Bull Math Biol (2014) 76:718–743
DOI 10.1007/s11538-014-9940-z

ORIGINAL ARTICLE

A Novel Approach to Evaluation of Pest Insect Abundance in the Presence of Noise

Nina Embleton · Natalia Petrovskaya

Received: 20 September 2013 / Accepted: 30 January 2014 / Published online: 12 March 2014

Abstract Evaluation of pest abundance is an important task of integrated pest management. It has recently been shown that evaluation of pest population size from discrete sampling data can be done by using the ideas of numerical integration. Numerical integration of the pest population density function is a computational technique that readily gives us an estimate of the pest population size, where the accuracy of the estimate depends on the number of traps installed in the agricultural field to collect the data. However, in a standard mathematical problem of numerical integration, it is assumed that the data are precise, so that the random error is zero when the data are collected. This assumption does not hold in ecological applications. An inherent random error is often present in field measurements, and therefore it may strongly affect the accuracy of evaluation. In our paper, we offer a novel approach to evaluate the pest insect population size under the assumption that the data about the pest population include a random error. The evaluation is not based on statistical methods but is done using a spatially discrete method of numerical integration where the data obtained by trapping as in pest insect monitoring are converted to values of the population density. It will be discussed in the paper how the accuracy of evaluation differs from the case where the same evaluation method is employed to handle precise data. We also consider how the accuracy of the pest insect abundance evaluation can be affected by noise when the data available from trapping are sparse. In particular, we show that, contrary to intuitive expectations, noise does not have any considerable impact on the accuracy of evaluation when the number of traps is small as is conventional in ecological applications.

Keywords Pest monitoring · Trap counts · Random error · Numerical integration

N. Embleton (✉) · N. Petrovskaya
School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK
e-mail: embleton@for.mat.bham.ac.uk

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

1 Introduction

Pest insect management in agriculture has the obvious goal of preventing or minimising the damage pests cause to crops. In past decades, the integrated pest management (IPM) approach emerged which incorporates several different tactics that work cooperatively together to protect crops from pest attack in a more sustainable way (Kogan 1998). An important part of any IPM programme is the monitoring of the pest insect abundance in an agricultural field. The decision of whether or not to implement a control action is then made by comparing the abundance of pests against some threshold level, i.e. the limit at which intervening becomes worth the effort or expense. Since the basic principle of IPM is that a control action is only used if and when it is necessary, accurate evaluation of pest insect abundance remains key to the decision process (Burn et al. 1987; Metcalf and Luckmann 1982).

Trapping is a widely used sampling technique for pest insect abundance evaluation (Alexander et al. 2005; Ferguson et al. 2000; Holland et al. 1999; Mayor and Davies 1976). Traps are installed in the field, exposed for a certain amount of time, after which the traps are emptied and the pests are counted. Under the assumption that trap counts can be converted into the pest population density at the trap locations it is possible to obtain an estimate of the total pest population size (Byers et al. 1989; Raworth and Choi 2001). However, optimising the accuracy of such an evaluation remains a complex and difficult problem where two main aspects must be kept in mind. First, the accuracy can be affected by how the sampled data are collected. There has been intensive research on what is the optimal number of sample units required to achieve a specified precision (e.g. see Binns et al. 2000; Dent 2000; Pedigo and Rice 2009). The sampling plan, i.e. the prescribed locations at which samples are to be taken, is also in the focus of ecological research (Ferguson et al. 2000; Holland et al. 1999), where comparison of various patterns of trap locations in the field has been made in order to understand how the sampling plan may affect the accuracy (Alexander et al. 2005).

The second, equally important aspect of the accuracy problem is how the collected data are processed. A conventional approach is to calculate the arithmetic mean number of pest insects from trap counts (Davis 1994). From the mean number of pests per unit area, an estimate of the number of pests in the entire agricultural field is obtained by scaling to the area of the agricultural field (Snedecor and Cochran 1980). Alternatively, the problem of pest abundance evaluation can be considered as a numerical integration problem and in recent years intensive study of numerical integration methods for ecological applications has been carried out (Embleton and Petrovskaya 2013; Petrovskaya and Embleton 2013; Petrovskaya et al. 2012, 2013; Petrovskaya and Petrovskii 2010; Petrovskaya and Venturino 2011). It was discussed in our recent paper (Petrovskaya and Embleton 2014) that the application of numerical integration techniques often results in a more accurate evaluation of pest abundance than straightforward statistical computation of the mean density. Since numerical integration methods have been emerging as a promising approach to evaluating pest abundance, in the present paper we focus our attention on them further. Namely, we consider the application of numerical integration techniques to the problem where the data used for evaluation are not exact values of the pest population density.

A standard assumption in numerical integration is that the method deals with exact data, i.e. an inherent random error is zero when data are collected. Meanwhile, an inherent random error is often present in field measurements and, along with evaluation error, contributes to the accuracy issues when the pest abundance is calculated. An evaluation error, also known as *an approximation error* in the theory of numerical integration is the error arising because a continuous density function is replaced in the evaluation procedure with a discrete function whose values are available at trap locations only. The approximation error depends on the number of traps used in monitoring and the theory states that the approximation error will be reduced to zero if we can hypothetically make the number of traps infinitely large (Davis and Rabinowitz 1975). At the same time the conventional definition of the approximation error implies that the data used for its computation are precise.

Inherent random errors are errors caused by unknown and unpredictable changes in data measurements (BIPM et al. 2008; Topping 1972). In ecological applications the source of that uncertainty can vary from a simple miscounting of the number of insects in a trap to some environmental conditions in an agricultural field that are responsible for generating an error in a trap count (e.g. a trap can undergo occasional interference from a bigger animal in the field). Trap counts are converted into the density values at the trap locations, and therefore the density values further used to evaluate pest abundance are also affected by the random error. Clearly, the impact of a random error on the accuracy of the evaluation of pest insect abundance should be taken into account to ensure that a correct pest management decision is made. Thus in our work we study the accuracy of evaluation of pest insect population size under the assumption that every trap count has a random error.

It is worth mentioning here that the problem of validation of the measured data has already received attention in the ecological literature. However, with regard to the trapping procedure, the mainstream of research has been focused on accurate conversion of the trap counts into the values of the true population density (Browde et al. 1992; Evans et al. 1983; Petrovskii et al. 2012). Meanwhile, once such a conversion has been made, the estimate of the pest abundance is assumed to be based on exact data and, to the best of our knowledge, no attempt has been made so far to incorporate the random measurement error into the evaluation procedure. In the discussion in this paper, we do not consider the problem of converting trap counts into a discrete population density function. In other words, further in the text we assume that the number of insects caught in each trap already represents the value of the absolute population density in its catchment area but each trap count has an inherent random error.

Numerical integration methods are convenient for the study of noisy data because their formulation allows one to easily control the contribution of the random error into the approximation of the pest insect abundance. It will be demonstrated in our paper how random error in collected trap counts can be converted into random error in a pest abundance estimate. We therefore explain how to calculate the mean as well as a credible interval of the evaluation error, when the discrete density function is randomly perturbed.

Another topic discussed in our paper is the impact of the error induced by noise on the accuracy of evaluation when the data are sparse. The problem of sparse data remains extremely important in IPM programmes, as a widespread situation is that financial, ecological and other restrictions do not allow for a large number of traps to be installed in an agricultural field. In routine pest monitoring programmes, the number of traps rarely exceeds 20 (Mayor and Davies 1976), while in some cases it can be as small as one or a few traps per field (Northing 2009). It has been discussed in Petrovskaya and Embleton (2013) and Petrovskaya and Petrovskii (2010) that an estimate of pest abundance can be very inaccurate on a coarse grid of traps, especially when pest abundance is evaluated from a heterogeneous density pattern. Hence the intuitive expectation is that an estimate of pest abundance based on noisy data will be even worse. However, it will be shown in the paper that, perhaps counter-intuitively, noise does not have a lot of impact on the accuracy of a pest abundance estimate when the number of traps is small.

2 Quantifying the Uncertainty in the Pest Abundance Evaluation Problem

In this section, we briefly recall a numerical integration technique for the problem of pest abundance evaluation. We consider a trapping procedure in an agricultural field and assume first that the trap counts are precise. We explain how exact information about the pest population density at trap locations can be transformed into a numerical integration problem. We then assume uncertainty in field measurements and incorporate a random error into the numerical integration problem.

2.1 Computation of Pest Abundance by Numerical Integration

For the sake of convenience, we focus the discussion in this paper on the one-dimensional case.[1] Let the domain D where the traps are installed be represented by the interval $[a, b]$. Since an obvious linear transformation maps the domain D onto the interval [0, 1], below we consider a total number N of traps installed across the unit interval. The location x_i of a trap is represented by the index i, thus f_i corresponds to the pest population density at that trap location.

Methods of numerical integration are applied when an integrand $f(x)$ defined over the interval [0, 1] is only available at points x_i, $i = 1, \ldots, N$. If we knew the pest insect spatial density distribution $f(x)$ at any point of the domain [0, 1], then the pest abundance I in the field would be computed as the integral of the continuous density function $f(x)$,

$$I = \int_0^1 f(x)\,dx.$$

However, the pest population density function is only given to us as a discrete set of data, that is, $f(x) \equiv f_i$, where $i = 1, \ldots, N$. Consequently, the above integral

---

[1] A detailed explanation of numerical integration techniques for two-dimensional problems with precise data can be found in Petrovskaya and Embleton (2014) and Petrovskaya et al. (2012).

cannot be evaluated exactly and must instead be approximated by means of numerical integration.

For the rest of Sect. 2.1, we assume that we know *precise* (i.e. unperturbed) values of the population density $f(x)$ at trap locations x_i, $i = 1, \ldots, N$. A general numerical integration formula is then written as (e.g. see Davis and Rabinowitz 1975)

$$I \approx I_a = \sum_{i=1}^{N} w_i f_i, \tag{1}$$

where I_a is an approximation of the exact integral I, and w_i, $i = 1, \ldots, N$, represent weight coefficients that define a particular method of integration. The values of the weights w_i are dependent on the number N of traps and on their location. In the case that the traps are located arbitrarily, there is no ready-to use formulas for the weight coefficients and they must be calculated in advance in order to employ the formula (1) (e.g. see Petrovskaya and Venturino 2011). When a systematic sampling plan is used whereby the traps have an equal distance between them, the problem of numerical integration is reduced to using a chosen method from the Newton–Cotes family of numerical integration methods and the weight coefficients are readily available in the literature. The trapezoidal rule is, perhaps, the most well-known member of the Newton–Cotes family with the weights defined as

$$w_i = h/2 \quad \text{for } i = 1 \text{ and } i = N \quad \text{and} \quad w_i = h \quad \text{for } i = 2, \ldots, N-1, \tag{2}$$

where $h > 0$ is the fixed distance between traps.

For any chosen method of numerical integration and any fixed number N of traps used to collect the data, the accuracy of an approximation I_a is assessed by analysing the approximation error. The relative approximation error E_{rel} is defined as

$$E_{\text{rel}}(N) = \frac{|I - I_a|}{|I|}, \tag{3}$$

where clearly a lower relative error means a more accurate estimation I_a of the pest abundance I. To ensure the correct pest management decision is made, e.g. whether or not to apply pesticides, the estimate should be sufficiently accurate. We therefore require the estimated pest abundance to be within a specified estimate tolerance τ of the true pest abundance, i.e. we require the relative error E_{rel} to satisfy the following condition:

$$E_{\text{rel}}(N) \le \tau. \tag{4}$$

Clearly, the approximation error (3) depends on the number N of traps where the values f_i are available. In ecological applications, the number N is usually small and that may result in a big approximation error $E_{\text{rel}}(N)$ (Petrovskaya and Embleton 2013, 2014). Hence an estimate tolerance of $\tau \sim 0.2$–0.5 is already considered acceptable in many ecological problems (Pascual and Kareiva 1996; Sherratt and Smith 2008). Furthermore, it has been shown in Petrovskaya and Petrovskii (2010), Petrovskaya and Embleton (2014) and Petrovskaya et al. (2012) that for any fixed N the error $E_{\text{rel}}(N)$ depends on the spatial pattern of the density function.

It is important to note here that in ecological problems an estimate of the pest abundance is very often obtained using the sample mean pest population density (Davis 1994) which we denote by $\bar{f}$. This is defined as follows (e.g. see Snedecor and Cochran 1980)

$$\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f_i .$$

An estimate I_a to the true number of pests I in the field is then given by

$$I \approx I_a = A\bar{f}, \tag{5}$$

where A is the area of the agricultural field.

Clearly, the method (5) can be incorporated into the general framework of numerical integration (1) with the weights given by $w_i = 1/N$ for $i = 1, \ldots, N$, if the integration is done over the unit interval (i.e. $A = 1$). Identification of (5) within the framework (1) allows us to compare it with other methods of numerical integration. While the statistical approach (5) provides a straightforward and convenient way to evaluate the pest abundance, it has been demonstrated in Embleton and Petrovskaya (2013), Petrovskaya and Embleton (2014) and Petrovskaya et al. (2012) that a different choice of weight coefficients in (1) gives us better accuracy than using the method (5) for the same number of traps. Meanwhile, we shall see later in the paper that considering the problem of pest abundance evaluation as one of numerical integration has another advantage. Namely, representation of the estimate I_a in the form (1) is extremely convenient when the evaluation of the pest population size is required based on perturbed data f_i. In the next section, we introduce the uncertainty of an approximation I_a generated by the uncertainty in the data f_i, $i = 1, \ldots, N$. The weight coefficients in a method of numerical integration given to us are then used in order to relate the uncertainty in the estimate I_a, and consequently in the error $E_{\mathrm{rel}}(N)$ to the uncertainty in trap counts.

2.2 The Uncertainty of Pest Abundance Evaluation from Noisy Measurements

As could be seen in the previous section, when the pest abundance is evaluated from trap counts, the evaluation error (3) is always present in the problem. This happens because we replace a continuous density function with a discrete set of function values f_i, $i = 1, 2, \ldots, N$. Our previous studies of estimating pest abundance by means of numerical integration (Petrovskaya and Embleton 2014; Petrovskaya et al. 2012) have been focused on how the error (3) can be controlled based on the assumption that the pest population densities provided by the trap counts are indeed equal to the true densities. However, this assumption is not entirely realistic, as measurements of the pest population density are subject to *measurement error*.

Let the measured pest population density at trap location x_i be denoted by $\tilde{f}_i$. Let also f_i refer to the exact density $f(x)$ at the point x_i, as discussed in Sect. 2.1. Applying a method of numerical integration (1) to the measured pest densities $\tilde{f}_i$, $i =$

$1, \dots, N$ gives the following estimate of the pest abundance:

$$\tilde{I} = \sum_{i=1}^{N} w_i \tilde{f}_i. \tag{6}$$

The relative error of an approximation based on measured data which we denote by $\tilde{E}_{\text{rel}}$ is then given by

$$\tilde{E}_{\text{rel}} = \frac{|I - \tilde{I}|}{|I|}. \tag{7}$$

The focus of our investigation is to establish how the introduction of noise to the data set $\{f_i\}$ affects the accuracy of the estimation, that is, to determine how $\tilde{E}_{\text{rel}}$ differs from E_{rel}.

The exact value of the pest density f_i at any location i is not known, hence the need to install traps. Nor can the exact value of the random measurement error be known. There is thus an *uncertainty* associated with the measured value $\tilde{f}_i$. In our work, we simulate the uncertainty by considering any measured value of the pest density $\tilde{f}_i$ to be a realisation of a normally distributed random variable F_i with mean μ_i, and standard deviation σ_i. The probability density function is (e.g. see Grimmett and Stirzaker 2001)

$$p(\tilde{f}_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{\tilde{f}_i - \mu_i}{\sigma_i} \right)^2 \right\}, \tag{8}$$

where we assume that the mean is equal to the true pest density, that is, $\mu_i = f_i$. The uncertainty in the measured value $\tilde{f}_i$, which we denote by $u(\tilde{f}_i)$ can be then quantified by the standard deviation σ_i of the random variable F_i,

$$u(\tilde{f}_i) = \sigma_i. \tag{9}$$

If a random variable has the normal distribution, then any single measurement $\tilde{f}_i$, i.e. a single realisation of the random variable F_i, lies in the range

$$\tilde{f}_i \in [f_i - z\sigma_i, f_i + z\sigma_i] \tag{10}$$

with probability

$$P(z) = \operatorname{erf}\left(\frac{z}{\sqrt{2}} \right), \tag{11}$$

where the error function $\operatorname{erf}(z)$ is given by

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp\left(-t^2\right) dt.$$

Let us assume that with the same probability, the pest population density obtained via a trap count is within a fixed percentage of the true density at the trap location. In other words, with probability $P(z)$ each measured pest population density f_i lies somewhere within the range,

$$\tilde{f}_i \in [f_i - \nu_m f_i, f_i + \nu_m f_i],$$

where we refer to $\nu_m \in [\nu_{m1}, \nu_{m2}] \subset (0, 1)$ as the *measurement tolerance*. Equating the interval above to that given by (10) gives the following relation between the standard deviation σ_i and the measurement tolerance ν_m:

$$\sigma_i = \frac{\nu_m f_i}{z}. \tag{12}$$

It is worth noting here that our definition of noise does not depend on the length of the time interval when traps are exposed in the field. Generally, a longer time of exposition can be thought of as collecting a bigger number of samples that, in turn, results in smaller uncertainty in data (i.e. a smaller value of the standard deviation σ_i in the normal distribution) (Steel and Torrie 1960). However, the measurement tolerance ν_m we use in the problem is always expressed as a percentage of the true value f_i at the trap location x_i. Hence a longer (shorter) time of traps exposition is already taken into account by considering larger (smaller) values f_i of the density function.

An example of the uncertainty associated with the function values is depicted in Fig. 1a. The ecologically relevant (i.e. non-negative) function $f(x)$ has been defined as

$$f(x) = \frac{1}{3}\sin\left(\frac{3\pi x}{2}\right) + \frac{2}{3}, \quad x \in [0, 1],$$

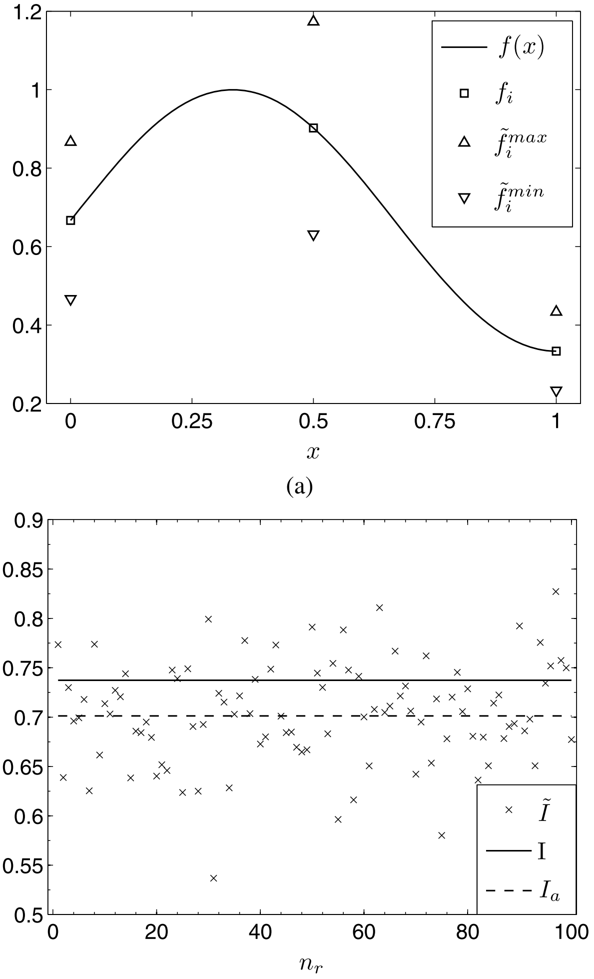
hence the pest abundance is $I = 0.737402$. The exact pest population densities f_i correspond to the function $f(x)$ evaluated at the trap locations x_i, $i = 1, \ldots, N$ which are regularly distributed on the interval [0, 1]. In the example shown in Fig. 1a, the number of traps has been fixed as $N = 3$ hence the traps are located at $x_1 = 0$, $x_2 = 0.5$ and $x_3 = 1$. The estimate I_a formulated by numerically integrating the exact data f_i, $i = 1, 2, 3$ via the trapezoidal rule (2) is $I_a = 0.701184$, while the error is $E_{\text{rel}} = 0.049115$ which is much lower than the required tolerance τ.

We then consider the perturbed data as shown in Fig. 1a. Sets of measured data values $\tilde{f}_i$ are generated by perturbing the function values f_i at each point x_i, $i = 1, 2, 3$, according to the transformation

$$\tilde{f}_i = f_i + \gamma\sigma_i, \tag{13}$$

where γ is a random variable taken from the standard normal distribution, and σ_i is defined according to (12). The measurement tolerance is set as $\nu_m = 0.3$. We also fix $z = 3$, therefore, the probability that each realisation $\tilde{f}_i$ lies within the range (10) is $P(z = 3) \approx 0.9973$. The transformation is applied $n_r = 100$ times to each value f_i to generate n_r sets of measured data for $i = 1, 2, 3$. These data sets are integrated for any fixed n_r using the same trapezoidal rule (2) to yield estimates of the pest abundance $\tilde{I}$.

The distribution of the estimate $\tilde{I}$ of pest abundance computed from the perturbed data $\tilde{f}_i$ on a grid of $N = 3$ traps is shown in Fig. 1b. It is clear from the figure that the introduction of noise can cause the estimate $\tilde{I}$ based on measured data to be further away from the true abundance I making the accuracy of evaluation very poor for some realisations of $\tilde{I}$. Hence we want to control a range of the error $\tilde{E}_{\text{rel}}$ induced by the noise in the data f_i, and in the next section we quantify the resulting uncertainty in the accuracy $\tilde{E}_{\text{rel}}$ of the approximated pest abundance.



(a)

(b)

Fig. 1 Evaluation of pest abundance from noisy data. (**a**) An example of the pest population density function $f(x)$. Three equidistant traps are installed over the unit interval to measure the density $f(x)$. The density value $\tilde{f}_i$, $i = 1, 2, 3$ measured at the position x_i of the trap lies within the range (10) with probability $P(z)$ as defined by (11). The lower and upper limits of this range are denoted $\tilde{f}_i^{\min}$ and $\tilde{f}_i^{\max}$, respectively. The measurement tolerance has been set as $\nu_m = 0.3$ and we have fixed $z = 3$. (**b**) The distribution of the estimate $\tilde{I}$ of pest abundance computed from the measured data $\tilde{f}_i$ on a grid of $N = 3$ traps. Each realisation is presented as a *skewed cross* in the figure, where $n_r = 100$ realisations of the estimate $\tilde{I}$ are shown. The values $\tilde{I}$ are compared with the exact value I of the pest abundance (*solid line*) and the estimate I_a computed from the exact data f_i (*dashed line*)

2.3 Calculation of the Evaluation Error $\tilde{E}_{\mathrm{rel}}$ from Noisy Data

Consider random perturbation (8) of the density function $f(x)$. It can be seen from (6) that an estimate $\tilde{I}$ of pest abundance is a linear combination of the measured pest densities $\tilde{f}_i$. Hence $\tilde{I}$ can in turn be considered as a realisation of a normally distributed random variable which we shall denote $\tilde{I}_F$ where

$$\tilde{I}_F = \sum_{i=1}^{N} w_i F_i. \tag{14}$$

The random variable $\tilde{I}_F$ has mean $\mu_{\tilde{I}} = I_a$, where I_a is the estimated abundance based on the exact pest densities. Furthermore, the standard deviation $\sigma_{\tilde{I}}$

is

$$\sigma_{\tilde{I}} = \sqrt{\sum_{i=1}^{N} w_i^2 u^2(\tilde{f}_i)} \tag{15}$$

(e.g. see Cox 2007).

We now determine the probability density function of the random variable $\tilde{E}_{\text{rel}}$. For the sake of convenience, let us first consider the following auxiliary quantity

$$E = \frac{I - \tilde{I}}{I}. \tag{16}$$

Since E is a linear function of $\tilde{I}$ which is a realisation of a normally distributed random variable, it can be considered as a realisation of a normally distributed random variable with mean $\mu_E = 1 - I_a/I$ and standard deviation $\sigma_E = \sigma_{\tilde{I}}/I$. We note that in ecological applications the true pest abundance I is always $I > 0$. The probability density function is described by

$$p(E) = \frac{1}{\sigma_E\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{E - \mu_E}{\sigma_E}\right)^2\right\}, \tag{17}$$

and the quantity E belongs to the range

$$E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E] \tag{18}$$

with probability $P(z)$ given by (11). Examples of the probability density function of E are shown in Fig. 2.

We have

$$\tilde{E}_{\text{rel}} = |E|,$$

and $\tilde{E}_{\text{rel}}$ becomes a realisation of a random variable with a folded normal distribution (e.g. see Leone et al. 1961). The probability density function of $\tilde{E}_{\text{rel}}$ is then formed from that of E by reflecting the negative contributions in the y-axis and is given by the following expression

$$\begin{aligned} p(\tilde{E}_{\text{rel}}) &= \frac{1}{\sigma_E\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{\tilde{E}_{\text{rel}} - \mu_E}{\sigma_E}\right)^2\right\} + \exp\left\{-\frac{1}{2}\left(\frac{\tilde{E}_{\text{rel}} + \mu_E}{\sigma_E}\right)^2\right\}\right] \\ &= \frac{I}{\sigma_{\tilde{I}}\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{I(1 - \tilde{E}_{\text{rel}}) - I_a}{\sigma_{\tilde{I}}}\right)^2\right\}\right. \\ &\quad \left. + \exp\left\{-\frac{1}{2}\left(\frac{I(1 + \tilde{E}_{\text{rel}}) - I_a}{\sigma_{\tilde{I}}}\right)^2\right\}\right], \end{aligned} \tag{19}$$

where the mean value is

$$\mu_{\tilde{E}_{\text{rel}}} = \left(1 - \frac{I_a}{I}\right)\left[1 - 2\Phi\left(\frac{I_a - I}{\sigma_{\tilde{I}}}\right)\right] + \frac{\sigma_{\tilde{I}}}{I}\sqrt{\frac{2}{\pi}} \exp\left\{-\frac{1}{2}\left(\frac{I_a - I}{\sigma_{\tilde{I}}}\right)^2\right\}, \tag{20}$$

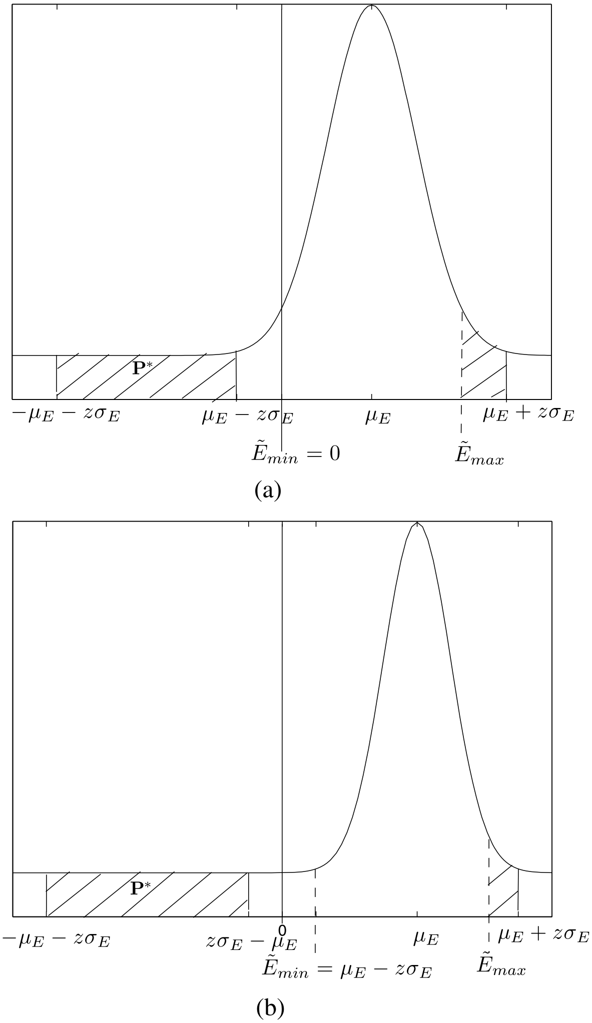


Fig. 2 The probability density function of the quantity E as described by (17). Reflecting the negative contributions in the y-axis yields the folded normal distribution of $\tilde{E}_{\text{rel}}$. The upper and lower limits of the interval $[\tilde{E}_{\text{min}}, \tilde{E}_{\text{max}}]$ to which $\tilde{E}_{\text{rel}}$ belongs with probability $P(z)$ are defined differently depending on the distance between the true pest abundance I and the estimate formulated on exact data I_a: (**a**) when $|I - I_a| \le z\sigma_{\tilde{I}}$ and (**b**) when $|I - I_a| > z\sigma_{\tilde{I}}$. See the Appendix for the details of how $\tilde{E}_{\text{min}}$ and $\tilde{E}_{\text{max}}$ are calculated

and the standard deviation is

$$\sigma_{\tilde{E}_{\text{rel}}} = \sqrt{\mu_E^2 + \sigma_E^2 - \mu_{\tilde{E}_{\text{rel}}}^2}. \tag{21}$$

We now seek a range $[\tilde{E}_{\text{min}}, \tilde{E}_{\text{max}}]$ to which $\tilde{E}_{\text{rel}}$ belongs with probability $P(z)$. It can be seen from (17) (see also Fig. 2) that the range of the error $\tilde{E}_{\text{rel}}$ depends on the quality of approximation I_a obtained from the exact values f_i of the pest population density. Two separate cases depending on the nature of the probability density function (17) should be considered.

The first case is when the mass to be reflected in the y-axis in order to obtain the folded normal distribution (19) contains part but not all of the range (18). That occurs when the distance between the true pest abundance I and the estimate I_a formed from exact data satisfies the condition $|I - I_a| \le z\sigma_{\tilde{I}}$ (see Fig. 2a). This condition requires a certain level of accuracy of the approximation formed from exact data (i.e. the approximation I_a is required to be sufficiently close to I).

We then consider the scenario when $|I - I_a| > z\sigma_{\tilde{I}}$, i.e. a poor approximation is obtained on integrating exact data. The mass to the left of the y-axis is either entirely exclusive of the interval (18) in the case that μ_E is positive (see Fig. 2b) or, when μ_E is negative, is entirely inclusive.

Combining the two cases above and making the calculations explained in the Appendix, we find that $\tilde{E}_{\mathrm{rel}} \in [\tilde{E}_{\min}, \tilde{E}_{\max}]$ with probability $P(z)$ when the lower limit is defined as

$$\tilde{E}_{\min} = \begin{cases} 0 & \text{for } |I - I_a| \le z\sigma_{\tilde{I}}, \\ E_{\mathrm{rel}} - \frac{z\sigma_{\tilde{I}}}{I} & \text{for } |I - I_a| > z\sigma_{\tilde{I}}, \end{cases} \tag{22}$$

and the upper limit is given by

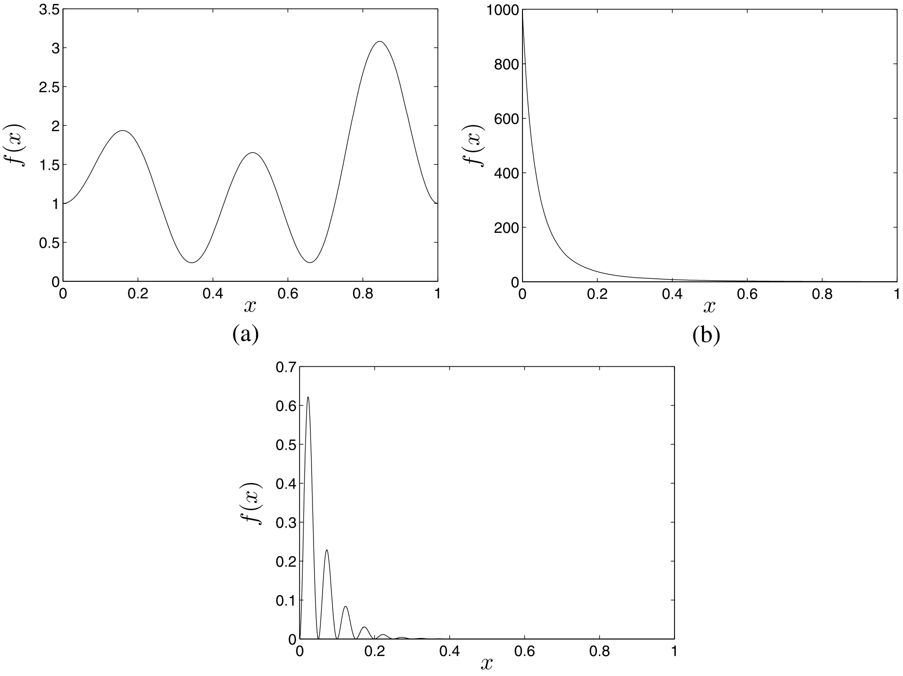
$$\tilde{E}_{\max} = \begin{cases} |\mu_E| + \sigma_E \Phi^{-1}[2\Phi(z) - \Phi(z + 2\frac{|\mu_E|}{\sigma_E})], & \text{for } |I - I_a| \le z\sigma_{\tilde{I}}, \\ |\mu_E| + \sigma_E \Phi^{-1}[\Phi(z) - \Phi(z - \frac{2|\mu_E|}{\sigma_E}) - \Phi(z + \frac{2|\mu_E|}{\sigma_E}) + 1], \\ \quad \text{for } |I - I_a| > z\sigma_{\tilde{I}}, \end{cases} \tag{23}$$

where Φ and Φ^{-1} are the standard normal cumulative distribution function and its inverse, respectively. We have thus constructed an α percent credible interval (e.g. see Bolstad 2007), where $\alpha = 100P(z)$, for the error $\tilde{E}_{\mathrm{rel}}$ of an estimate based on measured data. The quantities $\tilde{E}_{\min}$, $\tilde{E}_{\max}$ are the lower and upper limits of this credible interval, respectively.

It immediately follows from (22) and (23) that the impact noise in data makes on the approximation error is defined by the accuracy of the evaluation of pest abundance obtained from exact values of the pest population density, which in turn depends on the number N of traps where the data are available. In the next section, we illustrate this conclusion by various numerical examples.

3 Calculating the Pest Insect Abundance from the Noisy Density Function: Examples and Discussion

In this section, we perform some conventional numerical test cases to verify our approach. We then further investigate how introducing noise to the density function values affects the accuracy of the estimated pest abundance, and in particular we focus on the instance when the grid of traps is coarse. We follow the same methodology as used in (Petrovskaya and Petrovskii 2010) and begin by considering some continuous functions with various level of spatial complexity where we require that the exact pest abundance I is available in closed form. For each test case we generate a regularly spaced set of traps, and unless otherwise stated we take the unit interval [0, 1]



(a) (b)

(c)

Fig. 3 The test cases to validate the evaluation error $\tilde{E}_{\text{rel}}$. (**a**), (**b**), and (**c**) are defined over the unit interval [0, 1] by the functions given in (25), (29), and (30), respectively

to represent the agricultural field. Therefore, the traps are located as follows:

$$x_1 = 0, \qquad x_i = x_{i-1} + h, \quad \text{for } i = 2, \dots, N-1, \qquad x_N = 1, \tag{24}$$

where $h = (x_N - x_1)/(N-1)$ is the fixed distance between traps. The exact pest population densities are then given by $f_i \equiv f(x_i)$, $i = 1, \dots, N$.

Let us begin with a test case with simple behaviour whereby the function $f(x)$ has several wide peaks, as can be seen in Fig. 3a:

$$f(x) = \exp(x) \sin(3\pi x)^2 + \cos(\pi x)^2. \tag{25}$$

We fix the number N of traps and generate measured values of the pest density by perturbing each exact pest density f_i a total of $n_r = 100{,}000$ times according to the transformation (13). We therefore have n_r sets of measured values $\{\tilde{f}_i\}$. For each set of data an estimate of the pest abundance is obtained by implementing the compound trapezoidal rule (2) and the relative error is then calculated. To confirm that these $n_r = 100{,}000$ estimates of $\tilde{E}_{\text{rel}}$ are indeed realisations of a random variable with a folded normal distribution with mean $\mu_{\tilde{E}_{\text{rel}}}$ and standard deviation $\sigma_{\tilde{E}_{\text{rel}}}$, we calculate the sample mean

$$\bar{\mu}_{\tilde{E}_{\text{rel}}} = \frac{1}{N} \sum_{i=1}^{n_r} \tilde{E}_{\text{rel}_i}, \tag{26}$$

Table 1 Comparison between the theoretical mean and standard deviation of the quantity $\tilde{E}_{\mathrm{rel}}$ as defined by (20) and (21), and their numerical counterparts (26) and (27) over several grids of N traps. The theoretical means and standard deviations are shown in the columns labelled $\mu_{\tilde{E}_{\mathrm{rel}}}$ and $\sigma_{\tilde{E}_{\mathrm{rel}}}$, respectively, and the sample mean and standard deviations are labelled $\bar{\mu}_{\tilde{E}_{\mathrm{rel}}}$ and $s_{\tilde{E}_{\mathrm{rel}}}$. The relative difference between the theoretical quantity and its numerical counterpart is calculated in the last column of the table. Good agreement can be seen thus providing verification of our approach

N	$\mu_{\tilde{E}_{\mathrm{rel}}}$	$\bar{\mu}_{\tilde{E}_{\mathrm{rel}}}$	$\frac{\lvert\mu_{\tilde{E}_{\mathrm{rel}}}-\bar{\mu}_{\tilde{E}_{\mathrm{rel}}}\rvert}{\lvert\mu_{\tilde{E}}\rvert}$	$\sigma_{\tilde{E}_{\mathrm{rel}}}$	$s_{\tilde{E}_{\mathrm{rel}}}$	$\frac{\lvert\sigma_{\tilde{E}_{\mathrm{rel}}}-s_{\tilde{E}_{\mathrm{rel}}}\rvert}{\lvert\sigma_{\tilde{E}_{\mathrm{rel}}}\rvert}$
3	5.614872e–02	5.607518e–02	1.309661e–03	4.227882e–02	4.239840e–02	2.828365e–03
5	4.043406e–02	4.034606e–02	2.176191e–03	3.050063e–02	3.041125e–02	2.828365e–03
9	3.203438e–02	3.204198e–02	2.372352e–04	2.420232e–02	2.419244e–02	4.078940e–04
17	2.277488e–02	2.283417e–02	2.603297e–03	1.720666e–02	1.727488e–02	3.964279e–03
33	1.615665e–02	1.618614e–02	1.825433e–03	1.220652e–02	1.226183e–02	4.531462e–03
65	1.144294e–02	1.149041e–02	4.148032e–03	8.645263e–03	8.672099e–03	3.104113e–03

and the sample standard deviation

$$s_{\tilde{E}_{\mathrm{rel}}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{n_r}(\tilde{E}_{\mathrm{rel}_i} - \bar{\mu}_{\tilde{E}_{\mathrm{rel}}})^2}, \tag{27}$$

and make a comparison with the theoretical quantities given by (20) and (21), respectively.

We then establish the following proportion

$$P_{\mathrm{num}} = \frac{\tilde{n}_r}{n_r}, \tag{28}$$

where $\tilde{n}_r$ is the number of the relative errors $\tilde{E}_{\mathrm{rel}}$ which fall within the range $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ as defined by (22) and (23) in order to make a comparison with the theoretical probability $P(z)$. The number of traps is then increased as $2N - 1$ and the quantities (26)–(28) are recalculated.

We apply the above procedure to the test case (25), where the number of traps is subsequently increased to be $N = 3, 5, \ldots, 65$. We select the measurement tolerance as $\nu_m = 0.3$. As can be seen in Table 1, for each value of N we have good agreement between the sample mean $\bar{\mu}_{\tilde{E}_{\mathrm{rel}}}$ and the theoretical mean $\mu_{\tilde{E}_{\mathrm{rel}}}$, and likewise between the sample and theoretical standard deviations $s_{\tilde{E}_{\mathrm{rel}}}$ and $\sigma_{\tilde{E}_{\mathrm{rel}}}$. We fix $z = 3$. Therefore, we have the theoretical probability that $\tilde{E}_{\mathrm{rel}}$ lies within the range $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ as $P(z) \approx 0.9973$. It can be seen from Table 2 that the corresponding numerical probability P_{num} as given by (28) is indeed approximately 0.9973. We are therefore satisfied that the range given by (22) and (23) can be used to make reliable conclusions about the error $\tilde{E}_{\mathrm{rel}}$ of an estimated pest abundance based on measured data $\tilde{I}$.

We now directly compare the quantities E_{rel} and $\tilde{E}_{\mathrm{rel}}$ in order to understand how using noisy data rather than exact pest population densities impacts the accuracy of

Table 2 Comparison between the theoretical probability $P(z)$ as defined by (11) that $\tilde{E}_{\mathrm{rel}}$ lies within the range $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ and the numerical probability P_{num} computed according to (28) over a series of grids with N traps. We fix $z = 3$, thus $P(z) = P(3) \approx 0.9973$. The relative error between the two quantities is shown in the last column

N	P_{num}	$\frac{\lvert P(3)-P_{\mathrm{num}}\rvert}{\lvert P(3)\rvert}$
3	0.99732	1.984965e–05
5	0.99745	1.502016e–04
9	0.99722	8.042106e–05
17	0.99716	1.405835e–04
33	0.99739	9.003915e–05
65	0.99722	8.042106e–05

a pest abundance estimate. Let us introduce further test cases with an increased level of spatial complexity to consider alongside that prescribed by the function (25). The density is either concentrated in a narrow layer as defined by the following function (see Fig. 3b):

$$f(x) = (x + 0.1)^{-3}, \tag{29}$$

or is located within a small sub-domain of the unit interval and also exhibits oscillatory behaviour (see Fig. 3c):

$$f(x) = \exp(-20x)\sin(20\pi x)^2. \tag{30}$$

For an increasing number N of traps spaced regularly according to (24), the relative error $E_{\mathrm{rel}}(N)$ of an approximation based on exact data is calculated. The mean value $\mu_{\tilde{E}_{\mathrm{rel}}}$ of the error of an approximation based on measured values as well as the upper and lower bounds of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ are found from (20) and (22), (23), respectively, for the same set of values of N. The measurement tolerance is fixed as $\nu_m = 0.3$ throughout, and we set $z = 3$.

The corresponding graphs of the error as a function of the number N of traps (convergence curves) for each of the test cases are displayed in Fig. 4. An estimate of the integral I is considered to be accurate if it satisfies the condition (4). We select the tolerance $\tau = 0.25$ which lies within the acceptable range for ecological applications given in Sect. 2, and which has been recommended for routine monitoring (Robson and Regier 1964). The line $\tau = 0.25$ is therefore also plotted so as to determine when the estimates become sufficiently accurate.

It can be seen in Fig. 4a that for the spatially simpler test case (25), the estimates based on exact data are sufficiently accurate for the entire range of the number N of traps considered in the problem. The curve E_{rel} always lies below the line $\tau = 0.25$. It is also evident from the figure that the addition of noise to the data significantly slows the convergence of the pest abundance estimate to the exact value when we increase the number of traps. Clearly, the curve for the mean error based on perturbed data $\mu_{\tilde{E}_{\mathrm{rel}}}(N)$ has a less steep gradient than its $E_{\mathrm{rel}}(N)$ counterpart. This is because whilst the uncertainty associated with the estimate based on measured values decreases as the number of traps N increases, the contribution to the mean error $\mu_{\tilde{E}_{\mathrm{rel}}}$ from the noise is more dominant than that of the integration error E_{rel}. In other words, the uncertainty decreases at a slower rate than the integration er-

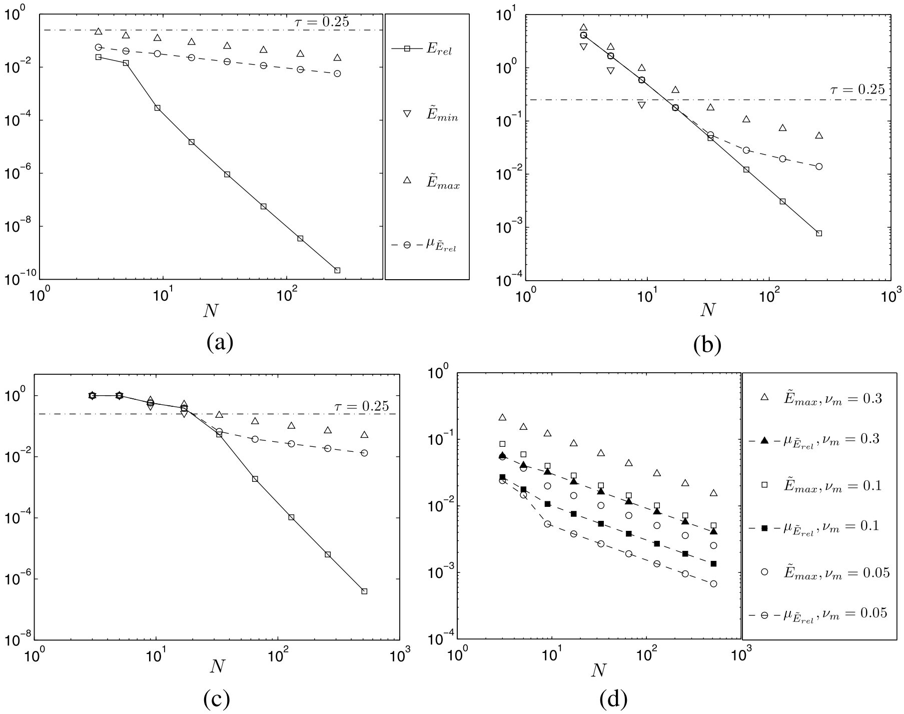


Fig. 4 (**a**)–(**c**) The error for the approximation based on exact data E_{rel} is compared with the mean error $\mu_{\tilde{E}_{\mathrm{rel}}}$ of an approximation based on noisy data alongside the limits of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ for the test cases (25), (29) and (30), respectively, as shown in Fig. 3a–3c. The measurement tolerance is fixed as $\nu_m = 0.3$ and $z = 3$ in each case. The *legend* for each figure is as shown in (**a**). (**d**) Mean error $\mu_{\tilde{E}_{\mathrm{rel}}}$ of an approximation based on noisy data and the upper limit of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ for the test case (25) as shown in Fig. 3a where values $\nu_m = 0.05, 0.1, 0.3$ of the measurement tolerance have been selected. We fix $z = 3$ as before

ror decreases. Meanwhile, it is important to note the mean error $\mu_{\tilde{E}_{\mathrm{rel}}}$ does converge to zero in the theoretical limit of an infinite number of traps (e.g. see Cox 2007).

For the test case above, the $\tilde{E}_{\max}$ curve entirely lies below the upper threshold $\tau = 0.25$ of the desired accuracy. The lower bound of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ is $\tilde{E}_{\min} \equiv 0$ as the estimate based on exact values $\tilde{I}$ is within $z\sigma_{\tilde{I}}$ of the exact pest abundance I right from the initial estimate, where we have chosen $z = 3$. The value $\tilde{E}_{\min} = 0$ is not displayed since the plots are given on a logarithmic scale.

Meanwhile, for the more spatially complex density distributions (29) and (30), the number of traps N has to be sufficiently increased before the desired level of accuracy $E \leq \tau = 0.25$ is obtained (see Figs. 4b and 4c). Similarly, there needs to be some level of grid refinement before the lower limit becomes $\tilde{E}_{\min} = 0$. Prior to this occurring, the mean error $\mu_{\tilde{E}_{\mathrm{rel}}}$ lies close to the error for the unperturbed data set E_{rel} as indeed does $\tilde{E}_{\max}$. After the lower limit of the credible interval for $\tilde{E}_{\mathrm{rel}}$ becomes $\tilde{E}_{\min} = 0$, a difference in the convergence rates becomes evident with the convergence of the perturbed data becoming much slower.

One feature of the graph in Fig. 4c has to be mentioned here. In the case of the initial estimates formulated from $N = 3$ and $N = 5$ trap counts, it can be seen that the upper and lower limits of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ lie extremely close to the error based on exact data E_{rel}. This is an artefact of the way in which each measured value of pest density $\tilde{f}_i$ is considered to be related to the true value f_i; each measured value is considered to be within some percentage of the true value. The function values at the initial $N = 3$ trap locations, which we recall are regularly distributed across the interval $[0, 1]$, are extremely small in magnitude meaning the resulting uncertainty is also very small. This is also the case on the subsequent grid of $N = 5$ traps, whereas, when the number of traps is increased to $N = 9$, some function values with a larger magnitude are detected and hence the uncertainty is larger in comparison to that associated with the previous estimate.

So far we have looked at how noise impacts the accuracy of an estimate of the pest abundance for a fixed measurement tolerance of ν_m. We now investigate the impact of noise on an estimate's accuracy as the quantity ν_m is varied. Let us again consider the simpler test case (25) as shown in Fig. 3a. Figure 4d shows the convergence curves for different values of the measurement tolerance: $\nu_m = 0.05, 0.1$ and 0.3 where z is fixed as $z = 3$. It can be seen that increasing the measurement tolerance causes the convergence curve to shift upwards; greater uncertainty associated with the set of measured values $\{\tilde{f}_i\}$ gives rise to greater uncertainty associated with the estimate formulated from this data set as one would expect. Obviously, the point at which the error becomes acceptable, that is, it falls below the upper threshold of $\tau = 0.25$, occurs later meaning a larger number of traps would be needed to acquire a sufficiently accurate estimate.

3.1 Ecological Test Cases

Although informative, the test cases above were chosen for their mathematically interesting characteristics rather than their direct relevance to the pest monitoring problem. Therefore, we now turn our attention to some ecologically meaningful test cases. We require the ability to repeat estimates of the pest abundance for the same density function for an increased number of traps. It is difficult to find field data in a one-dimensional domain which would be suitable for our purpose, so we simulate data using the spatially explicit form of what we consider the predator–prey model with the Allee effect (Murray 1989; Turchin 2003). The dimensionless form of the model is given by the following system of equations:

$$\begin{aligned} \frac{\partial f(x,t)}{\partial t} &= d\frac{\partial^2 f}{\partial x^2} + f(1-f) - \frac{fg}{f+p}, \\ \frac{\partial g(x,t)}{\partial t} &= d\frac{\partial^2 g}{\partial x^2} + k\frac{fg}{f+p} - mg, \end{aligned} \tag{31}$$

where $f(x,t)$ is the density of the prey which we consider to be the pest insect and $g(x,t)$ is that of some predatory species at position x and time $t > 0$, d is the diffusion coefficient, p is the half-saturation prey density, k is the food assimilation

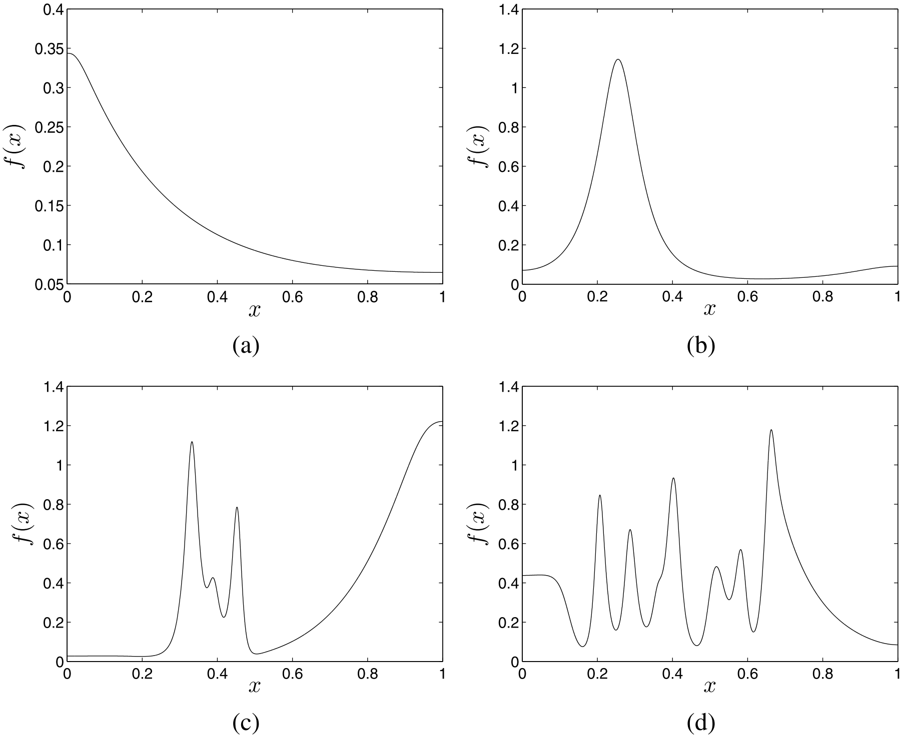
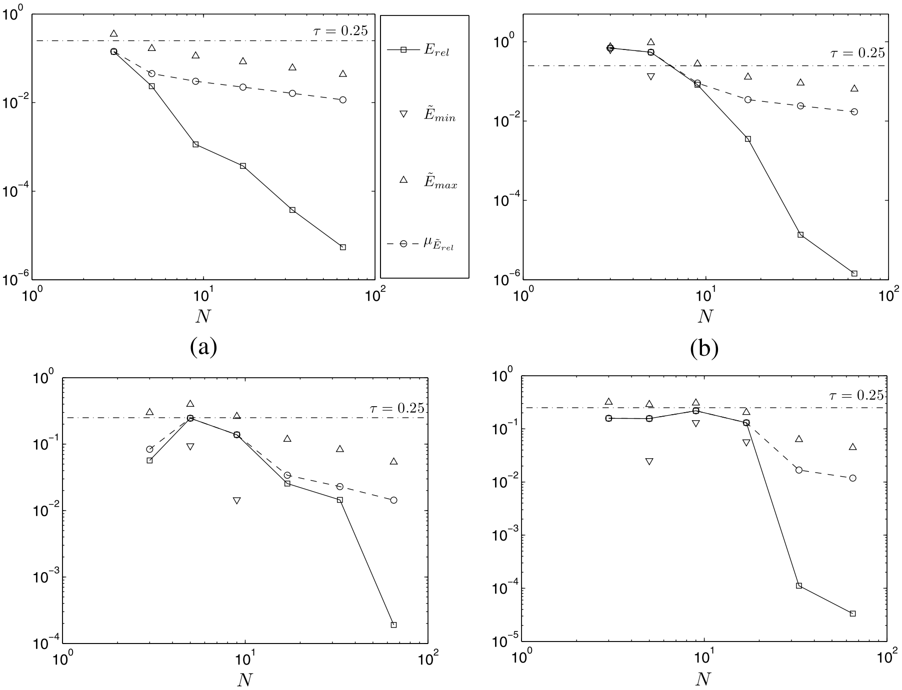


Fig. 5 Ecologically meaningful test cases as generated by the model (31) at different times t and for various choices of the dimensionless diffusion coefficient d: (**a**) $t = 5$, $d = 10^{-4}$ (**b**) $t = 50$, $d = 10^{-4}$ (**c**) $t = 100$, $d = 10^{-5}$ (**d**) $t = 400$, $d = 10^{-5}$. The reader is referred to Petrovskaya and Petrovskii (2010) for the choices of initial and boundary conditions

efficiency coefficient and m is the predator mortality. We fix the time as $t = \tilde{t} > 0$ and numerically solve the system of equations (31) to obtain the pest population density $f(x, \tilde{t})$. Since $\tilde{t}$ is fixed we shall henceforth denote this as simply $f(x)$. This is done for different values of the parameters in the model to generate four ecologically meaningful test cases which are shown in Fig. 5. The monotone test case as shown in Fig. 5a and the single peak test case (see Fig. 5b) are fairly simple in terms of spatial complexity. The pest density function shown in Fig. 5c, which we will refer to as the three peak test case, and the multi-peak test case (see Fig. 5d) are examples of more complex spatial heterogeneity. These test cases are the same as those discussed in Petrovskaya and Petrovskii (2010), therefore the interested reader is referred to this paper for the initial and boundary conditions that were used in their generation and for further details of the numerical solution.

The density $f(x)$ is found by numerically solving (31) at the positions of a large number N_f of regularly distributed traps; we take $N_f = 2^{15} + 1$. Since the pest density function for each of the ecological test cases is obtained as a result of numerical solution, the exact pest abundance I is not available. The 'exact' pest abundance I is then computed using the compound trapezoidal rule (2) from the exact data f_i obtained on a very fine grid of N_f traps. Once we have found the values of the pest



(a) (b) (c) (d)

Fig. 6 The error for the approximation based on exact data E_{rel} is compared with the mean error $\mu_{\tilde{E}_{\text{rel}}}$ of an approximation based on noisy data and the limits of the range $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ for the ecologically meaningful (**a**) monotone, (**b**) single peak, (**c**) three-peak and (**d**) multi-peak test cases as shown in Figs. 5a–5d, respectively. The measurement tolerance is fixed as $\nu_m = 0.3$ and $z = 3$ in each case. The *legend* for all figures is as shown in (**a**)

density function $f(x)$ at the trap locations $x_i, i = 1, \dots, N_f$, we can find estimates $I_a(N)$ of the pest abundance for any smaller number N of traps by extracting the relevant pest density function values from this data set and applying the same evaluation rule (2).

Let us fix the number of traps as $N = N_1$. As before an estimate I_a based on exact data is formed by employing the trapezoidal rule (2) and the relative error E_{rel} is calculated from (3). The mean error $\mu_{\tilde{E}_{\text{rel}}}$ of an estimate formed from noisy data is found from (20) and the limits of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ are calculated from (22) and (23). The number of traps is then increased as $2N_1 - 1$ and the above is repeated. This is done several times and the corresponding convergence curves are shown in Fig. 6. The measurement tolerance is fixed as $\nu_m = 0.3$ and we also set $z = 3$.

The results of the ecological test cases reconfirm our earlier findings. If the number N of traps installed can resolve the spatial pattern of the density function $f(x)$ and can therefore provide good accuracy of evaluation, then noise makes visible impact on the evaluation error. In other words, if for a given N the distance between the estimate based on exact data I_a and the exact abundance I remains within z multiples of the standard deviation $\sigma_{\tilde{I}}$, then the convergence curve for the error of the estimate based on exact data E_{rel} differs significantly from the mean error $\mu_{\tilde{E}_{\text{rel}}}$ of the estimate based on perturbed data. That can been seen in Fig. 6a where the results for the monotone density distribution of Fig. 5a are presented. For the monotone function the accuracy of evaluation is already good on coarse grids (e.g. see $N = 5$ in the graph) and the error E_{rel} obtained from exact data is several orders of magnitude smaller than the mean error $\mu_{\tilde{E}_{\text{rel}}}$ when N increases. However, it is important to emphasize here that (a) the mean error is already below the required tolerance even on very coarse grids and (b) as we already mentioned in our previous discussion, the mean error converges to zero as the number N of traps grows infinitely large.

On the other hand, if the estimate based on unperturbed data I_a is already poor, then the introduction of noise makes little difference to the accuracy of evaluation. This behaviour is shown in Figs. 6b–6d where the complex spatial density distributions are not well resolved on initial grids with a small number N of traps. As a result, the curves E_{rel} and $\mu_{\tilde{E}_{\text{rel}}}$ lie close to each other.

It should be mentioned that, as shown in Figs. 6c and 6d for both the three peak and multi-peak test cases, the quantity $\tilde{E}_{\min}$ on the initial grid of $N = 3$ traps is $\tilde{E}_{\min} = 0$ whereas for a number of subsequent grids it becomes nonzero before eventually returning to zero. It is by chance only that for these test cases the initial estimate on a grid of $N = 3$ nodes is sufficiently accurate to satisfy the condition $|I - I_a| \leq z\sigma_{\tilde{I}}$; see also our discussion of the test case (30). However, the distance between the estimate based on exact data I_a and the exact abundance I does not decrease fast enough to remain within z multiples of the standard deviation $\sigma_{\tilde{I}}$ until the grid of traps is sufficiently refined.

A generic behaviour of the approximation error is that the accuracy of approximation I_a worsens when the spatial complexity of the density function increases (Petrovskaya and Embleton 2014; Petrovskaya and Petrovskii 2010; Petrovskaya et al. 2012). Consequently the number of traps for which the error falls solidly below the required tolerance increases when the spatial density evolves from a monotone function to a multi-peak density distribution. It can be seen from Fig. 6d that for a multi-peak density function (i.e. the function that presents an ecologically important case of a patchy population density) the impact of noise is negligible when the number of traps is within the range $N \sim 10$ used in ecological applications. While this result should be further validated for two-dimensional density distributions, it may help ecologists to make a correct decision about accuracy of evaluation on coarse grids of traps.

4 Concluding Remarks

In our paper, the problem of pest insect abundance evaluation has been discussed. We have considered a trapping procedure where information about the pest population density $f(x)$ at trap locations is then used in a numerical integration problem in order to calculate an estimate of the total pest population size. Since a continuous density function $f(x)$ is replaced with a discrete set of function values f_i, $i = 1, 2, \ldots, N$, exact computation of the pest abundance is impossible and an evaluation (approximation) error is inevitably present in the problem.

The approximation error is the main indicator of the accuracy of an evaluation, and correct estimation of this error is extremely important in ecological problems. Accurate evaluation of the total pest population size remains a crucial requirement in any IPM programme, as it allows one to avoid making an unjustified decision about control action (e.g. application of pesticides). Generally, the approximation error depends on the number N of trap locations where the values f_i of the density function are available. Also, for any fixed N the approximation error depends on the spatial pattern of the density function.

The standard definition of the approximation error implies that an approximation of the pest abundance is based on exact data f_i, $i = 1, 2, \ldots, N$. However, random error (noise) should be expected when the information about the density function is collected. Thus in this paper the aim of our research was to incorporate noise into the evaluation procedure and further investigate the approximation error when the pest population density function is randomly perturbed at any trap location.

The main results of the paper are as follows:

1. We have suggested a novel approach to handling the approximation error when the pest abundance evaluation is based on randomly perturbed data. Evaluation is not based on statistical methods but is done using a numerical integration technique. An advantage of numerical integration methods over a standard statistical approach is that they offer better accuracy of evaluation for a wide range of spatial density distributions and are therefore considered as a promising alternative to the existing statistical methods of evaluation.
2. In the paper, we have first explained a numerical integration procedure under the assumption that the data used for evaluation are exact. We then incorporated noise in density measurements into the numerical integration formulation of the pest abundance problem. The mean approximation error has been obtained along with the range to which $\tilde{E}_{\text{rel}}$ belongs with probability $P(z)$. In other words, we have constructed an α percent credible interval $[\tilde{E}_{\text{min}}, \tilde{E}_{\text{max}}]$ for the error $\tilde{E}_{\text{rel}}$ of an estimate based on measured data, where $\alpha = 100P(z)$. The theoretical results obtained in the paper have been verified for various one-dimensional density distributions when a selected method of integration (the composite trapezoidal rule) is applied in the problem.
3. We have demonstrated that the error induced by noise in the pest population density data depends on the accuracy of evaluation obtained when exact density values are considered. In particular, the credible interval we have established for $\tilde{E}_{\text{rel}}$ contains zero if the estimate of pest abundance I_a formed in the absence of noise is sufficiently accurate. Otherwise the lower bound of this interval E_{min} will be greater than zero.
4. One ecologically important case studied in the paper is approximation on coarse grids where the number N of traps is small. It has been shown, perhaps contrary to intuitive thinking, that the impact of noise is negligible when the data available are sparse. In other words, the accuracy of evaluation on coarse grids can already be so poor that noise in field measurements of the pest population density does not make any significant contribution. This result has been numerically confirmed for ecologically meaningful data.

5. Numerical experiments also revealed that, when we increase the number of traps, noise becomes a dominant feature of the approximation and the mean error may differ from the approximation error obtained on exact values of the density function by several orders of magnitude. Our results confirm that the mean error converges to zero for an infinitely large number of traps. However, the convergence rate of the mean error is much slower than the convergence rate of the approximation error obtained when exact data are used for approximation. Some theoretical justification of this phenomenon has been provided in the literature (Cox 2007), but this issue requires further study with regard to ecological applications and should become the focus of our future research. In particular, we intend to compare the results obtained for uncorrelated noise (as discussed in this paper) with the case when the noise in neighbouring traps is correlated.

It is worth noting here that the approach developed in the paper is general enough and can be readily extended to multi-dimensional problems. As soon as the weight coefficients in the numerical integration method (1) are defined, our computation of the mean error along with the credible interval for $\tilde{E}_{\mathrm{rel}}$ does not rely upon the dimension of the physical space. Hence, our future work will be focused on two-dimensional problems where field data are available from real-life measurements. Another important direction of future work is to study the impact of noise when different methods are employed to evaluate the pest abundance. In our paper, we have only used the trapezoidal rule (2), while applying other methods of numerical integration (e.g. Simpson's rule) may give an estimate of pest abundance that is more accurate on coarse grids of traps. It has been shown in the paper that the accuracy of approximation on exact data is crucial when the ecologically relevant situation of sparse data is considered. Hence our research will be focused on further careful investigation of evaluation methods that can provide good accuracy on coarse grids of traps.

Appendix: Finding a Credible Interval for the Relative Error in the Presence of Noise

We seek the upper and lower limit of the interval $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ to which the quantity $\tilde{E}_{\mathrm{rel}}$ belongs with probability $P(z)$ given by (11) as discussed in Sect. 2.3. We recall that the estimate of pest abundance $\tilde{I}$ calculated from measured data is a realisation of a normally distributed random variable with mean $\mu_{\tilde{I}} = I_a$ and standard deviation $\sigma_{\tilde{I}}$ as defined by (15). Thus any realisation $\tilde{I}$ lies within the interval $[I_a - z\sigma_{\tilde{I}}, I_a + z\sigma_{\tilde{I}}]$ with probability $P(z)$. We use this credible interval for $\tilde{I}$ to construct a credible interval for $\tilde{E}_{\mathrm{rel}}$. We consider two cases based on the distance between the approximate integral formed from exact data I_a and the exact value of the integral I. Let us begin by finding the lower limit of the interval, $\tilde{E}_{\min}$.

Case 1. $|I - I_a| \leq z\sigma_{\tilde{I}}$. In this case, as can be seen from Fig. 7a, an estimate based on measured data $\tilde{I}$ which belongs to the range $[I_a - z\sigma_{\tilde{I}}, I_a + z\sigma_{\tilde{I}}]$ can coincide with the exact value of the integral. Therefore, the lower limit of the range $[\tilde{E}_{\min}, \tilde{E}_{\max}]$ is

$$\tilde{E}_{\min} = 0. \tag{32}$$

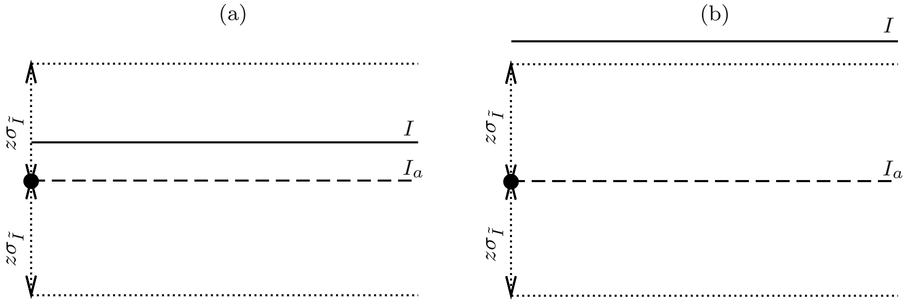


Fig. 7 Finding the interval $[\tilde E_{\text{rel}}, \tilde E_{\max}]$ to which $\tilde E_{\text{rel}}$ belongs with probability $P(z)$. (**a**) Case 1. $|I - I_a| \le z\sigma_{\tilde I}$. In this case, the exact value of the integral I lies within the credible interval for $\tilde I$ thus the lower limit of the credible interval for $\tilde E_{\text{rel}}$ is $\tilde E_{\min} = 0$. (**b**) Case 2. $|I - I_a| > z\sigma_{\tilde I}$. The exact value of the integral I lies outside, thus the interval $[\tilde E_{\min}, \tilde E_{\max}]$ does not include the zero value

Case 2. $|I - I_a| > z\sigma_{\tilde I}$. In this instance, from Fig. 7b we can see that the range $[I_a - z\sigma_{\tilde I}, I_a + z\sigma_{\tilde I}]$ does not include the exact value of the integral I. Either we have $I_a \le I$ in which case we can see that

$$\tilde E_{\min} = \frac{|I - I_a - z\sigma_{\tilde I}|}{|I|},$$

or we have $I_a > I$; therefore,

$$\tilde E_{\min} = \frac{|I - I_a + z\sigma_{\tilde I}|}{|I|}.$$

In both cases,

$$\tilde E_{\min} = E_{\text{rel}} - \frac{z\sigma_{\tilde I}}{I}, \tag{33}$$

which is a strictly positive quantity as the condition $|I - I_a| > z\sigma_{\tilde I}$, of course, means that $E_{\text{rel}} > z\sigma_{\tilde I}/I$, where we recall that $I > 0$.

It should be mentioned that a zero relative error is still possible in the second case, when the distance between the approximation based on exact data and the true value of the integral exceeds z multiples of the standard deviation $\sigma_{\tilde I}$, however, we choose to fix $\tilde E_{\text{rel}}$ as

$$\tilde E_{\min} = \begin{cases} \min\{E \ge 0 : E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E]\}, & \text{for } \mu_E \ge 0, \\ |\max\{E \le 0 : E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E]\}|, & \text{for } \mu_E < 0, \end{cases}$$

where E is defined by (16). In other words, we find the value of the quantity E closest to zero which lies within the range (18) and then take the absolute value as $\tilde E_{\min}$ (see Fig. 2).

Let us now consider the upper limit $\tilde E_{\max}$ of the credible interval of $\tilde E_{\text{rel}}$. To find $\tilde E_{\max}$, we use the condition that any single value of $\tilde E$ lies within the range $[\tilde E_{\min}, \tilde E_{\max}]$ with fixed probability $P(z)$ as defined by (11). As mentioned above,

$\tilde{E}_{\mathrm{rel}}$ is a realisation of a random variable with a folded normal distribution. This distribution is formed by reflecting the negative quantities of the distribution (17) of the auxiliary error E in the y-axis. Unless the mean value of this underlying normal distribution is $\mu_E = 0$, if we take $\tilde{E}_{\max} = \mu_E + z\sigma_E$ then the probability $\hat{P}$ that $\tilde{E}_{\mathrm{rel}}$ lies within the above range will exceed $P(z)$. We shall denote the additional contribution as P^*; therefore,

$$\hat{P} = P(z) + P^*.$$

We now seek the appropriate value of the upper limit $\tilde{E}_{\max}$ in order to satisfy the condition that $\hat{P} = P(z)$. Let us temporarily impose the restriction $\mu_E \geq 0$. As when constructing the lower limit $\tilde{E}_{\min}$, we consider the cases when the distance between the approximation based on exact data I_a and the true value of the integral I exceeds or is within z multiples of the standard deviation $\sigma_{\tilde{I}}$ separately.

Case 1. $|I - I_a| \leq z\sigma_{\tilde{I}}$. As shown in Fig. 2a, the probability P^* is given by

$$P^* = \int_{-\mu_E - z\sigma_E}^{\mu_E - z\sigma_E} p(E)\, dE. \tag{34}$$

In order to satisfy the condition $\hat{P} = P(z)$, we must then find $\tilde{E}_{\max}$ such that

$$\int_{\tilde{E}_{\max}}^{\mu_E + z\sigma_E} p(E)\, dE = P^*. \tag{35}$$

Using the transformation

$$E \rightarrow \frac{E - \mu_E}{\sigma_E}$$

from (34) and (35), we obtain the following in terms of the standard normal distribution function Φ:

$$\Phi(-z) - \Phi\left(\frac{-2\mu_E}{\sigma_E} - z\right) = \Phi(z) - \Phi\left(\frac{\tilde{E}_{\max} - \mu_E}{\sigma_E}\right).$$

Rearranging gives

$$\tilde{E}_{\max} = \mu_E + \sigma_E \Phi^{-1}\left[2\Phi(z) - \Phi\left(z + 2\frac{\mu_E}{\sigma_E}\right)\right]. \tag{36}$$

Case 2. $|I - I_a| > z\sigma_{\tilde{I}}$. Similar calculations for this case as illustrated in Fig. 2b yield

$$\tilde{E}_{\max} = \mu_E + \sigma_E \Phi^{-1}\left[\Phi(z) - \Phi\left(z - \frac{2\mu_E}{\sigma_E}\right) - \Phi\left(z + \frac{2\mu_E}{\sigma_E}\right) + 1\right]. \tag{37}$$

Earlier we assumed $\mu_E \geq 0$. Since the probability density function (19) for the folded normal distribution is the same for mean μ_E as it is for $-\mu_E$, we can replace the term μ_E for $|\mu_E|$ in Eqs. (36) and (37) so that they hold for arbitrary μ_E.

References

Alexander, C. J., Holland, J. M., Winder, L., Woolley, C., & Perry, J. N. (2005). Performance of sampling strategies in the presence of known spatial patterns. *Ann. Appl. Biol.*, *146*, 361–370.

Binns, M. R., Nyrop, J. P., & van der Werf, W. (2000). *Sampling and monitoring in crop protection: the theoretical basis for designing practical decision guides*. Wallingford: CABI Publishing.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML (2008). *Evaluation of measurement data—guide to the expression of uncertainty in measurement, JCGM 100: 2008 GUM 1995 with minor corrections* (1st ed.). Paris: Joint Committee for Guides in Metrology.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics* (2nd ed.). New York: Wiley.

Browde, J. A., Pedigo, L. P., Degooyer, T. A., Higley, L. G., Wintersteen, W. K., & Zeiss, M. R. (1992). Comparison of sampling techniques for grasshoppers (Orthoptera: Acrididae) in soybean. *J. Econ. Entomol.*, *85*, 2270–2274.

Burn, A. J., Coaker, T. H., & Jepson, P. C. (Eds.) (1987). *Integrated pest management*. New York: Academic Press.

Byers, J. A., Anderbrant, O., & Löfqvist, J. (1989). Effective attraction radius: a method for comparing species attractants and determining densities of flying insects. *J. Chem. Ecol.*, *15*, 749–765.

Cox, M. G. (2007). The area under a curve specified by measured values. *Metrologia*, *44*, 365–378.

Davis, P. M. (1994). Statistics for describing populations. In L. P. Pedigo & G. D. Buntin (Eds.), *Handbook of sampling methods for arthropods in agriculture* (pp. 33–54). Boca Raton: CRC Press.

Davis, P. J., & Rabinowitz, P. (1975). *Methods of numerical integration*. New York: Academic Press.

Dent, D. (2000). *Insect pest management* (2nd ed.). Wallingford: CABI Publishing.

Embleton, N. L., & Petrovskaya, N. B. (2013). On numerical uncertainty in evaluation of pest population size. *Ecol. Complex.*, *14*, 117–131.

Evans, E. W., Rogers, R. A., & Opfermann, D. J. (1983). Sampling grasshoppers (Orthoptera: Acrididae) on burned and unburned tallgrass prairie: night trapping vs. sweeping. *Environ. Entomol.*, *12*, 1449–1454.

Ferguson, A. W., Klukowski, Z., Walczak, B., Perry, J. N., Mugglestone, M. A., Clark, S. J., & Williams, I. H. (2000). The spatio-temporal distribution of adult Ceutorhynchus assimilis in a crop of winter oilseed rape in relation to the distribution of their larvae and that of the parasitoid Trichomalus perfectus. *Entomol. Exp. Appl.*, *95*, 161–171.

Grimmett, G. R., & Stirzaker, D. R. (2001). *Probability and random processes* (3rd ed.). Oxford: Oxford University Press.

Holland, J. M., Perry, J. N., & Winder, L. (1999). The within-field spatial and temporal distribution of arthropods in winter wheat. *Bull. Entomol. Res.*, *89*, 499–513.

Kogan, M. (1998). Integrated pest management: historical perspectives and contemporary developments. *Annu. Rev. Entomol.*, *43*, 243–270.

Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, *3*, 543–550.

Mayor, J. G., & Davies, M. H. (1976). A survey of leatherjacket populations in South-West England 1963–1974. *Plant Pathol.*, *25*, 121–128.

Metcalf, R. L. & Luckmann, W. H. (Eds.) (1982). *Introduction to insect pest management* (2nd ed.). London: Wiley.

Murray, J. D. (1989). *Mathematical biology*. Berlin: Springer.

Northing, P. (2009). Extensive field based aphid monitoring as an information tool for the UK seed potato industry. *Asp. Appl. Biol.*, *94*, 31–34.

Pascual, M. A., & Kareiva, P. (1996). Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology*, *77*, 337–349.

Pedigo, L. P., & Rice, M. E. (2009). *Entomology and pest management* (6th ed.). New York: Prentice Hall.

Petrovskaya, N. B., & Embleton, N. L. (2013). Evaluation of peak functions on ultra-coarse grids. *Proc. R. Soc. A.*, *469*. doi:10.1098/rspa.2012.0665.

Petrovskaya, N. B., & Embleton, N. L. (2014, submitted). Computational methods for accurate evaluation of pest insect population size. In W. A. C. Godoy & C. P. Ferreira (Eds.) *Ecological modelling applied to entomology*. Berlin: Springer.

Petrovskaya, N. B., & Petrovskii, S. V. (2010). The coarse-grid problem in ecological monitoring. *Proc. R. Soc. A*, *466*, 2933–2953.

Petrovskaya, N. B., & Venturino, E. (2011). Numerical integration of sparsely sampled data. *Simul. Model. Pract. Theory*, *19*, 1860–1872.

Petrovskaya, N. B., Petrovskii, S. V., & Murchie, A. K. (2012). Challenges of ecological monitoring: estimating population abundance from sparse trap counts. *J. R. Soc. Interface*, *9*, 420–435.

Petrovskaya, N. B., Embleton, N. L., & Petrovskii, S. V. (2013). Numerical study of pest population size at various diffusion rates. In M. A. Lewis, P. K. Maini, & S. V. Petrovskii (Eds.), *Lecture notes in mathematics: Vol. 2071. Dispersal, individual movement and spatial ecology, a mathematical perspective* (pp. 355–385). Berlin: Springer.

Petrovskii, S., Bearup, D., Ahmed, D. A., & Blackshaw, R. (2012). Estimating insect population density from trap counts. *Ecol. Complex.*, *10*, 69–82.

Raworth, D. A., & Choi, M.-Y. (2001). Determining numbers of active carabid beetles per unit area from pitfall-trap data. *Entomol. Exp. Appl.*, *98*, 95–108.

Robson, D. S., & Regier, H. A. (1964). Sample size in Petersen mark–recapture experiments. *Trans. Am. Fish. Soc.*, *93*, 215–226.

Sherratt, J. A., & Smith, M. J. (2008). Periodic travelling waves in cyclic populations: field studies and reaction–diffusion models. *J. R. Soc. Interface*, *5*, 483–505.

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames: Iowa State University Press.

Steel, R. G. D., & Torrie, J. H. (1960). *Principles and procedures of statistics*. CAB Direct.

Topping, J. (1972). *Errors of observation and their treatment* (4th ed.). London: Chapman & Hall.

Turchin, P. (2003). *Complex population dynamics: a theoretical/empirical synthesis*. Princeton: Princeton University Press.