’ of the probability graphs of figure 7a. The legend is the same as in figure 7a. (c) The threshold values h_t^{num} , computed for the function (5.4) when the peak width δ varies, are compared with the theoretical curve (5.3) (dashed line) obtained for a quadratic function.

entire probability graph $p(h)_{\text{num}}$ is now shifted with respect to the curve $p(h)_{\text{theor}}$. The maximum deviation is $d_p = 0.8129$, and the threshold value of the grid step size is $h_t^{\text{num}} \approx 0.5\delta$ instead of $h_t \approx 0.8\delta$ obtained for the quadratic function.

A smaller threshold grid step size h_t^{num} is a consequence of the interpolation error $e_{\text{int}} = O(\delta^3)$ that we introduce when we approximate the integrand function by a quadratic polynomial. As the interpolation error depends on the peak width δ , our probability estimate $p(h)_{\text{theor}}$ should be accurate enough when δ is small (a narrow peak), and it will differ from the actual probability $p(h)_{\text{num}}$ when we increase the peak width δ . Below, we study this dependence in more detail by considering the Lorentz distribution—another peak function that often appears in various fields of physics and interdisciplinary research [27],

$$u(x) = \left\{ \begin{array}{ll} \frac{\delta^2}{4} \frac{1}{4(x-x^*)^2 + \delta^2/4}, & x \in \left[x^* - \frac{\delta}{2}, x^* + \frac{\delta}{2} \right], \\ 0, & \text{otherwise,} \end{array} \right\} \quad (5.4)$$

Let us vary the peak width in (5.4) and compute the probability graph $p(h)_{\text{num}}$ for each fixed value δ by employing the procedure (5.1). A family of probability curves is shown in figure 7 for the function (5.4) with the peak width $\delta = 0.06$ (a baseline peak), the peak width $\delta = 0.1$ (a wide peak) and the peak width $\delta = 0.01$ (a narrow peak). For the sake of convenience, we show the

‘transition layer’ in figure 7a, while the ‘tails’ of the probability curves are shown in figure 7b. It can be seen from the figure that the computed probability curve $p(h)_{\text{num}}$ gets closer to the probability graph $p(h)_{\text{theor}}$ ¹ when we decrease the peak width δ . From the ecological viewpoint, the most important conclusion we derive from the consideration of the probability graphs shown in figure 7 is that the difference between the predicted value (5.3) and the actual threshold value h_t^{num} obtained by direct computation decreases when we decrease δ . In other words, we can rely upon the estimate (5.3) to determine how many grid nodes are required to guarantee the accuracy (2.6) when a narrow peak is integrated. Thus, our next test is to compute the function $h_t^{\text{num}}(\delta)$ and to compare it with the theoretical estimate (5.3).

The results of the computation of h_t^{num} for the function (5.4) where the peak width δ varies are shown in figure 7c. The value h_t^{num} is computed for each fixed peak width δ from the condition that $p(h_t^{\text{num}}) = 1$ and $p(h) < 1$ if $h > h_t^{\text{num}}$. The computed threshold grid step size is in very good agreement with the estimate (5.3) for narrow peaks ($\delta < 0.1$). At the same time, the theoretical curve (5.3) lies above the actual values when wide peaks are considered and therefore the estimate (5.3) can be used in ecological problems as the upper bound for the threshold value h_t .

The standard numerical test cases discussed in this section confirmed the following.

- Replacing the integrand function by a quadratic polynomial can be considered a reliable approximation when the probability of accurate integration is computed.
- The formula (5.3) gives us a good estimate of the threshold grid step size h_t when a narrow peak is integrated.
- The same formula (5.3) can be used as the upper bound for h_t if a wide peak is considered.

Once the estimate (5.3) of the threshold grid step size h_t has been verified for standard numerical test cases, our next goal is to check how formula (5.3) will work when ecologically meaningful peak functions are considered. This will be carried out in §5b.

(b) Ecological test cases

We use the pest density distributions $u(x)$ generated from the spatially explicit predator–prey model describing spatiotemporal dynamics of a pest insect population [28,29],

$$\left. \begin{aligned} \frac{\partial u(x,t)}{\partial t} &= d \frac{\partial^2 u}{\partial x^2} + u(1-u) - \frac{uv}{u+p} \\ \text{and} \quad \frac{\partial v(x,t)}{\partial t} &= d \frac{\partial^2 v}{\partial x^2} + k \frac{uv}{u+p} - mv. \end{aligned} \right\} \quad (5.5)$$

The model (5.5) presents a system of coupled diffusion–reaction equations written in dimensionless variables. The functions $u(x,t)$ and $v(x,t)$ are the densities of the pest insect (considered as the prey) and its consumer (the predator), respectively, at time $t > 0$ and position x . Coefficient d describes species diffusivity due to the movement of the individuals, p is the half-saturation prey density, k is the food assimilation efficiency coefficient and m is the predator mortality.

Ideally, our results on the accuracy of integration on ultra-coarse grids should be checked against appropriate ecological data. However, one serious obstacle in the problem is that our further discussion will require the handling of data on a sequence of refined grids, and it is not possible to fulfil this requirement when field data are considered. Meanwhile the model (5.5) and its two-dimensional counterpart [28] have been validated against experimental data to demonstrate that the model is in good agreement with the results of field measurements [5]. Hence, we believe the model (5.5) can be used for the generation of ecologically realistic data.

Equations (5.5) provide a rich variety of spatiotemporal distributions of the pest density function $u(x,t)$. In particular, it is well known [5] that the properties of the spatial distribution $u(x)$

¹We consider the theoretical curves $p(h)_{\text{theor}}$ instead of a single curve $p(\alpha)_{\text{theor}}$ in order to be able to compare them with the results of direct computation for each value of δ .

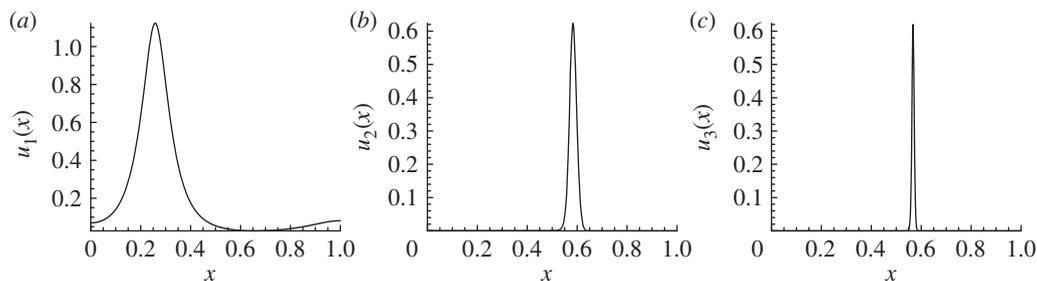


Figure 8. Ecological test cases. The spatial distribution of the pest population density $u(x)$ as predicted by the model (5.5) for different values of the diffusivity d : (a) the density distribution $u_1(x)$ has been obtained for $d = 10^{-4}$, (b) the pest population density $u_2(x)$ obtained for $d = 10^{-5}$, (c) the density $u_3(x)$ for $d = 10^{-6}$. An example of the system's parameters (the density distribution $u_2(x)$): $t = 50$, $k = 0.5$, $p = 2.0$, $m = 0.42$. The initial conditions are $u(x, 0) = u_0$, $0 < x < x_u$, $v(x, 0) = v_0$, $0 \leq x \leq x_v$, where $u_0 = 0.8$, $v_0 = 0.5$, $x_u = 0.6$, $x_v = 0.55$ and $u(x, 0) = 0$, $x > x_u$, $v(x, 0) = 0$, $x > x_v$.

Table 1. The integration error (2.5) for the density distribution $u_1(x)$ on a sequence of refined grids with grid step size h . N is the number of grid nodes on a uniform grid.

N	3	4	5	6	7	8	9	10
h	0.5	0.3333	0.25	0.20	0.1667	0.1429	0.125	0.1111
e	0.6948	0.1119	0.5459	0.07983	0.1699	0.02305	0.08228	0.001918

considered at a given time t are determined by the diffusion d in the problem and that the initial conditions $u(x, 0)$, $v(x, 0)$ can evolve into the one-peak spatial pattern if the diffusion is $d \ll 1$ (see [30] for a more detailed discussion on the pattern formation). Several examples of one-peak density distributions are shown in figure 8, where the functions $u_1(x)$, $u_2(x)$ and $u_3(x)$ have been obtained from the numerical solution of equation (5.5) with the boundary conditions $du/dx = 0$ at $x = 0$ and $x = 1$ for the diffusion coefficient $d = 10^{-4}$, $d = 10^{-5}$ and $d = 10^{-6}$, respectively.²

One important observation that can be made from figure 8 is that the peak width δ depends on the diffusion coefficient d . A simple estimate of the function $\delta(d)$ discussed in [20] can be written as

$$\delta = \omega\sqrt{d}, \quad (5.6)$$

where the coefficient ω depends on the system's parameters. An extensive numerical study performed in [30,31] revealed that, in the predator–prey system (5.5), the value ω is relatively robust to changes in the parameter values, and can typically be considered as $\omega \approx 25$. Hence, we can evaluate the upper bound for the threshold grid step size h_t as

$$h_t = \alpha_t \delta \approx C\sqrt{d}, \quad (5.7)$$

where the coefficient $C = \alpha_t \omega$.

Consider the function $u_1(x)$ shown in figure 8a. The diffusion coefficient is $d = 10^{-4}$, hence the estimate (5.7) gives us the grid step size as $h_t \approx 0.2$. The corresponding number of grid nodes is $N_t \approx 6$.

Let us compute the integration error (2.5) by the midpoint rule on a sequence of refined grids, where we add just one node to the grid at each consecutive refinement step. The results of numerical integration of the function $u_1(x)$ are shown in table 1.

It can be seen from the table that our estimate of the threshold grid step size is in surprisingly good agreement with the results of computation. The actual threshold number obtained in

²The other parameters of the one-dimensional system of equations as well as various initial conditions applied in numerical solution are discussed in detail in [19].

Table 2. The integration error (2.5) for the density distribution $u_2(x)$ on a sequence of refined grids with grid step size h . N is the number of grid nodes on a uniform grid.

N	17	18	19	20	21	22	23	24	25
h	0.0625	0.0588	0.0556	0.0526	0.05	0.0476	0.0455	0.0435	0.0416
e	0.4127	0.5412	0.5101	0.4124	0.1960	0.0028	0.1454	0.2234	0.2137

computations is $N_t = 6$ and, while the integration error keeps oscillating on grids with $N > N_t$ nodes, the condition (2.6) always holds on those grids.

Let us repeat the previous computational procedure for the function $u_2(x)$ shown in figure 8*b*. The diffusion coefficient $d = 10^{-5}$, for which the density distribution $u_2(x)$ has been generated, gives us the lower bound for the threshold number of nodes estimated as $N_t \approx 17$. The results of numerical integration of the function $u_2(x)$ are shown in table 2 where the threshold number obtained in computations is $N_t = 21$. This number of traps is not realistic in real-life computations, and other factors should be taken into account when a decision is made on the number of traps to be installed. One such factor that will be considered in future work is that the integral value depends obviously on the peak width, and the true value of the integral can be small if a narrow peak is considered. A small value of the integral (i.e. a small number of pest insects) means, in turn, that the risk of making a wrong decision about pesticide application is low. Hence, we can increase the tolerance τ_0 in the problem (e.g. $\tau_0 = 0.5$ or even $\tau_0 = 1$ in some cases) and use the technique discussed in §§3 and 4 in order to re-compute our estimate of the threshold number N_t for the new tolerance.

Finally, let us have a look at a very narrow peak $u_3(x)$ shown in figure 8*c*. The diffusion coefficient $d = 10^{-6}$ gives us the estimate $N_t \approx 51$. Clearly, this number of traps is beyond any reasonable range and cannot be used in real-life applications. However, as the integral is very small (for the density distribution $u_3(x)$, we have $I = \int_0^1 u_3(x) dx = 0.007161$), our recommendation to ecologists would probably be not to take any action until the time evolves and the peak function gets wider. On the other hand, an important conclusion derived from the extreme case of the distribution $u_3(x)$ is that installing a reasonable number of traps (e.g. $N \sim 10$) with the aim of trying to evaluate the total number of insects at a very early stage of the biological invasion is senseless, as the probability of a correct answer is very small (see the discussion of the case $h > \delta$ in §3).

6. Conclusions

We have considered a problem of numerical integration for high aggregation density distributions (peak functions) on coarse uniform grids. Our study has originally been prompted by the needs of ecological monitoring and control where minimization of the number of measurements made in order to accurately reconstruct a density distribution remains an important problem.

The main results of the paper are as follows.

- We have introduced the concept of ultra-coarse grids. An ultra-coarse grid is defined as a grid where the integration error is a random variable with high magnitude because of the insufficient information (uncertainty) about the integrand function.
- The definition of an ultra-coarse grid implies that we have to evaluate the probability p of achieving an integration error smaller than the given tolerance rather than to evaluate the error itself. In our paper, we have obtained a probability estimate based on a quadratic approximation of the integrand function. We then carried out a number of numerical test cases to demonstrate that our probability estimate gives us a reliable upper bound for the threshold grid step size h_t for which the probability $p(h_t) = 1$. Thus, the approach suggested in the paper can be used to evaluate the minimum number N_t , such that the desirable accuracy of integration is guaranteed on a grid of N_t nodes.

- It has been shown in the paper that conventional methods of error control and analysis do not work on ultra-coarse grids. It has been discussed in §2 that the uniform refinement of an ultra-coarse grid does not reduce the integration error, as the integration error on a refined grid can be even bigger than the error on the original grid unless the number N_t of nodes is reached. However, increasing the number N of grid nodes does increase the probability of an accurate integral estimate and is therefore desirable.

A general remark should be made here with regard to whether the information about the population density at a given time and location can be adequately obtained from trap counts at all, so that the approach developed in our paper can be applied. Indeed, as the traps are emptied only once in a while (e.g. weekly), the question is how much the population density distribution can possibly change over this time. The answer to this question depends, to a certain extent, on the biological and behavioural traits of the monitored species. However, we mention here that the spatial scale of variations in the population density distribution for walking insects, i.e. those that are usually sampled with pitfall traps, is known to be 30–40 m [32]. Meanwhile, typical dispersal distances for walking insects are estimated to be 1 m or less per day [33], which obviously corresponds to a diffusivity of the order of $1 \text{ m}^2 \text{ day}^{-1}$. Over 1 week of trap exposure, that results in a diffusion length of $\sqrt{7} \approx 2.6 \text{ m}$, which is about one order of magnitude less than the spatial scale of inherent variation.

One important direction of future research is to investigate the transition from ultra-coarse to coarse grids for other classes of functions. Peak functions present just one class of density distributions, and a similar problem of accurate integral evaluation arises when other spatially heterogeneous functions are considered. In particular, we are interested in rapidly oscillating functions that often appear in ecological applications. Also, our future work will be to extend our results to the two-dimensional case, as we want to apply our approach to real-life problems. Since field measurements are carried out on a regular basis in ecological applications, heuristic estimates are available for many common pest species distributions and practical recommendations with regard to the number N_t of traps that have already been obtained based on these heuristic estimates. We intend to check those recommendations against a theoretical probability estimate of the threshold number N_t after our evaluation technique is extended to two-dimensional problems.

The examples of integration of ecologically meaningful distributions we discussed in the paper can only be considered as the very first steps on our way to design a robust procedure for risk evaluation in agricultural applications. At the same time, our results demonstrate that a conventional approach that ecologists use in their problems has to be revisited. A standard procedure of the risk evaluation in pest management is to compare an estimate of the total number of pest insects with a certain critical number and to make a decision about the use of pesticides based on that comparison. Ecologists readily admit that there is uncertainty surrounding the estimation of the pest abundance, which may become worse as the number of samples decreases [34]. However, to the best of our knowledge, the error in the pest abundance estimation has never been considered in the ecological literature as a random variable in order to take into account the risk factor related to the uncertainty in integral evaluation when the number N of traps is small. Taking this risk factor into account when designing an appropriate methodology for decision making in pest insects management should constitute a topic of our future work.

References

1. Smallwood D. 1980 Integration of equally spaced sampled data using a generalized Whittaker's reconstruction formula. *IEEE Trans. Acoust. Speech Signal Process.* **28**, 341–343. (doi:10.1109/TASSP.1980.1163416)

2. Yaroslavsky L, Moreno A, Campos J. 2005 Frequency responses and resolving power of numerical integration of sampled data. *Opt. Express* **13**, 2892–2905. (doi:10.1364/OPEX.13.002892)
3. Liao SX, Pawlak M. 1996 On image analysis by moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 254–266. (doi:10.1109/34.485554)
4. Dunn SM, Constantinides A, Moghe PV. 2006 *Numerical methods in biomedical engineering*. New York, NY: Academic Press.
5. Malchow H, Petrovskii SV, Venturino E. 2008 *Spatiotemporal patterns in ecology and epidemiology: theory, models, and simulations*. London, UK: Chapman & Hall/CRC Press.
6. Engels H. 1980 *Numerical quadrature and cubature*. New York, NY: Academic Press.
7. Evans G. 1993 *Practical numerical integration*. New York, NY: Wiley.
8. Piessens R, deDoncker-Kapenga E, Uberhuber E, Kahaner D. 1983 *QUADPACK: a subroutine package for automatic integration*. Berlin, Germany: Springer.
9. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1988 *Numerical recipes in C*. Cambridge, UK: Cambridge University Press.
10. Shigesada N, Kawasaki K. 1997 *Biological invasions: theory and practice*. Oxford, UK: Oxford University Press.
11. Stern VM. 1973 Economic thresholds. *Annu. Rev. Entomol.* **18**, 259–280. (doi:10.1146/annurev.en.18.010173.001355)
12. Byers JA, 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)
13. Petrovskii S, Bearup D, Ahmed DA, Blackshaw R. 2012 Estimating insect population density from trap counts. *Ecol. Complex.* **10**, 69–82. (doi:10.1016/j.ecocom.2011.10.002)
14. Petrovskaya NB, Petrovskii SV, Murchie AK. 2012 Challenges of ecological monitoring: estimating population abundance from sparse trap counts. *J. R. Soc. Interface* **68**, 420–435. (doi:10.1098/rsif.2011.0386)
15. Ferreira PJSG. 2001 Iterative and noniterative recovery of missing samples for 1-D band-limited signals. In *Sampling theory and practice* (ed. Marvasti FA), pp. 235–282. New York, NY: Plenum Publishing Corporation.
16. Sava P. 2011 Micro-earthquake monitoring with sparsely-sampled data. *J. Petrol. Explor. Prod. Technol.* **1**, 43–49. (doi:10.1007/s13202-011-0005-7)
17. Yaroslavsky LP, Shabat G, Salomon BG, Ideses IA, Fishbain B. 2009 Nonuniform sampling, image recovery from sparse data and the discrete sampling theorem. *J. Opt. Soc. Am. A* **26**, 566–575. (doi:10.1364/JOSAA.26.000566)
18. Wilcox JZ, Wilcox TJ. 1995 Algorithm for extraction of periodic signals from sparse, irregularly sampled data. *Astron. Astrophys. Suppl. Ser.* **112**, 395–405.
19. Petrovskaya NB, Petrovskii SV. 2010 The coarse-grid problem in ecological monitoring. *Proc. R. Soc. A* **466**, 2933–2953. (doi:10.1098/rspa.2010.0023)
20. Petrovskaya NB, Embleton NL, Petrovskii SV. 2013 Numerical study of pest population size at various diffusion rates. In *Dispersal, individual movement and spatial ecology* (eds Lewis MA *et al.*). Lecture Notes in Mathematics, no. 2071, pp. 355–386. Berlin, Germany: Springer.
21. Davis PJ, Rabinowitz P. 1975 *Methods of numerical integration*. New York, NY: Academic Press.
22. Davis PM. 1994 Statistics for describing populations. In *Handbook of sampling methods for arthropods in agriculture* (eds Pedigo LP, Buntin GD), pp. 33–54. Boca Raton, FL: CRC Press.
23. Pascual MA, Kareiva P. 1996 Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology* **77**, 337–349. (doi:10.2307/2265613)
24. Sherratt JA, 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)
25. Mayor JG, Davies MH. 1976 A survey of leatherjacket populations in south-west England, 1963–1974. *Plant Pathol.* **25**, 121–128. (doi:10.1111/j.1365-3059.1976.tb01939.x)
26. Northing P. 2009 Extensive field based aphid monitoring as an information tool for the UK seed potato industry. *Asp. Appl. Biol.* **94**, 31–34.
27. Belkić D, Belkić K. 2010 *Signal processing in magnetic resonance spectroscopy with biomedical applications*. Boca Raton, FL: CRC Press.
28. Murray JD. 1989 *Mathematical biology*. Berlin, Germany: Springer.
29. Turchin P. 2003 *Complex population dynamics: a theoretical/empirical synthesis*. Princeton, NJ: Princeton University Press.

30. Petrovskii SV, 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. (doi:10.1016/S0092-8240(03)00004-1)
31. Petrovskii SV, Malchow H. 2001 Spatio-temporal chaos in an ecological community as a response to unfavorable environmental changes. *Adv. Complex Syst.* **4**, 227–250. (doi:10.1142/S0219525901000164)
32. Holland JM, Perry JN, 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)
33. Vinatier F, Chailleux A, Duyck P-F, Salmon F, Lescourret F, Tixier P. 2010 Radiotelemetry unravels movements of a walking insect species in heterogeneous environments. *Anim. Behav.* **80**, 221–229. (doi:10.1016/j.anbehav.2010.04.022)
34. Binns MR, Nyrop JP, Werf W. 2000 *Sampling and monitoring in crop protection: the theoretical basis for developing practical decision guides*. New York, NY: CABI Publishing.