Handling Sparse Spatial Data in Ecological Applications

by

Nina Lois Embleton

A thesis submitted to
The University of Birmingham
for the degree of
Doctor of Philosophy (PhD)

School of Mathematics
The University of Birmingham
September 2014

Abstract

Estimating the size of an insect pest population in an agricultural field is an integral part of insect pest monitoring. An abundance estimate can be used to decide if action is needed to bring the population size under control, and accuracy is important in ensuring that the correct decision is made. Conventionally, statistical techniques are used to formulate an estimate from population density data obtained via sampling.

This thesis thoroughly investigates an alternative approach of applying numerical integration techniques. We show that when the pest population is spread over the entire field, numerical integration methods provide more accurate results than the statistical counterpart. Meanwhile, when the spatial distribution is more aggregated, the error behaves as a random variable and the conventional error estimates do not hold. We thus present a new probabilistic approach to assessing integration accuracy for such functions, and formulate a mathematically rigorous estimate of the minimum number of sample units required for accurate abundance evaluation in terms of the species diffusion rate. We show that the integration error dominates the error introduced by noise in the density data and thus demonstrate the importance of formulating numerical integration techniques which provide accurate results for sparse spatial data.

Acknowledgments

I would like to take this opportunity to thank Dr Natalia Petrovskaya for her excellent supervision. I am extremely grateful for her expert advice and all of the support that she has provided throughout my studies. I would also like to thank Prof Sergei Petrovskii for his guidance and collaboration.

Thank you to my examining committee, Prof Rod Blackshaw, Dr David Leppinen, and Dr Sándor Németh for their time and expert advice. My gratitude goes to the School of Mathematics at the University of Birmingham for providing me with the scholarship which funded my research, and also to the Society of Mathematical Biology for providing me with additional conference funding. Thank you to the staff and students I have met during the course of my PhD and who have made the experience so enjoyable.

I give my thanks to the reviewers of my publications for the valuable feedback, and I am also grateful for all of the helpful comments and suggestions made by audience members of my research talks.

I would like to thank my family and friends who have been so supportive during my academic studies. This thesis is dedicated to my partner David, without whose love and encouragement, I would not have been able to write a single word.

Finally, thank you to those who take the time to read this thesis.

Publication List

Below is a list of publications which have been produced using work presented in this thesis.

- N. L. Embleton & N. B. Petrovskaya, On the Accuracy of Estimating Pest Insect Abundance from Data with Random Error, *Ecological Complexity*, DOI: 10.1016/j.ecocom.2014.05.006
- N. B. Petrovskaya & N. L. Embleton (in press), *Computational Methods for Accurate Evaluation of Pest Insect Population Size*, in W. A. C. Godoy & C. P. Ferreira (eds.), *Ecological Modelling Applied to Entomology*, Springer-Verlag, Berlin Heidelberg, pp. 171–218
- N. L. Embleton & N. B. Petrovskaya (2014), A Novel Approach to Evaluation of Pest Insect Abundance in the Presence of Noise, *Bulletin of Mathematical Biology*, vol.76, pp.718–743, DOI: 10.1007/s11538-014-9940-z
- N. L. Embleton & N. B. Petrovskaya (2013), On Numerical Uncertainty in Evaluation of Pest Population Size, *Ecological Complexity*, vol.14, pp.117–131, DOI: 10.1016/j.ecocom.2012.11.004
- N. B. Petrovskaya, & N. L. Embleton (2013), Evaluation of Peak Functions on Ultra-Coarse Grids, *Proceedings of the Royal Society A*, vol.469, DOI: 10.1098/rspa.2012.0665
- N. B. Petrovskaya, N. L. Embleton, & S. V. Petrovskii (2013), *Numerical Study of Pest Population Size at Various Diffusion Rates*, in M. A. Lewis et al. (eds.) *Dispersal, Individual Movement and Spatial Ecology, Lecture Notes in Mathematics, vol.2071*, Springer-Verlag, Berlin Heidelberg, pp. 355–385, DOI: 10.1007/978-3-642-35497-7_13

Contents

Nomenclature

Unless otherwise stated, the following notation is used throughout this thesis.

List of Acronyms

Acronym	**Description**
1D	One-Dimensional
2D	Two-Dimensional
IPM	Integrated Pest Management

List of Latin Symbols

Symbol	**Description**
d	Dimensionless diffusion rate of pest species
D	Domain representing the agricultural field
E_{rel}	Relative error of approximation formed from exact data
$\tilde{E}_{rel}$	Relative error of approximation formed from noisy data
f	Pest population density
h	Distance between sample units / grid step size
k	Degree of interpolating polynomial
L	Length of the agricultural field (one-dimensional problem)
I	Exact pest abundance
I_a	Estimate of pest abundance formed from exact data

$\tilde{I}$	Estimate of pest abundance formed from noisy data
M	Sample mean pest population density
$\bar{M}$	True mean pest population density
N	Number of sample units/grid nodes
p	Probability of achieving a sufficiently accurate estimate
$P_k(x)$	An interpolating polynomial of degree k
q	Order of convergence of a method of numerical integration
r	Degree of precision of a method of numerical integration
S	Area of the agricultural field (two-dimensional problem)
w	Weight of a numerical integration formula

List of Greek Symbols

Symbol	**Description**
β	Random variable perturbing the location of the grid
δ	Dimensionless width of a peak (local maximum)
Δx	Characteristic length of spatial heterogeneity
γ	Random variable perturbing the location of the peak maximum
ν	Measurement tolerance of population density data
τ	Accuracy tolerance of an estimate of pest abundance
ω	Factor of characteristic length of spatial heterogeneity

Chapter 1

Introduction

1.1 The Principles of Integrated Pest Management

Pests are a sustained and significant problem in the production of food across the globe. The term 'pest' can be used to describe any organism which is deemed to cause harm to mankind in some manner; in crop production this label is given to those which damage or destroy potential produce to an unacceptable extent. Crops are vulnerable to attack from pests both during the growing process and after they have been harvested. When pests of crops prior to harvest are considered, the focus is often predominantly on arthropods, plant pathogens and weeds (*e.g* [54, 88]). Estimates of the annual worldwide loss due to pests at this stage in the production process lie between 35 and 42% [61, 82]. In particular, the pre-harvest loss of 14-15% of the world's crops has been attributed to insect pests [81, 80]. Further losses are incurred after the crops have been harvested. This can be due to infestation of stored crops by pests such as insects, rodents, birds, as well as micro-organisms which cause damage both quantitative and qualitative in nature [41]. Such losses have been estimated to range from 10-25% [81].

In order to minimise these losses, the pest population must be managed in some way to control their abundance or density. Measures of so-called 'preventative pest management' can be put into practice; the idea being to try to stop the pest population from becoming a problem in the first place. Age-old examples of such a tactics are crop rotation and intercropping. In

[1] This chapter is an edited version of the introduction presented in [68].

a crop rotation, instead of an agricultural field consistently being used to grow the same crop, different crops which critically host different pests, are grown sequentially. Intercropping is the planting of different crops within the same field at the same time. Variety can also be introduced by planting several genotypes of the same crop species within a field. Introducing heterogeneity in such ways, either spatially, temporally, or genotypically, can destablise the life cycle of a pest and has been documented to help to control pest populations [53, 91]. A pest's preference for a certain plant can be exploited to the farmer's advantage using a technique called trap cropping. Here, crops are interspersed with plants that are more attractive to the pest and thus act as sacrificial decoys. This diversionary ploy can be sufficient to protect the crop in itself, otherwise it reduces the area of the field to be subjected to further management tactics should they be needed since the pests are then located in localised domains [44]. Another precautionary measure is to grow crops which have been cultivated to be resistant to pest attack. Grafting has been used for centuries to manage certain pathogens and it has also been deemed to be useful in the control of arthropod pests and weeds [54]. A more scientifically advanced means of pest resistant plant cultivation is genetic modification. This is a relatively recent initiative of which the risks are not yet fully understood, however, its potential to become the dominant pest management strategy has certainly been recognised and consequently it has become the focus of much research (*e.g* [7, 20, 38, 93]).

Another way of managing pests is to implement a control action, that is, to employ a means of killing the pest organisms. The most widely used control action is the application of pesticides. It has been estimated that around 3 million metric tonnes are used across the globe per year [80]. Biological control actions, *e.g* releasing a natural enemy of the targeted pest into the agroecosystem, provide an alternative to the use of chemicals. The indiscriminate use of control actions or using them as a preventative measure can have serious negative consequences. For instance the regular use of pesticides often leads to the pest becoming resistant making future management a more difficult task [3]. Another unwanted side effect can be that the pesticide has lethal or sub-lethal effects on natural enemies [95] which can cause a resurgence in the pest population or a secondary pest to emerge.

Recognition that precautionary tactics are rarely sufficient to manage pests alone and that relying entirely on control action is not a durable approach led to the emergence of 'integrated pest management' (IPM) [50]. IPM is the incorporation of several different tactics which work cooperatively together to protect crops from pest attack in a more sustainable way. It consists of the three phases. Firstly, preventative measures of pest management are put in to place. Subsequently, the pest abundance is monitored. 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. Such threshold values can be decided upon by taking a variety of factors into consideration, however, the most often used are economic thresholds in accordance with the work presented in [99] as often the overriding concern is that the pest management programme is financially viable (*e.g* see [43]). The principle of IPM is therefore that a control action is only used if and when it is necessary. Thus monitoring is key to the decision process and is considered an essential part of any integrated pest management and control programme [18, 57].

1.2 Methods of Estimating Pest Abundance

Since different pest types have different behaviours, the monitoring methodology varies accordingly. We thus limit our scope to the consideration of insect pests; henceforth in the text the generic term 'pest' is used synonymously with 'insect pest' unless otherwise stated. The procedure also depends on the environment to be monitored. Let us consider pest management of crops prior to harvest. Furthermore, let us take the spatial scale of the monitoring procedure to be that of an agricultural field. A complete census in this case is hardly practical or indeed possible, therefore the population abundance must instead be estimated. The data to form such an estimate is collected by sampling the pest population for which there exists a multitude of techniques (*e.g* see [4, 12, 46, 96]). A direct, in-situ count can be made of the number of pests in a sample unit *e.g* a plant or a unit area of habitat. For the more inconspicuous species, the counting process can be made easier by dislodging the pests from the plant using a practice known as 'knockdown'. In some instances a sample of the habitat itself may be carefully removed and taken to a laboratory where the count can then be made.

Once the data has been collected the arithmetic mean number of pests M per sample unit is calculated as follows:

$$M = \frac{1}{N} \sum_{i=1}^{N} f_i, \tag{1.2.1}$$

where f_i are the individual sample counts, and N is the number of sample counts taken [29]. 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 by its area [94]. A mean number of pests per plant can be converted to the mean per unit area by multiplying by the mean number of plants in such an area. Such an estimate of pest abundance is considered an 'absolute' estimate since the sample counts directly reflect the number of pests in the sample unit.

Although the sampling techniques outlined above are very effective, they are time consuming and costly to implement, particularly on a large scale. Thus, often quicker and more cost effective means of sampling are used instead. One such technique is netting. A net is swung into the crops for a prescribed time or number of sweeps. The pest insects caught inside are then counted (*e.g* see [66, 96]). A mean count per sample unit can then be calculated as above.

Another widely used sampling technique is trapping. Traps are installed in the field, exposed for a certain amount of time, after which the traps are emptied and the pests counted. The position of the traps can be arbitrary; some ecologists opt for random grids of traps or choose appropriate sampling patterns [2, 56], but in many cases they are placed at the nodes of a rectangular grid [36, 45]. The traps can either be active, whereby an attractant is used to draw the pests into the traps *e.g* bait or pheromones, or they can be passive where capture relies on the activity of the pest species. The trap counts provide information about the pest population density at the position of the traps [19, 84] and the sample mean density can then be calculated by scaling (1.2.1) with relation to the area of the agricultural field, where f_i are now the pest densities at the sample locations.

The above techniques yield a relative estimate of abundance rather than an absolute estimate. The counts are not a direct measure of pest abundance but are relative to the sampling technique and the conditions at the time of sampling. Therefore, only relative estimates which have been obtained via the same sampling technique and in the same conditions can be compared. It is

possible, however, to convert an estimate that is relative to one that is absolute using regression analysis [15] or through calibration using experimental data [35]. Steps to achieve this via mathematical modelling have also been made [73].

An estimate of the population abundance can also be achieved using Mark-Release-Recapture methods. Initial sampling is performed and the catch is counted and marked in some way (*e.g* see [42]). The marked population is then released back into the agroecosystem and another round of sampling is conducted. An estimate of the population size can then be formulated using the condition that the proportion of marked insects in the field is equal to the proportion of marked insects found in the second sample. That is, the following can be rearranged to solve for I

$$\frac{\hat{I}}{I} = \frac{\hat{C}}{C}, \tag{1.2.2}$$

where $\hat{I}$ is the total number of marked insects, I is the number of insects in the entire population, C is the number of insects caught in the second sample and $\hat{C}$ is the number of those which are marked. This method works well in scientific studies but is too labour intensive to be suitable for routine monitoring.

1.3 The Importance of Accurate Estimation

Once an estimate of the pest population size in an agricultural field has been acquired, a pest management decision is made by comparing it to some threshold value(s). Let us consider the simplest case where a single threshold value is used. If the estimate falls below the threshold the decision is to take no action, whereas if it exceeds the threshold the decision is to intervene and implement a control action (*e.g* see [9], Chapter 1). The decision can be considered to be correct if the same conclusion would have been reached if the true pest abundance had been known. However, by definition the true abundance is unknown, thus we require information about the reliability of the estimate in order to have confidence about the decision's validity. Suppose we can define the accuracy of an estimate I_a as being within some tolerance τ of the true abundance I. This subsequently means that an estimate I_a belongs to the range $[I - \tau, I + \tau]$. There is only a risk of an incorrect decision if the threshold value falls within this range. If the accuracy

of an estimate can be quantified, in turn, the risk can be quantified. Furthermore, clearly the smaller the tolerance τ, *i.e.* the more accurate the estimate, and the lesser the risk of making an incorrect pest management decision.

An incorrect decision could mean that action is not taken when it is needed leading to the loss of crops. The value of crops lost to pests in generic sense of the term (insects, plant pathogens and weeds) has been estimated to be \$2,000 billion per year even with the use of pesticides [80]. Obtaining a more accurate estimate of the pest abundance could lead to the more timely use of a control action and ultimately reduce crop loss. On the other hand, the use of a control action could be recommended unnecessarily. The most common means of control is the application of pesticides which is a costly procedure that can cause considerable damage to the environment [48]. Pesticides are known to contribute to air, soil and water pollution whilst there is growing evidence linking their use to human illnesses [1, 25]. It has been estimated that less than 0.1% of pesticides used reach their targeted pest, the remaining 99.9% is absorbed by some means into the environment [79]. Some of the loss occurs during application with the spray drifting outside of the intended area, however once applied to a crop, pesticides can then vaporise into the air, end up in surface or groundwater, be absorbed by plants or ingested by non-target species, or indeed remain in the soil. Furthermore, unnecessary application of pesticides is undesirable from an economic perspective; around \$40 billion is spent per year applying pesticides [80].

It is obvious from the above that there is a significant need for reliable methods to accurately evaluate the pest population size in order to avoid making an incorrect decision regarding the use of a control action. It is worth noting here that the accuracy required by pest monitoring is not always particularly demanding. It differs according to the monitoring purpose. In routine monitoring an error range can be $20\% - 100\%$ [65, 89], whereas monitoring for research purposes can demand a higher degree of accuracy of 10% (*e.g* see [66], p. 245).

Several means of optimising the accuracy of an estimate have been considered in the ecological literature. One way is to ensure that the size of the data set is large enough *i.e.* that enough sample units are taken. It follows from the equation (1.2.1) that the exact value of the population

size will be obtained for infinitely large number N. Hence we can expect better accuracy of the estimate when N gets larger. A pre-sample (or series of them) can be used to obtain a sample mean and sample variance from which an estimate of the number of sample units needed to achieve a specified precision can be calculated (*e.g* see [9, 31, 66]). However, there is a trade-off between the number of sample units needed to achieve sufficient accuracy and the number that can be practically afforded. For instance, if a trapping procedure is applied in ecological research, the number K of traps per given area can be made quite large, e.g. in the order of hundreds. Meanwhile in routine pest monitoring programmes K rarely exceeds twenty [56] per a typical agricultural field with a linear size of several hundred meters. In some cases, there may only be one or a few traps per field [60]. There are several practical reasons as to why the number of sample units cannot be made large. An increase in the number of sample units equates to an increase in the amount of labour and hence finances required. In any real-world scenario there is a limit to such resources. Also, sampling introduces a disturbance into the field and too intensive sampling can cause damage to the agricultural product. Furthermore, sampling also imposes a disturbance on the pests which can in turn affect the results of the sampling technique, therefore from this perspective the number of sample units should be minimised.

The efficacy of a sampling technique is also important to the accuracy of an estimate of the pest abundance. Means of sampling a pest population are constantly being reviewed leading to sampling equipment being developed and improved [11, 101]. Another key consideration is the sampling plan, that is, the prescribed locations at which samples are to be taken. For an estimate to be accurate the sample must capture sufficient information to adequately represent the true pest presence. In theory, if conditions are homogeneous across the field, insects could be randomly distributed. In reality, however, the population density distribution is never random as the conditions can never truly be homogeneous, and furthermore, the location of an insect is dependent on various factors where examples include where the egg it spawned from was laid, and the location of other members of the population. As such, insects exhibit an aggregated spatial distribution [36, 45] to varying degrees. The sampling plan thus becomes crucial; it is important to avoid bias stemming from samples being placed entirely in areas where the pests

are clustered, or likewise, entirely in areas of zero density. Comparisons of various patterns *e.g* random, transects, quadrats, *etc.* have been made in order to make recommendations [2].

1.4 Research Objectives

Although ensuring a sufficiently accurate estimate has been considered in the ecological literature as discussed in the previous section, to the best of our knowledge, the focus has predominantly been on how the data is collected. In this thesis, we instead look at the way in which the data is processed. We discuss numerical integration techniques as an alternative approach to the existing statistical methods.

We aim to demonstrate that numerical integration methods can be used to provide a reliable estimate of pest abundance. We explain the theory of numerical integration and how the techniques can be used in ecological applications. It will be shown that often numerical integration methods can provide more accurate results than statistical techniques which rely on the sample mean. However, restrictions imposed by the underlying ecological problem mean that the numerical integration methods may not perform according to the conventional theory. We go on to conduct an in depth study to establish which factors affect the accuracy of numerical integration techniques within the framework of the pest monitoring problem, and thus gain an understanding of how the accuracy of a pest abundance estimate may be controlled. Thus, we make some initial steps towards the development of numerical integration methods which could be implemented in routine pest monitoring.

We study the pest monitoring problem at the spatial scale of a single agricultural field, which we consider as a rectangular domain. To gain an understanding of the issues we first consider a simplified one-dimensional (1D) counterpart of the pest monitoring problem. We have then extended some of our results to handle the two-dimensional (2D) case. We primarily consider a regular grid of sample units, however we also carry out some investigation of quasi-regular and random sampling plans. Our study begins by considering the density data to be exact, we then in the later chapters consider noisy density data.

The structure of this thesis is as follows. In Chapter 2 we give an overview of the mathematical theory behind numerical integration and we outline how the techniques can be used to

evaluate pest abundance. The motive behind considering such methods as an alternative to the widely used statistical techniques is demonstrated. In Chapter 3 we explain how the conditions and restrictions imposed by ecological problem affect the performance of numerical integration methods. We introduce the concept of the coarse grid problem which arises due to the limited amount of data that can be collected. We show that the accuracy which can be achieved on coarse grids is strongly dependent on the spatial pattern of the density distribution and that any prior knowledge of this pattern should be used to its fullest extent. In Chapter 4 we show the link between the diffusion rate of a species and the spatial pattern of the pest population. In turn, we demonstrate that the diffusion is a controlling parameter of the accuracy of an abundance estimate. We construct a rudimentary estimate of the number of regularly (and quasi-regularly) spaced sample units required to achieve an estimate with a prescribed level of accuracy. In Chapter 5 we study a particular type of density distribution which is difficult to handle, namely a highly aggregated distribution. We introduce a new computational grid classification, ultra-coarse, where the accuracy of an estimate can only be described probabilistically. An improved estimate of the minimum number of regularly spaced sample units needed to guarantee accuracy is found. The results of this chapter are then extended in Chapter 6 to be applicable for a random sampling plan, which is often viewed favourably by ecologists. In Chapters 7 and 8 we investigate how noise in the density data affects the accuracy of an abundance estimate. Conclusions of the entire study are provided in Chapter 9.

Chapter 2

Numerical Integration Methods

In this chapter methods of numerical integration are discussed within the framework of the ecological problem of monitoring a pest population in a single agricultural field. First, the basic concepts of numerical integration are outlined in Section 2.1. Subsequently, in Sections 2.2 – 2.4 examples are given of specific methods which could be implemented to evaluate pest abundance. The potential benefits of using numerical integration techniques instead of a standard statistical approach are highlighted in Section 2.5. Chapter conclusions are given in Section 2.6.

2.1 Theory of Numerical Integration

Once information on the pest population in an agricultural field has been gathered by a chosen sampling technique, an estimate of the pest abundance can be formulated. Typically the estimate used within the ecological community depends on the sample mean [29]. Under the assumption that sample counts can be manipulated to give the pest density at each sample unit location [19, 84], an estimate I_a to the true abundance I can be calculated as

$$I \approx I_a = SM(N) = \frac{S}{N} \sum_{i=1}^{N} f_i, \tag{2.1.1}$$

where S is the area of the field, $M(N)$ is the sample mean pest density, N is the total number of sample units and f_i denotes the pest population density at the sample unit location x_i, $i = 1, \ldots, N$. By the *law of large numbers* (*e.g* see [40]), as N grows large, the sample mean density M tends to the true mean density $\bar{M}$. Thus, as N grows large the above estimate of the

pest abundance I_a tends to the exact pest abundance I.

The estimate of the pest abundance given by the formula (2.1.1) is simply a weighted sum of the available pest population density values. Generalising this formula leads us to consider numerical integration techniques as a means of estimating pest abundance. Let the agricultural field subjected to pest monitoring be represented by the domain D. If the pest population density function $f(x, y)$ were known to us almost everywhere across the domain D, then the exact pest abundance I could be calculated by analytically evaluating the integral

$$I = \iint_D f(x, y)\, dx\, dy. \tag{2.1.2}$$

However, the information on the pest population density is obtained by sampling and as such is only available to us at a finite number N of locations. Consequently, the pest population density function is discrete, namely, $f(x, y) \equiv f_i, i = 1, \ldots, N$. The above integral cannot be evaluated and thus we are forced to seek an approximation by means of numerical integration. The general formula for numerical integration is given by the weighted sum (*e.g* see [28])

$$I \approx I_a = \sum_{i=1}^{N} w_i f_i, \tag{2.1.3}$$

where the weights $w_i, i = 1, \ldots, N$ depend on the specific method of numerical integration. It is easy to see on comparison with the above that the formula (2.1.1) can be considered as a simple form of numerical integration where the weights are uniformly defined as

$$w_i = \frac{S}{N}, \qquad i = 1, \ldots, N. \tag{2.1.4}$$

There are, of course, many other possible combinations of weight coefficients other than that prescribed by (2.1.4) which can be used in the formula (2.1.3) to yield an estimate I_a. The weights must be chosen, however, such that the resulting estimate I_a satisfies the following condition

$$I_a(N) \to I, \text{as } N \to \infty. \tag{2.1.5}$$

In other words, as the data set $f_i, i = 1, \ldots, N$ increases in size, the estimate I_a produced increases in accuracy. To assess the accuracy we may consider the *absolute approximation error* which is a measure of the magnitude of the distance between an estimate and the exact solution and is defined as

$$E_{abs}(N) = |I - I_a(N)|. \tag{2.1.6}$$

Alternatively we may calculate the *relative approximation error* given by

$$E_{rel}(N) = \frac{E_{abs}}{|I|} = \frac{|I - I_a(N)|}{|I|}, \tag{2.1.7}$$

which quantifies the accuracy of an estimate I_a as a proportion of the exact solution.

Let us consider the absolute error (2.1.6). From (2.1.5) we obtain the following condition

$$E_{abs}(N) \to 0, \text{as } N \to \infty. \tag{2.1.8}$$

In order to calculate the absolute error (2.1.7) of an estimate I_a, we require the value I. We recall that in the pest monitoring problem I corresponds to the true pest abundance present in the agricultural field which in reality cannot be known. In such a situation the standard approach is to decide on the accuracy by considering the asymptotic error estimate of the employed numerical integration method. An error estimate often takes the following form (*e.g* see [83])

$$E_{abs}(N) \leq \frac{C}{N^q} \sup_{x \in [a,b]} |f^{(r+1)}(x)| \tag{2.1.9}$$

where C, q and r are constants dependent on the choice of weight combination used in formula (2.1.3). In particular, q is the *order of convergence i.e.* the rate at which E_{abs} converges to zero as N grows infinitely large. The *degree of precision*, that is the degree of polynomial for which the method is exact, is denoted by r. We typically expect that the higher the order of convergence of a method of numerical integration, the more accurate the estimate produced will be. Some examples of legitimate weight combinations such that condition (2.1.8) is satisfied are discussed in Sections 2.2 - 2.4 alongside their corresponding orders of convergence and degrees

of precision.

2.2 Methods of Numerical Integration on a Regular Grid

We first consider the 1D problem and discuss potential weight coefficient combinations to be used in the formula (2.1.3). 2D methods of numerical integration are discussed in Section 2.4. We reduce the problem to one dimension by considering the domain D in the focus of pest monitoring as a straight line. The domain D is therefore defined as the interval $D = [a, b]$ and the integral (2.1.2) becomes

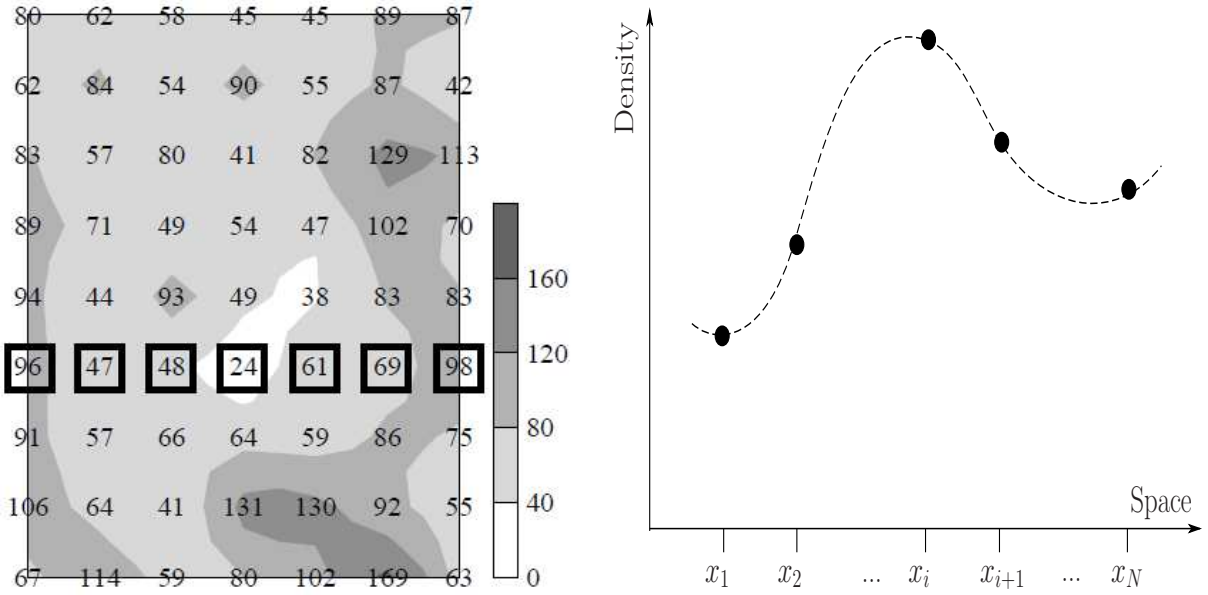
$$I = \int_a^b f(x)\,dx, \tag{2.2.1}$$

where $f(x) \equiv f_i, i = 1, \ldots, N$. This can be thought of as considering a single transect from the grid of data as depicted in Figure 2.1. It is worth noting that there is evidence to suggest that of the contiguous sampling formations, a straight line provides the most accurate estimations [2]. Therefore, 1D methods of numerical integration may also be useful in the practical problem of pest monitoring.

Numerical integration formulae are derived by somehow interpolating between the available data points $f_i, i = 1, \ldots, N$ (see Figure 2.1b). The weights $w_i, i = 1, \ldots, N$ in the formula (2.1.3) are then determined by integrating the resulting interpolating function. The definition of the weights therefore depend on the means of interpolation used, and the sampling plan *i.e.* the way the sample unit locations which correspond to the grid nodes $x_i, i = 1, \ldots, N$ are distributed. Let us first consider a regular sampling plan, whereby the sample units are evenly spaced over the domain D. Such a plan is often used in pest monitoring [36, 45]. In the 1D problem, this equates to each of the grid nodes x_i being a fixed distance h from its neighbouring grid nodes. We refer to h as the *grid step size*.

When the data points f_i are available at evenly spaced grid nodes x_i, commonly used weight choices in the formula (2.1.3) are those prescribed by the Newton-Cotes rules[1]. This family of numerical integration formulae finds a polynomial $P_k(x)$ of degree k which passes through the

[1]The Newton-Cotes formulae are the work of Isaac Newton (1642-1727) and Roger Cotes (1682-1716). For details of the history see [39], pp. 76-77.



(a) (b)

Figure 2.1: (a) Example of field data [45]. The numbers represent the sample unit counts from which the pest population density at the sample unit location can be obtained. Samples were obtained by means of pitfall traps. Data from a line of sample units is extracted from the grid therefore reducing the problem to one dimension; considered sample units are outlined. (b) Sketch of the pest population density function obtained from such counts [69]. The density $f(x) = f(x_i)$ represented by the filled circles is known only at the location of the sample units denoted by $x_1, x_2, ..., x_N$. The continuous distribution shown by the dashed curve is not known.

data points $f_i, i = 1, \ldots, k+1$. The polynomial $P_k(x)$ is defined by the Lagrange interpolation formula (*e.g* see [100])

$$P_k(x) \equiv \sum_{i=1}^{k+1} f_i L_i(x),$$

where L_i is the Lagrange polynomial given by

$$L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{k+1} \frac{x - x_j}{x_i - x_j}.$$

An approximation I_a to the exact integral I is then calculated by integrating analytically the interpolating polynomial $P_k(x)$ as follows

$$I \approx I_a = \int_a^b P_k(x)\, dx.$$

The degree k of the interpolating polynomial depends on the number of grid nodes where the pest population density values f_i are available. A polynomial of degree k requires $N = k+1$ data points f_i to be available. The weights in (2.1.3) as prescribed by the Newton-Cotes formulae are thus

$$w_i = \int_a^b L_i(x)dx.$$

There are two forms of Newton-Cotes formulae: open and closed (*e.g* see [17]). When the regular grid of nodes includes the limits a and b of the domain D, *i.e.* the grid nodes are defined as

$$x_1 = a, \qquad x_i = x_{i-1} + h, \quad i = 2, \ldots, N-1, \qquad x_N = b, \tag{2.2.2}$$

where the grid step size is $h = \dfrac{b-a}{N-1}$, then the weights produced correspond to the *closed Newton-Cotes formulae*. If, however, the endpoints of the grid do not coincide with the limits a, b and the grid nodes are instead defined as

$$x_1 = a + \frac{h}{2}, \qquad x_i = x_{i-1} + h, \quad i = 2, \ldots, N-1 \qquad x_N = b - \frac{h}{2}, \tag{2.2.3}$$

where the grid step size is defined as $h = \dfrac{b-a}{N}$, then the *open Newton-Cotes formulae* are derived. Regularly spaced nodes which include either the left endpoint a or the right endpoint b give rise to the *semi-open Newton-Cotes formulae* [30]. In the practical terms of pest monitoring, the implementation of formulae of the closed type require sample units to be located on the field boundary in order to obtain an estimate of the pest abundance in the whole field. The open type formulae, however, can be applied to obtain an estimate if the sample units are inset from the field's edge.

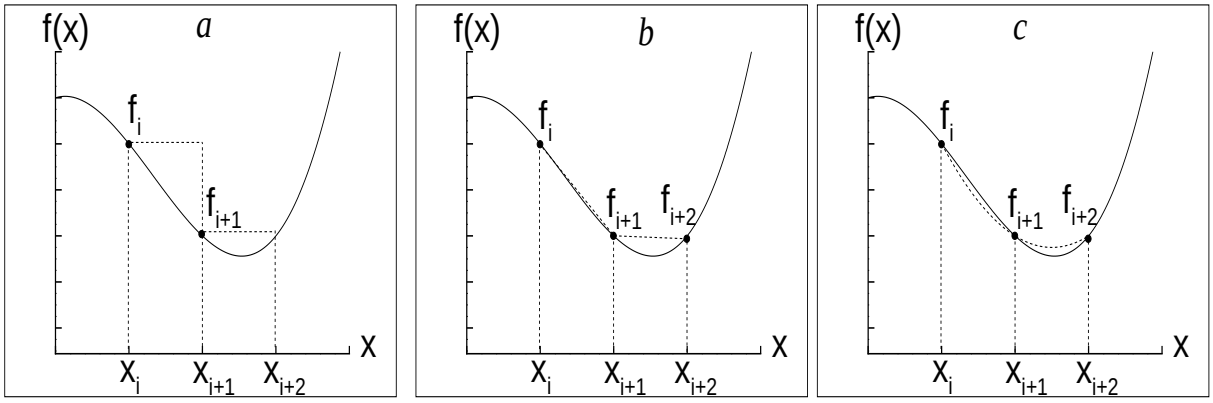


Figure 2.2: Piecewise approximation of the function $f(x)$ by polynomials of degree k. The figure is as given in [68]. (a) $k = 0$ (b) $k = 1$ (c) $k = 2$.

As implementation of the Newton-Cotes formulae requires the computational grid to be regular, their asymptotic error estimates can be expressed in the form

$$E_{abs} \leq Ch^q \sup_{x\in[a,b]} |f^{(r+1)}(x)|, \tag{2.2.4}$$

where $h > 0$ is the grid step size. The constants C, q and r have dependence on the degree k of the employed interpolating polynomial. Once again q represents the method's order of convergence and r its degree of precision. The accuracy of an estimate is clearly dependent on the grid step size h. Furthermore, a smaller grid step size corresponds to a larger number N of grid nodes and therefore a higher degree k of polynomial is required to interpolate the data over the entire interval $[a, b]$. As documented in many a textbook (*e.g* see [17]), higher order interpolating polynomials constructed on a regular grid exhibit oscillatory behaviour and therefore yield inaccurate estimates I_a of an integral I. As such, Newton-Cotes formulae are not usually applied to the entire interval of integration $[a, b]$. Instead, the Newton-Cotes formulae are used locally, *i.e.* on a sub-interval of $[a, b]$. Summing the local approximations then gives the approximation to the integral over $[a, b]$. This equates to applying the *compound* or *composite* Newton-Cotes formulae (*e.g* see [28, 100]).

We now give some examples of some well known composite Newton-Cotes formulae. Let us consider the sampling plan prescribed by (2.2.2). The simplest means of interpolating between data points f_i is using a constant function, that is, an interpolating polynomial $P_k(x)$ of degree

$k = 0$. An example of such interpolation is shown in Figure 2.2a where on each sub-interval $[x_i, x_{i+1}]$ for $i = 1, \ldots, N-1$ a rectangle has been constructed with the width $h = x_{i+1} - x_i$ and height f_i. Since integrating a function $f(x)$ can be thought of as calculating the area under the curve, an approximation to the integral

$$I_i = \int_{x_i}^{x_{i+1}} f(x)\, dx \tag{2.2.5}$$

is given by the area

$$\alpha_i = h f_i.$$

Computing the sum

$$I_a = \sum_{i=1}^{N-1} \alpha_i = \sum_{i=1}^{N-1} h f_i$$

gives an approximation to the exact integral I as defined by (2.2.1). The endpoint b of the interval $[a, b]$ has not been used as an interpolation point (the above sum does not contain the function value $f_N \equiv f(x_N) = f(b)$). Therefore, this method, which is commonly referred to as the *composite left rectangle rule*, is of the semi-open type of Newton-Cotes formula. Alternatively on each sub-interval $[x_i, x_{i+1}]$ for $i = 1, \ldots, N-1$, a rectangle with width $h = x_{i+1} - x_i$ and height f_{i+1} could be constructed. This means of interpolation yields another semi-open Newton-Cotes formula: the *composite right rectangle rule*. In both cases the weights of the formula (2.1.3) are uniformly defined as $w_i \equiv h$, and it is only the underlying computational grid of nodes x_i used in the calculation which differs. The error estimates for both composite rectangle methods are of the form (2.2.4) with order of convergence $q = 1$ and degree of precision $r = 0$. Generally, a more accurate estimate formed by piecewise constant interpolation can be obtained by instead locating the grid nodes at the midpoint of each sub-interval $[x_i, x_{i+1}]$. This is equivalent to using the computational grid (2.2.3). This open type formula is the *composite midpoint rule* and is known to have an improved order of convergence $q = 2$ and degree of precision $r = 1$.

Let us increase the degree of the interpolating polynomial $P_k(x)$ to $k = 1$, *i.e.* we interpolate by means of a linear function. Figure 2.2b shows such an interpolation where the points x_i and x_{i+1} are joined by a straight line. Thus, on each sub-interval $[x_i, x_{i+1}]$ a trapezium is constructed. The area of this trapezium α_i gives an approximation to the integral I_i given by (2.2.5) and we have

$$I_i \approx \alpha_i = \frac{h}{2}(f_i + f_{i+1}).$$

As before, we obtain an estimate I_a of the integral I by summing the areas α_i, where

$$I_a = \frac{h}{2}f_1 + \sum_{i=2}^{N-1} hf_i + \frac{h}{2}f_N.$$

This corresponds to the weights of the formula (2.1.3) being defined as

$$w_1 = \frac{h}{2}, \qquad w_i = h, \text{for } i = 2, \ldots, N-1, \qquad w_N = \frac{h}{2}. \tag{2.2.6}$$

This integration rule, for obvious reasons, is known as the *composite trapezium rule*. Since the construction of a straight line requires two points, the total number N of grid nodes is required to be $N \geq 2$. The order of convergence of the composite trapezium rule is $q = 2$ and the degree of precision is $r = 1$.

Let us now increase the degree of the interpolating polynomial to $k = 2$. We recall that a k degree interpolating polynomial requires the function value to be known at $k+1$ computational nodes x_i; thus to construct an interpolating polynomial of degree $k = 2$ (*i.e.* a quadratic function) we require the use of three grid nodes. Figure 2.2c shows a quadratic interpolating polynomial constructed using the consecutive grid nodes x_i, x_{i+1} and x_{i+2}. The area under the quadratic provides an estimate of the integral of the function $f(x)$ over the sub-interval $[x_i, x_{i+2}]$ and is calculated as follows:

$$I_i \approx \alpha_i = \frac{h}{3}(f_i + 4f_{i+1} + f_{i+2}).$$

The above is known as *Simpson's rule*. Let the interval $[a, b]$ be divided into sub-intervals $[x_{2i-1}, x_{2i+1}]$ where $i = 1, \ldots, \frac{N-1}{2}$ and Simpson's rule be applied to each sub-interval. Summing the resulting areas yields the following estimate I_a of the integral I of the function $f(x)$ over the entire interval $[a, b]$

$$I_a = \frac{h}{3}\left(f_1 + 4\sum_{i=1}^{\frac{N-1}{2}} f_{2i} + 2\sum_{i=1}^{\frac{N-3}{2}} f_{2i+1} + f_N\right),$$

which is the *composite Simpson's rule*. The order of convergence is $q = 4$ and the degree of precision is $r = 3$. It can be seen from the upper limits of the summations that the number N of grid nodes is required to be odd in order to apply the composite Simpson's rule. Furthermore, to be able to fit a quadratic to the data we also require that $N \geq 3$. The composite Simpson's rule corresponds to the weights of the formula (2.1.3) being defined as

$$w_1 = \frac{h}{3}, \quad w_i = \frac{4h}{3}, \text{for } i = 2, 4, \ldots, N-1, \quad w_i = \frac{2h}{3}, \text{for } i = 3, 5, \ldots, N-2, \quad w_N = \frac{h}{3}. \tag{2.2.7}$$

The error estimates of the composite $k+1$ point Newton-Cotes formulae of either closed or open type are known to be of the form (2.2.4) with order of convergence $q = k+1$ and degree of precision $r = k$ when k is odd, meanwhile $q = k+2$ and $r = k+1$ when k is even (*e.g* see [83]). From the error estimates, it appears that in general using a higher order interpolating polynomial will give rise to a more accurate numerical integration method. This, however, is not necessarily the case. It is well documented (*e.g* see [28, 100]) that for larger polynomial degree k, some of the weights of the corresponding $k+1$ point Newton-Cotes method become negative leading to some cancellation of terms when the estimate I_a is calculated. Therefore a less accurate estimate is produced. For this reason the $k+1$ point Newton-Cotes formulae are considered unsuitable when $k \geq 7$ in the case of the closed type, and when $k \geq 2$ in the case of the open type [83, 100]. Furthermore, whilst the composite Newton-Cotes methods which rely on constant interpolation (when $k = 0$) such as the aforementioned rectangle rules can be implemented when the total number N of grid nodes is entirely arbitrary, the implementation

of a composite $k+1$ point Newton-Cotes formulae for $k \geqslant 1$ require the the number of grid nodes to be of the form $N = mk + 1$, where $m \in \mathbb{N}$. For small k this condition is hardly a hindrance, *e.g* the only resulting requirement on the number N of grid nodes when $k = 1$ is that $N \geq 2$. However, the restriction makes implementation of Newton-Cotes formulae somewhat more awkward as k grows large. As such, numerical integration is most often conducted using methods which rely on lower order polynomial interpolation such as the composite trapezium or the composite Simpson's rule.

A regular computational grid, which corresponds to a regular sampling plan in the pest monitoring problem, is required to implement a numerical integration formula of the Newton-Cotes family. In the next section we discuss methods of numerical integration on irregular grids.

2.3 Methods of Numerical Integration on an Irregular Grid

In this section we discuss numerical integration methods on irregular computational grids. We continue to restrict our focus to the 1D problem, thus the agricultural field is still considered as the interval $D = [a, b]$. 2D numerical integration methods are discussed in the next section.

The Newton-Cotes methods are derived by integrating interpolating polynomials constructed over a regular grid. Different methods of numerical integration can be derived by selecting an irregular grid formation. In fact, the grid nodes can be chosen so as to maximise the degree of precision. This procedure leads to the family of *Gaussian integration* rules, where the grid nodes are located at the roots of orthogonal polynomials (*e.g* see [26, 100]). The weights are always positive thus convergence of the estimate I_a to the exact value of the integral I can be guaranteed as the number of nodes k increases. Furthermore, this class of k+1-point Gaussian integration rules is known to have degree of precision $r = 2k + 1$.

Another class of methods which use irregular computational grids are the *Clenshaw-Curtis formulae* [21]. The nodes of the computational grid are located at the extrema of a Chebyshev polynomial. A $k+1$ point formula has degree of precision $r = k$. It would appear, therefore, that the accuracy of such methods is much poorer than the Gaussian rules, though this is only the case if the integrand is well approximated by a polynomial. For a large class of functions the

accuracy of the two approaches is the same [103]. The Clenshaw-Curtis integration methods, however, can be implemented much more quickly than the Gaussian methods with the use of the fast Fourier Transform.

Both the Gaussian and Clenshaw-Curtis integration methods could be used to obtain an estimate of pest abundance if the sample units were located at the roots of the orthogonal polynomials. Even if collecting samples in a prescribed pattern is the intention, however, human error when collecting samples or obstructions present in the agricultural field may lead to the resulting grid of sample unit locations being different to that which was planned. We thus now design a family of numerical integration techniques which can handle an effectively arbitrary grid of sample units.

Let us define the sample unit locations, *i.e.* the computational grid of the numerical integration method, as follows:

$$x_1 = a, \qquad x_i = x_{i-1} + h_{i-1}, \quad i = 2, \ldots, N-1, \qquad x_N = b, \tag{2.3.1}$$

where $h_i > 0$ for $i = 1, \ldots, N-1$ is the distance between grid nodes x_i and x_{i+1}. The location of the interior nodes x_i for $i = 2, \ldots, N-1$ are therefore arbitrary. It is only the exterior nodes x_1 and x_N which are fixed to coincide with the endpoints of the interval $D = [a, b]$.

Once again, we seek an approximation to the integral

$$I = \int_a^b f(x)\, dx$$

where the population density function $f(x) \equiv f_i$ for $i = 1, \ldots, N$. As before, the index i is used to represent the sample unit location x_i, however, the layout of the sample units is now defined according to the sampling plan (2.3.1) rather than following a regular sampling plan. We first consider the integral of the population density function $f(x)$ over the sub-interval $[x_i, x_{i+1}]$, namely

$$I_i = \int_{x_i}^{x_{i+1}} f(x)\, dx, \qquad \text{for } i = 1, \ldots, N-1$$

An approximation I_{a_i} to the integral I_i is constructed by integrating an interpolating polynomial $P_i^k(x)$ of degree k which approximates $f(x)$ over the sub-interval $[x_i, x_{i+1}]$, that is we have,

$$I_i \approx I_{a_i} = \int_{x_i}^{x_{i+1}} P_i^k(x)\, dx, \qquad \text{for } i = 1, \ldots, N-1.$$

Summing the local approximations I_{a_i} gives an estimate I_a of the integral I:

$$I \approx I_a = \sum_{i=1}^{N-1} I_{a_i}.$$

The interpolating polynomials $P_i^k(x)$ are found by Lagrange interpolation

$$P_i^k(x) = \sum_{j=1}^{k+1} f(x_j) L_j(x),$$

where we recall that

$$L_j(x) = \prod_{\substack{l=1 \\ l \neq j}}^{k+1} \frac{x - x_l}{x_j - x_l}.$$

It can be seen from the above that each interpolating polynomial $P_i^k(x)$ depends on a set of points $x_j, j = 1, \ldots, k+1$. By the nature of the outlined problem, on each sub-interval $[x_i, x_{i+1}]$ the population density function $f(x)$ is known solely at the endpoints. Therefore, to construct an interpolating polynomial of degree $k > 1$, some additional points from outside the sub-interval $[x_i, x_{i+1}]$ must be used. We refer to the set of points $x_j, j = 1, \ldots, k+1$ as the *interpolation stencil*.

We use the interpolation stencil presented in [72] which we outline below. The stencil is presented for three cases: when the degree of the polynomial $P_i^k(x)$ is $k = 2$, $k = 3$, and $k = 5$. The aim is to keep the stencil as symmetric as possible about $[x_i, x_{i+1}]$ however the structure is dependent on where the sub-interval $[x_i, x_{i+1}]$ is positioned within the domain $[a, b]$.

For the case $k = 2$ the stencil consists of three points, x_i, x_{i+1}, and one other taken either immediately before or after the sub-interval $[x_i, x_{i+1}]$. The stencil used to determine $P_i^k(x)$ over

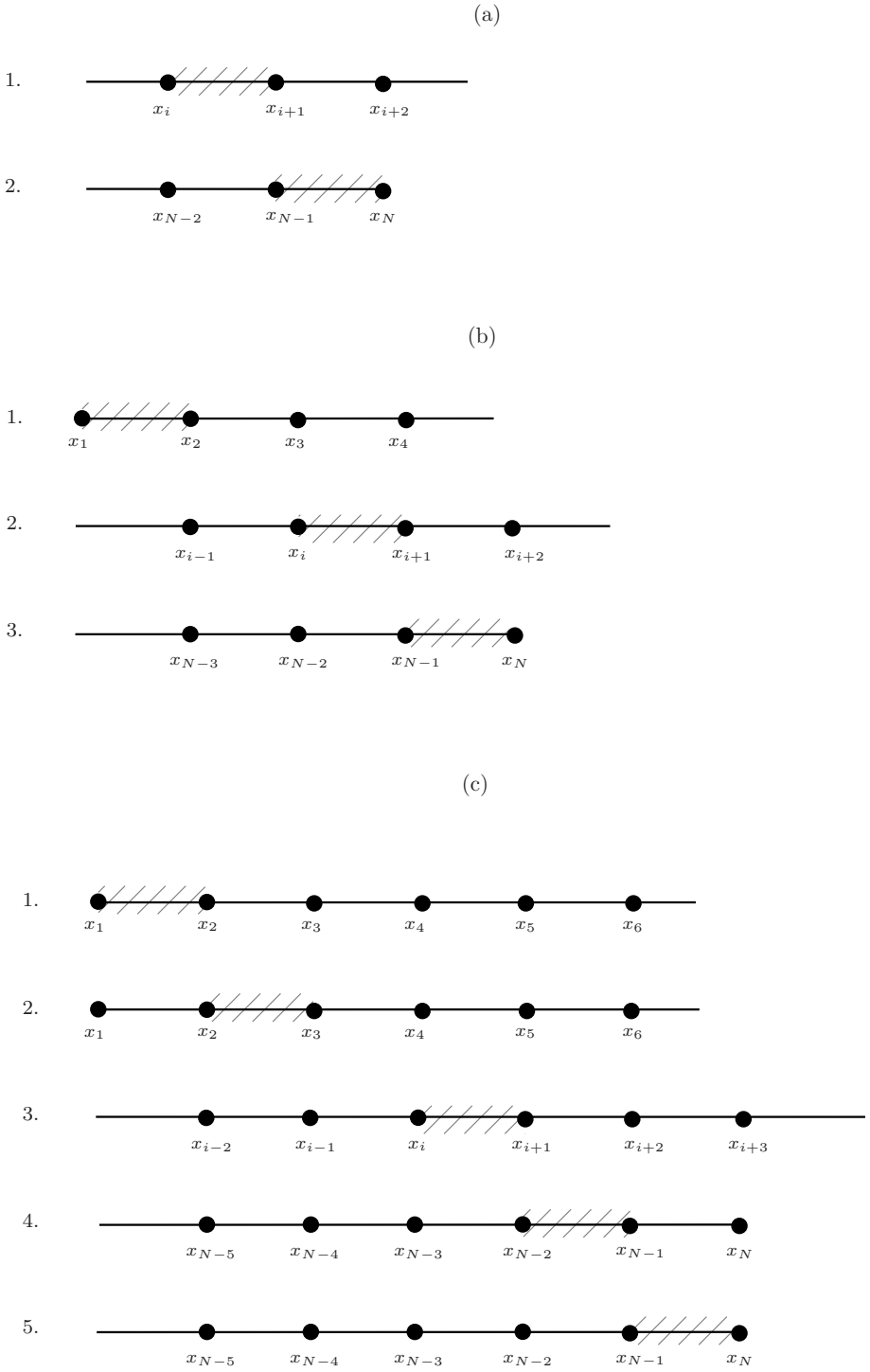


Figure 2.3: Interpolation stencil for polynomial approximation of degree (a) $k = 2$, (b) $k = 3$ and (c) $k = 5$. The shaded region indicates the interval over which the polynomial is to be integrated

each sub-interval $[x_i, x_{i+1}]$ when $k = 2$ is defined as follows

1. x_i, x_{i+1}, x_{i+2} for $i = 1, ..., N-2$

2. x_{i-1}, x_i, x_{i+1} for $i = N-1$.

This is illustrated in Figure 2.3a. To construct $P_i^k(x)$ when $k = 3$ we require four points including x_i and x_{i+1}. For the purpose of symmetry we aim to use x_{i-1} and x_{i+2} as the additional nodes however for the sub-intervals at the boundaries of the domain $[a, b]$ this is not possible. The stencil therefore takes three forms as outlined below and shown in Figure 2.3b.

1. $x_i, x_{i+1}, x_{i+2}, x_{i+3}$ for $i = 1$

2. $x_{i-1}, x_i, x_{i+1}, x_{i+2}$ for $i = 2, ..., N-2$

3. $x_{i-2}, x_{i-1}, x_i, x_{i+1}$ for $i = N-1$.

For $k = 5$ there are five forms of the interpolation stencil, namely,

1. $x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}, x_{i+5}$ for $i = 1$

2. $x_{i-1}, x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}$ for $i = 2$

3. $x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}, x_{i+3}$ for $i = 3, ..., N-3$

4. $x_{i-3}, x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}$ for $i = N-2$

5. $x_{i-4}, x_{i-3}, x_{i-2}, x_{i-1}, x_i, x_{i+1}$ for $i = N-1$

The stencil is symmetric for the interior sub-intervals and becomes asymmetric as $[x_i, x_{i+1}]$ approaches the left and right boundaries of $[a, b]$. See Figure 2.3c.

Clearly, increasing the number of stencil points allows a higher degree polynomial approximation to be used in the problem. This in turn increases the accuracy of the approximation, whether an approximation of the function itself or function derivatives is needed. Consequently, extended stencils are used in various numerical methods where accurate approximation of either temporal derivatives or spatial derivatives is required. Examples include multi-step ordinary

differential equation methods for temporal problems, and higher order approximation of the boundary value problem by Richardson extrapolation for spatial problems (*e.g* see [17]).

We shall refer to the above numerical integration technique as the *Lagrange Interpolation Integration* (LII) method. It provides more flexibility than the composite Newton-Cotes formulae. Firstly, it permits the use of an irregular computational grid. Secondly, the only condition for the total number N of grid nodes is that to implement polynomial interpolation of degree k we require $N \geq k+1$. This is less restrictive than the condition on N imposed by the composite Newton-Cotes formulae, where we recall that for $k \geq 1$ the $k+1$ point formulae require that the total number of grid nodes be of the form $N = mk + 1$, for some $m \in \mathbb{N}$.

Since the LII method comprises of integrating analytically local interpolating polynomials, the only source of error stems from the Lagrange interpolation. It can be shown (*e.g* see [83]) that the following provides an estimate of the error induced by Lagrange interpolation:

$$|f(x) - P_k(x)| \leq C h_{\max}^{k+1} \sup_{x \in [a,b]} |f^{(k+1)}(x)|, \qquad h_{\max} = \max_{i=1,\ldots,N-1} \{h_i\},$$

where $P_k(x)$ is the degree k polynomial approximation of $f(x)$ and the constant C is dependent on k. Thus an estimate of the error (2.1.6) obtained when approximating the abundance by the LII method is

$$E_{abs} \leq C h_{\max}^{k+1} \sup_{x \in [a,b]} |f^{(k+1)}(x)|. \tag{2.3.2}$$

The order of convergence is thus $q = k + 1$ as $h_{\max}$ tends to zero, and the degree of precision is $r = k$. When the degree of the interpolating polynomials $P_i^k(x)$ is $k = 1$, and a regular grid is considered, the method is simply the compound trapezium rule (2.2.6). The order of convergence for $k = 3$ on a regular grid corresponds to that of the compound Simpson's rule (2.2.7). This assertion is verified in Section 3.1 where this particular numerical integration method is employed (see Figure 3.2a).

This technique relies on the analytical integration of the interpolating polynomials. As such, implementation becomes awkward as k increases. Similar but more sophisticated numerical

integration techniques are given in [23, 72] where the piecewise interpolating polynomial of degree k is instead numerically integrated. In [72] a Gauss-Legendre formula is applied after the interpolation has been done, thus the procedure is named the Gauss-Legendre with interpolation (GLI) method. It is shown in [72] that it is the error induced by the Lagrange interpolation which is dominant thus the order of convergence q is the same as the above LII method *i.e.* $q = k+1$. In [23], this numerical integration is performed by a Clenshaw-Curtis rule. The order of convergence is again $q = k + 1$, however as mentioned above, the computation can be done with far less effort by using the fast Fourier Transform.

So far we have restricted our discussion to the 1D problem. We have outlined several 1D methods of numerical integration which could, in theory, be used to construct an estimate I_a of the pest abundance I in an agricultural field represented by the interval $D = [a, b]$. In the next section we discuss the 2D problem and correspondingly give the details of some 2D numerical integration methods.

2.4 Methods of Numerical Integration in Two Dimensions

We now look at the 2D problem and consider the agricultural field as a rectangular domain $D = [a, b] \times [c, d]$. The numerical integration techniques discussed in the previous section can be extended to form 2D formulae. Let the sample units be located at the nodes of a rectilinear grid with N_x nodes in the x direction and N_y nodes in the y direction, hence the total number of sample units is $N = N_x N_y$. The exact value of the pest abundance I is the evaluation of the integral (2.1.2) which can be expressed as

$$I = \int_c^d \left(\int_a^b f(x, y)\, dx \right) dy = \int_c^d F(y)\, dy,$$

where the pest population density function is discrete, namely $f(x, y) \equiv f(x_i, y_j)$, for $i = 1, \ldots, N_x$, and $j = 1, \ldots, N_y$. The integral $F(y)$ can be approximated by treating y as a constant and applying a 1D numerical integration formula

$$F(y) = \int_a^b f(x, y)\, dx \approx \sum_{i=1}^{N_x} w_i f(x_i, y),$$

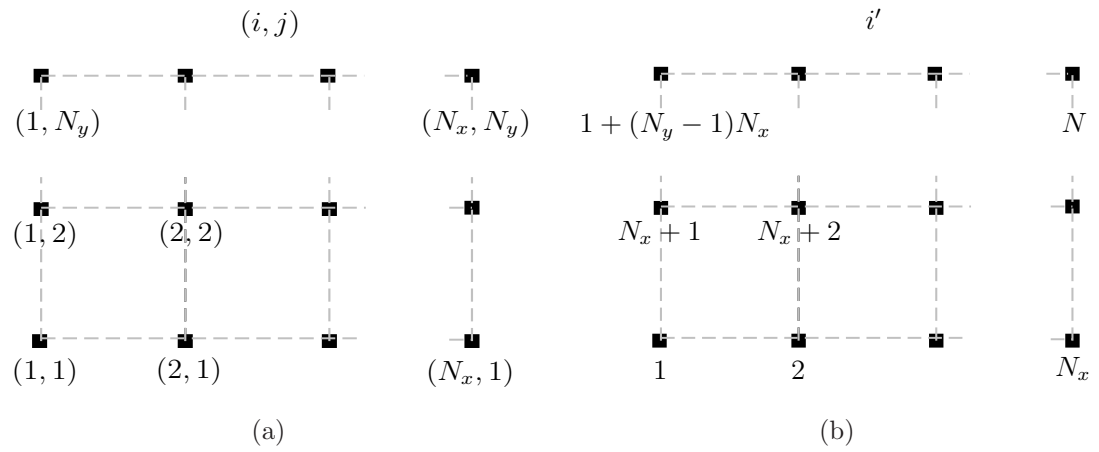


Figure 2.4: (a) The index pair (i, j) corresponding to the coordinate (x_i, y_j) of a sample unit. There is a total of N sample units, with N_x in the x direction and N_y in the y direction. (b) Applying the mapping (2.4.2), the index pair (i, j) is converted into a single index i' thus allowing the double summation (2.4.1) to be expressed in the form (2.1.3).

for some choice of weights w_i. An approximation I_a to the integral I can then be obtained by applying a 1D integration formula to $F(y)$ which is now the integrand

$$I \approx I_a = \sum_{j=1}^{N_y} w_j F(y_j).$$

The approximation I_a is then the evaluation of the double summation

$$I_a = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} w_i w_j f(x_i, y_j) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} W_{i,j} f_{i,j}, \tag{2.4.1}$$

where $f_{i,j} \equiv f(x_i, y_j)$, and we have let $W_{i,j} = w_i w_j$.

The above double summation can be converted to a single summation of the form (2.1.3). The index pair (i, j) corresponds to the coordinate (x_i, y_j) of a sample unit. Applying the following mapping, which is depicted in Figure 2.4, to the index pair (i, j), for $i = 1, \ldots, N_x$ and $j = 1, \ldots, N_y$,

$$(i, j) \to i', \qquad \text{where } i' = i + (j - 1)N_x, \tag{2.4.2}$$

the double summation (2.4.1) is transformed into the following single summation

$$I_a = \sum_{i'=1}^{N} W_{i'} f_{i'}.$$

The above formula is indeed of the form (2.1.3), where the $W_{i'}$ are equivalent to the w_i of (2.1.3). We shall thus revert to using w_i to denote the weights.

We now give the weights $w_i, i = 1, \ldots, N$ in formula (2.1.3) for some well known 2D numerical integration rules. We restrict our consideration to a regular sampling plan such that the grid nodes $(x_i, y_j), i = 1, \ldots, N$ are defined by

$$\begin{array}{lllll} x_1 = a & x_i = x_{i-1} + h_x & i = 2, \ldots, N_x - 1 & x_{N_x} = b, \\ y_1 = c & y_j = y_{j-1} + h_y & j = 2, \ldots, N_y - 1 & y_{N_y} = d, \end{array} \tag{2.4.3}$$

where $h_x = (b-a)/(N_x - 1)$ and $h_y = (d-c)/(N_y - 1)$ are the grid step sizes in the x and y directions respectively. On such a grid, the composite 2D Newton-Cotes formulae can be used to obtain an estimate I_a (*e.g* see [17, 26]). The 2D formulae are derived according to the procedure detailed above, whereby the 1D formula is implemented in each direction in turn. The weights w_i of the 2D composite trapezium rule are given below.

$$\begin{bmatrix} w_1 & \cdots & \cdots & \cdots & w_{N_x} \\ w_{N_x+1} & \cdots & \cdots & \cdots & w_{2N_x} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ w_{N-N_x+1} & \cdots & \cdots & \cdots & w_N \end{bmatrix} = \frac{h_x h_y}{4} \begin{bmatrix} 1 & 2 & 2 & \cdots & 2 & 1 \\ 2 & 4 & 4 & \cdots & 4 & 2 \\ 2 & 4 & 4 & \cdots & 4 & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 4 & 4 & \cdots & 4 & 2 \\ 1 & 2 & 2 & \cdots & 2 & 1 \end{bmatrix} \tag{2.4.4}$$

Meanwhile the 2D compound Simpson's rule has weights

$$\begin{bmatrix} w_1 & \cdots & \cdots & \cdots & w_{N_x} \\ w_{N_x+1} & \cdots & \cdots & \cdots & w_{2N_x} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ w_{N-N_x+1} & \cdots & \cdots & \cdots & w_N \end{bmatrix} = \frac{h_x h_y}{9} \begin{bmatrix} 1 & 4 & 2 & 4 & \cdots & 2 & 4 & 1 \\ 4 & 16 & 8 & 16 & \cdots & 8 & 16 & 4 \\ 2 & 8 & 4 & 8 & \cdots & 4 & 8 & 2 \\ 4 & 16 & 8 & 16 & \cdots & 8 & 16 & 4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 2 & 8 & 4 & 8 & \cdots & 4 & 8 & 2 \\ 4 & 16 & 8 & 16 & \cdots & 8 & 16 & 4 \\ 1 & 4 & 2 & 4 & \cdots & 2 & 4 & 1 \end{bmatrix} \tag{2.4.5}$$

where the application of this rule requires the number of sample units in each direction (*i.e.* N_x and N_y) to be odd.

The error (2.1.6) of an estimate obtained using a 2D composite Newton-Cotes rule is of the form (*e.g* see [17, 26])

$$E_{abs} = \mathcal{O}(h_x)^q + \mathcal{O}(h_y)^q. \tag{2.4.6}$$

For the 2D composite trapezium rule (2.4.4) it is known that $q = 2$, and for the 2D composite Simpson's rule (2.4.5), $q = 4$. The degree of precision, that is the maximum degree of the bivariate polynomial for which the formula produces an exact result, is $r = 1$ and $r = 3$ for the 2D composite trapezium and Simpson's rules respectively (*e.g* see [26]).

We have outlined how numerical integration formulae can be constructed in both one and two dimensions. In the following section we discuss how effective these methods could be in the pest monitoring problem of estimating pest abundance in an agricultural field.

2.5 On the Significance of the Asymptotic Convergence Rate

We have discussed methods of numerical integration along with asymptotic error estimates (2.1.9) which are conventionally used to assess a method's accuracy. In this section we explain how exploiting the convergence rate of a method could be useful in the evaluation of pest abundance. For simplicity, the examples given in this and the following section are 1D, however

the conclusions drawn are valid for the 2D problem which is considered later in the text.

It was outlined in the introduction how the ability to quantify the accuracy of an estimate of pest abundance translates to the ability to assess how much confidence there can be that the resulting pest management decision, *i.e.* whether or not to implement a control action, is indeed the correct one. More specifically, a more accurate estimate gives rise to greater confidence in the decision. Let us define an estimate to be sufficiently accurate if the relative error (2.1.7) satisfies the following condition

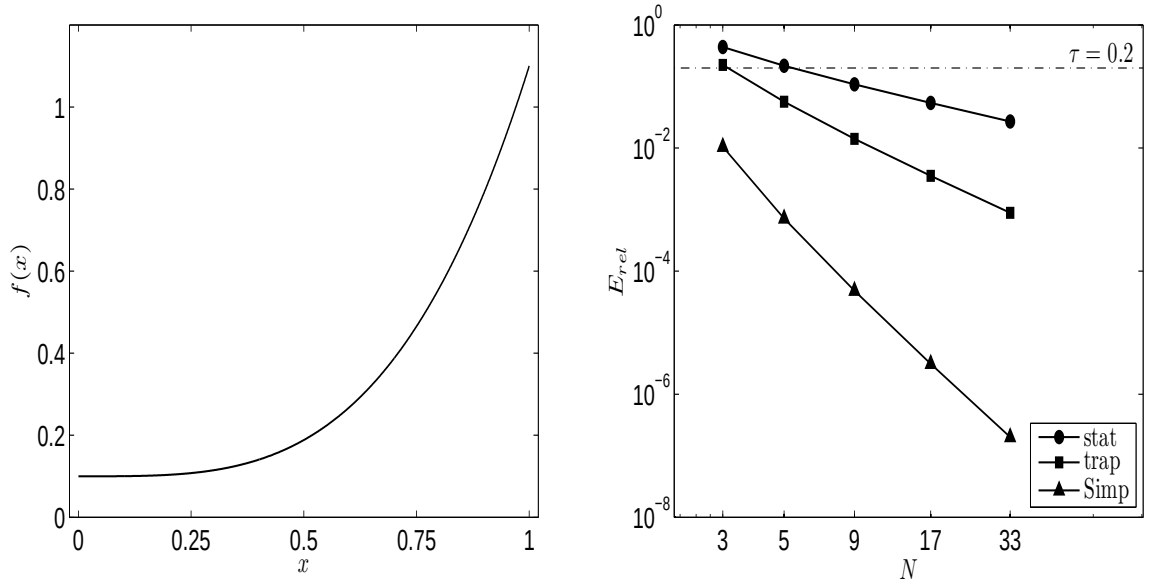
$$E_{rel} \leq \tau, \tag{2.5.1}$$

where we refer to τ as the *accuracy tolerance*. In ecological studies, $0.2 \leq \tau \leq 0.5$ can be thought of as a good level of accuracy while $\tau \sim 1$ may even be acceptable [65, 89]. It was explained in the introduction how the nature of the ecological problem means that the number N of sample units (grid nodes) is limited. It would thus be preferable to employ a means of estimation which achieves the required level of accuracy (2.5.1) for a smaller value of N. From the asymptotic error estimates (2.1.7), we expect a method with a faster convergence rate to be a better option.

To demonstrate this, let us consider an example to compare the accuracy of estimates I_a calculated by different numerical integration methods over a series of grids with an increasing number of N nodes. We assess the accuracy by calculating the relative error (2.1.7) and as such we require the the exact quantity I to be available. Consequently, let us take the pest population density to be mathematically defined by the integrable function $f(x)$ on the interval $[a, b] = [0, 1]$ as

$$f(x) = \sqrt{x^7} + 1/10 \qquad x \in [0, 1] \tag{2.5.2}$$

which is shown in Figure 2.5a. Integrating this function analytically gives the exact value $I = 0.3\dot{2}$. We consider a regular grid of sample unit locations defined according to the formula (2.2.2) for a fixed value of N. The sampling procedure is simulated by evaluating the function (2.5.2) at the computed points $x_i, i = 1, \ldots, N$ to produce a discrete set of data $f_i, i = 1, \ldots, N$.



(a) (b)

Figure 2.5: Example of evaluating the pest abundance by means of numerical integration as given in [34]. (a) A toy example of a pest population density function $f(x)$ defined by the equation (2.5.2). (b) Convergence of the relative approximation error E_{rel} (2.1.7) for estimates obtained using the numerical integration formula (2.1.3) with different weight choices. The abbreviations 'stat', 'trap' and 'Simp' indicate that the estimate was calculated by the statistical rule (2.1.4), the compound trapezium rule (2.2.6) and the compound Simpson's rule (2.2.7) respectively. The accuracy tolerance (2.5.1) is set as $= 0.2$ (dashed line). An estimate is considered to be sufficiently accurate when its relative error lies below this threshold.

An estimate I_a is found by means of formula (2.1.3) with the selected set of weights and the relative approximation error E_{rel} is calculated according to (2.1.7).

The convergence rate is established by computing E_{rel} over a series of grids. The first grid is formed by fixing $N = N_1$ for some initial value N_1. The subsequent grid is then generated by recomputing the number N as $N = 2N - 1$ and this process is repeated. Convergence curves for the error of estimates formed using the statistical method (2.1.4), the compound trapezium rule (2.2.6), and the compound Simpson's rule (2.2.7) are plotted in Figure 2.5b. It can be shown (*e.g* see [28]) that on a regular grid defined by (2.2.2), *i.e.* there are grid nodes at the endpoints of the interval $[a, b]$, the statistical rule has the order of convergence $q = 1$. Meanwhile, as mentioned in Section 2.2 the composite trapezium and composite Simpson's rule have orders of convergence $q = 2$ and $q = 4$ respectively. It can be seen that the relative error adheres to

the asymptotic error estimates and the gradient of the convergence curves correspond to the theoretical convergence rate.

The motivation for considering an estimate other than the sample mean population density is well illustrated by Figure 2.5b. Let us fix the accuracy tolerance (2.5.1) as $\tau = 0.2$. Whereas the statistical rule (2.1.4) requires a grid of $N = 9$ sample units to produce an estimate which satisfies the condition (2.5.1) that the relative error lies below the tolerance τ, the composite trapezium rule (2.2.6) requires a grid of only $N = 5$ units. Furthermore, the composite Simpsons rule (2.2.7) achieves the desired accuracy on the grid of only $N = 3$ sample units. In fact, for $N = 3$ we have $E_{rel} = 0.4363$ using the statistical rule and $E_{rel} = 0.0105$ when the composite Simpson's rule is employed. The estimate formed by Simpsons rule is over forty times more accurate than that provided by the statistical rule. This accuracy translates to greater confidence that pest management decision made using the estimate of pest abundance I_a is the same as the decision that would have been made had the true value of the abundance I been known.

2.6 Chapter 2 Conclusions

We have discussed in this chapter how numerical integration techniques (2.1.3) can be applied in the pest monitoring problem of estimating pest abundance and given examples of specific methods. Asymptotic error estimates (2.1.9) suggest that more accurate results can be obtained by using numerical integration rather than the standard statistical approach where in particular higher order methods seem preferable. This is illustrated by the example given in the previous section. A more accurate estimate of pest abundance is desirable as it translates to greater confidence that the management decision made from this estimate is correct. So far we have painted a rather promising picture for the use of numerical integration in the problem of pest monitoring. The matter is, as demonstrated in the next chapter, the asymptotic error estimates (2.1.9) do not always hold. Applying numerical integration within the framework of pest monitoring does not give rise to a standard problem.

Chapter 3

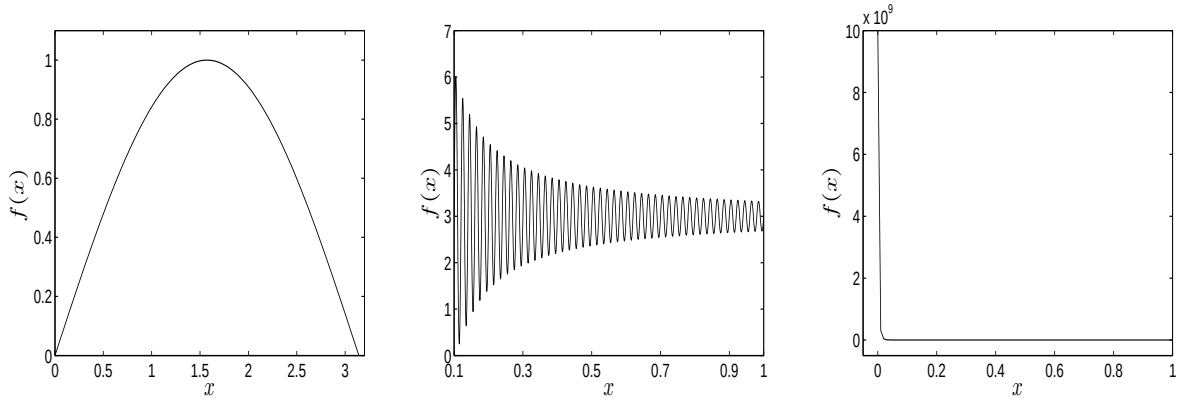
The Coarse Grid Problem of Estimating Pest Abundance

In this chapter it is demonstrated that the restrictions imposed by the ecological problem under consideration can affect the performance of a numerical integration technique. As such, estimating pest abundance gives rise to an atypical numerical integration problem which demands that the conventional theory be revisited. In Section 3.1 it is shown with the aid of mathematically significant test cases how the asymptotic error estimates given in the previous chapter do not always hold. In Section 3.2 we explain how ecologically significant test cases can be generated. Such test cases are studied in Section 3.3 on a regular computational grid. Results on irregular computational grids are given in 3.4. In Section 3.5, 2D data with ecological meaning are considered. Concluding remarks are made in Section 3.6.

3.1 The Coarse Grid Problem

In this section we explain how, in the matter of estimating pest abundance, the asymptotic error estimates (2.1.9) may not hold. This is called the *coarse grid problem* and was highlighted in [70, 72]. We follow the methodology of these papers in this and the following section, studying test case examples of the population density function to investigate the instances in which this problem arises.

An asymptotic error estimate uses the assumption that the distance between the sample units



(a) (b) (c)

Figure 3.1: Test case examples of the pest population density function. (a) The density is taken as the function $f(x)$ given by the equation (3.1.1) where the agricultural field is considered as the interval $[a, b] = [0, \pi]$, (b) the function $f(x)$ is defined by (3.1.2) and $[a, b] = [0.1, 1]$, (c) the equation (3.1.3) describes the function $f(x)$ and $[a, b] = [0, 1]$.

(grid nodes) is small. This means that the number N of sample units is required to be large. Thus, it is only reasonable to rely on an asymptotic error estimates when this condition on N is met. We, however, are proposing to obtain an estimate I_a of pest abundance I by numerically integrating the discrete pest population density function $f_i, i = 1, \ldots, N$ obtained from sampling data. A key restriction imposed by the pest monitoring problem is that the number N is small. In this scenario, an asymptotic error estimate may not hold and thus cannot be relied upon to conclude about the accuracy of an estimate I_a.

To demonstrate the issue let us consider some test cases. We follow the same procedure outlined in the previous section whereby the pest population density is considered to be mathematically defined by an integrable function $f(x)$. Thus the exact value of the abundance I is available to us by integrating the function $f(x)$ analytically. An estimate I_a is obtained over a series of regular grids (2.2.2) with an increasing number N of nodes by implementing a chosen method of numerical integration (2.1.3). The relative error (2.1.7) for each estimate is then calculated.

Let us first define $f(x)$ as follows, where the agricultural field is considered as the interval $[a, b] = [0, \pi]$

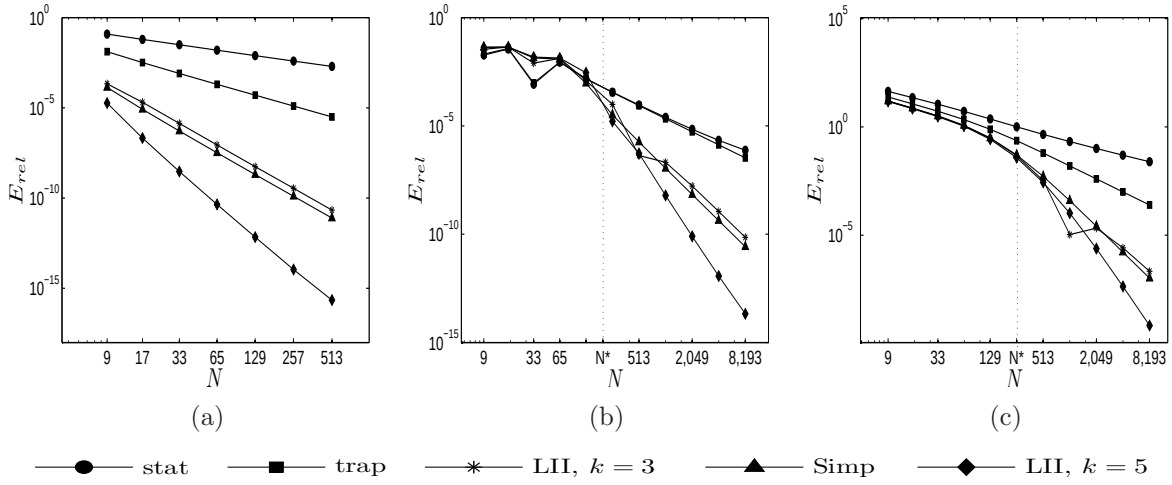


Figure 3.2: Convergence of the error E_{rel} of estimates formed by various methods of numerical integration. The legend in all graphs is as given above. The labels 'stat', 'trap' and 'Simp' indicate that the estimate was formed by the statistical rule (2.1.4), compound trapezium rule (2.2.6) or the compound Simpson's rule (2.2.7), while 'LII' indicates that the LII method outlined in Section 2.3 was employed for the specified interpolating polynomial degree k. The vertical dotted line marks the transitional number of nodes N^* when the asymptotic error estimates appear to begin to hold. (a) Convergence curves for the test case defined by the equation (3.1.1) as shown in Figure 3.1a. We note that the assertion made in Section 2.3 that the LII method when $k = 3$ converges at the same rate as the composite Simpson's rule is validated. (b) and (c) Convergence curves for the test case described by (3.1.2) and (3.1.3) respectively, as shwon in Figures 3.1b and 3.1c.

$$f(x) = \sin(x) \qquad x \in [0, \pi]. \tag{3.1.1}$$

The function is a single peak spanning the entire domain $[a, b]$ as shown in Figure 3.1a. The exact value of the integral is $I = 2$. In Figure 3.2a we compare the convergence of estimates formed by several methods of numerical integration, namely: the statistical rule (2.1.4), the composite trapezium rule (2.2.6), the composite Simpson's rule (2.2.7), as well as the so-called LII method from Section 2.3 for the interpolating polynomial degree $k = 3$ and $k = 5$. It can be seen that, like the example given in the previous section, the error E_{rel} behaves according to the asymptotic error estimates. For a fixed number N of nodes the method with the lowest order of convergence q, *i.e.* the statistical rule, produces the least accurate estimate. Meanwhile, the

method with the highest order of convergence *i.e.* the LII method with $k = 5$, produces the most accurate estimate. This is the case even when the number of grid nodes N is small. We recall that on a grid of type (2.2.2) according to the asymptotic error estimates the statistical rule has the order of convergence $q = 1$, whereas the LII method with $k = 5$ has order of convergence $q = 6$.

We now consider a different definition of the function $f(x)$, namely the following oscillating function

$$f(x) = \frac{\sin(100\pi x)}{\pi x} + 3, \qquad x \in [0.1, 1] \tag{3.1.2}$$

shown in Figure 3.1b, where we have $[a, b] = [0.1, 1]$ and analytical integration yields $I \approx 2.7091$. It can be seen from Figure 3.2b that convergence curves in this case do not always behave according to the asymptotic error estimates. A discernible difference between the methods in terms of accuracy does not present until the number of grid nodes is sufficiently increased. From visual inspection of the graph, a difference between the composite Simpson's method and the supposedly higher order LII method for $k = 5$ only occurs once the number of grid nodes reaches $N > N^* \approx 200$. Sufficient grid refinement is needed before the narrow peaks and troughs of the density function (3.1.2) can be resolved and the asymptotic error estimates begin to hold. Interestingly, on the grids with the smallest number of nodes (*e.g* $N = 9, 17, 33, 65$) it is actually the method with the lowest asymptotic convergence rate, the statistical rule, which produces the most accurate estimates. The conventional approach of using a higher order method to yield a more accurate estimate is not effective on coarse grids.

Finally, let us investigate the convergence of estimates when $f(x)$ is defined as the following function

$$f(x) = \frac{1}{(x + 0.01)^5}, \qquad x \in [0, 1] \tag{3.1.3}$$

where we have $[a, b] = [0, 1]$. A plot of this function is shown in Figure 3.1c. Integrating analytically we obtain the exact value of the abundance

$$I = \frac{(0.01)^{-4} - (1.01)^{-4}}{4} \approx 2.5 \times 10^{7}.$$

Convergence curves for the same set of numerical integration methods considered above are plotted in Figure 3.2c. We see from Figure 3.2c that once again the asymptotic error estimates do not always hold. For the smaller values of N the difference in the performance between the different numerical integration methods is not as distinct as in the case of the examples (2.5.2) and (3.1.1). In particular, there is little difference between the accuracy of the estimates yielded by the composite Simpson's rule and the LII method for $k = 5$ until the number N of nodes in the grid is sufficiently increased. It appears that it is only when $N > N^* \approx 257$ that the performance of the two methods become differentiable and the higher order LII method for $k = 5$ begins to become distinguishably superior. The matter is that the function $f(x)$ is such that the density is largely contained in a very narrow sub-interval. It thus takes sufficient refinement of the computational grid to detect and adequately resolve this region. For grids with a small number N of nodes, this region is entirely missed and the estimate I_a obtained by numerical integration is inaccurate. Note that the jump in accuracy visible in Figure 3.2c for convergence curve corresponding to the the LII method approximation with interpolating polynomial degree $k = 3$, is due to the computational grid geometry (see Appendix A).

Asymptotic error estimates are the usual means of drawing conclusions about the accuracy of an estimate I_a when the exact value I is not known. However, the above examples demonstrate that such estimates do not hold when the number N of grid nodes is too small, *i.e.* the grid is too coarse, to effectively resolve the density function $f(x)$. This leads us to define a computational grid on which the asymptotic error estimates do not hold as a *coarse grid*. Meanwhile, we will use the term *fine grid* to describe a computational grid where the asymptotic error estimates do hold. We denote the number of grid nodes at which the transition occurs as N^*. A grid with $N < N^*$ is therefore considered as coarse, whilst a grid with $N > N^*$ will be defined to be fine.

The examples considered above alongside the wider class of test cases considered in [72] show that the number of nodes N^* at which the asymptotic error estimates begin to hold *i.e.*

the number of nodes at which the computational grid transitions from coarse to fine depends on the spatial pattern of the pest population density function. When the population is spread homogeneously across the entire domain as in Figure 3.1a the asymptotic error estimates hold even for small N. Meanwhile, when the density distribution becomes patchy which is comparable to the test case shown in Figure 3.1b, an increased number N^* of nodes is required for the asymptotic error estimates to hold. The situation becomes more difficult when the density is confined to a small region (see Figure 3.1c). A heuristic estimate of the value of N^* for an integral calculated on an interval of length 1 was given in [70] as

$$N^* \approx s\frac{1}{\Delta x}, \tag{3.1.4}$$

where Δx is defined as the width of a characteristic of spatial heterogeneity in the density function *e.g* the width of a sub-region of steep gradient, or the width of a single peak. The coefficient $s \geq 1$ corresponds to the number of points needed to resolve the spatial heterogeneity characteristic, therefore it depends on its nature. This coefficient is set as $s = 1$ if the characteristic is a sub-region of steep gradient (as shown in Figure 3.1c) as only one point needs to be located within this sub-region for it to be detected [70]. Meanwhile, it is considered in [70] that when the spatial heterogeneity takes the form of a peak (*e.g* as in Figure 3.1b) that there needs to be three grid nodes located within each peak to resolve it, *i.e.* $s = 3$. The estimate (3.1.4) can be applied to an interval of arbitrary length $[a, b]$ by multiplying by $(b - a)$. For the rapidly oscillating test case of Figure 3.1b we thus arrive at the estimate $N^* \approx 135$ where $\Delta x = 0.02$ is the width of a single peak and we recall that $[a, b] = [0.1, 1]$. Meanwhile, for the test case shown in Figure 3.1c we have $[a, b] = [0, 1]$ and $\Delta x = 0.05$ as the width of the region of steep gradient so the estimate is $N^* \approx 200$. Whilst these estimates are not precise, they give a good indication of the order of the number N^* of nodes required to ensure that the asymptotic error estimates hold.

We emphasise that the definitions of coarse and fine grids depend on the behaviour of the error E_{rel} of an estimate rather than being definitively defined by the number N of grid nodes. A grid with a fixed number N of nodes may be considered fine for one function $f(x)$, and coarse

for another. For example it can be seen by comparing Figures 3.2a with 3.2b that a grid with $N = 9$ nodes can be described as fine when the function (3.1.1) is considered but coarse for the function (3.1.2). This means that it cannot be determined from the knowledge of the number N alone whether or not the corresponding computational grid is coarse or fine. What can be said, however, is that when N is small the coarse grid problem may be encountered. Thus, for small N the asymptotic error estimates cannot be relied upon to conclude about the accuracy of an estimate. In terms of the pest monitoring problem, if we are not able to quantify the accuracy of an estimate of pest abundance, then we cannot say how much confidence can be placed in the management decision it implicates. An alternate means of assessing the accuracy of estimates on coarse grids is needed.

In this section we have given examples of mathematically defined functions $f(x)$ such that the coarse grid problem arises. Whilst these test cases were demonstrative, they are of no real ecological significance. We now go on to consider ecologically meaningful test cases to investigate if and when the same problem is likely to occur in the problem of pest abundance evaluation.

3.2 Generating Ecologically Significant Data

We continue our study of the coarse grid problem by turning our attention to data sets $f_i, i = 1, \ldots, N$ which are ecologically meaningful. We want to again look at the convergence rates of estimates formed by methods of numerical integration. To produce convergence curves akin to those shown in Figure 3.2 in the last section, we need to be able to evaluate the relative error E_{rel} over a series of computational grids. Using simulated data allows us to fulfill these requirements.

The model used to generate the data is the spatially explicit form of a predator-prey model with the Allee effect [1] also known as the *Rosenzweig-MacArthur model* [87]. It consists of a set of coupled reaction-diffusion equations which describes the spatio-temporal dynamics of the pest species i.e. how the pest species moves within a closed space over time. The prey species is considered to be the pest attacking the crops, while the predator is a species which in turn

[1] The Allee effect (*e.g* see [59, 104]) is a nonlinear positive relationship between the population density and the per capita growth rate. It means that if the population size falls below a threshold value, the per capita growth rate will decrease.

feeds on the pests. The spatially explicit form of the Rosenzweig-MacArthur model is as follows

$$\frac{\partial F(X,Y,T)}{\partial T} = D_1\left(\frac{\partial^2 F}{\partial X^2}+\frac{\partial^2 F}{\partial Y^2}\right)+F(F-F_T)(K-F)\frac{4\alpha}{(K-F_T)^2}-\frac{AFG}{F+B} \quad (3.2.1)$$

$$\frac{\partial G(X,Y,T)}{\partial T} = D_2\left(\frac{\partial^2 G}{\partial X^2}+\frac{\partial^2 G}{\partial Y^2}\right)+\kappa\frac{AFG}{F+B}-MG \quad (3.2.2)$$

(*e.g* see [59, 104]). The prey and predator population densities are denoted F and G respectively and are functions of position (X, Y) and time T, where $T > 0$ and we take $0 < X < L$, $0 < Y < L$. The parameter L represents the lengths of the closed square space, which represents an agricultural field. The first terms in both equations represent the dispersal of the species. In (3.2.1) the second is the combined births and deaths of the prey, and the third term represents the prey lost to the predator. In (3.2.2), the second term signifies the growth of the species due to predation, and the final term represents the death of the predator.

The above system depends on eight parameters: D_1 and D_2 denote the diffusion rates of populations F and G respectively, K represents the carrying capacity for the prey, F_T is the Allee threshold density where $0 < F_T < K$, α is the maximum prey per capita growth rate, A is the rate at which the predator attacks, B is the half-saturation prey density, the food assimilation efficiency coefficient is denoted by κ, and M signifies the death rate of the predator. The model described by equations (3.2.1) and (3.2.2), has been rigorously tested against field data [55] and shown to be a reliable within a certain parameter range. Therefore we believe the test cases subsequently generated from this model to be ecologically realistic.

In order to make the implementation of the model simpler, the number of parameters upon which the system depends is reduced. Firstly let $D_1 = D_2 = D$. The number of parameters can then be further reduced by writing the equations (3.2.1) and (3.2.2) in dimensionless form. The non-dimensionalisation process is as described in [71]. The variables X and Y represent the species position in space and hence they are lengths. Therefore, X can be non-dimensionalised by scaling by the length of the agricultural field in the X direction, namely L. That is, the dimensionless variable $x = X/L$ is introduced. Correspondingly, the variable $y = y/L$ is also introduced where we recall that L is also the length of the agricultural field in the y direction.

The carrying capacity K is measured in inverse area units, as is the prey density F, therefore $f = F/K$ is the dimensionless prey density. The dimensionless predator density is defined as $g = G/(\kappa K)$. Finally, the attack rate A has units in terms of inverse time, and the parameter B is measured in inverse area units. Hence $t = \kappa AKT/B$ is a dimensionless time variable. Substituting these dimensionless variables into the equations (3.2.1) and (3.2.2) gives a dimensionless form of the system (*e.g* see [71])

$$\frac{\partial f(x,y,t)}{\partial t} = d\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) + \beta f(f-\chi)(1-f) - \frac{fg}{1+\Lambda f} \tag{3.2.3}$$

$$\frac{\partial g(x,y,t)}{\partial t} = d\left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) + \frac{fg}{1+\Lambda f} - mg \tag{3.2.4}$$

where $d = BD/(\kappa AKL^2)$, $m = MB/(\kappa AK)$, $\Lambda = K/B$, $\beta = 4\alpha BK/(A\kappa(K-F_T)^2)$, and $\chi = F_T/K$ are dimensionless parameters. Numerical solution of the equations (3.2.3) and (3.2.4) at a fixed time $t = \hat{t} > 0$ and for chosen values of the other parameters yields a discrete density distribution $f(x,y) \equiv f_i, i = 1, \ldots, N$.

Before we consider the 2D problem, we first gain some insight by studying some 1D test cases with ecological significance. These test cases were generated in [70] from the numerical solution of the 1D counterpart of the system of equations (3.2.1) and (3.2.2), namely

$$\frac{\partial F(X,T)}{\partial T} = D_1\frac{\partial^2 F}{\partial X^2} + \alpha F\left(1 - \frac{F}{K}\right) - A\frac{FG}{F+B} \tag{3.2.5}$$

$$\frac{\partial G(X,T)}{\partial T} = D_2\frac{\partial^2 G}{\partial X^2} + \kappa A\frac{FG}{F+B} - MG \tag{3.2.6}$$

where $T > 0$ and $0 < X < L$ and the parameters have the same meaning as described earlier. Matters are simplified by again assuming $D = D_1 = D_2$. The non-dimensionalisation procedure is as described in [70] whereby the variables are redefined as $x = X/L$, $t = \alpha T$, $f = F/K$ and $g = AG/(\alpha K)$. The dimensionless form of the above 1D system is thus

$$\frac{\partial f(x,t)}{\partial t} = d\frac{\partial f^2}{\partial x^2} + f(1-f) - \frac{fg}{f+\Lambda} \tag{3.2.7}$$

$$\frac{\partial g(x,t)}{\partial t} = d\frac{\partial g^2}{\partial x^2} + k\frac{fg}{g+\Lambda} - mg \tag{3.2.8}$$

where $d = D/(\alpha L^2)$, $\Lambda = B/L$, $k = \kappa A/\alpha$, and $m = M/\alpha$ are dimensionless parameters. Solving the equations (3.2.7) and (3.2.8) at some fixed time t on a regular grid (2.2.2) of nodes $x_i, i = 1, \ldots, N$ gives a discrete density function $f(x_i) \equiv f_i, i = 1, \ldots, N$.

3.3 Numerical Integration of Ecological Data on Regular Grids

In this section 1D ecologically significant test cases are studied to gain insight, 2D examples are considered later in the chapter in Section 3.5. Three 1D test case examples as shown in Figure 3.3 are considered. These test cases were generated in [70] by numerically solving the system of equations (3.2.7 –3.2.8) on a very fine regular grid of $N_f = 32,769$ nodes across the unit interval $[0, 1]$. The datasets were supplied by the authors of [70] and subsequently used within this study. The system of equations were solved by the method of finite differences where the results of a explicit finite difference scheme was also validated by a more advanced alternate directions scheme to avoid any numerical artefact (see [78]). The boundary conditions were taken as $\partial f/\partial x = \partial g/\partial x = 0$ at any boundary of the domain (i.e. the no-flux conditions). Varying the dimensionless time t and diffusion d parameters produces a differing number of peaks: a single peak function shown in Figure 3.3a, a three peak function shown in Figure 3.3b, while Figure 3.3c shows a multi-peak density function. The values of Λ, k and m are the same for each test case. The specific parameter values, along with the initial conditions used are given in the caption of Figure 3.3.

We now investigate the rate at which estimates of pest abundance converge for each of the test cases shown in Figure 3.3. We again follow the procedure previously outlined. That is, we calculate estimates I_a by a means of numerical integration (2.1.3) and then the relative error E_{rel} as defined by (2.1.7) over a series of regular grids (2.2.2) with an increasing number N of nodes. Since the pest population density function was constructed by numerical solution of

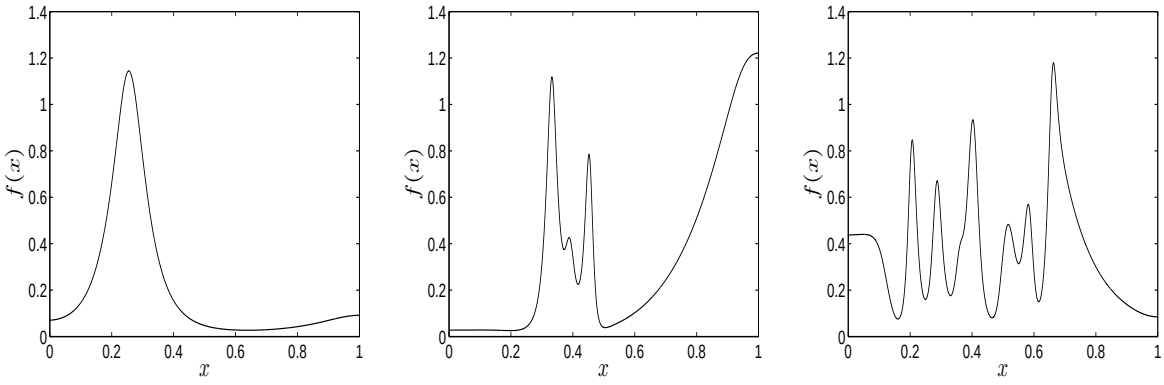


Figure 3.3: Test case examples of the pest population density function as used in [70] generated from the 1D spatially explicit form of the Rosenzweig-MacArthur model given by equations (3.2.7) and (3.2.8). The dimensionless time and diffusion parameters are set as (a) $t = 50, d = 10^{-4}$ (b) $t = 100, d = 10^{-5}$ and (c) $t = 400, d = 10^{-5}$. The remaining parameters are fixed as $\Lambda = 0.3$, $k = 2$ and $m = 0.7$ for each test case. The initial conditions used in the system are $f(x, 0) = \hat{f}$ and $g(x, 0) = \hat{g} + x + 0.01$, where $(\hat{f}, \hat{g})$ is the unique steady state of the non-spatial form of the system, *i.e.* the equations (3.2.7) and (3.2.8) without the first term. The population density function is discrete available only at the nodes of a fine, regular, computational grid.

a mathematical model, it is discrete. Hence, we cannot integrate the function analytically to obtain the true value of the pest abundance I. However, the population density function is available to us at the nodes of a very fine regular grid of $N_f = 32,769$. We thus integrate the function numerically on this fine grid and treat this value which we denote I_{N_f} as the 'exact' value. We choose the compound trapezium rule (2.2.6) to generate this 'exact' value. A simple, low order method is used in order to limit the effect of the round off error incurred over such a fine mesh. An estimate I_a on a coarser grid of $N < N_f$ nodes is found by extracting the relevant pest density values from the already established set $f_i, i = 1, \ldots, N_f$ and applying the chosen method of numerical integration.

Figures 3.4a–c show the convergence curves of estimates formed by the composite Simpson's rule (2.2.7) and the LII method of Section 2.3 with interpolating polynomial degree $k = 5$ for the test cases shown in Figure 3.3a–c respectively. We compare the performance of these two methods to determine if the coarse grid problem described in the previous section arises when ecologically significant data is considered. If the coarse grid problem does not present itself, and

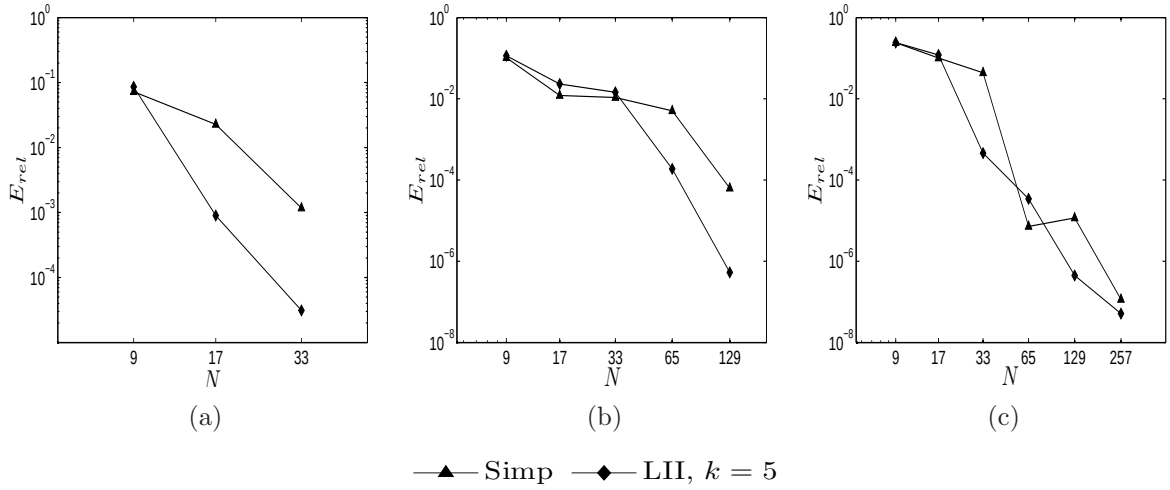


Figure 3.4: Convergence of the error E_{rel} of estimates formed by the the compound Simpson's rule (2.2.7) or the LII method outlined in Section 2.3 where the interpolating polynomial degree is set as $k = 5$. The legend in all graphs is as given above. (a)–(c) Convergence curves for the test case shown in Figure 3.3a–c respectively.

the asymptotic error estimates (2.1.9) hold, then we would expect the LII method with $k = 5$ which has order of convergence $q = 6$ to produce more accurate estimates than the composite Simpson's rule which we recall has order of convergence $q = 4$.

It can be seen from Figure 3.4 that this is not always the case; the number N of grid nodes must be sufficiently increased before the accuracy of the estimates behave in accordance with the asymptotic error estimates and the accuracy of the LII method with $k = 5$ becomes superior. For instance, Figure 3.4a shows that on a grid of $N = 9$ nodes, the accuracy of the two methods is very similar for the one peak density function of Figure 3.3a. In fact, the estimate produced by the composite Simpson's rule is slightly more accurate. However, once the grid is refined to have $N > N^* \approx 9$ nodes, the LII method with $k = 5$ produces a more accurate estimate. By the time the grid is refined to have $N = 17$ nodes, the LII, $k = 5$ estimate is more accurate by several orders of magnitude than the composite Simpson's rule. Grids of $N > N^* \approx 9$ nodes can be described as fine since the asymptotic error estimates hold. Figures 3.4b and c show that further grid refinement, *i.e.* a higher number N of grid nodes is required in order to sufficiently resolve the increased level of spatial heterogeneity of the three peak and multi-peak density functions such that the error behaves according to the error estimates. The grid appears to be

sufficiently refined when $N = 65$ for the three peak distribution. On grids with $N = 9, 17$ and 33, the difference between the accuracy of the two numerical integration methods is small. For the multi-peak density function, the grid must be refined to $N = 129$ nodes before the asymptotic error estimates hold reliably. Looking at the behaviour of the error prior to this level of grid refinement, there is at first little difference between the accuracy of the two methods, *cf.* the errors on the grids with $N = 9$ and $N = 17$. There is then some alternation between which method produces the most accurate estimate until the grid is refined to have $N = 129$ nodes.

These results demonstrate that the coarse grid problem arises when ecologically significant data is considered. Furthermore, they confirm the conclusion of the previous section that the number N^* of grid nodes required for the asymptotic error estimates to hold depends on the spatial pattern of the density distribution. The value of N^* grows larger as the density distribution becomes more patchy. This was shown to be the case for a more comprehensive set of 1D test cases with ecological significance in [70].

It should be pointed out that sufficiently accurate estimates may still be obtained on coarse grids; however that the accuracy cannot be predicted by the error estimates. For instance, on a grid with just $N = 9$ nodes, the relative error of estimates obtained via the compound Simpson's rule and the LII method for $k = 5$ are $E_{rel}^{Simp} \approx 0.10267$ and $E_{rel}^{LII,k=5} \approx 0.11483$ respectively for the three peak test case shown in Figure 3.3b. For the multi-peak test case the corresponding errors are $E_{rel}^{Simp} \approx 0.24011$ and $E_{rel}^{LII,k=5} \approx 0.24135$. These values are within the tolerance range $0.2 \leq \tau \leq 0.5$ considered as good accuracy in ecological applications. Before we move on to study ecologically significant data in 2D, we first investigate numerical integration on irregular grids in the next section which follows the discussion given in [68].

3.4 Numerical Integration on Irregular Grids

So far we have considered using methods of numerical integration to evaluate pest population abundance when the sampling plan is a regular grid, *i.e.* the samples are taken at regular spatial intervals. However, it may be that an irregular grid is prescribed in a pest monitoring programme. Furthermore, even if a regular grid has been selected as the intended sampling plan, taking samples at precisely regular intervals may not be possible in practice. The landscape of

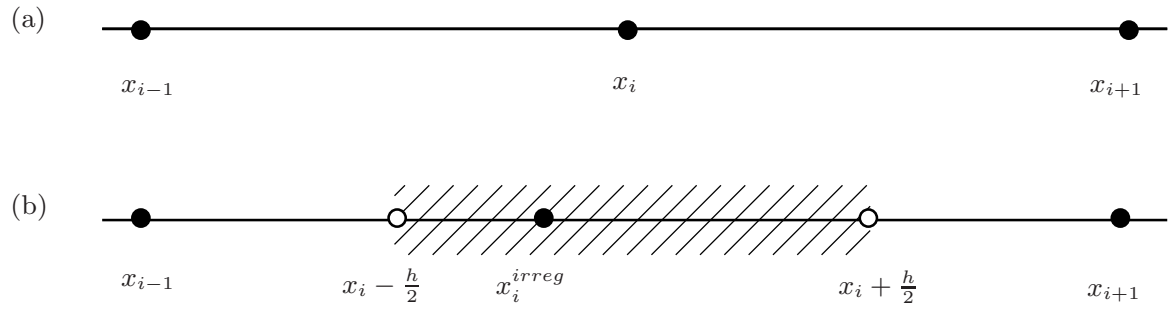


Figure 3.5: *(a) An interior grid node x_i for some $i = 2, \ldots, N-1$ is a fixed distance h from its neighbouring grid nodes in accordance with a regular sampling plan. (b) An interior grid node x_i^{irreg} which has been perturbed according to the transformation (3.4.1). The shaded region shows the possible locations for x_i^{irreg}, where this node is no longer an equal distance from its neighbouring nodes.*

an agricultural field may have natural obstacles (e.g. a bush or a tree) that mean one or many of the samples may then have to be taken at a location shifted from that which was intended, hence the resulting grid of samples is irregular. We thus now investigate the accuracy of numerical integration methods formulated on an irregular grid. Our analysis is focused on a 1D problem for the sake of simplicity.

We have seen that estimates on coarse grids can have poor accuracy *e.g* see the integration error for small values of N in Figure 3.2b. One way to improve the accuracy of integration on coarse grids would to use an irregular grid where most of the grid nodes are concentrated in sub-regions that present difficulties in their numerical integration (i.e. peaks or sub-regions with a steep function gradient). However, in pest monitoring it often is not possible to use irregular grids adapted to a spatial pattern of the density distribution because that pattern is usually not known a priori. We assume we have no such existing knowledge.

We consider several types of grids with varying degrees of irregularity: a slightly irregular grid, a quasi-random grid, and a random grid. We use the term *slightly irregular* to refer to a simple example of an irregular grid, whereby a single sampling location is shifted from the position prescribed by a regular sampling plan. We generate such a grid by first constructing a regular grid (2.2.2) of N nodes. A single interior node x_i, for some $i = 2, \ldots, N-1$ is then

perturbed according to the following transformation:

$$x_i^{irreg} = x_i + h\left(\gamma - \frac{1}{2}\right), \tag{3.4.1}$$

where x_i is a node location on a regular grid, and $\gamma \in (0,1)$ is a uniformly distributed random variable. The transformation (3.4.1) is further illustrated in Figure 3.5.

A *quasi-random grid* has an increased level of irregularity whilst preserving some structure. Such grids are generated in a similar way to the method discussed above for the slightly irregular grids. The difference is that instead of a single interior node being perturbed, *all* interior nodes are perturbed. That is, the transformation (3.4.1) is applied to all interior nodes $x_i, i = 2, \ldots N-1$ of the regular grid. This form of grid is closely related to the so called *centric systematic* sampling plan (*e.g* see [58]) whereby the field is divided into sections and a sample is taken from a random location within each section. Our version differs only in that we have fixed the boundary points so as to preserve the interval of integration as $[a, b]$. A random sampling plan is often viewed favourably from a theoretical viewpoint as it is considered to avoid introducing bias into the estimate [13, 51, 85, 92], the concern being that a systematic distribution of samples will somehow coincide with the distribution of the pests. We therefore take into consideration such a distribution of samples in our investigation and generate the points $x_i, i = 1, \ldots, N$ as follows:

$$x_i = a + \gamma(b-a), \qquad i = 1, \ldots, N, \tag{3.4.2}$$

where $\gamma \in (0,1)$ is a uniformly distributed random variable. The points $x_i, i = 1 \ldots N$ are then sorted into ascending order and the endpoints on a random grid are then replaced as

$$x_1 = a, \qquad x_N = b. \tag{3.4.3}$$

We now look at the accuracy of pest abundance estimates obtained by methods of numerical integration on the grids outlined above. We will be using the statistical rule, and forms of the trapezium rule and Simpson's rule to evaluate the pest abundance. Since the statistical rule (2.1.4) has no spatial dependence it can be applied to regular and irregular grids alike.

Meanwhile, we must use different forms of the trapezium and Simpson's rules to (2.2.6) and (2.2.7) in order to be able to apply them to irregular grids. The idea remains the same as discussed in Sections 2.2 and 2.3: replace the integrand by a polynomial function and integrate the polynomial. The generalised trapezium rule on irregular grids is given by

$$I \approx \tilde{I} = \sum_{i=1}^{N-1} h_i \frac{(f_i + f_{i+1})}{2}, \tag{3.4.4}$$

where the grid step size $h_i = x_{i+1} - x_i$ is variable rather than fixed as in the formula for regular grids. We use the following adapted version of Simpson's rule to handle irregular grids

$$I \approx \tilde{I} = \sum_{i=1}^{\frac{N-1}{2}} \frac{h_{2i-1} + h_{2i}}{6} \left(f_{2i-1} + 4f_{2i} + f_{2i+1}\right), \tag{3.4.5}$$

which also relies on the variable grid step size $h_i = x_{i+1} - x_i$. As with the conventional Simpson's rule (2.2.7), the number of grid nodes N is required to be odd.

We illustrate the convergence on irregular grids by considering a sequence of grids as was done above for regular grids. In the case of the slightly irregular grids, we want to determine how perturbing a single node affects the convergence rate of a method of numerical integration, rather than how the position of the grid node which is perturbed affects the accuracy. As such, in each generation of the slightly irregular grids, the same interior grid node is perturbed. We will begin all of our calculations on a grid of 3 grid nodes which has only one interior node. The unperturbed position of this node lies at $x = (a + b)/2$, therefore, it will always be this central node which is perturbed in the generation of each slightly irregular grid. For grids with a more significant level of irregularity *i.e.* the quasi-regular and random grids, each grid generation is repeated a total of n_r times thus providing n_r values of the relative error (2.1.7) for any given grid of N nodes. The mean error on a grid of N nodes is then calculated as

$$\mu(E_{rel}) = \frac{1}{n_r} \sum_{i=1}^{n_r} E_{rel_i} \ . \tag{3.4.6}$$

We first return to a standard mathematical test case considered in Section 3.1 and evaluate

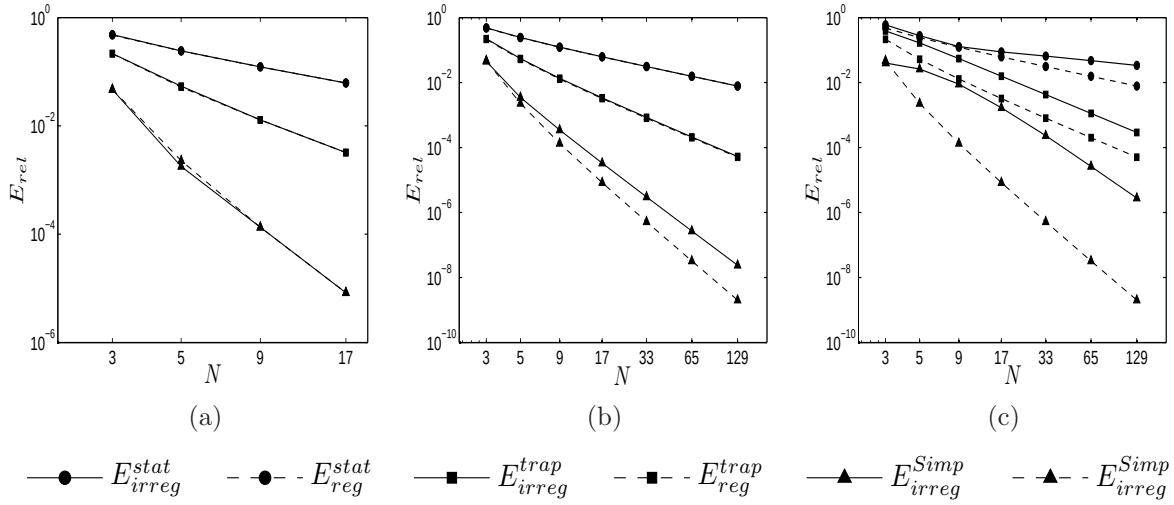


Figure 3.6: Numerical integration of the function (3.1.1) on irregular grids. Computations are made using the statistical rule (2.1.4) (solid line, closed circle), the trapezium rule (3.4.4) (solid line, closed square), and Simpson's rule (solid line, closed triangle). (a) The relative error (2.1.7) is calculated on slightly irregular grids. Meanwhile the mean (3.4.6) of $n_r = 10,000$ evaluations of the relative error is obtained on (b) quasi-random and (c) random grids. Each convergence curve is compared with the convergence on regular grids (dashed line). The legend for each figure is as given above.

the integral of the function 3.1.1 shown in Figure 3.1a over a sequence of increasingly refined irregular grids. Evaluations of the error are shown in Figure 3.6. The convergence rate of errors calculated over increasingly refined regular grids have also been plotted in each graph for comparison purposes (see the dashed line in the figure). It can be seen from Figure 3.6a that very little difference is made to the accuracy by perturbing a single node as the results for the regular and slightly regular grids are close to each other.

For the quasi-random and random grids, the mean (3.4.6) of $n_r = 10,000$ evaluations of the error have been plotted in Figure 3.6b and Figure 3.6c respectively. Again the convergence curves of errors calculated over regular grids have also been plotted. It can be seen from Figure 3.6b that on average the accuracy on the quasi-random grids is similar to that on regular grids for the statistical rule and trapezium rule. A more distinctive difference is evident in the case of Simpson's rule, where the increased level of irregularity leads to a more inaccurate estimate. Using a random computational grid affects the convergence rate with varying degrees

of prominence depending on the method of numerical integration employed as can be seen in the Figure 3.6. The behaviour of the convergence curve for the statistical rule is different from the convergence for the trapezium rule and Simpson's rule. On average the randomness introduced to the computational grid causes the convergence curves of the trapezoidal and Simpson's rules to be shifted upwards, that is, the resulting estimates are less accurate although they begin to converge at a similar rate to those formulated on regular grids as N increases. The higher the degree of method applied, the more prominent the effect seems to be, although it should be noted that on average the accuracy still improves when a higher degree method is used. Let us now consider the accuracy of the numerical integration of ecologically significant data on irregular grids. Since we are required to perform repeated calculations over increasingly refined grids, we again use simulated data. As earlier explained, the simulated ecological population density functions were obtained through numerical solution of the 1D system (3.2.7 –3.2.8) on an extremely fine, regular grid of $N_f + 1 = 2^{15} + 1$ nodes on the interval $[a, b] = [0, 1]$. Since the density functions are thus discrete rather than continuous, the method for generating the slightly irregular computational grid is now different to that outlined above although the fundamental ideas are the same. We have available a fine grid of points $x_i^f, i = 1, \ldots, N_f$ where

$$x_1^f = a = 0, \qquad x_i^f = x_{i-1} + \frac{b-a}{N_f - 1}, \quad i = 2, \ldots, N-1, \qquad x_{N_f}^f = b = 1.$$

To generate a slightly irregular grid of N nodes, a regular grid is first obtained by extracting the required N nodes from the available fine grid as

$$x_i = x_j^f, \qquad j = 1 + (i-1)\left(\frac{N_f - 1}{N-1}\right), \quad i = 1, \ldots, N. \tag{3.4.7}$$

A single interior node must then be perturbed, however, it must be perturbed to a value for which the population density is available. This is achieved by replacing an interior grid node as

$$x_i = x_{j+r}^f, \qquad \gamma \in \left[-\frac{N_f - 1}{2(N-1)}, \frac{N_f - 1}{2(N-1)}\right] \tag{3.4.8}$$

for some $i = 2, \ldots, N-1$, where j is as given in (3.4.7) and γ is a uniformly distributed random

integer.

The generation of quasi-random grids for use with simulated ecological data is as follows. The endpoints are fixed as

$$x_1 = x_1^f, \qquad x_N = x_{N_f}^f, \tag{3.4.9}$$

and the interior points are defined as

$$x_i = x_{j+r}^f, \qquad r \in \left[-\frac{N_f - 1}{2(N-1)}, \frac{N_f - 1}{2(N-1)} - 1 \right], \quad i = 2, \ldots, N-1. \tag{3.4.10}$$

Note that here the upper limit of the interval to which r belongs is one less than that in (3.4.8) so as to avoid any nodes coinciding.

To extract a random grid from the available data, the grid nodes of the fine grid $x_i^f, i = 1, \ldots, N_f$ are first permuted randomly. We shall denote the resulting points as $\tilde{x}_i^f, i = 1, \ldots, N_f$. We begin to form a random grid of N nodes by selecting the first N nodes from the permuted fine grid so we have

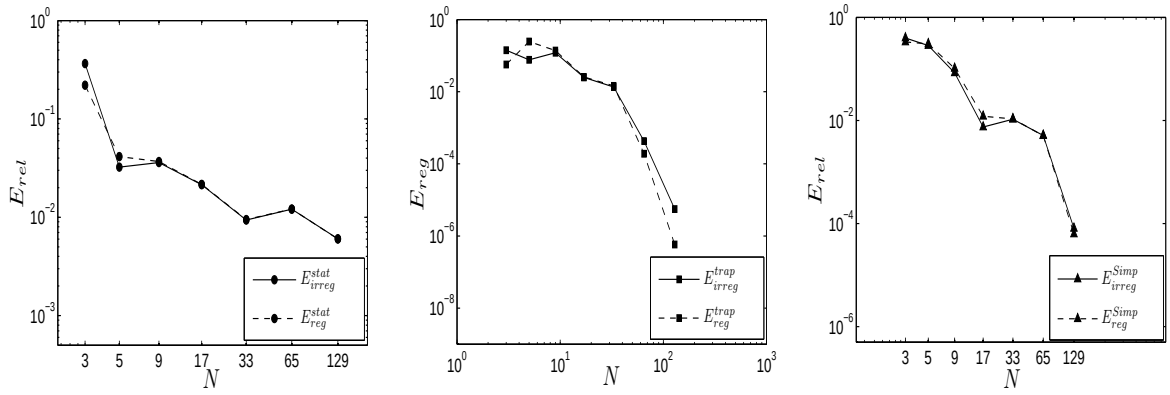
$$x_i = \tilde{x}_i^f, \qquad i = 1, \ldots, N. \tag{3.4.11}$$

The nodes $x_i, i = 1, \ldots, N$ are then sorted into ascending order and the endpoints are replaced as

$$x_1 = a = 0, \qquad x_N = b = 1. \tag{3.4.12}$$

Let us now consider the three-peak simulated ecological test case from the previous section as shown in Figure 3.3b. As above, we generate a sequence of increasingly refined grids and the relative errors are calculated according to (2.1.7). The value I has been calculated by applying the composite trapezium rule on the extremely fine, regular grid of N_f nodes. For the quasi-random and random grids, $n_r = 10,\,000$ of each grid are generated and a mean of the errors is calculated.

Convergence curves for the slightly irregular grids, where one node is randomly shifted from its original location on a regular grid, are shown in Figure 3.7. The integration error (2.1.7) computed for the statistical rule (2.1.4) is presented in Figure 3.7a, while the error for the



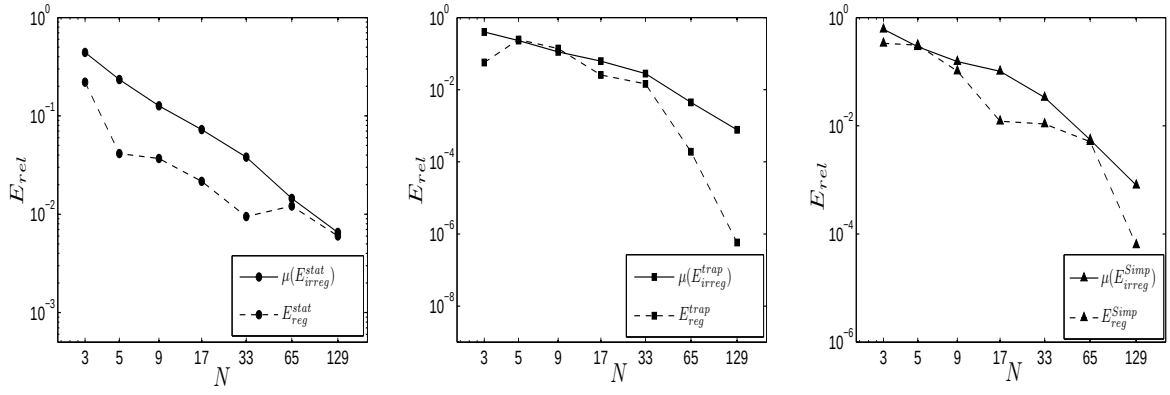
(a) (b) (c)

Figure 3.7: Convergence curves on slightly irregular grids for the ecologically meaningful density distribution of Figure 3.3b. Convergence on a sequence of grids where a central grid node is randomly shifted is compared to the convergence on regular grids. The method of integration is (a) the statistical rule (2.1.4), (b) the trapezium rule (3.4.4), (c) Simpson's rule (3.4.5).

trapezoidal rule (3.4.4) and the Simpson rule (3.4.5) is shown in Figure 3.7b and Figure 3.7c, respectively. The convergence results in the figure confirm our previous conclusion made for the function (3.1.1). A slight perturbation of grid regularity results in a slight perturbation in the integration error.

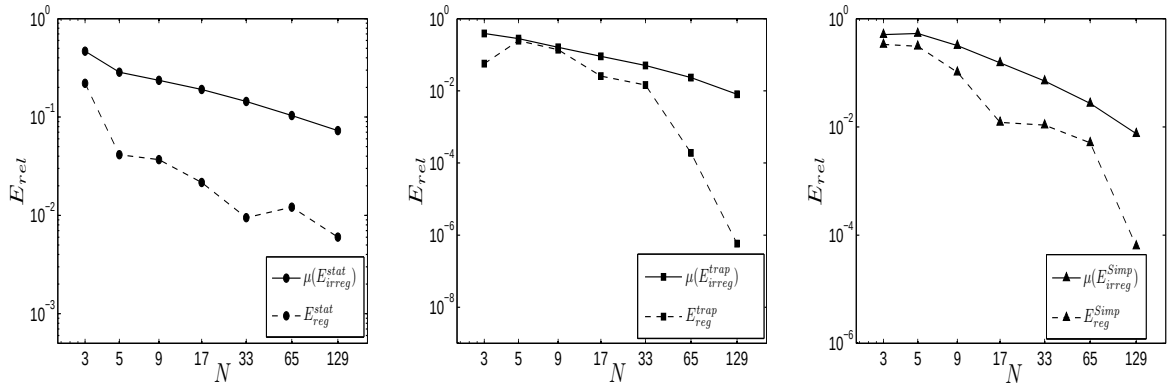
Let us now make a stronger perturbation of a regular grid and consider numerical integration on a sequence of quasi-random grids where each interior grid node is randomly shifted around its position on a regular grid. The corresponding convergence curves are shown in Figure 3.8. It can be seen from the figure that increasing the degree of grid randomness in the problem results in a bigger integration error, no matter what integration method is used. This conclusion is further illustrated by consideration of integration error on truly random grids; see Figure 3.9. Again, the convergence curves shown in Figure 3.9 for integration on regular grids always lie below convergence curves obtained for random grids for any integration rule employed in the problem.

The results of our study demonstrate that grid randomisation leads to a bigger integration error on coarse and fine grids alike. Surprisingly, this conclusion is true even for the statistical method which has no spatial dependence. While further careful study of this issue is required, our first experience with a problem of numerical integration on random grids demonstrates that



(a) (b) (c)

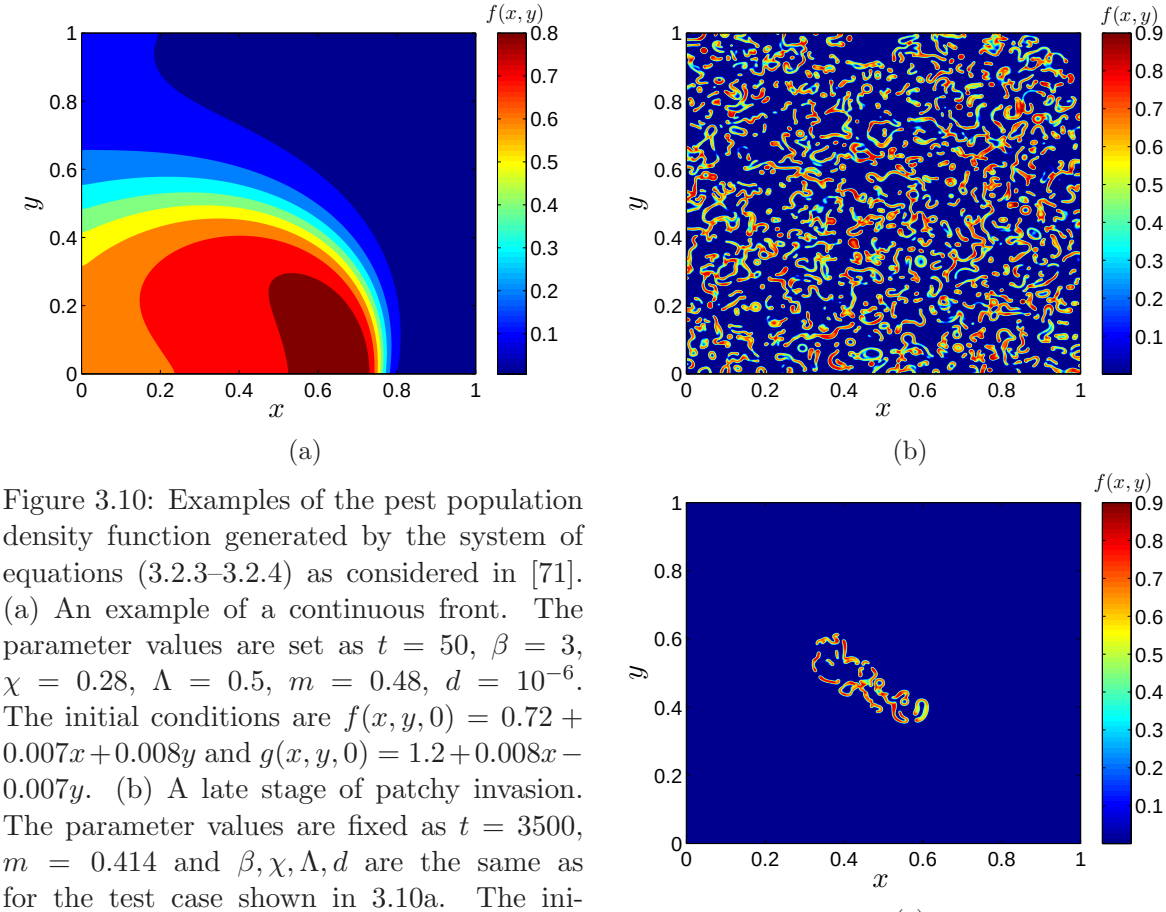
Figure 3.8: Convergence curves of the mean error on quasi-random grids for the ecologically meaningful density distribution of Figure 3.3b. Convergence on a regular grid is also plotted for comparison purposes (dashed line). The method of numerical integration used is (a) the statistical rule (2.1.4), (b) the trapezoidal rule (3.4.4), (c) the Simpson rule (3.4.5).



(a) (b) (c)

Figure 3.9: Convergence curves of the mean error on random grids for the ecologically meaningful density distribution of Figure 3.3b. This is compared with the convergence on regular grids (dashed line). The method of numerical integration used is (a) the statistical rule (2.1.4), (b) the trapezoidal rule (3.4.4), and (c) the Simpson rule (3.4.5).

an equidistant distribution of traps produces more accurate results than a random distribution. For an irregular distribution to be beneficial, prior knowledge of the spatial pattern of the pest population would be needed.



(a) (b)

(c)

Figure 3.10: Examples of the pest population density function generated by the system of equations (3.2.3–3.2.4) as considered in [71]. (a) An example of a continuous front. The parameter values are set as $t = 50$, $\beta = 3$, $\chi = 0.28$, $\Lambda = 0.5$, $m = 0.48$, $d = 10^{-6}$. The initial conditions are $f(x, y, 0) = 0.72 + 0.007x + 0.008y$ and $g(x, y, 0) = 1.2 + 0.008x - 0.007y$. (b) A late stage of patchy invasion. The parameter values are fixed as $t = 3500$, $m = 0.414$ and β, χ, Λ, d are the same as for the test case shown in 3.10a. The initial conditions in this case are $f(x, y, 0) = 1$ if $0.42 < x < 0.53, 0.45 < y < 0.55$ and $f(x, y, 0) = 0$ otherwise, $g(x, y, 0) = 1$ if $0.42 < x < 0.48, 0.45 < y < 0.51$ and $g(x, y, 0) = 0$ otherwise. (c) An early stage of patchy invasion where the population density is located in a small sub-domain of the field. The parameter values and initial conditions are as for the test case shown in 3.10b except for the time parameter which is set as $t = 450$.

3.5 Numerical Integration of 2D Data

Let us now consider some examples of a 2D pest population density distribution. We return our attention to a regular sampling plan, however the study of the previous section could readily be extended to 2D. We consider test cases presented in [71]. The 2D system of equations (3.2.3–3.2.4) are solved at some fixed time $t = \hat{t}$ to produce a discrete density function $f(x, y) \equiv f_i, i = 1, \ldots, N$. As in the 1D case, an explicit finite differences scheme validated by a more advanced alternate directions scheme (see [78]) was used to solve the coupled partial differential equations, and the no-flux boundary conditions $\partial f/\partial \mathbf{n} = \partial g/\partial \mathbf{n} = 0$ were imposed at each boundary of the

domain, where **n** is the normal vector. The authors of [71] supplied the solution of the coupled partial differential equations on a very fine regular grid (2.4.3) on the unit square $[0,1] \times [0,1]$ with the number of nodes in the x and y direction set as $N_x = N_y = 1025$. Thus the total number of grid nodes representing sample units is $N_f = 1,050,625$. The test cases shown in Figure 3.10 were generated using the parameter values and initial conditions given in the figure caption. Further details of the numerical solution of the system of equations (3.2.3–3.2.4) can be found in [71]. The spatial pattern of the three density functions are very different. The density function of Figure 3.10a is an example of a continuous front, whereas that of Figure 3.10b can be considered an example of a late stage of patchy invasion [77, 78]. Meanwhile Figure 3.10c shows an early stage of patchy invasion where the pest population is concentrated in a small sub-area of the field.

The same computational procedure that was carried out for the 1D ecological test cases is applied for the 2D pest population density functions shown in Figure 3.10. The 2D composite trapezium rule (2.4.4) is applied to the fine mesh of data $f_i, i = 1, \ldots, N_f$ and this is taken as the value of I. A series of increasingly refined regular grids with $N < N_f$ nodes are extracted from the existing data set and a method of numerical integration (2.1.3) is applied. Table 3.1 compares the error of estimates formed by the statistical rule (2.1.4), the 2D composite trapezium rule (2.4.4) and the 2D composite Simspon's rule (2.4.5) for each test case.

We have seen that the level of grid refinement required for the asymptotic error estimates (2.1.9) to hold depends on the spatial pattern of the pest population density function. In Section 3.3 it was shown that sufficiently accurate estimates are possible on coarse grids, *i.e.* on grids where the asymptotic error estimates do not hold. This raises the question of what is the minimum number N of sample units (grid nodes) required to obtain sufficient accuracy. From Table 3.1 it can be seen that this is again dependent on the spatial pattern of the population density distribution. The grids displayed in Table 3.1 are considered to be coarse as the highest order numerical integration method under consideration, *i.e.* the composite Simpson's rule, does not yet consistently produce the most accurate estimates. In the case of the continuous front depicted in Figure 3.10a where the population density is fairly spread across the domain

N	E_{rel}^{stat}	E_{rel}^{trap}	E_{rel}^{Simp}
9	0.1383	0.05065	0.02558
25	0.06410	0.01421	0.02210
81	0.03230	6.517e-004	0.003895
289	0.01763	2.869e-004	9.557e-005

(a)

N	E_{rel}^{stat}	E_{rel}^{trap}	E_{rel}^{Simp}
9	0.4216	0.4964	0.4929
25	0.1798	0.2632	0.1798
81	0.1124	0.1115	0.06742
289	0.08671	0.06473	0.05380

(b)

N	E_{rel}^{stat}	E_{rel}^{trap}	E_{rel}^{Simp}
9	0.9999979	0.999995	0.9999915
25	0.9999992	0.999999	0.9999995
81	0.9999020	0.999876	0.9997859
289	0.1685830	0.061408	0.1136888
1089	0.1386613	0.210940	0.4078478
4225	0.0766259	0.047545	0.1275943

(c)

Table 3.1: Relative approximation error (2.1.7) of estimates generated by the statistical rule (2.1.4), the compound trapezium rule (2.4.4) and the compound Simpson's rule (2.4.5) on a series of regular coarse grids with an increasing number N of nodes. The population density function under consideration is: (a) the continuous front given in Figure 3.10a, (b) a late stage of patchy invasion as shown in Figure 3.10b, and (c) an early stage of patchy invasion as shown in Figure 3.10c whereby the population is concentrated in a small sub-area of the field.

considered as the agricultural field, it can be seen from Table 3.1a that a high level of accuracy ($E_{rel} \leq \tau = 0.5$) is already achieved on a grid with $N = 9$ nodes, *i.e.* three nodes in both the x and y directions. Table 3.1b shows that the estimates produced when the patchy distribution shown in Figure 3.10b is considered are not as accurate in comparison. However, the accuracy still remains reasonable by ecological monitoring standards ($E_{rel} \leq \tau = 0.5$) even on a grid with a total number of nodes as small as $N = 9$. The accuracy of estimates obtained for the test case shown in Figure 3.10c where the population is concentrated in a small sub-area of the field is much poorer. Errors of the order of $E_{rel} \sim 1$ are evident on grids with $N \leq 81$ nodes. There is little difference between the accuracy of the three numerical integration methods. Furthermore, whereas Tables 3.1a and b show that increasing the number N of grid nodes leads to a more accurate estimate, oscillatory behaviour is evident in Table 3.1c. For example, the trapezium and Simpson's rule estimates on a grid with a total number of $N = 1089$ nodes are less accurate than the corresponding estimates which were obtained on a grid with the smaller number of $N = 289$ nodes.

The above examples demonstrate that accuracy of estimates depends on the spatial pattern of the population density distribution. A more comprehensive set of 2D test cases with ecological significance were studied in [71] where it was shown that, like in the 1D problem, density

distributions where the population are spread homogeneously over the entire domain are easiest to handle. A higher number of nodes is required to achieve comparable levels of accuracy when the density distribution is patchy. The most problematic case is when the density is concentrated within a small sub-region. On grids with a small number of nodes this region may be missed leading to very inaccurate estimates. We reiterate that it is still possible to achieve an acceptable level of accuracy can on coarse grids *i.e.* when the asymptotic error estimates do not hold, however the accuracy is not reliable as the error may oscillate.

3.6 Chapter 3 Conclusions

Conclusions about the accuracy of an estimate formed by numerical integration are usually made from the asymptotic error estimates (2.1.9). Conditions imposed by the ecological scenario, however, mean that this is not a standard numerical integration problem. The asymptotic estimates cannot be relied upon under the restriction imposed that the number N of sample units (grid nodes) is small.

Since asymptotic error estimates do not hold on coarse grids, the conventional approach of using a higher order method to improve the accuracy of an estimate may not be effective. It was demonstrated that on coarse grids the error may oscillate and it can be difficult to differentiate between the accuracy of different numerical integration methods. It should be noted that other conventional techniques used in numerical integration problems to improve accuracy are not available to us. For instance grid refinement would correspond to repeating the sampling procedure with an increased number of samples, however of course the distribution of the pest population would change in the time it takes to do this. We do not consider the time dependent problem and instead are concerned with how the accuracy may be quantified and optimised at a fixed point in time. For the same reason adaptivity with moving grids (e.g see [16]), *i.e.* rearranging the same number of traps and repeating the process is also not viable. Furthermore, we cannot distribute the sample units (grid nodes) such that more traps are installed local to areas of pest population that may require higher resolution without prior knowledge of the density distribution.

It was shown that whilst the error does not behave according to the asymptotic error esti-

mates on coarse grids, sufficiently accurate results may still be obtained. The question of what is the minimum number N of sample units (grid nodes) required to obtain sufficient accuracy needs to be addressed. The findings of [70, 71, 72] have been confirmed, namely that the accuracy of an estimate depends on the spatial pattern of the pest population density distribution. In the next chapter we identify a particular parameter which has control over the spatial pattern and hence construct an estimate of the minimum number of sample units required to achieve accuracy within a desired tolerance.

Chapter 4

Effects of Species Diffusion on Approximation Accuracy

It is essential to gain an understanding into which factors affect numerical integration accuracy on coarse grids since the asymptotic error estimates (2.1.9) cannot be relied upon. It was discussed in the previous chapter that the spatial heterogeneity of the integrand function is one such factor. This in itself is affected by a range of parameters. Here, we focus on the role of the rate at which a pest species diffuses.

In Section 4.1 we establish a link between the diffusion rate and the spatial pattern of the population density. We then investigate how the accuracy of numerical integration on regular grids is affected by diffusion in Section 4.2, where in particular we construct a lower bound for the number of sample units necessary to obtain a sufficiently accurate approximation. This lower bound is then validated for ecological significant data in Section 4.3. In Sections 4.4 and 4.5 we look at how sparsity of the spatial data and grid irregularity affects integration accuracy respectively, and we establish further recommendations for the minimum number of grid nodes required in each case. Discussion and conclusions of this chapter's results are given in Section 4.6. For simplicity, we limit our study to the 1D problem within this chapter and the agricultural field is considered as the unit interval $[0, 1]$. A simple linear transformation can be applied to recover an interval $[a, b]$ of arbitrary length. This chapter follows the work we presented in [69].

4.1 Spatial Heterogeneity and the Effect of Diffusion

It is well known that ecological populations often exhibit significant spatial heterogeneity [63, 104]. We now investigate the effect of species diffusion on the spatial heterogeneity of pest population distribution as discussed in [69] where we assume that each insect performs Brownian motion. Let us first consider a very simple model of the spatial temporal dynamics of a pest species, ignoring the impact of population multiplication and interspecific interactions. That is, the pest population density distribution is described by the scalar diffusion equation, namely

$$\frac{\partial F(X,T)}{\partial T} = D\frac{\partial^2 F}{\partial X^2}\ , \tag{4.1.1}$$

where D is the diffusion coefficient caused by the self-movement of individuals [63]. Let us consider a population in the unbounded domain, $-\infty < X < \infty$. An initial population of size I_0 is introduced by point-source release at the position $X = X_0$. The solution of equation 4.1.1 is then given by (*e.g* see [24])

$$F(X,T) = \frac{I_0}{\sqrt{4\pi DT}}\exp\left(-\frac{(X-X_0)^2}{4DT}\right)\ . \tag{4.1.2}$$

The characteristic width of the distribution (4.1.2), that is the characteristic length of the spatial heterogeneity, is therefore given as

$$\Delta \sim \sqrt{DT}\ , \tag{4.1.3}$$

where the sign $\sim$ means 'up to a constant coefficient'.

It has been shown (see [74], section 9.3) that when other ecologically meaningful initial conditions are considered instead, *e.g* an initial population is introduced over a finite domain, the characteristic length of the spatial heterogeneity is also given by (4.1.3). Dimensional analysis shows that this is a generic property of the diffusion equation (4.1.1). This equation contains

a single parameter, namely the diffusion coefficient D which has the dimension [distance]^2 $\cdot$ [time]$^{-1}$. Therefore, for any given time T, the only quantity with the dimension of length is $\sqrt{DT}$. Such analysis is explained in [6], in particular see Chapter 2 where it is performed on the heat equation. This is essentially the same equation as 4.1.1 therefore the calculations can be easily repeated.

Let us now look at a slightly more complex model, a single-species model with multiplication, in other words a reaction-diffusion equation. We consider a particular case where the multiplication is described by the logistic function, therefore the model is given by the following equation which is also known as Fisher's equation [37]:

$$\frac{\partial F(X,T)}{\partial T} = D\frac{\partial^2 F}{\partial X^2} + \alpha F\left(1 - \frac{F}{K}\right), \tag{4.1.4}$$

where α is the per capita growth rate and K is the carrying capacity. The dimension of α is time^{-1} and K is distance^{-1} . Hence the only quantity with the dimension of length that can be generated from the parameters of equation (4.1.4) is as follows,

$$\Delta x \sim \sqrt{D/\alpha}. \tag{4.1.5}$$

As $T \to \infty$, for a wide class of initial conditions equation (4.1.4) describes a travelling front [59]. Consequently, the characteristic length Δx of the spatial heterogeneity is then given by the width of the front.

Such dimensional analysis is not as effective when a system of reaction-diffusion equations is considered, that is when we have a multi-species system. Such systems have more than one parameter with the dimension of time or inverse time, and more than one diffusion coefficient, therefore there are multiple ways of generating a quantity with the dimension of length. An alternative approach is needed. Let us assume for simplicity that all diffusion coefficients have the same value D. This system is known to develop complex, chaotic spatio-temporal patterns sometimes referred to as 'biological turbulence' when the corresponding non-spatial system has

a unique positive state namely an unstable focus [55]. The characteristic length Δx of the emerging multi-peak spatio-temporal pattern, that is the width of a single peak, is then given as [75]

$$\Delta x \;=\; 2\pi c^* \left(\frac{D}{\max \mathrm{Re}(\lambda)} \right)^{1/2}, \tag{4.1.6}$$

where $\max \mathrm{Re}(\lambda)$ is the maximum real part of the eigenvalues of the linearised non-spatial system and c^* is a numerical coefficient of the order of one. It is worth noting that since $\max \mathrm{Re}(\lambda)$ has the dimension of time^{-1}, Δx as defined by (4.1.6) indeed has the dimension of length. Therefore equation (4.1.6) is in a good agreement with the previous dimensional analysis performed for the single-species model. In fact, it can be regarded as a generalisation of equation (4.1.5).

In all three cases (4.1.3), (4.1.5) and (4.1.6), the characteristic length of the spatial heterogeneity is proportional to $\sqrt{D}$, i.e.

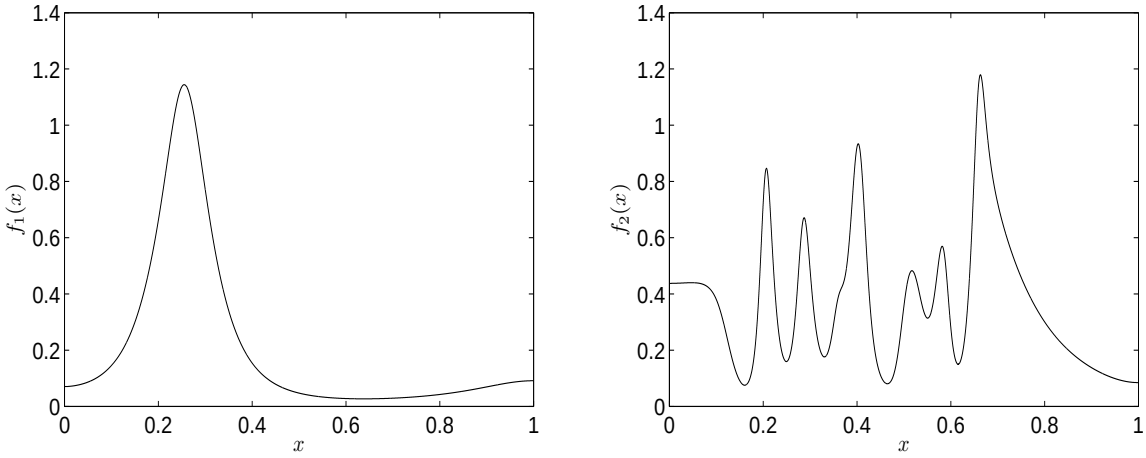
$$\Delta x \;=\; \omega \sqrt{D}, \tag{4.1.7}$$

where ω is a factor that can depend on the parameters of the interactions between the individuals of a single species (intraspecific) and those between individuals of different species (interspecific), but is independent of the diffusion coefficient. Extensive numerical study [75, 76] has shown ω to be relatively robust to changes in these parameters and typically has a value of $\omega \approx 25$. Non-dimensionalising (4.1.7) we obtain the same relationship between the dimensionless forms of the diffusion coefficient and the characteristic length of spatial heterogeneity, namely d and δ, as follows

$$\delta = \frac{\Delta x}{L} = \omega \sqrt{d}, \tag{4.1.8}$$

where we recall that L is the length of the agricultural field.

To further understand how the diffusion coefficient controls the spatial pattern of the population density let us re-consider two of the ecologically significant test cases from the previous chapter generated by the 1D Rosenszweig-Macarthur model (3.2.7–3.2.8), namely the single peak and multi-peak distributions of Figure 3.3a and Figure 3.3c respectively. For convenience the



(a) (b)

Figure 4.1: Test case examples of the pest population density function generated from the 1D spatially explicit form of the Rosenzweig-MacArthur model (3.2.7–3.2.8). The dimensionless diffusion parameter is set as (a) $d = 10^{-4}$ and (b) $d = 10^{-5}$. The reader is referred to Figure 3.3 for the details of the other parameters.

test cases are presented again in Figure 4.1. Throughout this chapter, the single peak population density function is denoted by $f_1(x)$ and the multi-peak density function by $f_2(x)$. It can be seen from the figure that the number of peaks present in the density function increases as the diffusion coefficient d decreases. [1]. In practical terms this means the slower a species diffuses, the more likely it is to form complex spatial patterns.

4.2 The Impact of the Diffusion Rate on the Accuracy of Integration

We have seen from previous chapter that the number of grid nodes N required to achieve a sufficiently accurate estimate of pest abundance depends on the spatial pattern of the density distribution and in the previous section it was discussed how the spatial pattern is controlled by the diffusion coefficient. We now derive the relationship between the diffusion coefficient d and the number of grid nodes N required for accurate evaluation of the population size. We assume we have no prior knowledge of the spatial pattern. As such, we consider a regular grid

[1] This is the case so long as the time t is sufficiently large such that that impact of the initial conditions is no longer an issue. For the time parameters used to generate the test cases of Figure 4.1 please see the caption of Figure 3.3.

formation (2.2.2). Thus the grid step size, *i.e.* the distance between grid nodes, is fixed. The effects of deviating from this regular sampling plan is investigated later in the chapter in Section 4.5. We previously established that using a higher order numerical integration method does not necessarily improve the accuracy of the resulting estimate on coarse grids. Hence, in our following analysis we use a simple means of numerical integration namely the composite trapezium rule (2.2.6). That is, the population density function is interpolated by linear polynomials.

Our analysis focuses on a component of spatial heterogeneity, namely a single peak. Let the population density function $f(x)$ which by definition is non-negative have a peak (local maximum) somewhere on the unit interval $[0, 1]$ (*e.g* see Figure 4.1a). Further let the domain of the peak be the sub-interval $[x_{i-1}, x_{i+1}]$ of length $2h$, where the peak maximum is located at the sub-interval midpoint x_i. We Taylor expand the density function $f(x)$ about the location of the maximum x_i as follows

$$f(x) = f(x_i) + \frac{1}{2}\frac{d^2 f(x_i)}{dx^2}(x - x_i)^2 + R(x). \tag{4.2.1}$$

We now assume that the remainder term $R(x)$ can be ignored and thus consider the density function in the vicinity of the peak as the following quadratic function

$$f(x) \approx Q(x) = B - A(x - x_i)^2, \qquad x \in [x_{i-1}, x_{i+1}], \tag{4.2.2}$$

where $A = -\frac{1}{2}\frac{d^2 f(x_i)}{dx^2} > 0$ and $B = f(x_i) > 0$. The maximum of the quadratic coincides with the maximum of the peak and is thus symmetric about the location of the maximum namely x_i. We require the quadratic $Q(x)$ to be non-negative over the sub-interval $[x_{i-1}, x_{i+1}]$, *i.e.* $Q(x_{i-1}) = Q(x_{i+1}) = B - Ah^2 \geq 0$. This provides us with the following relationship between h, A, and B:

$$h^2 \leq \frac{B}{A}. \tag{4.2.3}$$

Examples of replacing a peak with a quadratic function are shown in Figure 4.2 and the details are discussed in the Appendix B. Clearly, by ignoring the remainder term of the Taylor expansion (4.2.1) and considering the peak as the quadratic function $Q(x)$ as given in (4.2.2)

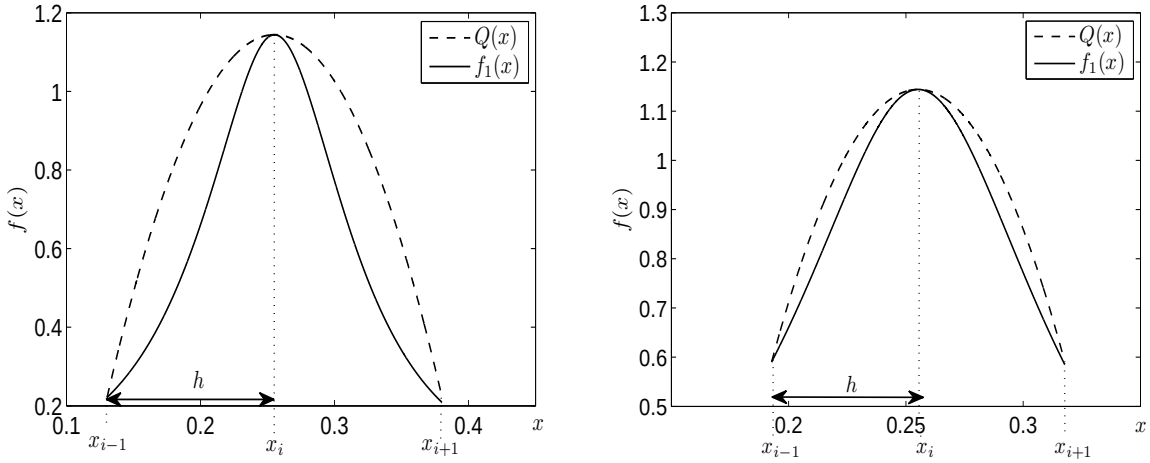


Figure 4.2: The peak of the population density distribution $f_1(x)$ as shown in Figure 4.1a is replaced with a quadratic function on the interval $[x_{i-1}, x_{i+1}]$ of width $2h$. The quadratic function $Q(x)$ is as defined by 4.2.2 where h is taken as (a) $h = 0.125$, and (b) $h = 0.0625$.

we have introduced the error $R(x)$, however we consider this error to be sufficiently small (see Appendix B).

We now proceed to use the approach of considering a peak as a quadratic function to obtain a recommendation regarding the grid step size needed for accurate estimation of the pest abundance. We consider the pest abundance in the vicinity of the peak, *i.e.* on the interval $[x_{i-1}, x_{i+1}]$ of width $2h$. Now let us consider a local regular computational grid of three nodes x_{i-1}, x_i, and x_{i+1} where the distance between neighbouring nodes is h. An estimate I_a of the integral of the peak is obtained by integrating a piecewise linear interpolating polynomial of the peak which is depicted in Figure 4.3 *i.e.* we apply the composite trapezium rule 2.2.6 on the local grid

$$\begin{aligned} I_a &= \frac{h}{2} f(x_{i-1}) + h f(x_i) + \frac{h}{2} f(x_{i+1}) \\ &= \frac{h}{2} Q(x_{i-1}) + h Q(x_i) + \frac{h}{2} Q(x_{i+1}) \\ &= 2Bh - Ah^3 \end{aligned} \tag{4.2.4}$$

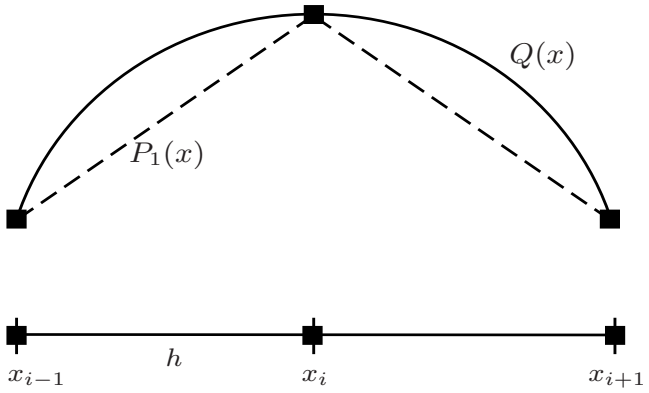


Figure 4.3: Approximation of a peak by a piecewise linear polynomial $P_1(x)$ on a grid of three regularly spaced nodes. The peak is treated as the quadratic function $Q(x)$ as defined by (4.2.2). The location of the peak maximum coincides with the central grid node. We consider the integral of the quadratic as the exact integral I of the peak, while integrating $P_1(x)$ yields the approximation I_a.

Treating the peak as a quadratic allows us to calculate the exact abundance I as

$$I = \int_{x_{i-1}}^{x_{i+1}} Q(x)\,dx = 2Bh - \frac{2Ah^3}{3}. \tag{4.2.5}$$

We want to ensure that the estimate I_a is sufficiently accurate. Thus, we require that the relative error satisfies the condition (2.5.1) namely that $E_{rel} \leq \tau$ for some chosen value of τ. By definition $I > 0$, therefore this condition implies that

$$|I - \tilde{I}| \leq \tau I. \tag{4.2.6}$$

Let us set the tolerance as $\tau = 0.25$, which is within the acceptable range of error for ecological applications given in Chapter 2 (see Section 2.5). Such accuracy has been previously recommended for ecological monitoring [86]. From (4.2.5) and (4.2.4) it can be seen that $|I - \tilde{I}| = \dfrac{Ah^3}{3}$. The expression (4.2.6) then becomes

$$\frac{Ah^3}{3} \leq \frac{1}{4}\left(2Bh - \frac{2Ah^3}{3}\right),$$

which after some rearrangement and taking into consideration $A > 0$, gives us the condition on the grid step size $h \leq \sqrt{B/A}$ to ensure the relative error is $E \leq \tau = 0.25$. We define the limiting grid step size where $E_{rel} = \tau = 0.25$ to be $h_0 = \sqrt{B/A}$.

We now wish to link this to the species diffusion. As discussed in Section 4.1, the diffusion

can be written in terms of the characteristic length of the spatial heterogeneity. In the case of the quadratic function $Q(x)$, this equates to the width of the peak, which we define as the distance between the two roots. The roots of (4.2.2) are given as

$$x_I = x_1 - \sqrt{B/A}, \qquad x_{II} = x_1 + \sqrt{B/A},$$

hence the peak width which we denote δ is given as $\delta = 2\sqrt{B/A}$. The grid step size restriction to ensure $E_{rel} \leq \tau = 0.25$ on a regular mesh can then be written as $h \leq h_0 = \delta/2$. Since the spatial heterogeneity can be expressed in terms of the diffusion coefficient d, we have

$$h \leq h_0 = \frac{\omega\sqrt{d}}{2}. \tag{4.2.7}$$

where we recall that typically $\omega \approx 25$. Hence if the species diffusion d is known, we can determine the grid step size h_0 necessary to ensure accurate evaluation of the pest abundance. Estimates of the diffusion coefficient exist for a variety of species (*e.g* see [49, 90]). The number of grid nodes *i.e.* sample units corresponding to the grid step size h_0 is given by $N_0 = 1 + 1/h_0$ thus we arrive at the following condition on the number N of grid nodes to ensure that $E_{rel} \leq \tau = 0.25$

$$N \geq N_0 = 1 + \frac{2}{\omega\sqrt{d}}. \tag{4.2.8}$$

4.3 Application of Approach to Ecological Data

In this section we seek to validate the recommendation (4.2.8) by considering the single peak and multi-peak population density functions, $f_1(x)$ and $f_2(x)$ shown in Figure 4.1. We follow the procedure outlined in the previous chapter in Section 3.3 whereby we integrate the density function on the unit interval $[0, 1]$ over a series of grids with an increasing number N of nodes and calculate the relative error (2.1.7). The chosen method of numerical integration is the composite trapezium rule (2.2.6). We recall that the density distributions $f_1(x)$ and $f_2(x)$ were obtained via a mathematical model (3.2.7–3.2.8) and are thus discrete. As in Section 3.3, the approximation yielded by the composite trapezium rule (2.2.6) on the finest available regular

N	3	5	9	17
h	0.5	0.25	0.125	0.0625
E_{rel}	0.6948	0.5459	0.0823	0.0036

Table 4.1: Integration error (2.1.7) for the population density distribution $f_1(x)$ shown in Figure 4.1a where the composite trapezium rule (2.2.6) has been implemented on the unit interval $[0, 1]$. The error is given for increasingly small grid step size h.

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

Table 4.2: Integration error (2.1.7) for the population density distribution $f_2(x)$ shown in Figure 4.1b where the composite trapezium rule (2.2.6) has been implemented on the unit interval $[0, 1]$. The error is given for increasingly small grid step size h.

grid of $N = 32,796$ nodes is taken to be the exact integral I.

First we focus on the single peak population distribution $f_1(x)$ of Figure 4.1a. The diffusion coefficient for this distribution is $d = 10^{-4}$. Hence from (4.2.7) and (4.2.8) we have the recommendation that the grid step size should be $h \leq h_0 = 0.125$ *i.e.* we require $N \geq N_0 = 9$ grid nodes in order to ensure an error of $E_{rel} \leq \tau = 0.25$. The relative error produced on applying the composite trapezium rule (2.2.6) to $f_1(x)$ is given in Table 4.1, where it can be seen that for grids with $N < 9$ nodes, the relative error E_{rel} exceeds the desired tolerance $\tau = 0.25$. However, once the grid is refined to $N = 9$ nodes, the accuracy becomes much higher and is well within this limit.

Now we consider the population density distribution $f_2(x)$ shown in Figure 4.1b. The density coefficient is $d = 10^{-5}$ which means that according to (4.2.7) and (4.2.8) we require a grid step size of $h \leq h_0 = 0.0395$ *i.e.* the number of grid nodes is required to be $N \geq N_0 = 27$ to achieve an error of $E \leq \tau = 0.25$. Table 4.2 shows the integration error 2.1.7 for the composite trapezium rule (2.2.6) applied to $f_2(x)$. In fact the desired level of accuracy is achieved on all grids. We recall that the composite trapezium rule approximation is founded on piecewise linear interpolation of the population density function. We suspect that this higher level of accuracy may be due to the cancellation that occurs as a result of the piecewise linear polynomial

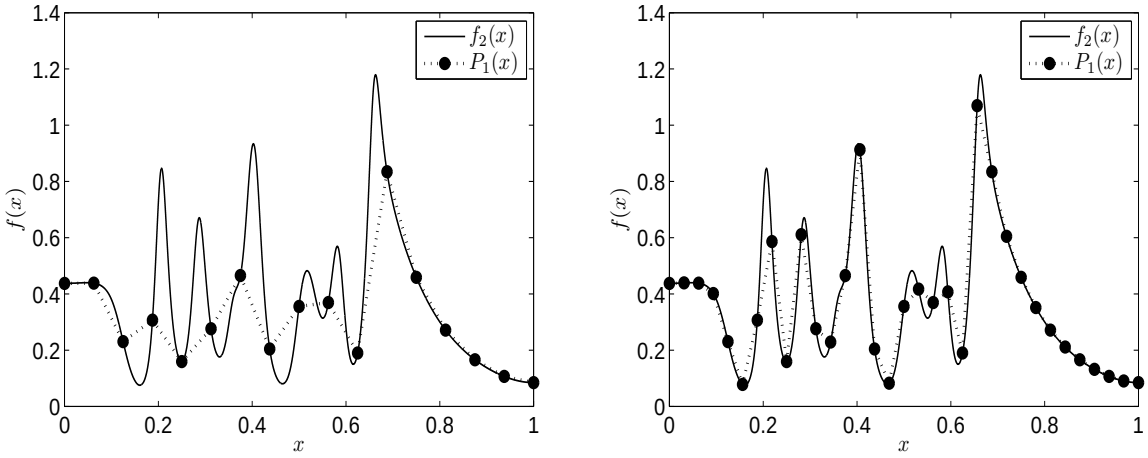


Figure 4.4: Piecewise linear polynomial approximation $P_1(x)$ for the multi-peak population density distribution $f_2(x)$. (a) Approximation on a grid of $N = 17$ nodes, (b) Approximation on a grid of $N = 33$ nodes.

approximation of the population density function $f_2(x)$ being too high on one sub-interval and then too low on another. Therefore it is evident that whilst the grid step size recommendation (4.2.7) ensures the error is within the desired limit $E_{rel} \leq \tau = 0.25$, it does not give an indication of the error on coarser grids. It could be, as is the case in this instance, that sufficient accuracy has already been achieved on a grid with fewer nodes than recommended by (4.2.8).

It can also be seen in Table 4.2 that a significant jump in accuracy occurs when the grid is refined from $N = 17$ to $N = 33$ grid nodes. Figure 4.4, which shows the piecewise linear polynomial approximation of $f_2(x)$ over the two aforementioned grids, illustrates why this is the case. When the grid has $N = 17$ nodes, the majority of the peaks in $f_2(x)$ are approximated by a single linear polynomial. This is equivalent to approximating a quadratic polynomial on a local grid of just 2 nodes, which we discuss later. However, when the grid is refined to $N = 33$ nodes, all but two of the peaks are approximated by at least two linear polynomials. It is clear that a reduction in the number of peaks that are approximated by a single linear polynomial causes a significant reduction in the integration error (2.1.7).

4.4 Arbitrary Location of Peak Maximum

The recommendation (4.2.8) was found by assuming that there are three grid nodes in the vicinity of the peak, namely x_{i-1}, x_i, x_{i+1}, where the central node coincides with the location of the peak maximum. By the nature of the ecological monitoring problem, the location of the local maxima of the population density function $f(x)$ are generally unknown. Hence in practice, it is likely that the sample locations, *i.e.* the grid nodes will not coincide with the position of the local maxima. Thus we now need to investigate how the integration error is affected by the peak maximum not being captured by a grid node. There is also the possibility that there are not three grid nodes local to the peak. It could be that a peak falls between two grid nodes and so is approximated by a single linear polynomial, a problem that was touched upon in the previous section.

Let us again consider a mesh of three regularly spaced nodes $x_{i-1} = x_i - h, x_i$ and $x_{i+1} = x_i + h$, and assume that the population density function $f(x)$ has a peak somewhere on the interval $[x_{i-1}, x_{i+1}]$ which is of width $2h$. As before, we approximate the peak by a quadratic function. Let us define the quadratic approximation of the peak $Q(x)$ as follows

$$Q(x) = \begin{cases} B - A(x - x^*)^2 & \text{when } x \in [x_I, x_{II}], \\ 0 & \text{otherwise,} \end{cases} \tag{4.4.1}$$

where $A, B > 0$, x^* is the location of the maximum point which is now distinct from the grid node x_i. The values x_I and x_{II} are the roots of quadratic $Q(x)$. We can express the location of the maximum x^* in relation to the position of the grid nodes as

$$x^* = x_i + \gamma h = x_{i-1} + h(\gamma + 1),$$

where because of the symmetry of the quadratic it is sufficient to consider $\gamma \in [0, 1/2]$. The roots x_I and x_{II} can then be written in terms of x_{i-1}, A and B as

$$x_I = x_{i-1} + h(\gamma + 1) - \sqrt{B/A} \qquad x_{II} = x_{i-1} + h(\gamma + 1) + \sqrt{B/A}. \tag{4.4.2}$$

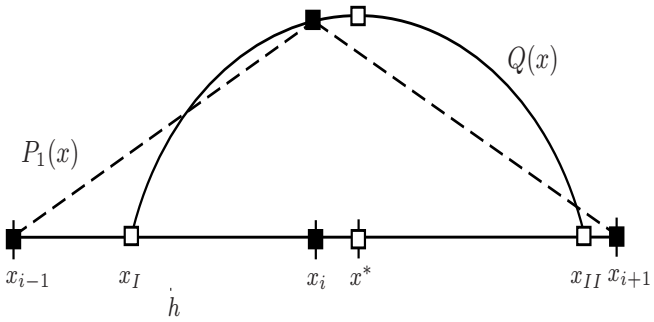


Figure 4.5: Approximation of a peak by a piecewise linear polynomial $P_1(x)$ on a grid of three regularly spaced nodes. The peak is treated as the quadratic function $Q(x)$ as defined by (4.4.1). The location of the peak maximum x^* is arbitrary and is no longer restricted to coincide with a grid node. We consider the integral of the quadratic as the exact integral I of the peak, while integrating $P_1(x)$ yields the approximation I_a.

The exact integral of the peak is considered to be

$$I = \int_{x_{i-1}}^{x_{i+1}} Q(x)\ dx = \int_{x_I}^{x_{II}} Q(x)\ dx = \frac{4}{3}B\sqrt{B/A} = \frac{2}{3}B\delta, \tag{4.4.3}$$

where $\delta = 2\sqrt{B/A}$ is the distance between the roots of the quadratic $Q(x)$.

We follow the same approach used in Section 4.2 and construct an estimate I_a of the integral of the peak by integrating a piecewise linear interpolating polynomial of the peak *i.e.* we apply the composite trapezium rule 2.2.6 on the local grid of the three regularly spaced nodes x_{i-1}, x_i, x_{i+1}. This is depicted in Figure 4.5. The approximated value of the integral is thus

$$\begin{aligned} I_a &= \frac{h}{2}f(x_{i-1}) + hf(x_i) + \frac{h}{2}f(x_{i+1}) = hQ(x_i) \\ &= h(B - A\gamma^2 h^2). \end{aligned} \tag{4.4.4}$$

To ensure a sufficiently accurate estimate we apply the condition (2.5.1), that $E_{rel} \leq \tau$, where we again set $\tau = 0.25$. Since by definition $I > 0$, this condition on E_{rel} is equivalent to the following condition on the estimate I_a

$$0.75I \leq I_a \leq 1.25I. \tag{4.4.5}$$

We wish to establish how far away the location of the peak maximum can be from a grid node whilst still achieving an acceptable error of $E_{rel} \leq \tau = 0.25$. In other words we seek the values of γ for which the above inequality is satisfied. Let us first consider the lower limit of (4.4.5),

that is when $I_a \geq 0.75I$. From (4.4.3) and (4.4.4) this inequality becomes

$$Bh - A\gamma^2 h^3 \geq \frac{B\delta}{2},$$

which after rearrangement gives us

$$\gamma \leq \gamma_{II} = \frac{\delta}{2h}\sqrt{1 - \frac{\delta}{2h}}, \tag{4.4.6}$$

where γ_{II} is the upper limit for γ. For $\gamma_{II} \in \mathbb{R}$ to exist, we require the grid step size to be such that $h \geq \delta/2$.

We now consider the upper limit of (4.4.5), when $I_a \leq 1.25I$. Substituting (4.4.3) and (4.4.4) into this inequality we arrive at

$$Bh - A\gamma^2 h^3 \leq \frac{5B\delta}{6}.$$

Rearranging the above we find a lower limit for γ which we denote γ_I:

$$\gamma \geq \gamma_I = \frac{\delta}{2h}\sqrt{1 - \frac{5\delta}{6h}}, \tag{4.4.7}$$

where $\gamma_I \in \mathbb{R}$ exists for $h \geq 5\delta/6$. Consequently, even when the peak lies between two grid nodes, there is a parameter range where sufficiently accurate results are achieved. Furthermore, for any fixed h, there exists the value of γ for which the approximation is exact, that is $I_a = I$. Setting (4.4.3) and (4.4.4) equal to each other and solving for γ we find that this value is given by

$$\gamma(h) = \frac{\delta}{2h}\sqrt{\frac{3h - 2\delta}{3h}}, \quad h \geq \frac{2\delta}{3}. \tag{4.4.8}$$

Clearly $\gamma(h) > 0$, meaning that the location of the grid node which leads to zero error is not that which coincides with the location of the peak maximum.

The curves γ_I and γ_{II} are shown in Figure 4.6 for the diffusion coefficients $d = 10^{-4}$ and $d = 10^{-5}$, therefore corresponding to distributions $f_1(x)$ and $f_2(x)$ of Figures 4.1a and b respectively. Let us fix $h = h^*$. From (4.4.6) and (4.4.7) we can calculate the limits $\gamma_I(h^*) = \gamma_I^*$ and $\gamma_{II}(h^*) = \gamma_{II}^*$. Thus, for any value of γ within the range $\gamma_I^* \leq \gamma \leq \gamma_{II}^*$ the relative error of

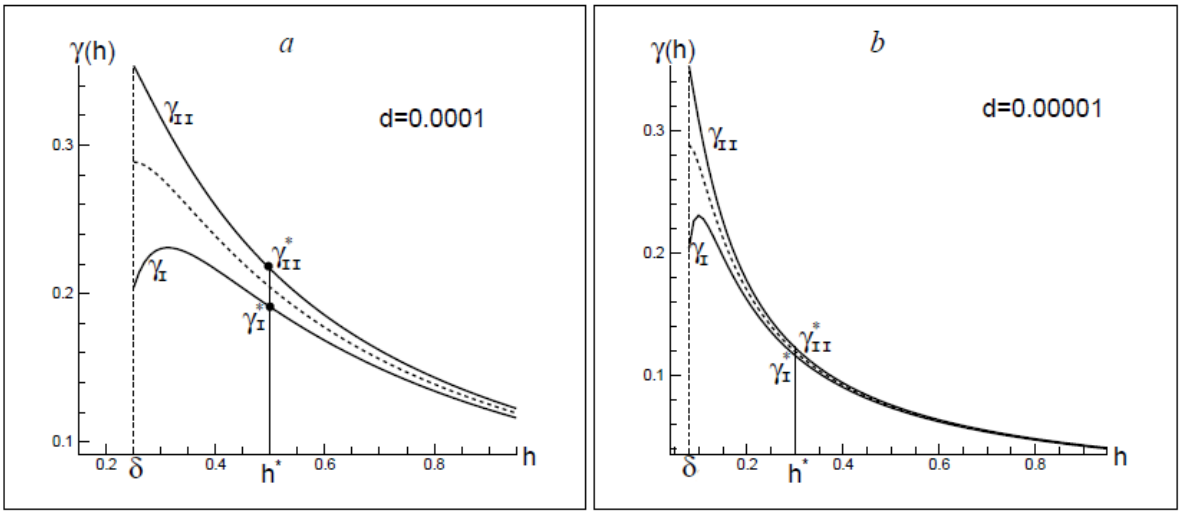


Figure 4.6: The function $\gamma(h)$ for different values of the dimensionless diffusivity d when the grid step size is larger than the peak width *i.e.* $h > \delta$. The part of the (h, γ) plane between the two solid curves (4.4.6) and (4.4.7) gives the parameter range where the integration is performed with the required accuracy $E_{rel} \leq \tau = 0.25$. The dashed curve represents where the integration error is $E_{rel} = 0$. The diffusion coefficient is (a) $d = 10^{-4}$ and (b) $d = 10^{-5}$.

integration will satisfy the condition $E_{rel} \leq \tau = 0.25$. The dashed line represents where $I = I_a$. An obvious, but nonetheless important observation that can be made from this figure is that as the diffusion coefficient d decreases, the admissible range for the error becomes increasingly narrow. Effectively the peak is lost on the computational mesh.

4.5 Analysis on a Non-Uniform Grid

So far our analysis has been performed on a regular computational grid. As previously mentioned in Sections (2.3) and (3.4) the sampling plan used in practice may not be regular. Thus, we now analyse the effect grid irregularity has on the integration error (2.1.7). We still assume that we have no prior knowledge of the spatial pattern of the pest population. As such we are not able to purposely choose the grid node (sample unit) locations such that more grid nodes are situated local to areas of pest population density which require a higher resolution *e.g* where there is a peak or a steep function gradient. Instead our analysis focuses on the type of grid irregularity whereby the interior grid nodes of a regular grid have become locally shifted. This

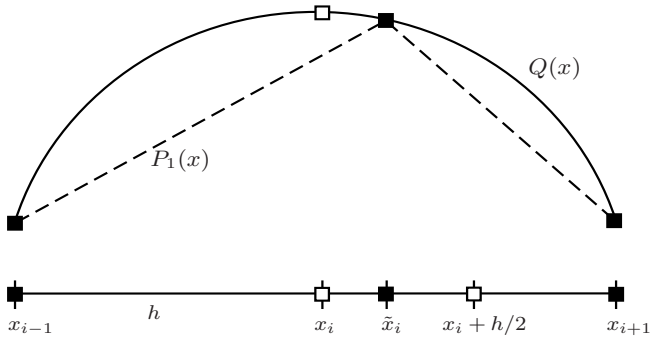


Figure 4.7: Approximation of a peak by a piecewise linear polynomial $P_1(x)$ on a grid of three irregularly spaced nodes. The peak is treated as the quadratic function $Q(x)$ as defined by (4.2.2). The location of the central computational grid node $\tilde{x}_i$ is defined by (4.5.1) and is distinct from the location of the peak maximum x_i. We consider the integral of the quadratic as the exact integral I of the peak, while integrating $P_1(x)$ yields the approximation I_a.

grid type was referred to as quasi-random in Section 3.4 of the previous chapter.

We return to the case when the peak is well resolved, *i.e.* that there are three grid nodes in the vicinity of the peak and follow a very similar approach to that used for the analysis on a regular mesh in Section 4.2. We again assume that the population density function $f(x)$ has a peak on the interval $[x_{i-1}, x_{i+1}]$ which is of length $2h$, and that the maximum is located at the interval midpoint $x_i = x_{i-1} + h$. The grid used for integration, however, is the set of three points $\{x_{i-1}, \tilde{x}_i, x_{i+1}\}$, where $\tilde{x}_i$ is a perturbation of x_i according to the following mapping:

$$x_i \to \tilde{x}_i = x_i + \beta h, \tag{4.5.1}$$

and is shown in Figure 4.7. The parameter β is chosen such that $\beta \in (0, 1/2)$. Moving the grid node in this fashion is representative of the situation when a sample unit can not be installed at the desired location x_i due to an obstacle and is instead placed nearby. Hence the movement is restricted by the upper limit $\beta = 1/2$ which ensures $\tilde{x}_i$ is strictly less than halfway between the location of the maximum x_i and grid node x_{i+1}. Setting $\beta = 0$ would result in a uniform mesh where the grid midpoint coincides with the location of the maximum, as was the case in Section 4.2.

In the vicinity of the peak, *i.e.* on the interval $[x_{i-1}, x_{i+1}]$ the peak is replaced by the quadratic as defined by (4.2.2). Since the interval of integration remains the same as that for the regular grid analysis, namely $[x_i, x_{i+1}]$, the exact integral is given by (4.2.5). We obtain an approximation to the integral by applying the composite trapezium rule (2.2.6) on the local grid

of nodes $\{x_{i-1}, \tilde{x}_i, x_{i+1}\}$ as follows

$$\begin{aligned} I_a &= \frac{h}{2} f(x_{i-1}) + h f(\tilde{x}_i) + \frac{h}{2} f(x_{i+1}) \\ &= \frac{h}{2} Q(x_{i-1}) + h Q(\tilde{x}_i) + \frac{h}{2} Q(x_{i+1}) \\ &= 2Bh - Ah^3(1+\beta^2). \end{aligned} \tag{4.5.2}$$

The underlying piecewise linear polynomial interpolation which gives rise to this estimate is shown in Figure 4.7.

We again require that the relative approximation error (2.1.7) is such that $E_{rel} \leq \tau = 0.25$. Thus, from the condition (4.4.5) and the expressions (4.2.5) and (4.5.2) we obtain the following:

$$Ah^3 \left(\frac{1}{3} + \beta^2 \right) \leq \frac{1}{4} \left(2Bh - \frac{2Ah^3}{3} \right).$$

The above can be rearranged to give

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

By recalling that the peak width is $\delta = 2\sqrt{B/A}$ and implementing the relation (4.1.8), we arrive at a grid step size recommendation for the underlying regular mesh from which the irregular grid is formed by perturbing the interior nodes. In terms of the diffusion coefficient this recommendation is

$$h \leq \frac{\omega\sqrt{d}}{2\sqrt{1+2\beta^2}}. \tag{4.5.3}$$

which ensures that the error (2.1.7) is $E_{rel} \leq \tau = 0.25$ on the irregular mesh under consideration. The upper limit for h is a monotone decreasing function which achieves its minimum value at the upper limit of β which is $\beta = 1/2$. Setting $\beta = 1/2$ in (4.5.3) gives us the upper limit for the grid step size which can be expressed in terms of the recommended grid step size h_0 for a uniform grid (4.2.7)

$$h \leq \omega\sqrt{\frac{d}{6}} = h_0\sqrt{\frac{2}{3}}. \tag{4.5.4}$$

Thus it can be seen that a finer grid is required, *i.e.* a higher number of grid nodes, in order to obtain sufficiently accurate results if the interior nodes are locally shifted from a regular formation.

4.6 Chapter 4 Conclusions

In this Chapter we have shown that the rate at which a species diffuses is one of many factors which define the spatial heterogeneity of the pest population density function. Coupled with the results of the previous chapter this means that the diffusion rate affects the accuracy of the numerical integration of the pest density data.

By considering a component of spatial heterogeneity, namely a peak (local maximum) and treating it as a quadratic function we have been able to obtain a recommendation for the number N of regularly spaced grid nodes required to ensure the population abundance estimate I_a achieves a chosen level of accuracy $E_{rel} \leq \tau$ where for calculation purposes we fixed $\tau = 0.25$. This recommendation is given in terms of the diffusion coefficient d. Whilst we clearly introduce an error by treating a peak as a quadratic function, the analytical prediction (4.2.7) was shown to be in good agreement with numerical results. We considered the effect introducing irregularity to the computational grid has on the integration error and provided the recommendations (4.5.3–4.5.4) for the necessary average grid step size to ensure sufficient accuracy.

Furthermore, we studied the case when the grid is so coarse that only a single grid node lies within the vicinity of the peak. In ecological terms this case corresponds to the distribution being aggregated such that it is localised to one (or several) small patch(es) of the field. In this scenario we established that there is a parameter range (4.4.6–4.4.7) such that the condition on the relative error $E_{rel} \leq \tau = 0.25$ is satisfied. Outside of this range, however, the resulting estimate is not sufficiently accurate. We emphasise again that the location of the local maxima of a pest population density function are generally not known a priori. Thus, obtaining an accurate estimate of pest abundance for density distributions aggregated in such a way becomes a matter of chance. This indicates that in this scenario a different approach, namely probabilistic rather than deterministic, is needed for assessing integration accuracy. This issue is investigated in more detail in the next chapter.

Chapter 5

Numerical Integration on Ultra-Coarse Grids

In this chapter we follow the work we presented in [67]. We improve on the estimate from the previous chapter of the minimum number of grid nodes (sample units) needed to achieve a sufficiently accurate estimate of the pest abundance. When the number of traps is less than this recommended number, we cannot guarantee accuracy, therefore the error is in this sense probabilistic rather than deterministic. We show that by considering the error as a random variable, we are able to quantify the chance of achieving an accurate estimation in this instance.

The analysis is founded on considering highly aggregated density distributions (peak functions) which are discussed in Section 5.1. It is shown that achieving a sufficiently accurate estimate when integrating such density distributions is a matter of chance. This leads us to introduce a new grid classification; a grid is considered to be *ultra-coarse* when the probability of achieving an accurate estimate is $p < 1$. In Section 5.2 we obtain an estimate for the probability of achieving an accurate estimate on ultra-coarse grids. We go on to study the transition from ultra-coarse to coarse grids *i.e.* when sufficient accuracy becomes guaranteed in Section 5.3, and find an estimate of the threshold number N_t of grid nodes where this transition occurs. Numerical verification of our theoretical predictions is provided in Sections 5.4 and 5.5 where standard and ecological significant test cases are considered respectively. In Section 5.6 we explain how the probabilistic approach can be used to compare different numerical integration methods on

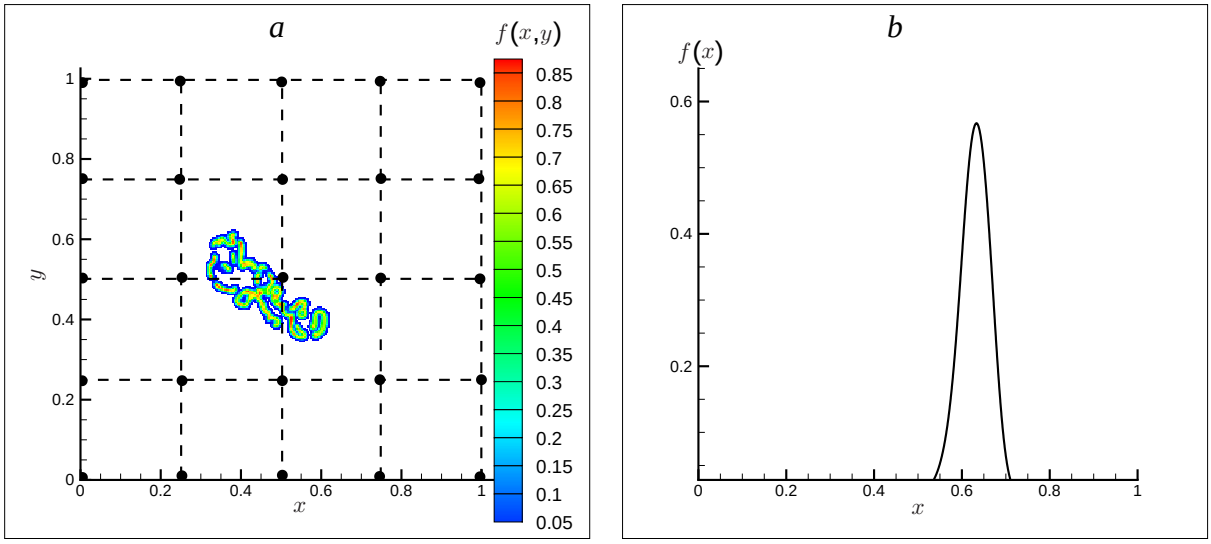


Figure 5.1: (a) The pest population density distribution $f(x, y)$ at an early stage of patchy invasion. The density distribution is as given in Figure 3.10c, and has been obtained by numerically solving the system of equations (3.2.3–3.2.4). The filled circles represent grid node (sample unit) locations. It can be seen that the $N = 25$ grid nodes have entirely missed the patch where the population is located. (b) A 1D counterpart of the density distribution shown in (a). It comprises of a single peak and is thus referred to as a *peak function*.

coarse grids. A discussion of the results and concluding remarks are provided in Section 5.7.

5.1 Integration of Severely Aggregated Density Distributions (Peak Functions)

We have demonstrated how the spatial heterogeneity of a pest population density distribution, which is in turn controlled by the species diffusion, affects the number of sample units necessary to achieve an accurate estimation of the pest population size. It has also been discussed how accurate estimations can be achieved on very coarse grids when the pest population is distributed across the whole field, even when the spatial structure is strongly heterogeneous. The accuracy deteriorates, however, when the distribution is what we will refer to as *highly aggregated*, that is, the pest population is located within a relatively small area of the field.

Such distributions may indeed arise in ecological applications. For example in the case of biological invasion, the pest species can be initially located in a very small area (see Figure 5.1a), and then as time passes spread over the entire field [78]. The scenario of a spreading pest species

forming a strongly heterogeneous patchy spatial distribution is referred to as *patchy invasion*. It would be desirable to accurately evaluate the pest population before the patch spreads to limit the damage caused to the crop. The problem is, however, that the location of this patch of pest population is unknown. We cannot therefore improve the accuracy of the estimation by placing traps local to where the population is concentrated. Instead we consider a regular grid of traps that covers the entire domain. It can be that the grid is so coarse that the sample units miss the population entirely as shown in Figure 5.1a.

A 1D counterpart of an early stage of patchy invasion is what we will refer to as a *peak function*. This is where the population density function on the interval $D = [a, b]$ is such that there is a single peak located in a sub-domain $D_u = [x_I, x_{II}]$ and elsewhere the density is zero. Thus, the population density distribution function $f(x)$ is of the following form:

$$\begin{cases} f(x) > 0 & \text{for } x \in (x_I, x_{II}) \\ f(x) = 0, & \text{otherwise} \end{cases} \tag{5.1.1}$$

where the maximum of the peak is assumed to be located at $x^* = 0.5(x_I + x_{II})$. An example of a 1D peak function is shown in Figure 5.1b. Throughout this chapter we shall fix the interval representing the agricultural field as $D = [0, 1]$, since a simple linear transformation can be applied to scale the unit interval to the arbitrary interval $[a, b]$.

We investigate the accuracy of numerical integration of peak functions. We begin our study by considering standard examples of peak functions for which the integral is available in closed form, and as such the exact value I can be calculated. Test cases with ecological significance are then considered later in the text. One standard example of a peak function is the normal distribution given by (*e.g* see [40])

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x - x^*)^2}{\sigma^2}\right), \tag{5.1.2}$$

where the location of the peak maximum x^* is the mean of the distribution and σ is the standard deviation. The density of the normal distribution is a bell shaped curve, symmetrical about the

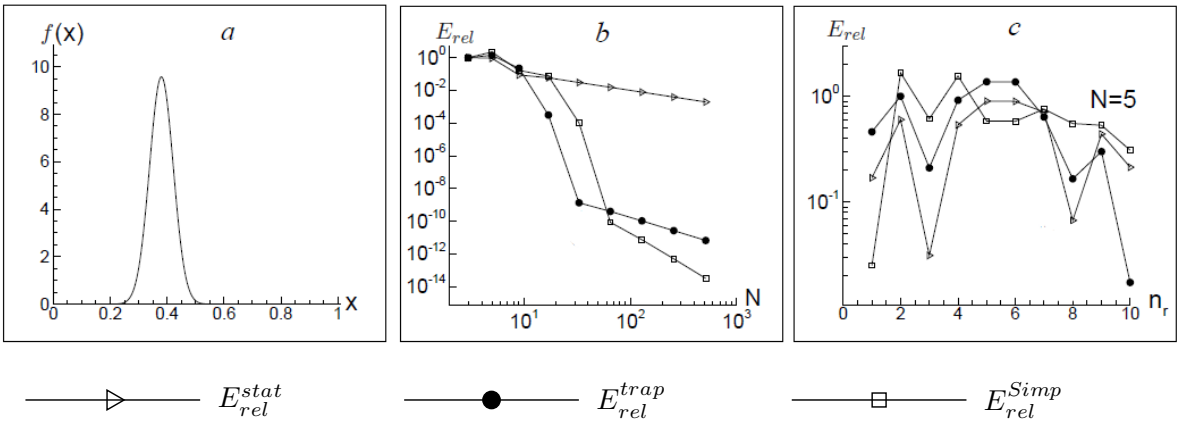


Figure 5.2: Numerical integration of a peak function. (a) The peak function (5.1.2) where the location of the peak maximum is $x^* = 0.38$ and the peak width is $\delta = 0.25$. (b) The relative integration error (2.1.7) for the peak function shown in (a) over a series of uniformly refined grids. The methods of numerical integration which are applied are the statistical rule (2.1.4), the composite trapezium rule (2.2.6), and Simpson's rule (2.2.7) where the legend is as shown above. (c) The relative error when the peak is located randomly on a uniform grid of $N = 5$ nodes for each of the aforementioned numerical integration methods. The location of the peak maximum x^* is considered as a random variable and the relative error is shown for 10 realisations n_r of x^*. The legend is again as shown above.

mean and it is known from the Empirical Rule that 99.7% of the total area under the curve lies within 3 standard deviations of the mean (*e.g* see [64] p. 93). Therefore we define the peak width to be $\delta = 6\sigma$.

Let us consider the peak function (5.1.2) where we fix $\delta = 0.25$ as shown in Figure 5.2a. We numerically integrate the function over a series of increasingly refined regular grids (2.2.2) and calculate the relative error (2.1.7). The following numerical integration methods are applied: the statistical rule (2.1.4), the composite trapezium rule (2.2.6), and the composite Simpson's rule (2.2.7). Convergence curves of the relative error are shown in Figure 5.2b. We use the same criterion (2.5.1) to assess accuracy that was used in previous chapters, *i.e.* that the relative error is such that $E_{rel} \leq \tau$ and recall that in ecological applications a range of $0.2 \leq \tau \leq 0.5$ is considered to be a good level of accuracy.

Figure 5.2b displays three distinct regions where the relative error behaves in a different manner. When the number N of grid nodes is very small an accurate estimate is not achieved, instead the relative error is $E_{rel} \sim 1$. Furthermore, the accuracy between the three numerical

integration methods cannot be differentiated. The method with the lowest order of convergence *i.e.* the statistical rule (2.1.4) performs at the same level of accuracy as the composite Simpson's method (2.2.7) which has the highest order of convergence. Grids with such a small number of nodes are the focus of the ecological monitoring problem since a realistic number of sample units used in practice is $N \sim 10$ [56, 60]. Once the grid is refined to have $N \approx 17$ nodes the accuracy improves and the methods differentiate themselves in terms of performance. The computational grid is still considered to be coarse, however, until the grid has $N \approx 65$ nodes where the asymptotic error estimates (2.1.9) begin to hold and the Simpson's rule has superior accuracy.

Clearly the lack of sufficient accuracy on very coarse grids is due to the grid nodes (sample units) being situated such that the peak of the pest population density distribution is missed. We recall that the location of the peak sub-domain D_u is unknown. In order to understand how the integration error (2.1.7) is affected by the relationship between the location of the peak and the position of the grid nodes, let us consider the location of the peak maximum x^* as a uniformly distributed random variable. We fix the number of grid nodes to be $N = 5$, and randomly move the location of the peak (5.1.2) over the domain $D = [0, 1]$. For each realisation n_r of the random variable x^* the peak (5.1.2) was integrated using the three methods outlined above and the relative integration error (2.1.7) calculated. This was done 10 times and the results can be seen in Figure 5.2c. For instance, when $n_r = 3$ the location of the peak maximum is $x^* = 0.7013$, while for $n_r = 7$ the same peak is located at $x^* = 0.4188$ *etc.* . There is significant variation in the accuracy achieved for all methods, with the error ranging from $E_{rel} \sim 0.01$ to $E_{rel} \sim 1$. When the number of grid nodes is very small, it is clear that the accuracy is probabilistic in nature, *i.e.* it becomes a matter of chance as to whether an acceptable level of accuracy is achieved. We define such grids to be *ultra-coarse*.

We have thus expanded on the grid classification used in previous chapters such that there are now three types of computational grid: fine, coarse, and ultra-coarse. The definitions are as follows:

Fine Grid: The asymptotic error estimates (2.1.9) hold.

Coarse Grid: The asymptotic error estimates do not hold, however, the error is deterministic in the sense that it can be described as being within a certain tolerance. Namely, the error is guaranteed to satisfy the condition (2.5.1) that $E_{rel} \leq \tau$ for a chosen tolerance τ.

Ultra-Coarse Grid: The asymptotic error estimates do not hold and the accuracy is not deterministic. The accuracy can only be described in terms of the probability of achieving an error within a prescribed tolerance. The probability that $E_{rel} \leq \tau$ for a chosen tolerance τ is $p < 1$.

In the next two sections we seek the threshold number of grid nodes N_t where the transition between ultra-coarse and coarse grids takes place, that is, when the error changes from being probabilistic to deterministic. We also aim to find an expression to describe the probability p of the integration error being within a chosen level of tolerance when the number of grid nodes N is such that $N < N_t$. As in the previous chapter, we focus on the composite trapezium rule of integration (2.2.6).

5.2 Analysis on Ultra-Coarse grids

Our analysis follows the approach of the previous chapter, whereby we Taylor expand the peak function (5.1.1) about the location of the maximum x^*

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

and we ignore the remainder term $R(x)$. Thus, we model the peak by the quadratic function (4.4.1) which we restate below for convenience:

$$\begin{cases} f(x) \approx Q(x) = B - A(x - x^*)^2 & \text{for } x \in [x_I, x_{II}] \\ f(x) = 0, & \text{otherwise} \end{cases} \tag{5.2.1}$$

where $A = -\frac{1}{2}\frac{d^2 f(x^*)}{dx^2} > 0$ and $B = f(x^*) > 0$. The exact integral I of the peak is then given by (4.4.3), that is we have

$$I = 2/3B\delta \tag{5.2.2}$$

where the peak width δ is the distance between the roots of the quadratic $Q(x)$ namely $\delta = 2\sqrt{B/A}$.

We reiterate the remarks we made in the previous chapter that by considering the peak as a quadratic function we introduce an interpolation error of the order of δ^3. However, it will be shown via numerical computations in Sections 5.4 and 5.5 that the theoretical predictions gained as a result of this approach are reasonably accurate for any kind of peak function.

We consider a uniform grid (2.2.2) of N nodes over the domain $D = [0, 1]$, thus $x_1 = 0$. $x_i = x_{i-1} + h, i = 2, \ldots, N$ where the grid step size is $h = 1/(N-1)$. Let us define the grid step size h in terms of the peak width δ, that is let

$$h = \alpha\delta. \tag{5.2.3}$$

We also express the location of a grid node x_i in terms of its relation to the position of the peak maximum x^*. In other words we paramaterise x_i such that

$$x_i = x^* + \gamma h, \qquad \gamma \in [0, 1/2] \tag{5.2.4}$$

where γ is a uniformly distributed random variable. It is sufficient to consider $\gamma \in [0, 1/2]$ due to the symmetry of the quadratic peak $Q(x)$. The aim is to find where the condition $E_{rel} \leq \tau$ is satisfied where the approximate integral I_a is calculated via the composite trapezium rule (2.2.6). We choose the tolerance to be $\tau = 0.25$, therefore this condition on the error is equivalent to

$$0.75I \leq I_a \leq 1.25I. \tag{5.2.5}$$

Let us begin our study of ultra-coarse grids by considering the case when the grid is so coarse that there is *at most* a single grid node located within the peak sub-domain $D_u = [x_I, x_{II}]$. Thus

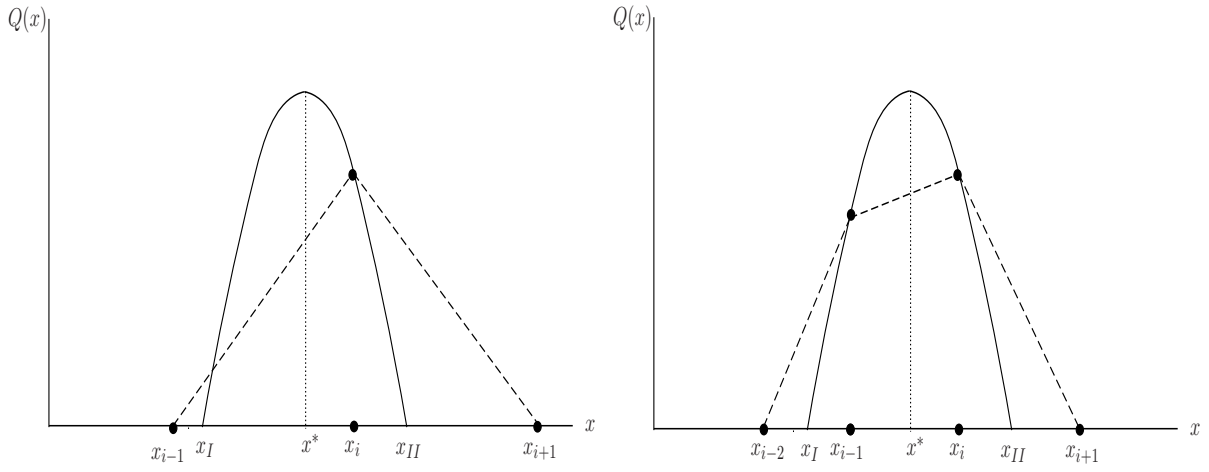


Figure 5.3: The trapezium rule (2.2.6) approximation of a peak function. (a) One grid node is located within the peak sub-domain $D_u = [x_I, x_{II}]$. (b) Two grid nodes are located within the peak sub-domain D_u.

we require that the grid step size is such that $h > \delta$. From (5.2.3) this in turn means that we consider $\alpha > 1$. For there to be at most one grid node in the vicinity of the peak means that we either have no grid nodes within the region D_u, or precisely one. The absence of grid nodes in the region D_u means that the peak is entirely lost on the computational grid and thus is a degenerate case. Consequently, we now focus on the scenario where a single grid node x_i lies within the region D_u and the composite trapezium rule (2.2.6) is employed to obtain an estimate I_a of the integral of the peak as shown in Figure 5.3a.

We have, in fact, already studied this case in Section 4.4 of the previous chapter and we established a range for the parameter γ (which we recall from (5.2.4) controls the location of the grid node x_i) such that the condition (5.2.5) is satisfied. Hence from (4.4.6),(4.4.7) and using the condition (5.2.3) above, we have the following condition for γ in terms of the parameter α,

$$\gamma_I(\alpha) \leqslant \gamma(\alpha) \leqslant \gamma_{II}(\alpha), \tag{5.2.6}$$

where

$$\gamma_I(\alpha) = \frac{1}{2\alpha}\sqrt{1 - \frac{5}{6\alpha}}, \quad \gamma_{II}(\alpha) = \frac{1}{2\alpha}\sqrt{1 - \frac{1}{2\alpha}}. \tag{5.2.7}$$

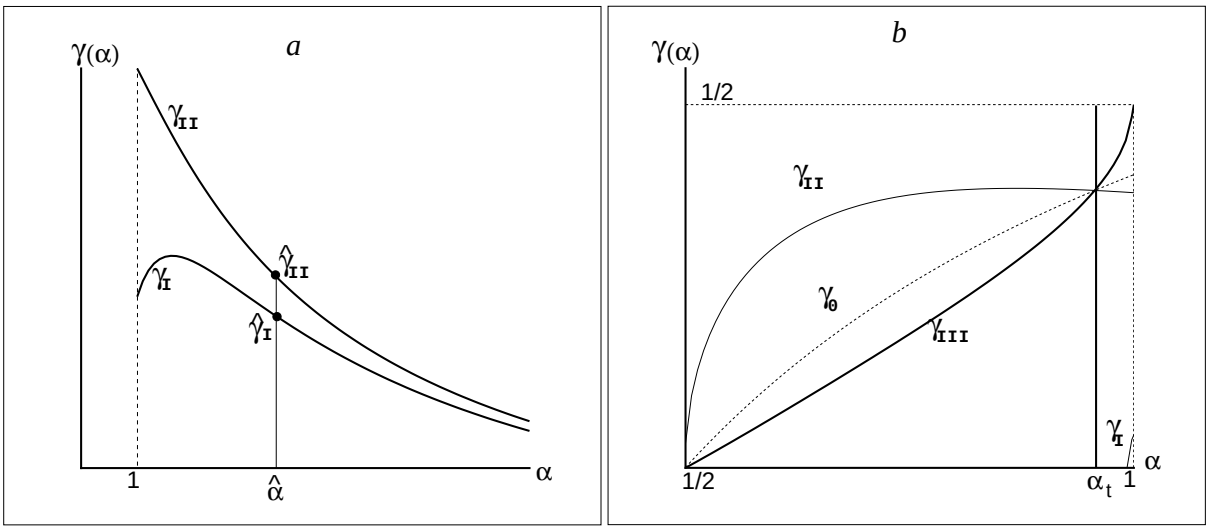


Figure 5.4: The range of parameter $\gamma = \gamma(\alpha)$ for which the integration error is $E_{rel} \leqslant \tau$ where $\tau = 0.25$. (a) The grid step size is $h > \delta$ where δ is the peak width. (b) The grid step size is $\delta/2 \leqslant h \leqslant \delta$.

The lower limit $\gamma_I(\alpha)$ exists only for $\alpha \geq \alpha_I = 5/6$, and the upper limit $\gamma_{II}(\alpha)$ exists only for $\alpha \geq \alpha_{II} = 1/2$. Since for the time being we are considering the case where $\alpha > 1$, both limits exist.

The inequalities (5.2.6) define the parameter range where integral is computed with the required accuracy $E_{rel} \leq \tau = 0.25$, this range can be seen in Figure 5.4a. Let us choose $\alpha = \hat{\alpha} > 1$. Then for a peak of width δ we have the grid step size defined as $h = \hat{h} = \hat{\alpha}\delta$. The error is then within the prescribed tolerance for $\hat{\gamma}_I \leqslant \gamma \leqslant \hat{\gamma}_{II}$, where $\hat{\gamma}_I = \gamma_I(\hat{\alpha})$ and $\hat{\gamma}_{II} = \gamma_{II}(\hat{\alpha})$.

We assume that the peak is as likely to be situated in one location as another within the domain $D = [0, 1]$. In other words, we assume the location of the maximum x^* to be a uniformly distributed random variable. Using the fact that $\alpha = h/\delta$, the probability of achieving an error $E_{rel} \leq \tau = 0.25$ on a regular grid with N nodes is then given by

$$p(E_{rel} \leq \tau, \alpha)_{theor} = \frac{(\gamma_{II}(\alpha) - \gamma_I(\alpha))}{(\gamma_{max} - \gamma_{min})} = 2(\gamma_{II}(\alpha) - \gamma_I(\alpha)), \tag{5.2.8}$$

where $\gamma_{min} = 0$, $\gamma_{max} = 1/2$. Since we have fixed the tolerance as $\tau = 0.25$ we shall henceforth

write the probability as $p = p(\alpha)_{theor}$. From (5.2.7) it can be seen that the probability of achieving an error $E_{rel} \leq \tau = 0.25$ is $p < 1$ when we have $\alpha > 1$ (*i.e.* $h > \delta$).

By considering the relative integration error (2.1.7) as a random variable and modelling a peak function by a quadratic, we have obtained the probability (5.2.8) of achieving a sufficiently accurate estimate when the grid step size h exceeds the width δ of the peak. In the next section we use the same approach to determine the transitional number of grid nodes N_t required to guarantee sufficient accuracy, *i.e.* the probability is $p = 1$.

5.3 On the Transition from Ultra-Coarse to Coarse grids

We now look at the transition from ultra-coarse grids where the error E_{rel} is a random variable, to coarse grids where the error is deterministic in the sense that the accuracy $E_{rel} \leq \tau$ is guaranteed. We again fix $\tau = 0.25$. We thus consider a finer computational grid such that there is a minimum of one grid node, and a maximum of two grid nodes within the peak sub-domain $D_u = [x_I, x_{II}]$ and the composite trapezium rule (2.2.6) is applied to approximate the integral of the peak (see Figure 5.3b). Therefore the grid step size h is such that

$$\delta/2 \leq h \leq \delta \tag{5.3.1}$$

and subsequently from (5.2.3) this means that $1/2 \leq \alpha \leq 1$.

Consider the location (5.2.4) of the grid node x_i which lies in the peak sub-domain $D_u = [x_I, x_{II}]$. Let us consider another grid node x_{i-1} which from the definition of the computational grid is defined as $x_{i-1} = x_i - h$. For x_{i-1} to also lie within the peak sub-domain, we require it to be within half the peak width of the location of the peak maximum x^*, since the peak is symmetric. Therefore we have the lower bound, $x_{i-1} \geq x^* - \delta/2$. This condition, in combination with the definition of x_{i-1}, and (5.2.4), gives

$$x^* + h(\gamma - 1) \geq x^* - \frac{\delta}{2}.$$

After rearrangement and using the paramaterisation (5.2.3) we obtain a minimum value of γ

namely γ_0 which ensures there are two grid nodes within the peak sub-domain D_u,

$$\gamma \geq \gamma_0 = 1 - \frac{\delta}{2h} = 1 - \frac{1}{2\alpha}.$$

For $\gamma \in [0, \gamma_0)$ there is only one grid node in the peak subinterval $D_u = [x_I, x_{II}]$. Let us therefore return to the condition (5.2.6) from the previous section. As previously mentioned, there are restrictions as to when the limits $\gamma_I(\alpha)$ and $\gamma_{II}(\alpha)$ exist. The upper limit requires $\alpha \geq \alpha_{II} = 1/2$. Since we are considering $1/2 \leq \alpha \leq 1$, this condition is satisfied. However the lower limit requires $\alpha \geq \alpha_I = 5/6$. Therefore for $\alpha \geq \alpha_I$, condition (5.2.6) holds and we can apply the result (5.2.7) to compute the probability p of achieving a sufficiently accurate estimate. On the other hand, for $\alpha < \alpha_I$, the lower limit $\gamma_I(\alpha)$ does not exist and we instead have to replace the inequalities (5.2.6) as follows

$$0 \leq \gamma(\alpha) \leq \gamma_{II}(\alpha). \tag{5.3.2}$$

Let us now investigate the case where $\gamma \in [\gamma_0, 1/2]$, thus there are two grid nodes in the peak sub-domain as shown in Figure 5.3b. We use the same analysis approach as before whereby we replace the peak function by the quadratic approximation (5.2.1). The exact integral I is thus given by (5.2.2) and we obtain the estimate I_a using the composite trapezium rule (2.2.6) on the local grid of nodes $\{x_{i-2}, x_{i-1}, x_i, x_{i+1}\}$

$$\begin{aligned} I_a &= \frac{h}{2} f(x_{i-2}) + h f(x_{i-1}) + h f(x_i) + \frac{h}{2} f(x_{i+1}) \\ &= \frac{h}{2} Q(x_{i-2}) + h Q(x_{i-1}) + h Q(x_i) + \frac{h}{2} Q(x_{i+1}) \\ &= 2Bh - A((\gamma - 1)^2 + \gamma^2) h^3. \end{aligned} \tag{5.3.3}$$

The tolerance is set as $\tau = 0.25$ and so we again have to solve for the inequality (5.2.5). Let us first look at the lower bound $I_a \geq \frac{3}{4} I$. From (5.2.2), the parameterisation (5.2.3), and the

fact that $B/A = \delta^2/4$ we obtain the following

$$\gamma^2(\alpha) - \gamma(\alpha) + C(\alpha) \leq 0, \tag{5.3.4}$$

where $C(\alpha) = \dfrac{1}{2} - \dfrac{1}{16\alpha^3}(4\alpha - 1)$. The equation $\gamma^2 - \gamma + C = 0$ has roots

$$\gamma_{III} = \frac{1 - \sqrt{1 - 4C(\alpha)}}{2}, \qquad \gamma_{IV} = \frac{1 + \sqrt{1 - 4C(\alpha)}}{2}.$$

The inequality (5.3.4) is satisfied in the range $\gamma \in [\gamma_{III}, \gamma_{IV}]$ which is only non-empty if the above roots exist. This requires $4C(\alpha) \leqslant 1$. From the definition of $C(\alpha)$, this condition is equivalent to

$$4\alpha^3 - 4\alpha + 1 \leqslant 0. \tag{5.3.5}$$

Solving the equation $4\alpha^3 - 4\alpha + 1 = 0$ numerically gives the roots $\alpha_1 \approx -1.1072$, $\alpha_2 \approx 0.2696$ and $\alpha_3 \approx 0.8376$. The inequality (5.3.5) is then satisfied for $\alpha \in (-\infty, \alpha_1] \cup [\alpha_2, \alpha_3]$. Since, however, we are concerned with $1/2 \leq \alpha \leq 1$ we have that $\alpha \in [1/2, \alpha_3]$. Hence, if $\alpha \in [1/2, \alpha_3]$, the range $\gamma \in [\gamma_{III}, \gamma_{IV}]$ will provide us with the integration error $E_{rel} \leqslant \tau = 0.25$. We should also take into account the restriction $\gamma \in [\gamma_0, \dfrac{1}{2}]$. It readily follows from the above computation that $\gamma_{III}(\alpha) < \dfrac{1}{2}$ and $\gamma_{IV}(\alpha) > \dfrac{1}{2}$ for any $\alpha \in [1/2, \alpha_3]$.

Let us now look at the lower boundary γ_0. Setting $\gamma_0(\alpha_t) = \gamma_{III}(\alpha_t)$ gives the equation

$$8\alpha_t^3 - 8\alpha_t^2 + 1 = 0. \tag{5.3.6}$$

The above equation has roots at $\alpha = 1/2$ and $\alpha_t \approx 0.8090$ in the subinterval $[1/2, \alpha_3]$. Hence $\gamma_{III}(\alpha) \leq \gamma_0(\alpha)$ for $\alpha \in [1/2, \alpha_t]$ and $\gamma_{III}(\alpha) \geq \gamma_0(\alpha)$ when $\alpha \in [\alpha_t, \alpha_3]$ (see Figure 5.4b). Let us also note that $\alpha_t < \alpha_I < \alpha_3$.

Finally, we consider the upper bound of inequality (5.2.5), that is $I_a \leqslant \dfrac{5}{4} I$. From (5.2.2), the parameterisation (5.2.3), and the fact that $B/A = \delta^2/4$, after some rearrangement we arrive at

$$\gamma^2(\alpha) - \gamma(\alpha) + D(\alpha) \geqslant 0, \tag{5.3.7}$$

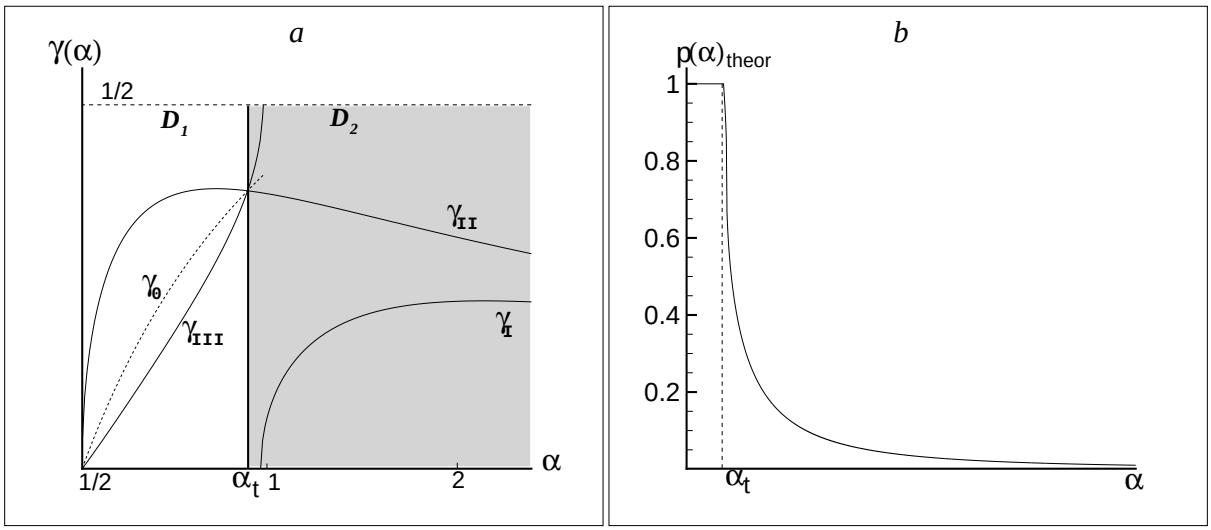


Figure 5.5: (a) The set of parametric curves defining the admissible range of node location γ on grids with the grid step size $h \geqslant \delta/2$. In the domain $D_1 : \alpha \in [1/2, \alpha_t]$ the condition $e \leqslant \tau$ holds for any γ, while in the domain $D_2 : \alpha > \alpha_t$ the integration error becomes a random variable. (b) The probability $p(\alpha)$ of having the integration error $E_{rel} \leqslant \tau = 0.25$ for a peak of the width δ integrated on a grid with the grid step size $h = \alpha\delta$.

with $D(\alpha)$ given by $D(\alpha) = \dfrac{1}{2} - \dfrac{1}{48\alpha^3}(12\alpha - 5)$. Roots of the equation $\gamma(\alpha)^2 - \gamma(\alpha) + D(\alpha)$ exist if $4D(\alpha) \leqslant 1$. Substituting the given value for $D(\alpha)$ results in the inequality

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

which does not have any real roots for $\alpha > 0$. Hence the inequality (5.3.7) holds for any value of γ.

We now compute the probability $p = p(\alpha)_{theor}$ of the event that the error is $E_{rel} \leqslant \tau = 0.25$ on a grid with fixed grid step size $h = \alpha\delta$, where $\alpha \geqslant 1/2$. The entire domain $\alpha \geqslant 1/2$ is shown in Fig. 5.5a, where the curves $\gamma(\alpha)$ of Figure 5.4a and b are now 'glued' together. There are four subintervals of α to look at, and for each we need to consider the case when there is one node within the peak sub-domain ($\gamma \in [\gamma_{min} = 0, \gamma_0)$), and when there are two ($\gamma \in [\gamma_0, \gamma_{max} = 1/2]$). The probability p of achieving an 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 sub-domain, and p_2 is the probability of an accurate estimate computed when two nodes belong to the peak

sub-domain.

Case 1: $\alpha \in [1/2, \alpha_t]$

(i) First we consider the range $\gamma \in [0, \gamma_0)$, thus is there is only one grid node in the peak sub-domain $D_u = [x_I, x_{II}]$. Since $\alpha_t < \alpha_I$, the admissible range of α is given by inequality (5.3.2). However we now need to investigate the position of the curve $\gamma_{II}(\alpha)$ in relation to the curve $\gamma_0(\alpha)$. In order to find the point of intersection we set the two expressions to be equal to each other and solve for α. Rearranging the equation $\gamma_0(\alpha) = \gamma_{II}(\alpha)$ gives the same equation as (5.3.6). Therefore we have the same points of intersection, $\alpha = 1/2$ and $\alpha = \alpha_t$. For $\alpha \in [1/2, \alpha_t]$ we have $\gamma_0(\alpha) \leq \gamma_{II}(\alpha)$, and for $\alpha \in [\alpha_t, 1]$ we have $\gamma_0(\alpha) \geq \gamma_{II}(\alpha)$. Thus for $\alpha \in [1/2, \alpha_t]$ the upper bound of (5.3.2) is replaced by $\gamma_0(\alpha)$ so we have the admissible range of γ as $0 \leq \gamma(\alpha) \leq \gamma_0(\alpha)$. The probability of achieving an accurate estimate I_a is then calculated as $p_1(\alpha) = (\gamma_0(\alpha) - 0)/(\gamma_{max} - \gamma_{min}) = 2\gamma_0(\alpha)$.

(ii) Now let us consider the range $\gamma \in [\gamma_0, 1/2]$ where there are two nodes in the peak sub-domain D_u. In this case the admissible range is $\gamma \in [\gamma_{III}(\alpha), \gamma_{IV}(\alpha)] \cap [\gamma_0, 1/2]$. In the working above we found that, for $\alpha \in [1/2, \alpha_t]$ we have $\gamma_{III}(\alpha) \leq \gamma_0(\alpha)$ and $\gamma_{IV} > 1/2$. Therefore the condition on γ becomes $\gamma_0(\alpha) \leq \gamma(\alpha) \leq 1/2$ and we find the probability of achieving an accurate estimate I_a to be $p_2(\alpha) = (1/2 - \gamma_0(\alpha))(\gamma_{max} - \gamma_{min}) = 1 - 2\gamma_0(\alpha)$.

So for $\alpha \in [1/2, \alpha_t]$, where $\alpha_t \approx 0.8090$, the probability of obtaining the error $E_{rel} \leq \tau = 0.25$ for any $\gamma \in [0, 1/2]$ is then $p(\alpha)_{theor} = p_1(\alpha) + p_2(\alpha) = 1$.

Case 2: $\alpha \in (\alpha_t, \alpha_I]$

(i) For the range $\gamma \in [0, \gamma_0)$, the inequality (5.3.2) holds. In Case 1(i) we established that for $\alpha \in (\alpha_t, 1]$ we have $\gamma_0(\alpha) > \gamma_{II}(\alpha)$. Therefore the admissible range of γ remains as (5.3.2) and we have the probability of achieving acceptable accuracy as $p_1(\alpha) = (\gamma_{II}(\alpha) - 0)/(\gamma_{max} - \gamma_{min}) = 2\gamma_{II}(\alpha)$.

(ii) Now we consider $\gamma \in [\gamma_0, 1/2]$. As in Case 1(ii) the admissible range is $\gamma \in [\gamma_{III}(\alpha), \gamma_{IV}(\alpha)] \cap [\gamma_0, 1/2]$. In the range $\alpha \in (\alpha_t, \alpha_3]$ we have $\gamma_{III}(\alpha) > \gamma_0(\alpha)$ and $\gamma_{IV} > 1/2$. Since $\alpha_I \leq \alpha_3$,

the admissible range of γ becomes $\gamma \in [\gamma_{III}(\alpha), 1/2]$. The probability of achieving an accurate answer is $p_2(\alpha) = (1/2 - \gamma_{III}(\alpha))/(\gamma_{max} - \gamma_{min}) = 1 - 2\gamma_{III}(\alpha)$.

Thus for $\alpha \in (\alpha_t, \alpha_I]$, where $\alpha_I = 5/6 \approx 0.8333$, the probability of obtaining the error $E_{rel} \leq \tau = 0.25$ for any $\gamma \in [0, 1/2]$ is then $p(\alpha)_{theor} = p_1(\alpha) + p_2(\alpha) = 1 - 2\gamma_{III}(\alpha) + 2\gamma_{II}(\alpha)$. For this range of α, we have $\gamma_{III} > \gamma_{II}$, therefore $p(\alpha)_{theor} < 1$.

Case 3: $\alpha \in (\alpha_I, \alpha_3]$

(i) For $\gamma \in [0, \gamma_0)$, since we now have $\alpha > \alpha_I$, the inequality (5.2.6) holds for any $\gamma \in [0, \gamma_0(\alpha))$. Therefore $p_1(\alpha) = (\gamma_{II}(\alpha) - \gamma_I(\alpha))/(\gamma_{max} - \gamma_{min}) = 2(\gamma_{II}(\alpha) - \gamma_I(\alpha))$.

(ii) For $\gamma \in [\gamma_0, 1/2]$, as $(\alpha_I, \alpha_3] \subset (\alpha_t, \alpha_3]$, from Case 2 (ii) we have $p_2(\alpha) = 1 - 2\gamma_{III}(\alpha)$

The resulting probability for $\alpha \in (\alpha_I, \alpha_3]$, where $\alpha_3 \approx 0.8376$, the probability of obtaining the error $E_{rel} \leq \tau = 0.25$ for any $\gamma \in [0, 1/2]$ is then $p(\alpha)_{theor} = p_1(\alpha) + p_2(\alpha) = 1 - 2\gamma_{III}(\alpha) + 2(\gamma_{II}(\alpha) - \gamma_I(\alpha))$. For this range of α, we have $\gamma_{III} > \gamma_{II}$, therefore $p(\alpha)_{theor} < 1$.

Case 4: $\alpha > \alpha_3$

(i) For $\gamma \in [0, \gamma_0)$, as $\alpha_3 > \alpha_I$, the probability is as in Case 3(i), that is $p_1(\alpha) = (\gamma_{II}(\alpha) - \gamma_I(\alpha))/(\gamma_{max} - \gamma_{min}) = 2(\gamma_{II}(\alpha) - \gamma_I(\alpha))$.

(ii) In the case of $\gamma \in [\gamma_0, 1/2]$, the admissible range $\gamma \in [\gamma_{III}(\alpha), \gamma_{IV}(\alpha)] \cap [\gamma_0, 1/2]$ is the empty set since the existence of $\gamma_{III}(\alpha), \gamma_{IV}(\alpha)$ requires $\alpha \in [1/2, \alpha_3]$. Hence $p_2(\alpha) = 0$.

For $\alpha > \alpha_3$, we therefore have the probability of achieving the accuracy of $E_{rel} \leq \tau = 0.25$ as $p(\alpha)_{theor} = p_1(\alpha) + p_2(\alpha) = 2(\gamma_{II}(\alpha) - \gamma_I(\alpha))$, where $p < 1$.

The function $p(\alpha)_{theor}$ is shown in Figure 5.5b. For the fixed width δ of the peak, the parameter α_t used in (5.2.3) yields the threshold value h_t of the grid step size where the transition from ultra-coarse grids to coarse grids takes place, namely,

$$h_t = \alpha_t \delta, \tag{5.3.8}$$

where $\alpha_t \approx 0.8090$. On any grid with $\alpha \leq \alpha_t$ (*i.e.* $h \leq h_t$) represented by domain D_1 in Figure 5.5a, the error (2.1.7) is deterministic in the sense the probability of achieving an error

$E_{rel} \leq \tau = 0.25$ is $p(\alpha)_{theor} = 1$. In other words we are guaranteed to achieve sufficient accuracy. When $\alpha > \alpha_t$ (*i.e.* $h > h_t$) represented by the shaded domain D_2 in Figure 5.5b, the computational grid is ultra-coarse. We cannot say what level of accuracy is achieved. Instead we can only say what the probability is of achieving a desired level of accuracy, and this probability is $p < 1$. Since $h = 1/(N-1)$, from (5.3.8) we obtain the transitional number N_t of grid nodes required to guarantee an error of $E_{rel} \leq \tau = 0.25$

$$N_t = 1 + \frac{1}{h_t} = 1 + \frac{1}{\alpha_t \delta}. \tag{5.3.9}$$

By considering the error as a random variable and considering a peak as a quadratic function, we have obtained an estimate of the number N_t of grid nodes required to guarantee a sufficiently accurate estimate I_a. We have also made a theoretical prediction of the probability p of achieving sufficient accuracy when the number of grid nodes is such that $N < N_t$. In the following two sections we test our predictions for a variety of test cases where the shape of the peak differs from a quadratic function.

5.4 Numerical Verification of Approach: Standard Test Cases

We begin testing our theoretical predictions by considering standard examples of peak functions for which we can calculate the exact integral I. We then turn our attention to ecologically meaningful data sets in the next section.

For each test case, the peak width δ is fixed and the location of the peak maximum x^* is considered as a uniformly distributed random variable. Since we wish to ensure the entire peak is always situated within the interval $D = [0, 1]$, and we consider both symmetric and asymmetric cases, x^* is randomised over the interval $[\delta, 1-\delta]$. The number of grid nodes is set to be N_l, the location of the peak maximum x^* is randomly selected and the peak function is then integrated via the composite trapezium rule (2.2.6). This is done for $n_r = 10,000$ realisations of x^*. The probability $p(h_l)_{num}$ of accurately evaluating the integral over a grid of N_l nodes is then given by

$$p(h_l)_{num} = \frac{\hat{n}_r}{n_r}, \tag{5.4.1}$$

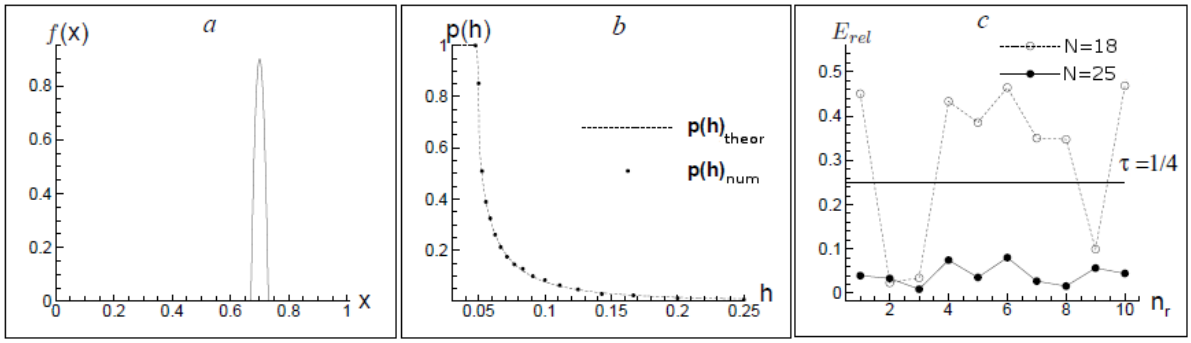


Figure 5.6: (a) A quadratic function (5.2.1). The peak width is $\delta = 0.06$ and the parameters are $A = 1000$, $B = 0.9$. (b) The probability $p(h)_{num}$ obtained by direct computation agrees with the theoretical results $p(h)_{theor}$ for the function (5.2.1) of Fig. 5.6a. (c) The integration error for the function (5.2.1) on an ultra-coarse grid and a coarse grid with a fixed number of nodes. The error (2.1.7) is shown for the ten realisations n_r of the random variable x^*, where x^* is uniformly distributed. The probability of achieving an error $E_{rel} \leq \tau = 0.25$ 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).

where $h_l = 1/(N_l - 1)$ is the grid step size and $\hat{n}_r$ is the number of realisations of x^* for which the error is $E_{rel} \leq \tau$ for $\tau = 0.25$. The number of grid nodes is then increased to $N_{l+1} = N_l + 1$ and we repeat the process, stopping when we reach the number of grid nodes N_L such that the grid step size is $h_L \leq \delta/2$.

Our approach to obtaining the theoretical probability $p(h)_{theor}$ was to treat the peak function as a quadratic. We thus first verify this theoretical prediction by finding $p(h)_{num}$ for the quadratic peak (5.2.1). Let us set the peak width to be $\delta = 0.06$ and choose $A = 1000, B = 0.9$, as shown in Figure 5.6a. The probability (5.4.1) is calculated for a series of grids starting with $N_1 = 5$ nodes and ending in $N_{18} = 22$ for which condition (5.3.1) holds. From Figure 5.6b it can be seen that the values of the numerical probability $p(h_l)_{num}$ for $l = 1, ..., 18$ lie very close to the theoretical curve $p(h)_{theor}$. Let us define the maximum deviation d_p of the numerical results from the theoretical curve as

$$d_p = \max_{l=1,...,N_L} |p(h_l)_{theor} - p(h_l)_{num}|, \tag{5.4.2}$$

where $p(h_l)$ is the numerical probability (5.4.1) and $p(h_l)_{theor}$ the theoretical probability on a

grid of step size h_l. For the above quadratic test case with $n_r = 10,000$, the maximum deviation was found to be $d_p = 0.0087$. Therefore we consider the number of realisations $n_r = 10,000$ to provide sufficiently reliable results, and we continue to use this value in the study of other test cases.

As discussed in the previous section, on coarse grids the error is deterministic, that is we can guarantee accuracy within a tolerance τ. However on ultra-coarse grids the error is probabilistic where we can only describe the probability of achieving an accuracy of $E_{rel} \leq \tau$. This is demonstrated in Figure 5.6c. Here, the location of the maximum x^* of the quadratic peak (5.2.1) has been moved randomly $n_r = 10$ times on the interval $[\delta, 1-\delta]$. For each realisation the integration error (2.1.7) has been calculated both on an ultra-coarse and a coarse grid. Applying our theoretical predictions (5.3.8) and (5.3.9) from the previous section we anticipate that the transitional grid step size will be $h_t = 0.04854$, which corresponds to $N_t = 22$ grid nodes. Therefore we choose grids with $N = 18$ and $N = 25$ nodes as examples of ultra-coarse and coarse grids respectively. As shown in Figure 5.6c, the error on the ultra-coarse grid depends on the location of the peak. The theoretical prediction of the probability of achieving an error of $E_{rel} \leq \tau = 0.25$ for a grid with $N = 18$ nodes is $p \approx 0.3$ (see Figure 5.6b) and this matches with the numerical experiment shown in Figure 5.6c. Meanwhile, it can be seen in Figure 5.6c that on the coarse grid the error is deterministic, and the desired accuracy is always achieved regardless of the positioning of the peak. Again, this corroborates our theoretical prediction.

We now wish to ascertain how effective our predictions are for peak functions other than quadratics. We first consider another example of a symmetric peak, namely a quartic function as defined by

$$f(x) = \begin{cases} A\left(\left(\frac{\delta}{2}\right)^4 - (x - x^*)^4\right), & \text{for } x \in [x^* - \frac{\delta}{2}, x^* + \frac{\delta}{2}], \\ 0, & \text{otherwise,} \end{cases} \tag{5.4.3}$$

where $A = 1,200,000$ and we again choose the peak width to be $\delta = 0.06$. The resulting peak is shown in Figure 5.7a. A comparison of the numerical probability function $p(h)_{num}$, as calculated according the method outlined at the beginning of the section, and the theoretical curve $p(h)_{theor}$ is shown in Figure 5.8a. Since this peak is not a quadratic, we expect some discrepancy with

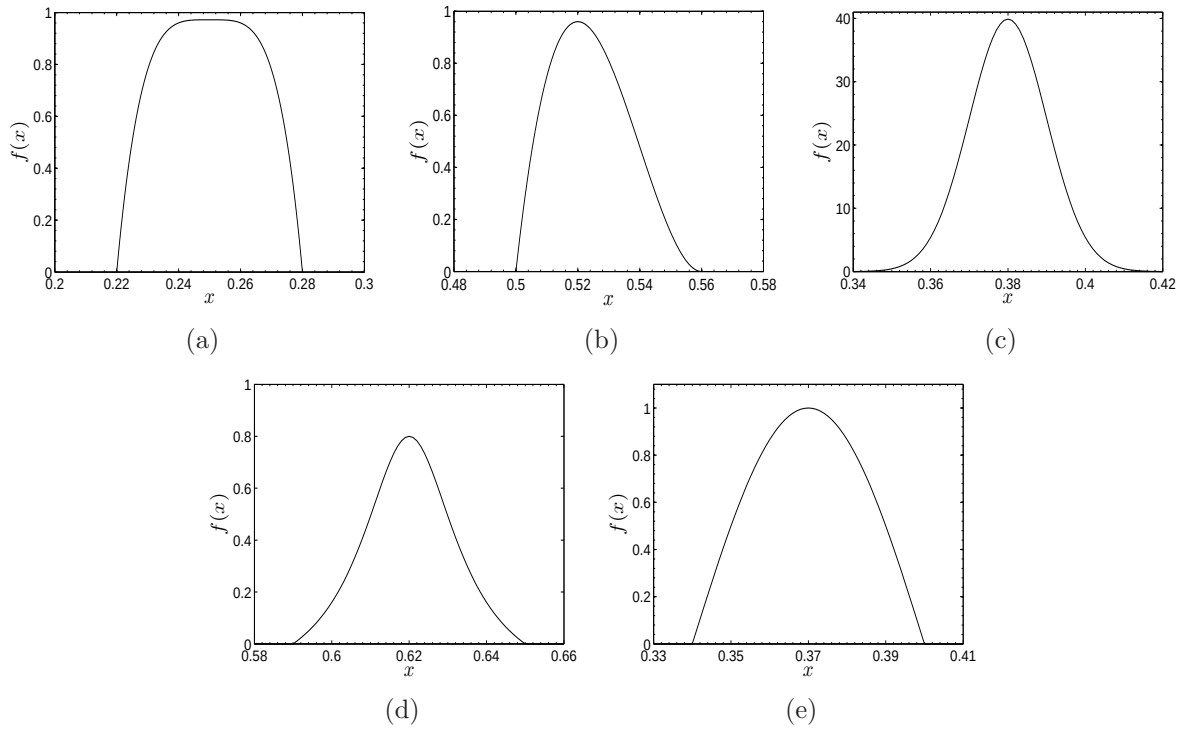


Figure 5.7: Standard numerical test cases: (a) quartic function (5.4.3), (b) cubic function (5.4.4), (c) normal distribution (5.1.2), (d) Lorentz distribution (5.4.5), (e) sine peak (5.4.6). In all cases the peak width is set as $\delta = 0.06$. In order to show the shape of the peak clearly the functions are plotted local to the peak, however, the interval of integration remains the entire unit interval $D = [0, 1]$.

the theoretical curve. However, it can be seen there is good agreement between the two curves, except in what we shall refer to as the *transition layer*, that is when $\alpha_t \leq \alpha_3$ (*i.e.* $h_t \leq \alpha_3\delta$). It is in this region that the maximum deviation $d_p = 0.498764$ occurs, as calculated by (5.4.2).

The next test case is constructed from a cubic function therefore providing us with an example of a peak which is asymmetric. The density distribution $f(x)$ is defined as follows

$$f(x) = \begin{cases} A\left(x - x^* + (\delta/3)\right)\left(x - x^* - (2\delta/3)\right)^2, & \text{for } x \in [x^* - \delta/3, x^* + 2\delta/3] \\ 0, & \text{otherwise.} \end{cases} \tag{5.4.4}$$

We set $A = 30,000$ and use the same peak width as before that is $\delta = 0.06$ (see Figure 5.7b). Once again the numerical and theoretical probability curves are plotted, as shown in Figure 5.8b. When $h > \alpha_3\delta$ the numerical results match well with the theoretical predictions. Here

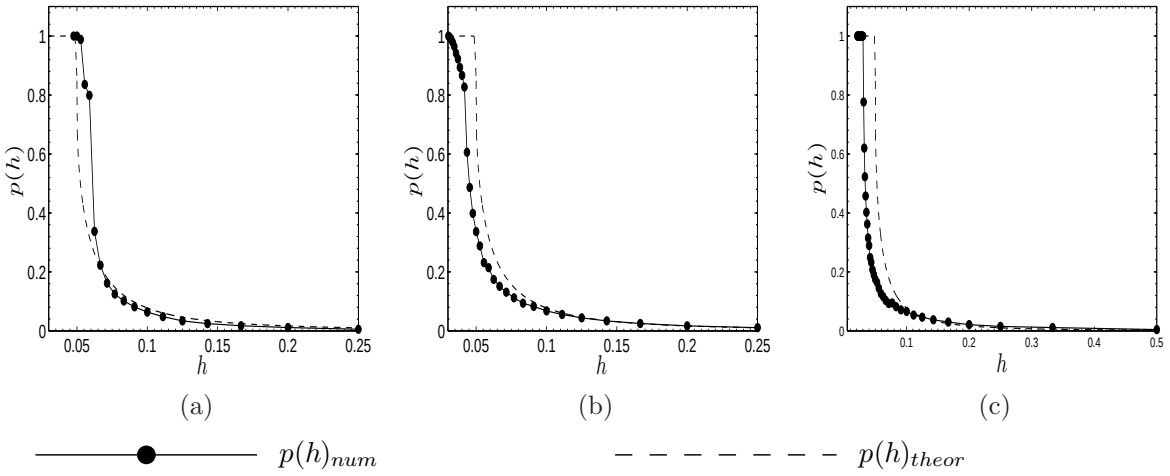


Figure 5.8: The numerically computed probability (5.4.1) (solid line) of obtaining an accurate estimate, that is, the relative error (2.1.7) satisfies the condition $E_{rel} \leq \tau$ where we have fixed $\tau = 0.25$. It is compared with the theoretical curve $p(h)_{theor}$ (dashed line) obtained for the quadratic function. The probability graph is shown for the (a) quartic function (5.4.3) (b) cubic function (5.4.4), (c) normal distribution (5.1.2). The peak width is $\delta = 0.06$ in each case.

the probability of achieving an accurate estimate is small. In the transition layer, the numerical probability is smaller than predicted by the theory. In particular, the grid step size h_t^{num} for which the transition between ultra-coarse and coarse grids occurs, is smaller than the theoretical estimate (5.3.8) obtained for the quadratic approximation of the peak.

So far the test cases we have considered have been constructed such that population density $f(x)$ is zero outside of the peak sub-domain D_u. The normal distribution (5.1.2) already considered in Section 5.1 provides an example of a peak function that is non-zero over the entire domain $D = [0, 1]$. Although outside of the peak of width $\delta = 6\sigma$ the function (5.1.2) is comparatively close to zero, and therefore we still expect there to exist ultra-coarse grids where the error is a random variable. The peak width is fixed as $\delta = 0.06$ (see Figure 5.7c), and the theoretical and numerical curves can be seen in Figure 5.8c. Still there is good agreement between the curves on very coarse grids, however the whole of the numerical curve $p(h)_{num}$ is now shifted with respect to the theoretical curve $p(h)_{theor}$. The maximum deviation (5.4.2) is $d_p = 0.8129$ and the transitional value of the grid step size according to the numerical results is $h_t^{num} \approx 0.5\delta$, where as the theoretical prediction is $h_t \approx 0.8\delta$.

This discrepancy in the values of the transitional grid step size is due to the fact that the theoretical prediction was constructed by considering the peak as a quadratic function. Therefore as earlier discussed in Section 5.2 an interpolation error of the order δ^3 is introduced for peaks that are not quadratic. Consequently for small δ *i.e.* a narrow peak, we expect our theoretical probability function $p(h)_{theor}$ to be accurate, diverging from the numerically computed probability $p(h)_{num}$ as the peak width δ increases. We investigate this for another peak function, constructed from the Lorentz distribution, which is well known in fields of physics and interdisciplinary research (*e.g* see [8]).

We define the test case as

$$f(x) = \begin{cases} \dfrac{\delta^2}{4}\dfrac{1}{4(x-x^*)^2+\delta^2/4} - \dfrac{1}{5}, & \text{for } x \in [x^* - \delta/2, x^* + \delta/2], \\ \\ 0, & \text{otherwise}, \end{cases} \tag{5.4.5}$$

where the peak is given by a Lorentz distribution. Let us now select three peak widths, the baseline width of $\delta = 0.06$ which we have used in the previous test cases, a narrow peak width $\delta = 0.01$, and a wider peak with $\delta = 0.1$. The peak function (5.4.5) is shown for peak width $\delta = 0.06$ in Figure 5.7d. The theoretical and numerical probability curves are computed for each peak width. The 'transition layer' of the curves can be seen in Figure 5.9a, whilst the 'tails' are shown in Figure 5.9b. It can be seen that as the peak becomes narrower, the numerically computed probability curve $p(h)_{num}$ gets closer to the theoretical curve $p(h)_{theor}$. This means that the estimate (5.3.8) of the actual transitional grid step size h_t^{num} becomes more reliable as δ decreases.

We investigate this further by calculating h_t^{num} for the peak (5.4.5) at various peak widths δ and comparing the values to those produced by the theoretical estimate (5.3.8). For each δ the values of h_t^{num} were calculated according to the condition that $p(h_t^{num}) = 1$ and $p(h) < 1$ when $h > h_t^{num}$. The results are shown in Figure 5.10a. It can be seen that for narrow peaks, in particular where $\delta < 0.1$, the numerically computed threshold value h_t^{num} is close to the theoretical predictions h_t^{theor}. The numerical values all lie below the theoretical curve, therefore

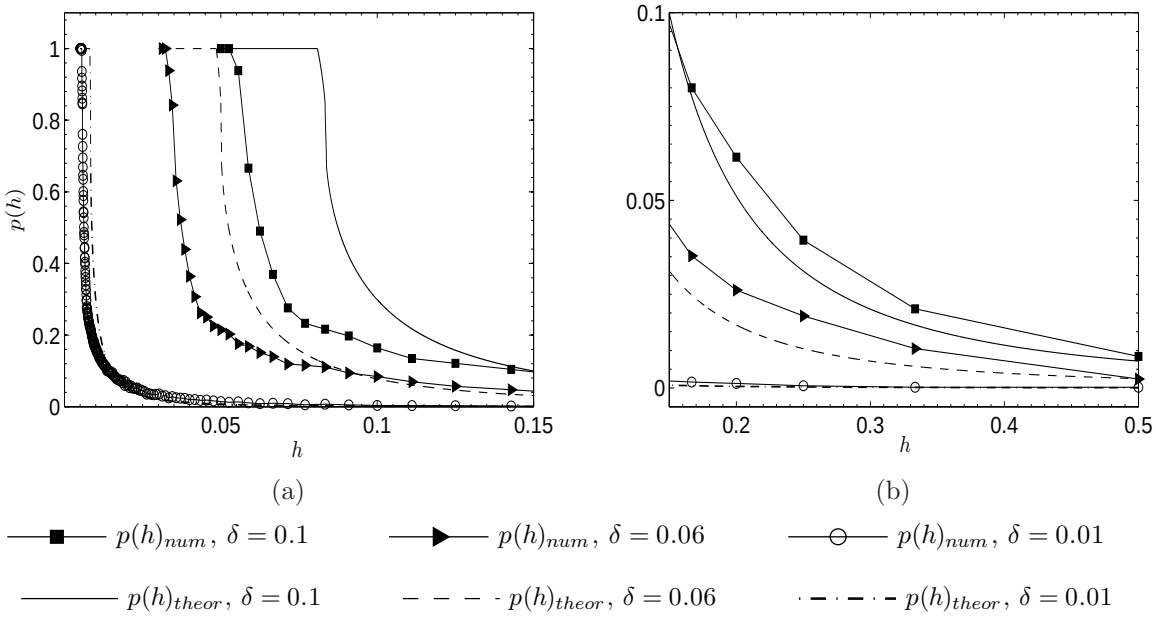


Figure 5.9: The probability graphs $p(h)_{num}$ computed for the function (5.4.5) for different values of the peak width namely $\delta = 0.1, 0.06$, and 0.01. The computed probability $p(h)_{num}$ is compared with the theoretical estimate $p(h)_{theor}$ made for a quadratic function. The probability graph is shown for (a) the transition layer, and (b) the tail region. The legend for both figures is given above.

the theoretical prediction (5.3.8) can be used as an upper bound to the transitional grid step size h_t. An estimate of the minimum number of grid nodes necessary to ensure an accuracy $E_{rel} \leq \tau = 0.25$ can thus be calculated from (5.3.9).

We now repeat the process detailed above for another example of a peak function. This test case is formed from a sine function in the following manner

$$f(x) = \begin{cases} \sin\left(\dfrac{\pi}{\delta}\left(x - x^* + \dfrac{\delta}{2}\right)\right), & \text{for } x \in [x^* - \delta/2, x^* + \delta/2], \\ \\ 0, & \text{otherwise.} \end{cases} \tag{5.4.6}$$

This function is shown in Figure 5.7e for the peak width $\delta = 0.06$. The numerically calculated transitional grid step h_t^{num} size for various values of the peak width δ are plotted alongside the theoretical curve (5.3.8) in Figure 5.10b. Again it can be seen that the theoretical curve provides

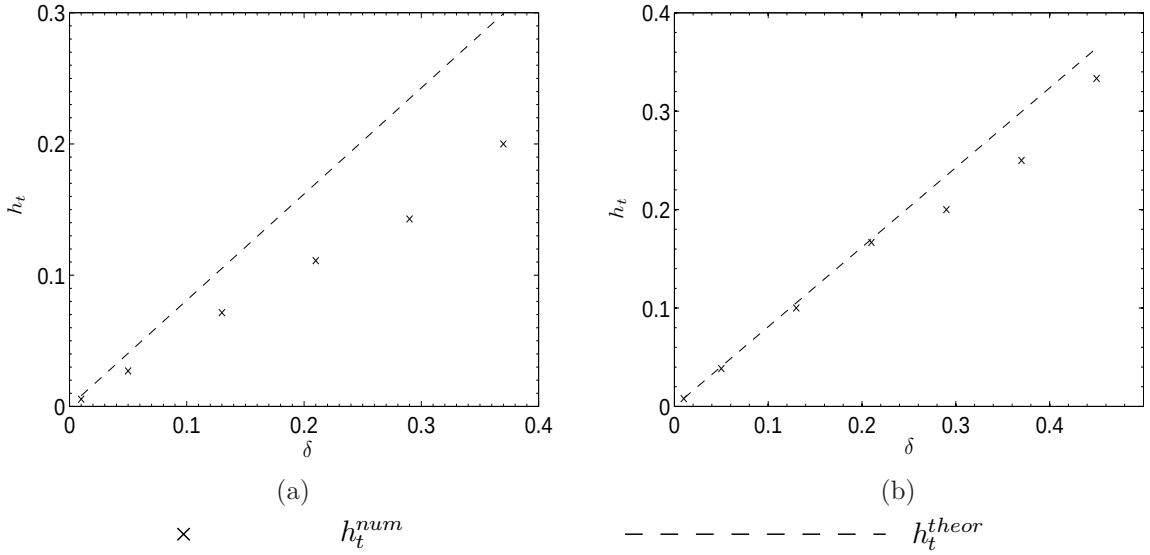


Figure 5.10: The threshold grid step size h_t when the peak width δ varies. The threshold values h_t^{num} obtained by direct computation are compared with the theoretical curve h_t^{theor} defined by (5.3.8). The calculations are made for (a) the Lorentz peak (5.4.5), and (b) the sine peak (5.4.6).

an upper bound for h_t, with the estimation becoming more reliable for narrower peaks.

We have shown by considering a variety of standard examples of peak functions that (5.3.8) provides a good estimate of the grid step size h_t where the probability p of obtaining an error $E_{rel} \leq \tau = 0.25$ transitions from $p < 1$ (sufficient accuracy is a matter of chance), to $p = 1$ (sufficient accuracy is guaranteed). This estimate is more reliable for narrower peaks, however it can be considered as an upper bound for the grid step size h when wider peaks are considered. In turn, (5.3.9) provides a lower bound for the number of grid nodes N required. In the next section we proceed to apply our theoretical predictions to ecologically meaningful data.

5.5 Numerical Verification of Approach: Ecological Test Cases

We now consider test cases of ecological significance, generated by the 1D spatially explicit form of the Rosenzweig-MacArthur model (3.2.7–3.2.8). It was demonstrated in Section 4.1 of the previous chapter that the spatial distribution of the pest population density function $f(x)$ at a fixed time t is controlled by the diffusion coefficient d. For $d \ll 1$ the initial condition $f(x, 0)$ can evolve into a function with a single peak (*e.g* see [75]). We thus use three ecologically meaningful

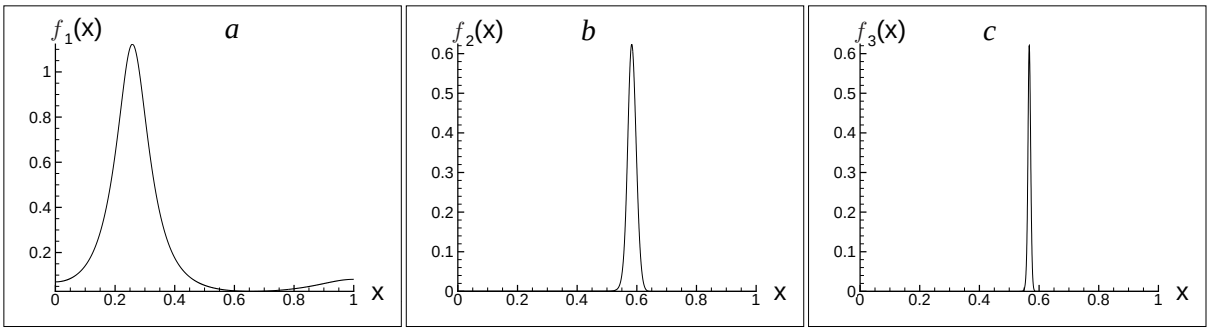


Figure 5.11: Ecological test cases. The spatial distribution of the pest population density $f(x)$ as predicted by the 1D Rosenzweig-MacArthur model (3.2.7–3.2.8) for different values of the diffusion coefficient d; (a) the density distribution $f_1(x)$ has been obtained for for $d = 10^{-4}$ (b) the density distribution $f_2(x)$ obtained for $d = 10^{-5}$ (c) the density distribution $f_3(x)$ where $d = 10^{-6}$. An example of the system's parameters (the density distribution $f_2(x)$): $t = 50$, $k = 0.5$, $\Lambda = 2.0$, $m = 0.42$. The initial conditions are $f(x,0) = 0.8, 0 < x < 0.6$, $g(x,0) = 0.5, 0 \leqslant x \leqslant 0.55$, and $f(x,0) = 0, x > 0.6$, $g(x,0) = 0, x > 0.55$.

examples of peak functions genereated by numerically solving on a very fine, regular grid, the system of equations (3.2.7–3.2.8) for the diffusion coefficient values $d = 10^{-4}$, $d = 10^{-5}$, and $d = 10^{-6}$. The numerical solutions were supplied by the authors of [70]. The resulting one-peak density distributions $f_1(x)$, $f_2(x)$, and $f_3(x)$ are shown in Figure 5.11a–c respectively. It is clear from the figures that the diffusion d affects the peak width δ. We recall the relationship (4.1.8) formulated in Section (4.1), namely

$$\delta = \omega\sqrt{d}.$$

Using this estimate of the peak width, we can rewrite the estimate for the transitional grid step size (5.3.8) in terms of the diffusion coefficient

$$h_t = \alpha_t \delta \approx \alpha_t \omega \sqrt{d}, \tag{5.5.1}$$

where typically $\omega \approx 25$ (see Section 4.1).

Let us first consider the distribution $f_1(x)$ shown in Figure 5.11a. The diffusion coefficient is $d = 10^{-4}$, thus according to (5.5.1) the estimated grid step size for which we are guaranteed an accuracy of $E_{rel} \leq \tau = 0.25$ is then $h_t \approx 0.2$. From (5.3.9), the estimate for the minimum number of grid nodes needed for an accurate estimation is then $N_t \approx 6$. We now integrate the

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_{rel}	0.6948	0.1119	0.5459	0.07983	0.1699	0.02305	0.08228	0.001918

Table 5.1: The integration error (2.1.7) for the density distribution $f_1(x)$ on a sequence of refined regular grids with grid step size h (*i.e.* the number N of grid nodes).

density function $f_1(x)$ using the compound trapezium rule (2.2.6), and compute the relative integration error (2.1.7) over a series of refined grids[1]. At each stage of refinement a single grid node is added. The results displayed in Table 5.1 show that the estimate of the threshold number of grid nodes N_t is in surprisingly good agreement with the numerical results, where the actual value of is indeed $N_t = 6$ as predicted. Whilst the error (2.1.7) does continue to oscillate on grids with $N > N_t$ nodes, it remains within the limit $E_{rel} \leq \tau = 0.25$.

We now turn our attention to the distribution $f_2(x)$ of Figure 5.11b, where the diffusion coefficient is $d = 10^{-5}$, and repeat the process outlined above. The threshold number of grid nodes is now estimated as $N_t \approx 17$. The integration error obtained on implementing the midpoint rule is shown in Table 5.2 and it can be seen that the actual minimum number of grid nodes which guarantees a sufficiently accurate integral evaluation is $N_t = 21$. In terms of the ecological problem of pest management, this number of sample units may not be realistic. As previously discussed, a typical number of sample units used in practice is $N \sim 10$. Thus, other factors need to be considered in order to decide on the number of sample units that should be used. A factor that should be considered in future work is that in the case of narrow peaks such as $f_2(x)$, it may be that the value of the integral is small. That is if the pest population density is located to a small area, the total number of pests may also be small. This means that the risk of the pest abundance being at a harmful level is smaller, and we can therefore afford to be less accurate in our approximation and instead consider a tolerance of $\tau > 0.25$ *e.g* $\tau = 0.5$, or perhaps even $\tau = 1$. The same approach as described in Sections 5.2 and 5.3 could then be used to obtain an estimate of the threshold number of traps N_t.

Finally we consider the extremely narrow peak $f_3(x)$ with diffusion coefficient $d = 10^{-6}$

[1]In order to calculate the relative error (2.1.7), we apply the approach from previous chapters and use the trapezium rule approximation (2.2.6) calculated on the finest available grid as the value of the exact integral I. This is also done for the density distributions $f_2(x)$ and $f_3(x)$.

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_{rel}	0.4127	0.5412	0.5101	0.4124	0.1960	0.0028	0.1454	0.2234	0.2137

Table 5.2: The integration error (2.1.7) for the density distribution $f_2(x)$ on a sequence of refined regular grids with grid step size h (*i.e.* the number N of grid nodes).

shown in Figure 5.11c. The estimate of the threshold number of grid nodes in this instance is $N_t \approx 51$. This number of sample units is too large to be used in routine pest monitoring. Since the integral is very small (in the case of the distribution $f_3(x)$ we have $I = 0.007161$), our recommendation to ecologists would be to wait until time evolves and the peak gets wider especially since, as the results in the previous section show, when the peak is very narrow the probability of achieving a sufficiently accurate estimation from a more reasonable number of sample units, *i.e.* $N \sim 10$, is very small.

Now that we have validated our theoretical results, in the next section we go on to investigate how we might apply this probabilistic approach to assessing the integration error in order to compare the accuracy of numerical integration methods on ultra-coarse grids.

5.6 Comparing Numerical Integration Methods on Ultra-Coarse Grids

In Section 5.1 we demonstrated that on ultra-coarse grids, methods of numerical integration cannot be differentiated in terms of their performance by conventional convergence analysis. We propose that on such grids, the efficiency of a numerical integration method should instead be assessed by the probability of achieving a sufficiently accurate estimate. That is, on comparing two methods on ultra-coarse grids, the one with the highest probability of achieving an accurate evaluation of the integral would be recommended. We shall now use the approach detailed in Sections 5.2 and 5.3 to compare the performance on ultra-coarse grids of the composite trapezium rule (2.2.6) with the statistical method (2.1.4) often used in ecological applications. For the purposes of this investigation, we restrict our consideration of the grid step size to the range $h > \delta$ (*i.e.* $\alpha > 1$). We use the superscripts 'trap' and 'stat' to distinguish between the approximations

obtained via the composite trapezium rule and the statistical method respectively.

The composite trapezium rule (2.2.6) can be rewritten as

$$I_a^{trap} = \frac{1}{2(N-1)} f_1 + \frac{1}{N-1} \left(\sum_{i=2}^{N-1} f_i \right) + \frac{1}{2(N-1)} f_N.$$

We consider a peak function (5.1.1), therefore we have $f_1 = f_N = 0$ and the approximation obtained via the statistical method can be expressed in terms of the composite trapezium rule approximation

$$I_a^{stat} = \frac{N-1}{N} I_a^{trap} = \kappa I_a^{trap}, \tag{5.6.1}$$

where $0 < \kappa = (N-1)/N < 1$.

Let us impose the condition on the relative integration error (2.1.7) used throughout this chapter, that $E_{rel} \leq \tau$ where we again fix $\tau = 0.25$. This is equivalent to the approximate integral satisfying the inequalities $0.75I \leq I_a^{stat} \leq 1.25I$ which using (5.6.1) become

$$0.75I \leq \kappa I_a^{trap} \leq 1.25I.$$

Since we consider the grid step size $h > \delta$ the term I_a^{trap} is given by (4.4.4) calculated in Section 4.4 of the previous chapter. We also recall that the exact integral is given by (5.2.2). Using these expressions in the above set of inequalities yields

$$\frac{B\delta}{2} \leq \kappa h (B - A\gamma^2 h^2) \leq \frac{5B\delta}{6}.$$

Solving for γ we have

$$\gamma_I^{stat}(\alpha) \leq \gamma(\alpha) \leq \gamma_{II}^{stat}(\alpha),$$

where

$$\gamma_I^{stat}(\alpha) = \frac{1}{2\alpha} \sqrt{1 - \frac{5}{6\alpha\kappa}}, \quad \gamma_{II}^{stat}(\alpha) = \frac{1}{2\alpha} \sqrt{1 - \frac{1}{2\alpha\kappa}}, \tag{5.6.2}$$

The probability of achieving sufficient accuracy when using the statistical method (2.1.4) on

regular grids with grid step size $h > \delta$ is then

$$p^{stat}(\alpha) = \frac{\gamma_{II}^{stat} - \gamma_{I}^{stat}}{\gamma_{max} - \gamma_{min}} = 2(\gamma_{II}^{stat} - \gamma_{I}^{stat}) = \frac{1}{\alpha}\left(\sqrt{1 - \frac{1}{2\alpha\kappa}} - \sqrt{1 - \frac{5}{6\alpha\kappa}}\right). \tag{5.6.3}$$

The equivalent probability when implementing the composite trapezium rule (2.2.6) is given by (5.2.8) and can be written as

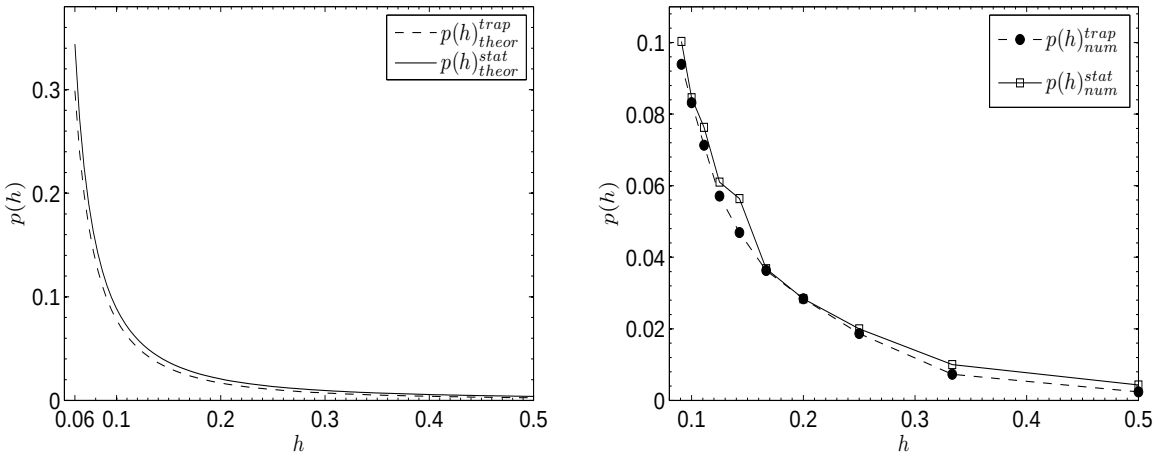
$$p^{trap}(\alpha) = \frac{1}{\alpha}\left(\sqrt{1 - \frac{1}{2\alpha}} - \sqrt{1 - \frac{5}{6\alpha}}\right) \tag{5.6.4}$$

Let us consider the ratio

$$\frac{p^{trap}(\alpha)}{p^{stat}(\alpha)} = \frac{\sqrt{1 - \frac{1}{2\alpha}} - \sqrt{1 - \frac{5}{6\alpha}}}{\sqrt{1 - \frac{1}{2\kappa\alpha}} - \sqrt{1 - \frac{5}{6\kappa\alpha}}},$$

where $\kappa\alpha < \alpha$ as $\kappa < 1$. The function $\rho(\alpha) = \sqrt{1 - \frac{1}{2\alpha}} - \sqrt{1 - \frac{5}{6\alpha}}$ is monotonically decreasing for $\alpha > 1$ (*i.e.* for $h > \delta$). Furthermore, since $\kappa\alpha < \alpha$, the ratio $p^{trap}/p^{stat} < 1$. This means that the probability of achieving an accurate evaluation of the integral via the statistical method is greater than that for the composite trapezium rule.

This result can be seen in Figure 5.12a, where the theoretical probability curves $p(h)_{theor}^{trap}$ and $p(h)_{theor}^{stat}$ have been found from (5.6.4) and (5.6.3) respectively using the substitution $h = \alpha\delta$, for $\alpha > 1$. The curve representing the statistical method is above that corresponding to the composite trapezium rule. Figure 5.12b meanwhile shows the probability curves obtained numerically according to the computation (5.4.1) for the Lorentz distribution (5.4.5) with peak width $\delta = 0.06$. The difference between the numerical curves is small but the probability of achieving an accurate answer via the statistical method remains greater the trapezium rule probability. Therefore for a small number of grid nodes such that the grid step size satisfies the condition $h > \delta$, the statistical method can be considered the more reliable of the two methods to produce an accurate estimate. This is in spite of the fact that the composite trapezium rule has a higher asymptotic convergence rate than the statistical method. The analysis should, however, be conducted over the entire range of ultra-coarse grids (*i.e.* for $h \geq \delta/2$) in order to make a proper comparison between the two methods.



(a) (b)

Figure 5.12: Comparison of the composite trapezium rule (2.2.6) and the statistical method (2.1.4) on ultra-coarse regular grids with $h > \delta$. (a) The probability of achieving sufficient accuracy $E_{rel} \leqslant \tau = 0.25$ is higher for the statistical method, as the curve obtained for the composite trapezium method $p^{trap}_{theor}(h)$ lies below the curve obtained for the statistical method $p^{stat}_{theor}(h)$. (b) The probability curves $p(h)^{trap}_{num}$ and $p(h)^{stat}_{num}$ obtained by direct computation (5.4.1) for the Lorentz distribution (5.4.5) with the peak width $\delta = 0.06$.

5.7 Chapter 5 Conclusions

In this chapter we have introduced the concept of ultra-coarse computational grids where the number N of grid nodes is too small to provide sufficient information about the pest population density distribution, thus the accuracy of an estimate cannot be described deterministically as being within a certain tolerance. Instead, the relative integration error behaves as a random variable of high magnitude and achieving a prescribed level of accuracy becomes a matter of chance. The accuracy on ultra-coarse grids should therefore be assessed by the probability p of achieving a sufficiently small error, rather than considering the error itself.

The problem of ultra-coarse grids arises when the spatial pattern of the population is aggregated, and as such in this chapter we have focused on single peak density functions *i.e.* when the population is concentrated within a single patch. By modelling a peak function as a quadratic, we have obtained a theoretical prediction for the probability $p < 1$ of accurate evaluation on ultra-coarse grids. We have also constructed an estimate of the threshold number N_t of grid nodes required for the probability to become $p = 1$. Numerical experiments showed this estimate

to be a reliable lower bound thus N_t can be considered an estimate of the minimum number of grid nodes required for sufficient accuracy to be guaranteed.

We showed that on ultra-coarse grids, conventional convergence analysis is not conducive to distinguishing between the performance of numerical integration methods. We have therefore proposed that on such grids the accuracy should be described in terms of the probability that the estimate is within a certain tolerance of the true pest abundance, rather than the error itself. This then permits the comparison of methods where the method with the highest probability p would be recommended. It should be mentioned that since our analysis relied on a quadratic approximation of the peak, we could only compare numerical integration methods with an order of convergence $q \leq 2$. In order to compare higher order methods *e.g* the composite Simpson's rule (2.2.7), we would need to use a higher order approximation of the peak. The analysis would then be conducted in the same way.

The study conducted in this chapter has focused on a regular grid of sample units and implementing the composite trapezium rule (2.2.6) to estimate the pest abundance. The statistical rule (2.1.4) often used in ecological applications does not, however, require the sample units to adhere to a specific formation. In the next chapter we investigate the accuracy of pest abundance evaluation when the number of sample units is small, and they are located randomly.

Chapter 6

Evaluating Pest Abundance Using Random Sampling

In the previous chapter we constructed an estimate of the threshold number N_t of grid nodes such that, for any regular grid with $N \geq N_t$, the estimate obtained via the composite trapezium rule (2.2.6) achieves a prescribed level of accuracy. Meanwhile, we developed an approach which allowed us to evaluate the probability $p < 1$ of obtaining an accurate estimate on regular grids with $N < N_t$ nodes. This chapter consists of our work presented in [32] where we extend the results to the statistical counterpart of numerical integration, namely the statistical method (2.1.4). This method is space-implicit, and the location of the grid nodes (sample units) can be arbitrary. A common recommendation is to locate the sample units randomly to avoid biasing [13, 51, 85, 92]. The concern is that a non-random sample may coincide with the pattern of the population. In this chapter we show that whilst such a sampling plan works well when the population is spread across the entire domain, it may not be the best approach when the density distribution is highly aggregated. Furthermore we show that if a random sampling plan is used, there is an optimal number of sample units to provide accurate evaluation.

In Section 6.1 we outline how we formulate the problem. We present the analysis approach in Section 6.2 where we focus on the 1D problem. The approach is then verified for both standard test cases and ecologically significant 1D data in Section 6.3. In Section 6.4 we extend our analysis to the 2D problem. A comparison between a random and regular sampling plan is then

made in Section 6.5 and it is shown that regular plan is more likely to produce accurate results when a highly aggregated density distribution is the focus of the monitoring procedure. The conclusions and discussion of the results of this chapter are provided in Section 6.6.

6.1 Evaluating the Mean Density for Highly Aggregated Distributions

We have already identified in previous chapters that the risk of inaccurate estimation of abundance is highest when the spatial distribution of the population is highly aggregated, *i.e.* the entire pest population is located with in a single sub-domain (patch) of the field and the population density is zero elsewhere. Therefore, we continue the theme of the previous chapter and concentrate on such distributions. As mentioned in Section 2.1 of Chapter 2, the statistical method (2.1.4) is a commonly used means of forming an abundance estimate I_a. We recall that this estimate is defined as

$$I_a = SM(N),$$

where S is the area of the agricultural field and $M(N)$ is the sample mean density, namely

$$M(N) = \frac{1}{N}\sum_{i=1}^{N} f_i. \tag{6.1.1}$$

Within this chapter we consider the agricultural field as the unit square $D = [0,1] \times [0,1]$ thus the area $S = 1$. Consequently, we focus our attention on the estimate (6.1.1).

An obvious, yet important observation made in the previous chapter was that the accuracy of the evaluation of pest abundance depends on how many sample units (grid nodes) are located within the sub-domain D_u of non-zero density. It may be that the small number N of sample units miss the patch entirely thus generating an estimate of the mean density of $M(N) = 0$. Meanwhile, if one or more sample units fall within the patch a better estimate of $M(N) > 0$ is produced. The members of the Newton-Cotes family of numerical integration (*e.g* the composite trapezium rule and the composite Simpson's rule), which have so far largely been the focus of our attention, require the sample units to be regularly spaced. The estimate (6.1.1), however, is

space-implicit and permits a random sampling plan. While an increase in the number N of nodes in a uniform grid implies an increase in the number of nodes located within the patch of non-zero density, this is not the case for a random distribution of nodes. As we already mentioned in the introduction to this chapter, random (or pseudo-random) positioning of sample units is a widespread technique widely recommended in the literature because it allows one to eliminate a bias-related error. For highly aggregated density distributions, however, sampling in a random way means that detecting the patch of pest insects is a matter of chance and the trade-off between the bias error and the approximation error may simply not exist. Below we consider this issue in more detail.

If there are N sample units located randomly across the field of the unit area $S = 1$, then the probability p_m that there are m units within the sub-domain D_u of zero-density is defined by the binomial distribution (*e.g* see [97]):

$$p_m = \frac{N!}{m!(N-m)!}\delta^m(1-\delta)^{N-m}, \tag{6.1.2}$$

where δ is the area of the sub-domain D_u and the following condition holds:

$$\sum_{m=0}^{N} p_m = 1. \tag{6.1.3}$$

Consider the probabilities $p_0, p_1, \ldots, p_N$ defined by the formula (6.1.2). Obviously, the first meaningful case is that we have one sample unit within the sub-domain D_u as a result of a random installation of sample units. Let us compute the probability $\hat{p}(N)$ given by the sum

$$\hat{p}(N) = p_0 + p_1 = (1-\delta)^N + N\delta(1-\delta)^{N-1}, \tag{6.1.4}$$

where p_0 is the probability that we have no sample units within the sub-domain D_u, and p_1 is the probability of having a single sample unit within that sub-domain. It follows from the expression (6.1.4) that the probability $\hat{p}(N)$ is dominant in the sum (6.1.3) when a narrow sub-domain δ is considered. In other words, for small δ the number of sample units m located within the

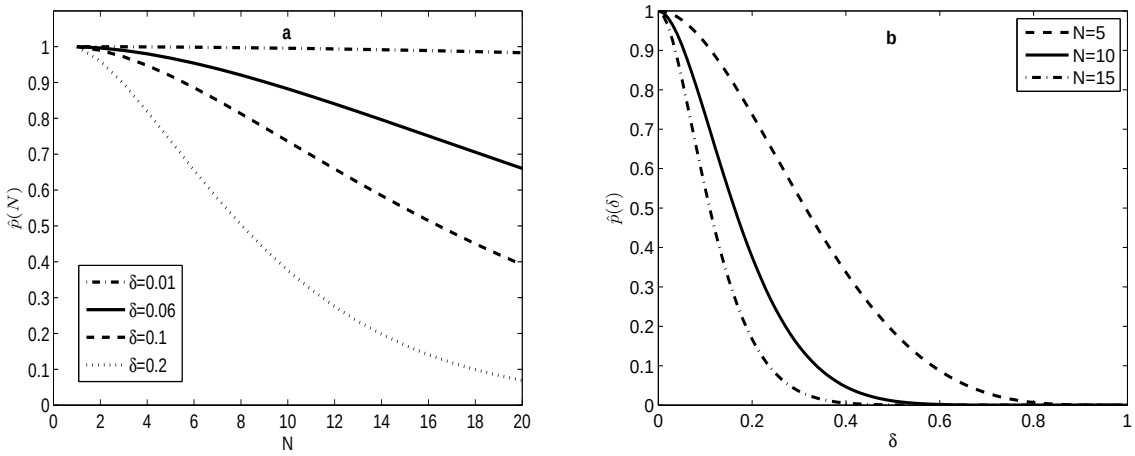


Figure 6.1: The probability $\hat{p}(N)$ given by the formula (6.1.4). (a) The value of the total number of sample units N varies and the area δ of the peak sub-domain is fixed. (b) N is fixed and δ varies.

population of pests is more likely to be either $m = 0$ or $m = 1$ than $m \geqslant 2$. This statement is illustrated in Figure 6.1a where the probability $\hat{p}(N)$ has been plotted for an increasing number N of sample units. The area δ for each graph shown in Fig. 6.1(a) has consequently been fixed as $\delta = 0.01, 0.06, 0.1$, and 0.2, respectively. It can be seen from the figure that for the two smaller areas $\delta = 0.01$ and $\delta = 0.06$ the probability $\hat{p}$ remains greater than 50% for the entire range of N displayed. For the larger choices of δ, that is the patch takes up 10 or 20% of the field, we have $\hat{p}(N) > 0.5$ for $N \leq 16$ and $N \leq 8$ respectively. In Figure 6.1b a similar graph has been plotted, except this time the area of the patch of pests is varied for fixed values of the total number of sample units N. In each case there is a range of sub-domain areas δ for which the condition $\hat{p}(N) > 0.5$ is satisfied. Thus the probability $\hat{p}(N)$ is dominant when the patch is small in comparison to the area of the field, a situation which corresponds to an early stage of biological invasion. The observations made above will be further discussed in Section 6.5.

The conclusion that we make here is that the bias problem is not very important when a single peak distribution is considered, as the most likely scenario is that we lose either all or all but one of the sample units outside the sub-domain D_u of non-zero density. Hence, the question we would like to investigate is whether a random distribution is still better than a grid of equidistant sample units for highly aggregated density of pest insects. Clearly, a complete answer

to this question would require investigation of all cases described by the formula (6.1.2), that is one sample unit within the sub-domain D_u, two sample units within D_u, *etc.* . However, in the present paper we restrict our discussion to the case of a single sample unit placed within the sub-domain D_u even when the total number N of sample units is large. Despite this case not giving us a complete answer to the question of accurate estimation of the pest abundance, its study will allow us to identify and resolve several very important issues related to handling strongly localised density distributions. In particular, it will be revealed that a standard approach in the evaluation of the pest population size should be revisited when a highly aggregated density distribution is considered. In the next section we follow a similar approach to that used in the previous chapter whereby we handle the relative approximation error (2.1.7) as a random variable and compute the probability of obtaining an accurate estimate of the pest population size.

6.2 Probability Analysis: 1D Case

Let us first consider the 1D problem; the 2D problem is considered later in the chapter in Section 6.4. We recall from the previous chapter that when the pest population is highly aggregated, *i.e.* the population is restricted to a single patch, the population density function $f(x)$ is a peak function as shown in Figure 6.2a. The spatial heterogeneity of a one-peak distribution $f(x)$ can be thought of as being constructed of two components - a peak region and a tail region. Clearly, the peak is a dominant feature of the density $f(x)$ and it provides the main contribution to the mean density $M(N)$. Hence for the sake of our further discussion we again use the simplified version of the distribution $f(x)$ from Chapters 4 and 5. In the peak region we consider $f(x)$ as a quadratic function of the width δ. Elsewhere we set the population density to be zero, therefore the tail region is essentially 'cut off'. We thus have

$$f(x) \approx \begin{cases} Q(x) = B - A(x - x^*)^2, & x \in [x_I, x_{II}], \\ 0, & \text{otherwise}, \end{cases} \tag{6.2.1}$$

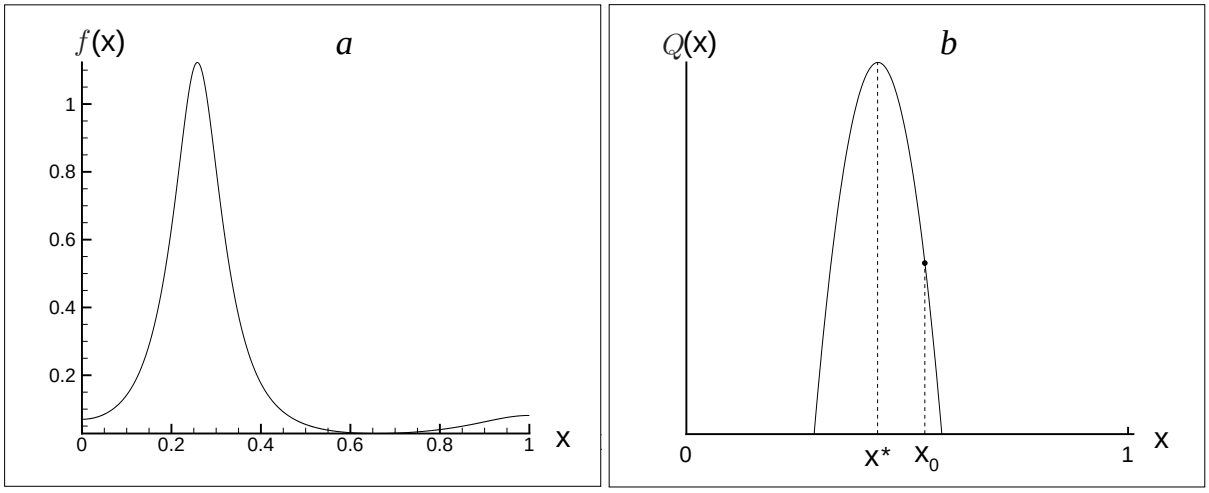


Figure 6.2: (a) An example of high aggregation density distribution in a 1D ecological system. (b) Approximation of the one peak density distribution by a quadratic function. One sample unit at the location x_0 lies within the peak sub-domain.

where x^* is the location of the maximum of the peak, $A = -\frac{1}{2}\frac{d^2 f(x^*)}{dx^2}$, $B = f(x^*)$. The peak width is $\delta = 2\sqrt{B/A}$ and the roots are $x_I = x^* - \delta/2$, $x_{II} = x^* + \delta/2$. While such an approximation of the peak introduces an error of the order δ^3, the numerical study conducted in the previous chapter showed that our conclusions drawn from considering a quadratic peak can be extended to peak functions of arbitrary shape. Hence we consider the approximation (6.2.1) to be reliable and use it in our further analysis.

As we already mentioned in the previous section, we intend to consider the limiting case of one sample unit being located in the peak sub-domain $D_u = [x_I, x_{II}]$, where we have non-zero density $f(x)$. In other words, if the sample unit numeration is $i = 1, 2, \ldots, N$, then we have $f_{i_0} \neq 0$ for fixed $i = i_0$ and $f_i = 0$ for any $i \neq i_0$. Let us re-define the index i_0 as $i_0 = 0$ for the sake of convenience. The sample unit location $x_{i_0} \equiv x_0$, where the density $f_{i_0} \equiv f_0 \neq 0$, is defined as

$$x_0 = x^* + \gamma\frac{\delta}{2},$$

where the parameter $\gamma \in [0, 1]$ as we only consider the right half of the peak sub-domain, because of the obvious symmetry of the peak (see Figure 6.2b).

The density f_0 is computed as

$$f_0 \approx Q(x_0) = B - A(x_0 - x^*)^2 = B(1 - \gamma^2), \quad \gamma \in [0, 1], \tag{6.2.2}$$

where we take into account that $\dfrac{\delta}{2} = \sqrt{\dfrac{B}{A}}$.

Let us now define $\bar{M}$ to be the true mean density and we require the approximation error to be small when $\bar{M}$ is approximated by the sample mean density $M(N)$:

$$E_{rel} = \frac{|\bar{M} - M(N)|}{\bar{M}} < \tau, \tag{6.2.3}$$

where τ is a prescribed tolerance, $0 < \tau < 1$. Hence the mean value $M(N)$ computed when we use N sample units should be within the range

$$(1 - \tau)\bar{M} \leqslant M(N) \leqslant (1 + \tau)\bar{M}. \tag{6.2.4}$$

We have already mentioned in the previous section that in ecological applications the exact location of the peak sub-domain cannot be predicted for a high aggregation density distribution. We now make an assumption that is crucial for our further discussion. Since the location of the sub-domain $D_u = [x_I, x_{II}]$ is not known, and any location is as likely to occur as another, the location x^* of the peak maximum can be considered to be a uniformly distributed random variable. Under the requirement that only one sample unit lies in the domain $D_u = [x_I, x_{II}]$ the assumption about uniformly random location of the peak sub-domain can be re-formulated in terms of the location of the point x_0. Namely, we fix the point x^* and then consider $\gamma \in [0, 1]$ as a uniformly distributed random variable in order to randomise the location of x_0.

We now solve the inequalities (6.2.4) in order to see whether any location x_0 of a sample unit within the peak sub-domain can provide the desirable accuracy (6.2.4). From (6.2.2), the sample mean density (6.1.1) is

$$M(N) = \frac{f_0}{N} = \frac{B(1 - \gamma^2)}{N}. \tag{6.2.5}$$

Meanwhile, by considering the peak as the quadratic density distribution $Q(x)$ defined by (6.2.1), the true mean density $\bar{M}$ is calculated as

$$\bar{M} = \int_0^1 Q(x)\,dx = \frac{2}{3}B\delta. \tag{6.2.6}$$

Substituting the above expressions for $M(N)$ and $\bar{M}$ in (6.2.4) we arrive at

$$(1-\tau)B\hat{\delta} \leqslant \frac{B(1-\gamma)^2}{N} \leqslant (1+\tau)B\hat{\delta}, \tag{6.2.7}$$

where $\hat{\delta} = \dfrac{2}{3}\delta$.

Consider the inequality

$$\frac{B(1-\gamma)^2)}{N} \leqslant (1+\tau)B\hat{\delta}. \tag{6.2.8}$$

We have

$$1-\gamma^2 \leqslant (1+\tau)N\hat{\delta} \quad \Rightarrow \quad \gamma \geqslant \gamma_I = \sqrt{1-(1+\tau)N\hat{\delta}}, \tag{6.2.9}$$

where we have to choose a positive root, as $\gamma \in [0,1]$. It follows from the solution of (6.2.9) that the number N of sample units should be restricted as

$$N \leqslant N^* = \frac{1}{\hat{\delta}(1+\tau)}. \tag{6.2.10}$$

If the restriction (6.2.10) breaks, then the inequality (6.2.8) holds for any $\gamma \in [0,1]$.

Let us now solve

$$\frac{B(1-\gamma^2}{N} \geqslant (1-\tau)B\hat{\delta}. \tag{6.2.11}$$

Similar computation results in

$$1-\gamma^2 \geqslant (1-\tau)N\hat{\delta} \quad \Rightarrow \quad \gamma \leqslant \gamma_{II} = \sqrt{1-(1-\tau)N\hat{\delta}}, \tag{6.2.12}$$

where the number of sample units is restricted as

$$N \leqslant N^{**} = \frac{1}{\hat{\delta}(1-\tau)}. \tag{6.2.13}$$

If the restriction (6.2.13) does not hold then the accuracy (6.2.4) is never achieved. In other words, if we have a large number $N > N^{**}$ of sample units, but they are randomly distributed over the entire domain $D = [0, 1]$, so that only one sample units is positioned within the peak sub-domain, then the accuracy of the sample mean density evaluation will always be poor, as $M(N)$ will not be within the range (6.2.4).

Let us note that the number $N^* < N^{**}$ for any peak width δ and tolerance τ. Hence we have to consider the following cases:

Case 1: $N \leqslant N^*$.

For any number of sample units that is smaller than N^*, the admissible range of the sample unit location x_0 where we can guarantee prescribed accuracy (6.2.4) is given by $\gamma_I \leqslant \gamma \leqslant \gamma_{II}$. In other words, we require that $x_{0_I} \leqslant x_0 \leqslant x_{0_{II}}$, where $x_{0_I} = x^* + \gamma_I \frac{\delta}{2}$ and $x_{0_{II}} = x^* + \gamma_{II} \frac{\delta}{2}$.

The same result holds when we consider a sample unit location at the left-hand side of the peak, $x_0 = x^* + \gamma \frac{\delta}{2}$, where $\gamma \in [-1, 0]$. We therefore have two subintervals $[-\gamma_{II}, -\gamma_I]$ and $[\gamma_I, \gamma_{II}]$ where the sample unit location within each of those subintervals will give us the accuracy required by (6.2.4). As the length of the entire interval is $\gamma \in [-1, 1]$ and a sample unit is randomly placed at any point of the peak sub-domain, then the probability $p(N)$ of obtaining a value $M(N)$ that meets the condition (6.2.4) is given by

$$p(N) = \frac{2(\gamma_{II} - \gamma_I)}{\gamma_{max} - \gamma_{min}}, \tag{6.2.14}$$

where $\gamma_{min} = -1$, $\gamma_{max} = 1$ and we multiply the range $\gamma_{II} - \gamma_I$ by 2 as we now consider the left-hand side and the right-hand side of the peak. Substituting γ_I and γ_{II} from (6.2.9) and (6.2.12) respectively in the equation (6.2.14) we arrive at

$$p_I(N) = \sqrt{1 - (1-\tau)N\hat{\delta}} - \sqrt{1 - (1+\tau)N\hat{\delta}}. \tag{6.2.15}$$

The probability $p_I(N)$ for $N < N^*$ is shown as branch I of the graph in Figure 6.3a. It can be seen from the graph as well as from the analytical expression (6.2.15) obtained for the probability $p(N)$ that the maximum value $p_{max} = p_{max}(\tau) = \sqrt{1 - \frac{1-\tau}{1+\tau}}$ of the probability $p(N)$ is achieved when $N = N^*$. It is important to note here that the maximum probability is always $p_{max} < 1$. At the same time the probability $p_{max}(\tau)$ predictably grows when we make the tolerance τ bigger, that is $p_{max} \to 1$ as $\tau \to 1$.

Case 2: $N^* < N \leqslant N^{**}$.

For any number of sample units $N > N^*$ the inequality (6.2.8) always holds. Hence we only have the restriction (6.2.13) and the admissible range of γ becomes $\gamma \in [0, \gamma_{II}]$. The probability of obtaining an accurate estimate (6.2.4) is given by

$$p_{II}(N) = \sqrt{1 - (1-\tau)N\hat{\delta}}. \tag{6.2.16}$$

The probability $p_{II}(N)$ defined for the number of sample units $N^* < N \leqslant N^{**}$ is shown as curve II in Figure 6.3a.

Case 3: $N > N^{**}$.

In the case that N is sufficiently large, the probability of the event that the error is within the range (6.2.4) is $p_{III}(N) = 0$ as we cannot meet the condition (6.2.11) (see branch III in Figure 6.3a).

Let us note again that the results above are entirely based on the assumption that only one sample unit belongs to the peak sub-domain. However, as explained in Section 6.1 if a random distribution of sample units over the domain is applied, then we cannot guarantee that more than one sample unit will be located in the peak sub-domain even when the total number N of sample units is large. The branches II and III of the curve $p(N)$ in Figure 6.3a will exist as long as we have a single sample unit within the peak sub-domain, no matter how large the total number N of sample units becomes. Hence if we want to keep a random distribution of sample units, our recommendation would be to restrict the number of sample units as $N \approx N^*$ as this

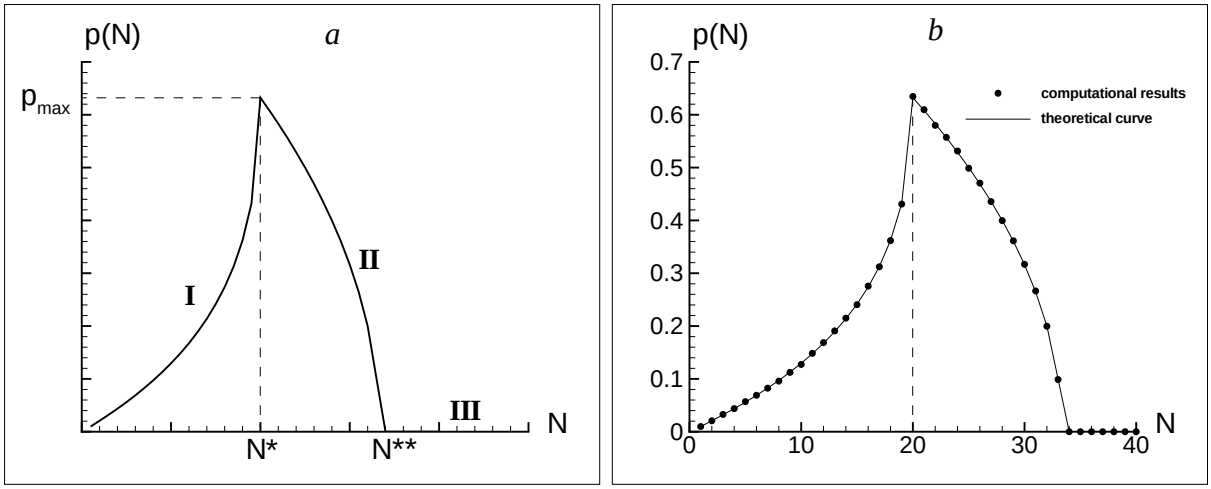


Figure 6.3: The probability of obtaining an accurate estimate $M(N)$ of the true mean density $\bar{M}$ in the case when a single sample unit is located within the peak sub-domain D_u where the peak is formed by the quadratic function (6.2.1). (a) The theoretical curve. (b) Comparison of the theoretical curve and computational results. The probability is computed for the peak width $\delta = 0.06$ and the tolerance $\tau = 0.25$.

number of sample units provides the greatest chance of obtaining an accurate estimate $M(N)$ of the mean density.

6.3 Numerical Verification: 1D Case

In this section the probability $p(N)$ will be obtained in several test cases by direct computation and compared with a theoretical curve obtained for the quadratic function (6.2.1). The first test case is to confirm that our theoretical results derived for a quadratic function are correct. Let us fix the peak width δ, the tolerance τ and the location x^* of the peak maximum. We then consider the location x_0 of a sample unit as a random variable that is uniformly distributed over the interval $[x^*, x^* + \delta/2]$. In our computations we provide $n_r = 100,000$ realisations of the random variable x_0 for the fixed total number N of sample units, compute $M(N)$ and check the condition (6.2.4) for each realisation of x_0. The probability $p(N)_{num}$ of accurate evaluation (6.2.4) of the mean density is then computed as

$$p(N)_{num} = \frac{\hat{n_r}}{n_r}, \tag{6.3.1}$$

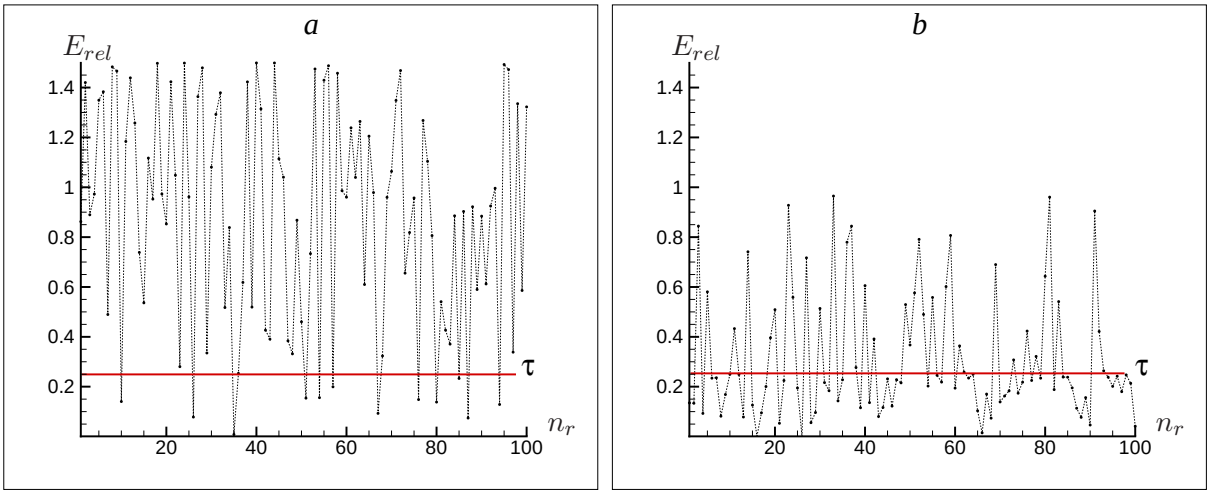


Figure 6.4: Computation of the error (6.2.3) for a random peak location. The peak is given by a quadratic function (6.2.1) of width $\delta = 0.06$. The tolerance in the formula (6.2.3) is set as $\tau = 0.25$. The location of the peak maximum is randomly generated 100 times and the error value is computed for each realisation $n_r = 1, 2, \ldots, 100$ of the peak location. (a) The number of sample units is $N = 10$. The probability of getting the error $E_{rel} \leq \tau$ is low and most of the error values are beyond the required range. (b) The number of sample units is $N = N^* = 20$. The probability of getting an accurate result $E_{rel} \leq \tau$ achieves its maximum when $N = N^*$ and most of the error values are within the required range.

where $\hat{n}_r$ is the number of realisations for which the condition (6.2.4) holds. We then increase the number of sample units by one and repeat computation (6.3.1) for $N+1$ total sample units. We stop increasing the number N, when the number N_L of sample units becomes so large that the condition (6.2.13) breaks and we have $p(N_L) = 0$.

The probability $p(N)_{num}$ of the accurate evaluation of the mean density is shown in Figure 6.3b for the peak width $\delta = 0.06$ and the tolerance $\tau = 0.25$. We start from $N_1 = 1$ sample unit and then increase the number of sample units until $N_L = 40$. It can be seen from the figure that all values of the probability $p(N)_{num}$, $N = 1, \ldots, 40$, computed by direct evaluation (6.3.1) belong to the theoretical curve $p(N)$.

The probability (6.2.15)–(6.2.16) is further illustrated for a quadratic function (6.2.1) in Figure 6.4. Again, we assume that only one sample unit is located within the peak sub-domain D_u and the location of that sample unit is random with respect to the position of the peak maximum. We make 100 random realisations n_r of the sample unit location x_0 and compute

the error (6.2.3) for each realisation when the total number of sample units is fixed as $N = 10$. The integration error (6.2.3) computed for the function (6.2.1) is shown in Figure 6.4a. The theoretical value of the probability $p(N)$ is $p(N) = 0.12$ when $N = 10$. This is well illustrated by the results shown in Figure 6.4a where approximately 10% of the error values belong to the range (6.2.4). Clearly, the value $p \approx 0.1$ must tend to the theoretical probability $p(N) = 0.12$ when we increase the number of realisations n_r (cf. Figure 6.3b). Consider now $N = N^*$, where the optimal number N^* of sample units is defined from (6.2.10) as $N^* = 20$ for the peak width $\delta = 0.06$ and the tolerance $\tau = 0.25$. The probability of an accurate estimate is $p(N^*) = 0.66$ and this result is confirmed by the error distribution shown in Figure 6.4b where most of the error values (every 2 out of 3) lie within the required range.

Let us now consider several standard test cases where the highly aggregated density distributions (peak functions) are different from the quadratic function (6.2.1). The test cases below are taken from those considered in the previous chapter, where they were investigated for the composite trapezium integration rule (2.2.6). For convenience we restate the equations defining the test cases below, meanwhile the reader is referred to Figure 5.7 in Section 5.4 of the previous chapter for plots of the peaks. Our first test case is to consider the cubic function

$$f(x) = \begin{cases} A(x - x^* + (\delta/3))(x - x^* - (2\delta/3))^2, & x \in [x^* - \delta/3, x^* + 2\delta/3], \\ 0, & otherwise, \end{cases} \tag{6.3.2}$$

where the peak width is $\delta = 0.06$ and $A = 30,000$ (see Figure 5.7b). We apply the same computational procedure (6.3.1) as for the quadratic function discussed above to obtain the probability $p(N)_{comput}$ for various N. The probability graph for the function (6.3.2) is shown in Figure 6.5a. Obviously, the probability graph obtained for a cubic function cannot coincide with the theoretical curve (6.2.15)–(6.2.16) (a dashed line in Figure 6.5a). In particular, the critical number $N^* = 24$ is now different from the theoretical value $N^* = 20$ computed from (6.2.10) for $\tau = 0.25$ and $\delta = 0.06$ Nevertheless, it can be seen from the figure that the theoretical curve obtained for a quadratic function is a good approximation of the probability $p(N)$ computed for a cubic function (6.3.2).

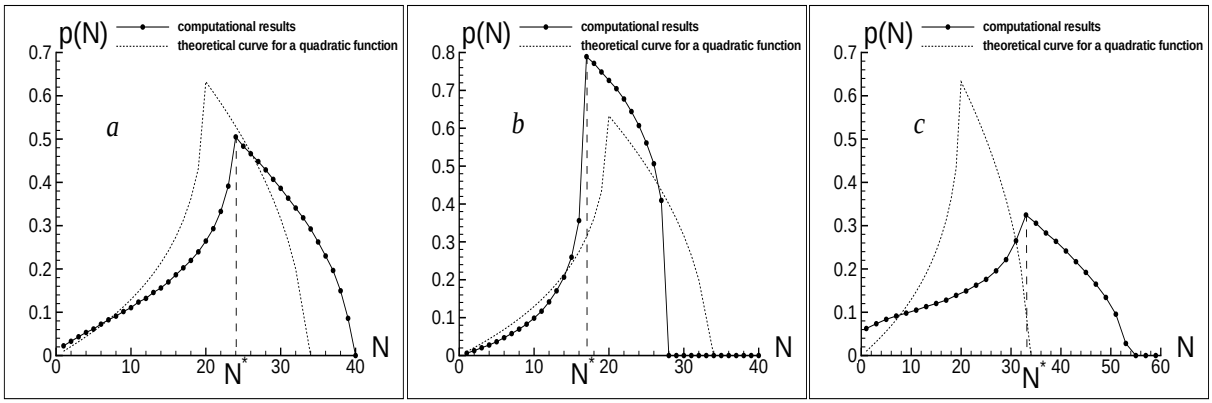


Figure 6.5: Numerical test cases. The probability (6.3.1) (solid line) of achieving sufficient accuracy (6.2.4) computed for (a) a cubic function (6.3.2) (b) a quartic function (6.3.3) and (c) a normal distribution (6.3.4). For the functions (a)-(c) the peak width is chosen as $\delta = 0.06$. For each function (a)–(c) the probability (6.3.1) is compared with the theoretical curve obtained for a quadratic function (dashed line).

The next test case is a quartic function defined as

$$f(x) = \begin{cases} A\left(\left(\frac{\delta}{2}\right)^4 - (x - x^*)^4\right), & x \in [x^* - \frac{\delta}{2}, x^* + \frac{\delta}{2}], \\ 0, & otherwise, \end{cases} \tag{6.3.3}$$

where $A = 1,200,000$ and the peak width is again taken as $\delta = 0.06$ (see Figure 5.7a). The probability graph for the function (6.3.3) is shown in Figure 6.5b. It can be seen from the figure that the graph has a similar shape to the theoretical graph for the quadratic function, but the critical number $N^* = 17$ is again different from the number $N^* = 20$ obtained from the analysis of a quadratic distribution.

Finally, we consider a normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x - x^*)^2}{\sigma^2}\right), \tag{6.3.4}$$

which gives us an example of a peak function that is different from zero everywhere in the domain $x \in [0, 1]$. The peak width is defined by the parameter σ as $\delta = 6\sigma$ and we again consider $\delta = 0.06$ (see Figure 5.7c). The probability graph computed from (6.3.1) for the function (6.3.4) is shown

in Figure 6.5c. It can be seen from the figure that the critical number $N^* = 33$ strongly differs from the number of sample units obtained for a quadratic function with the same peak width. However, the shape of the graph is still similar to the theoretical curve (a dashed line in the figure) and the critical value N^* of sample units provides the maximum probability $p(N^*)$. The presence of the critical value N^* in each graph in Figure 6.5 remains the most essential feature of our analysis.

We now test our estimate (6.2.10) of the critical number N^* of sample units for some ecologically meaningful test cases. We consider two of the peak functions from the previous chapter which were generated from the 1D Roseznweig-Macarthur model (3.2.7–3.2.8), namely the peak functions generated for the diffusion coefficient $d = 10^{-4}$ and $d = 10^{-5}$. We refer to the density distributions as $f_1(x)$ and $f_2(x)$ respectively and they are shown in Figure 6.6.

We recall from Section 4.1 of Chapter 4 that the diffusion coefficient d is a controlling parameter of the spatial heterogeneity of the population density function [55, 75], where a simple estimate of the peak width δ is

$$\delta = \omega\sqrt{d},$$

and numerical experiments [75, 76] have shown that typically $\omega \approx 25$. Using this estimate of the peak width, the critical number (6.2.10) of sample units can be evaluated as

$$N^* = \frac{1}{\hat{\delta}(1+\tau)} \approx \frac{C}{\sqrt{d}}, \tag{6.3.5}$$

where the coefficient $C(\tau) = \dfrac{3}{2\omega(1+\tau)}$.

Consider the density distribution $f_1(x)$ shown in Figure 6.6a. Since the diffusion coefficient is $d = 10^{-4}$, the estimate (6.3.5) gives us the number $N^* \approx 5$ for the tolerance $\tau = 0.25$. The probability graph obtained by direct computation is shown in Figure 6.6b, where the number $N^* = 7$ taken from the graph is in good agreement with the theoretical estimate.

Let us now evaluate the number N^* in the case that we have the density distribution $f_2(x)$ shown in Figure 6.6c. For the diffusion coefficient $d = 10^{-5}$ we have $N^* \approx 16$. The direct computation gives us $N^* = 23$ (see Figure 6.6d) which is greater than the theoretical value of

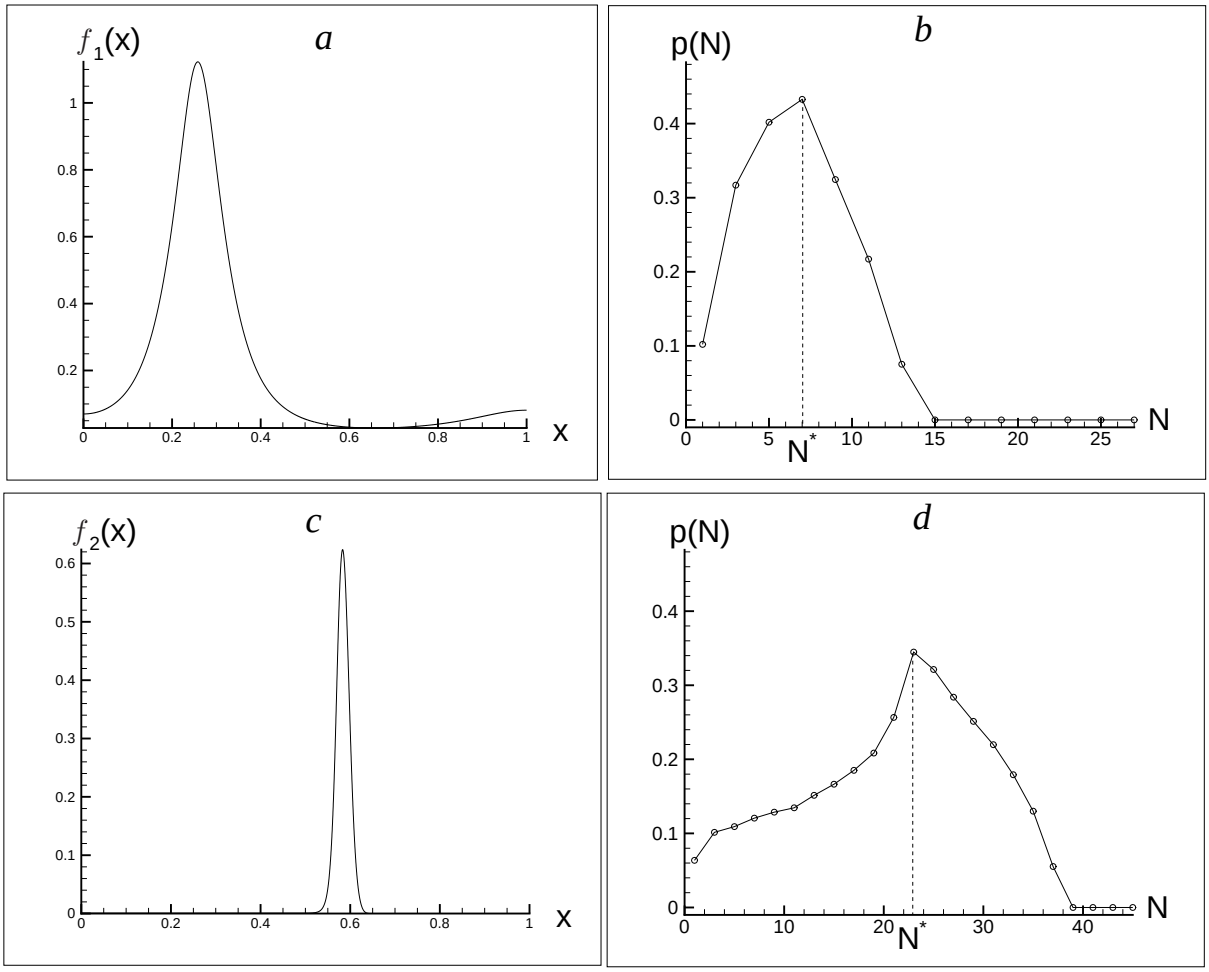


Figure 6.6: Ecological test cases. (a) The spatial distribution $f_1(x)$ of the pest population density $f(x)$ for the diffusivity $d = 10^{-4}$. Other parameters along with the initial and boundary conditions used to generate this distribution are given in the caption of Figure 3.3 in Chapter 3. (b) The probability (6.3.1) of an estimate achieving the required accuracy (6.2.4) computed for the density distribution $f_1(x)$ under the condition that a single sample unit is located within the peak sub-domain. (c) The pest population density $f_2(x)$ obtained for the diffusivity $d = 10^{-5}$; the other parameters are as for the distribution $f_1(x)$. (d) The probability (6.3.1) computed for the density distribution $f_2(x)$.

N^* obtained for a quadratic function. However, the results obtained for a quadratic function are still true for an ecologically meaningful density distribution. Namely, if sample units are randomly located over the monitored area and we cannot guarantee that more than one sample unit will fall within the peak sub-domain, then the best chance to get an accurate estimate of the mean density is when we use the number of sample units close to the critical number N^*.

Further increase in the number of sample units reduces our chance for an accurate estimate.

6.4 Probability Analysis: 2D Case

In this section we expand the results obtained from the analysis of the $1D$ problem to the more realistic $2D$ problem. We again focus on a highly aggregated density distribution where there is a single peak in the domain. The domain of interest is now represented by the unit square $D = [0,1] \times [0,1]$. As in the analysis for the $1D$ problem, we consider the peak as a quadratic function, and ignore the tail region by setting the population density function $f(x,y)$ to be zero outside of the peak domain. That is we consider the population density function to be as follows:

$$f(x,y) \approx \begin{cases} Q(x,y) = B - A\left((x-x^*)^2 + (y-y^*)^2\right), & (x,y) \in D_u, \\ 0, & \text{otherwise}, \end{cases} \tag{6.4.1}$$

where (x^*, y^*) is the location of the peak maximum. The peak sub-domain D_u is a circular disc of radius R, where $R = \sqrt{B/A}$, and is centred at (x^*, y^*). This region can be seen in Figure 6.7a. We define the peak width as $\delta = 2R$.

Performing similar analysis to the $1D$ case, we arrive at the following probability function $p(N)$ (The details of the calculation are given in Appendix C):

$$p(N) = \begin{cases} \sqrt{1 - \frac{N(1-\tau)\pi R^2}{2}} - \sqrt{1 - \frac{N(1+\tau)\pi R^2}{2}}, & N \le N^*(\Delta), \\ \sqrt{1 - \frac{N(1-\tau)\pi R^2}{2}}, & N^*(\Delta) < N \le N^{**}(\Delta), \\ 0 & N > N^{**}(\Delta), \end{cases} \tag{6.4.2}$$

The probability $p(N)$ is shown in Figure 6.7b. It can be seen from the that the shape of the graph $p(N)$ computed for a 2D quadratic distribution is identical to the probability graph generated for a 1D quadratic function (the dashed line in Figure 6.7b; see also Figure 6.3), except the critical number of sample units N^*_{2D} is different from the number of sample units N^* obtained in the $1D$ case. This is a consequence of the definition (C.0.2) where the function $f(x,y)$ is effectively a function of a single variable, $f(x,y) \equiv f(r)$. Thus both probability functions can

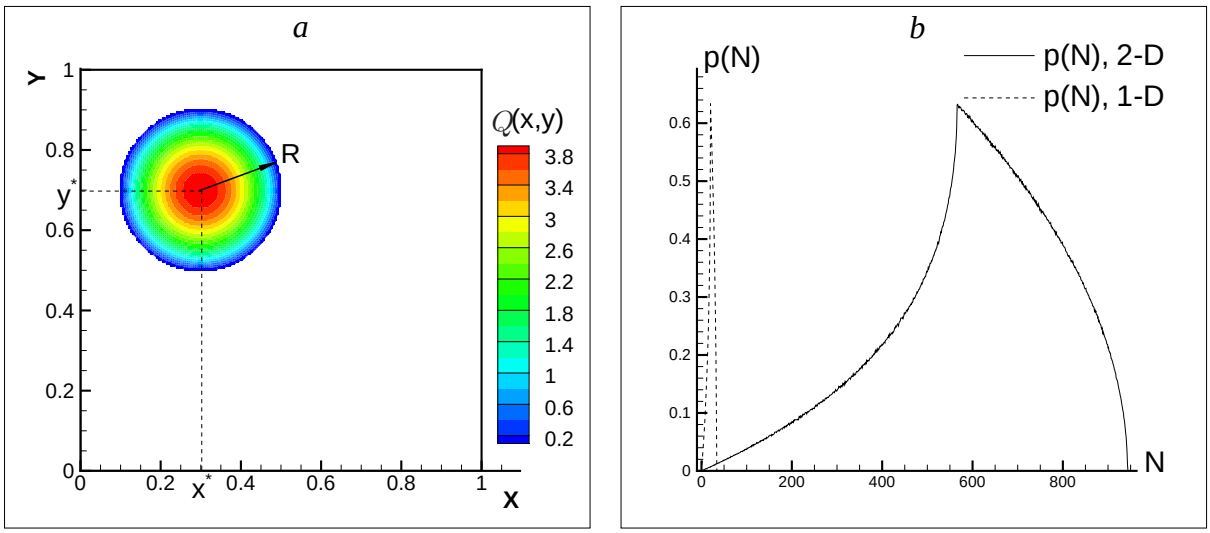


Figure 6.7: (a) Quadratic distribution (6.4.1) where the location of the peak maximum is chosen to be $(x_0, y_0) = (0.3, 0.7)$ and the peak width is $\delta = 0.4$. (b) The probability curve $p(N)$ obtained for a $2D$ quadratic peak with the peak width $\delta = 0.06$. The probability curve for a $1D$ quadratic peak with the same peak width is shown as a dashed line.

be written in the following form:

$$p(N) = \begin{cases} \sqrt{1 - N(1-\tau)\Delta} - \sqrt{1 - N(1+\tau)\Delta}, & N \le N^*(\Delta), \\ \sqrt{1 - N(1-\tau)\Delta}, & N^*(\Delta) < N \le N^{**}(\Delta), \\ 0 & N > N^{**}(\Delta), \end{cases} \tag{6.4.3}$$

where we now use a uniform notation $N^*(\Delta)$ and $N^{**}(\Delta)$ for the critical number of sample units and the definition of the parameter Δ varies according to the number of dimensions in which we are working.

In the $1D$ case we have

$$\Delta_{1D} = 2\delta_{1D}/3 \tag{6.4.4}$$

and in the $2D$ case

$$\Delta_{2D} = \pi R^2/2 = \pi\delta_{2D}^2/8, \tag{6.4.5}$$

where δ_{1D} and δ_{2D} are the peak widths for the dimension denoted by the subscript.

It is clear that the theoretical probability curves will be the same when $\Delta_{1D} = \Delta_{2D}$. We

can write the $1D$ peak width δ_{1D} in terms of δ_{2D} as

$$\delta_{1D} = \frac{3\pi\delta_{2D}^2}{16}. \tag{6.4.6}$$

Hence the probability of achieving an error (6.2.3) within a prescribed tolerance τ for a $2D$ peak can be calculated using the $1D$ theory. The critical number N_{2D}^* in (C.0.6) can then be computed as

$$N_{2D}^* = \frac{3}{(1+\tau)2\delta_{1D}}. \tag{6.4.7}$$

For the $1D$ quadratic peak (6.2.1) with $\delta_{1D} = 0.06$ and the tolerance $\tau = 0.25$ we have that the number $N_{2D}^* = 566$ when a $2D$ counterpart with the same peak width $\delta_{2D} = 0.06$ is considered (see Figure 6.7b). On the other hand, if we want to obtain the same critical number $N_{2D}^* = 20$ as in the $1D$ case, we have to set the peak width $\delta_{2D} = \sqrt{\frac{16\delta_{1D}}{3\pi}} = 0.3192$.

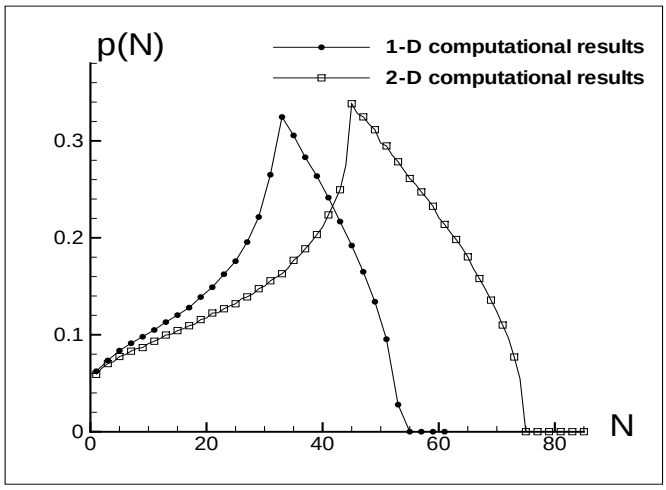
At the same time it is worth noting that the relation (6.4.6) between $1D$ and $2D$ problems is accurate for a quadratic function only. For a spatial distribution different from a quadratic function the equation (6.4.6) gives us an approximate estimate of the peak width and therefore an approximate value of the number N_{2D}^* of sample units. Consider for example, a 2D counterpart of the normal distribution (6.3.4). The function $f(x, y)$ is given by

$$f(x,y) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\;\exp\left(-\frac{(x-\hat{x}_1)^2+(y-\hat{y}_1)^2}{2\sigma^2}\right), \tag{6.4.8}$$

where the peak width is $\delta = 6\sigma$. Let us set $\delta = 0.06$ for a $1D$ distribution (6.3.4). It has been discussed above that the estimates (6.4.6) and (6.4.7) give us the peak width $\delta_{2D} = 0.3192$ for which the critical number N_{2D}^* in the $2D$ case should be the same as in the $1D$ case. The probability graphs for a $1D$ distribution (6.3.4) with $\delta = 0.06$ and a $2D$ distribution (6.4.8) with $\delta = 0.3192$ are shown in Figure 6.8, where we expect the two graphs to be the same. However, it can be seen from Figure 6.8 that the probability graph obtained for the normal distribution (6.4.8) is shifted from the graph $p(N)$ obtained for the $1D$ normal distribution (6.3.4).

We conclude this section by considering a simple yet ecologically meaningful example of a highly aggregated density distribution in $2D$. Namely, we focus our attention on the pest pop-

Figure 6.8: Probability curves for the 1D and 2D normal distributions (6.3.4) and (6.4.8). We set $\Delta_{1D} = \Delta_{2D} = 0.04$ thus the peaks have 'equivalent' peak widths namely $\delta_{1D} = 0.06$, $\delta_{2D} = 0.3192$.



ulation density distribution supplied by the authors of [71] which was obtained from numerical solution of the 2D Rosenzweig-MacArthur model (3.2.3–3.2.4) as shown in Figure 6.9. The details of the parameters used to generate the density distribution are given in the figure caption. We consider a distribution $f_1(x, y)$ where the peak is wide, that is it takes up a large portion of the entire domain (see Figure 6.9a). We also look at a second distribution $f_2(x, y)$ for which the peak is restricted to a much smaller sub-domain (see Figure 6.9c). The distribution $f_2(x, y)$ was formed by placing the peak from $f_1(x, y)$ on a domain ten times larger in each direction. This is essentially the same as considering a peak with width δ ten times smaller than the original distribution.

In each case, the peak sub-domain is defined as the region in which the pest population density is such that $f(x, y) \geq 10^{-4}$. The region outside of D_u, i.e. the tail region is then ignored. Let $(\tilde{x}, \tilde{y})$ denote the points which belong to the peak sub-domain D_u. The width of the peak in the x and y directions, δ_x and δ_y, are calculated as $\delta_x = \max(\tilde{x}) - \min(\tilde{x})$, $\delta_y = \max(\tilde{y}) - \min(\tilde{y})$. We then define the peak width δ to be $\delta = \min(\delta_x, \delta_y)$. The distributions $f_1(x, y)$ and $f_2(x, y)$ were found to have peak widths of $\delta = 0.848541$ and $\delta = 0.0848541$ respectively.

An estimate of the point (x^*, y^*) is given by $x^* \approx (\max(\tilde{x}) + \min(\tilde{x}))/2$, $y^* \approx (\max(\tilde{y}) + \min(\tilde{y}))/2$. The random location (x_0, y_0) of a sample unit within the peak sub-domain is generated as

$$x_0 = r\cos\theta + x^*, \quad y_0 = r\sin\theta + y^*,$$

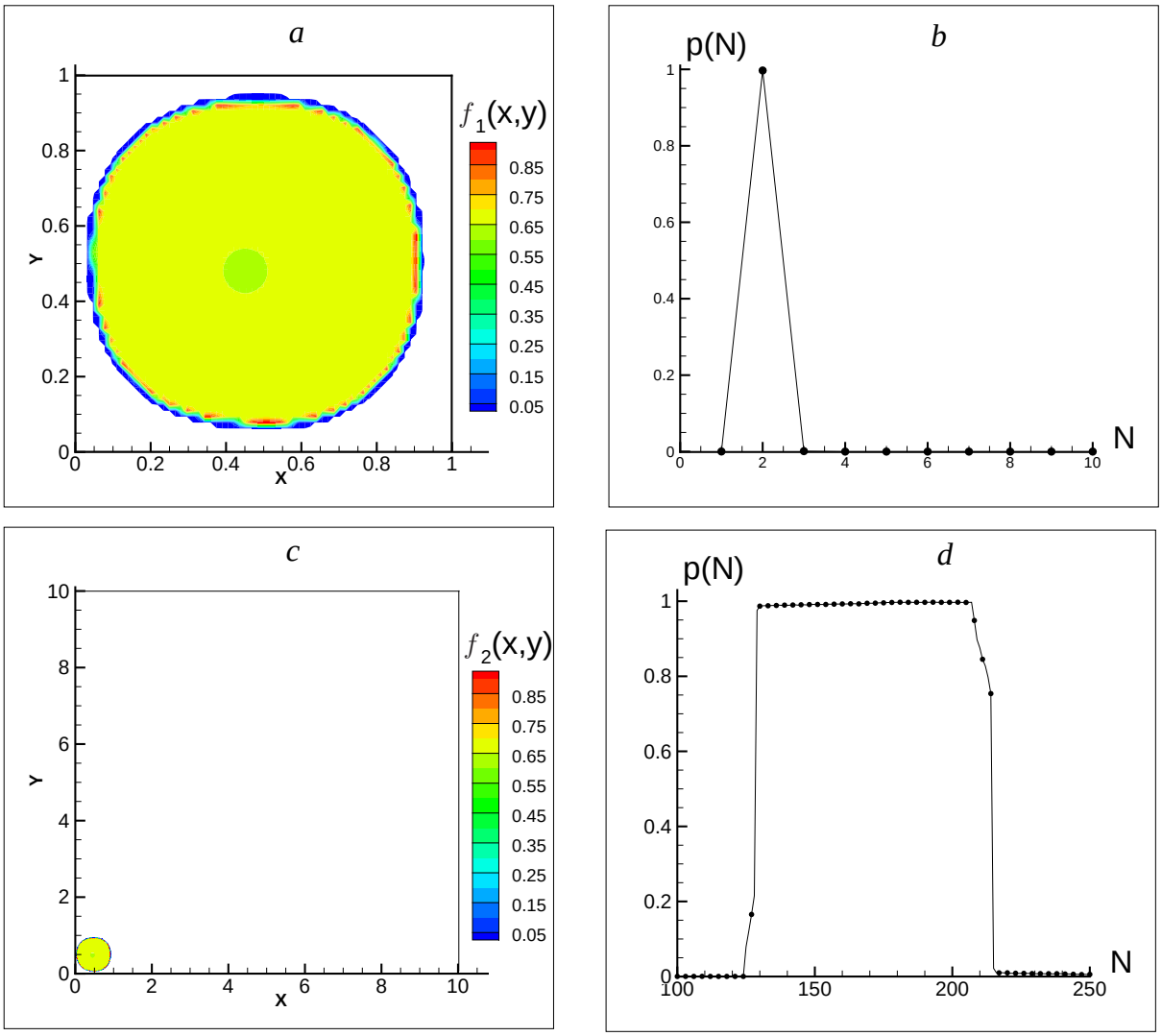


Figure 6.9: $2D$ ecological test cases. (a) The spatial distribution $f_1(x, y)$ of the pest population density generated by the system of equations (3.2.3–3.2.4) on the unit square. The test case is as considered in [71] where the parameters used in the system are: $t = 800$, $d = 10^{-6}$, $m = 0.5$, $\Lambda = 0.5$, $\beta = 3$, and $\chi = 0.28$. The initial conditions are $f(x, y, 0) = 1$ for $0.42 < x < 0.53$ and $0.45 < y < 0.55$, and $f(x, y, 0) = 0$ otherwise, and $g(x, y, 0) = 1$ for $0.42 < x < 0.48$ and $0.45 < y < 0.51$, and $g(x, y, 0) = 0$ otherwise. (b) The probability (6.3.1) of an an estimate achieving sufficient accuracy (6.2.4) computed for the density distribution $f_1(x, y)$ under the condition that a single sample unit is located within the peak sub-domain. (c) The pest population density $f_2(x, y)$ considered in the domain $D : x \in [0, 10], y \in [0, 10]$. (d) The probability (6.3.1) computed for the density distribution $f_2(x, y)$.

where $r \in [0, R]$ and $\theta \in [0, 2\pi]$ are uniformly distributed random variables. As before, we consider $n_r = 100,000$ realisations of the sample unit location.

We assume there is only one sample unit in the peak sub-domain D_u. In accordance with the procedure previously outlined, we now calculate $p(N)_{num}$ for the population distributions $f_1(x, y)$ and $f_2(x, y)$. The results are shown in Figure 6.9b and Figure 6.9d respectively. It can be seen from the figure that the probability curves obtained for density distributions $f_1(x, y)$ and $f_2(x, y)$ differ from the graphs $p(N)$ computed for $1D$ ecologically meaningful density distributions (*cf.* Figure 6.6). The difference can be explained by the fact that the functions $f_1(x, y)$ and $f_2(x, y)$ present the simplest case of a peak function when a highly aggregated density distribution is almost constant in the peak sub-domain. Hence the value of the mean density does not depend on a random location of the point (x_0, y_0) and the value of N in the expression (6.1.1) can be considered as a scaling coefficient. Nevertheless, this simple test case confirms our conclusions made in Section 6.1 that random installation of a large number of sample units does not result in an accurate estimate of the mean population density, as we have $p(N) = 0$ for a large number N of sample units in both cases (see Figure 6.9b and Figure 6.9d).

6.5 Comparison of Sampling Plans: Random vs Regular Grid

The analysis made in the previous sections for the 1D and 2D cases revealed that there exists a critical number N^* of sample units for which the probability of an accurate answer achieves its maximum value. The estimate of N^*, however, does not take into account the whole complexity of the problem when we have to deal with a random sampling plan. First of all, let us note that the probability $p(N)$ should be scaled by the probability $p_1(N)$ of the event that exactly one sample unit is installed within the peak sub-domain. According to the formula (6.1.2)the probability $p_1(N)$ is calculated as

$$p_1(N) = N\delta(1 - \delta)^{N-1}. \tag{6.5.1}$$

The probability $\tilde{p_1}(N)$ of having the error with the given range (6.2.4) when a single sample unit falls into the peak sub-domain is then given by $\tilde{p_1}(N) = p_1(N)p(N)$. The functions $p_1(N)$ and $\tilde{p_1}(N)$ are shown in Figures 6.10a and b, respectively. It can be seen from Figure 6.10b that the resulting probability $\tilde{p}_1(N)$ is much smaller than $p(N)$. For a quadratic function with the

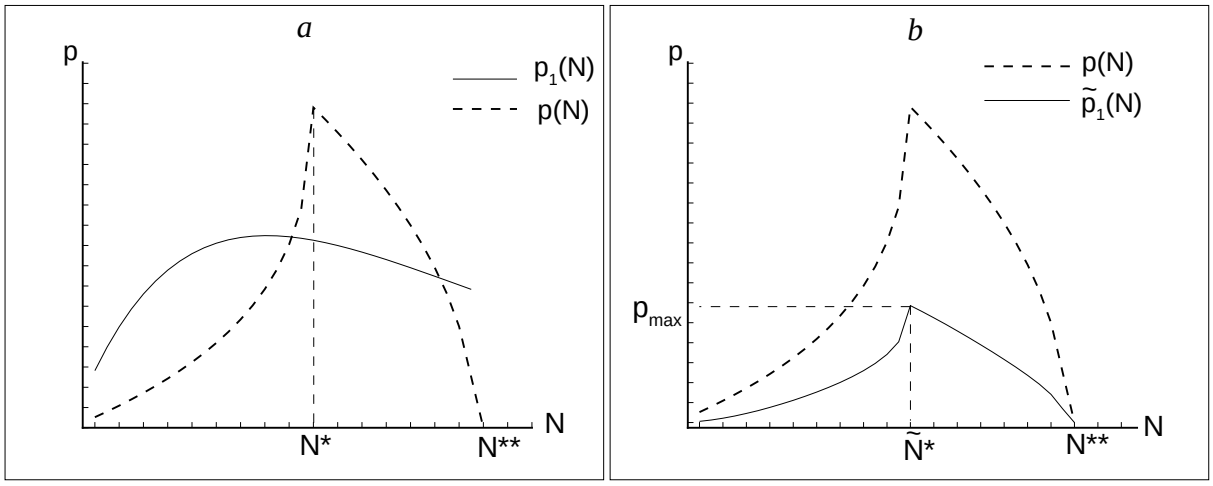


Figure 6.10: (a) The probability $p_1(N)$ of having a single sample unit located within the peak sub-domain when the sampling plan across the unit interval is random. (b) The resulting probability $\tilde{p}_1(N)$ of the estimate being with the given range (6.2.4) when a single sample unit falls into the peak sub-domain. The probability $p(N)$ is shown as a dashed line in both figures.

peak width $\delta = 0.06$ and the tolerance $\tau = 0.25$ the critical number $\tilde{N}^*$ for which the resulting probability $\tilde{p}(\tilde{N}^*)$ has its maximum is $\tilde{N}^* = 20$ and the probability is $\tilde{p}(\tilde{N}^*) \approx 0.23$.

On the other hand, it was established in the previous chapter that a uniform grid of equidistant sample units over the unit interval $[0, 1]$ provides the desirable accuracy of the mean density evaluation with the probability $p(N) = 1$ when the distance between sample units is $h = \alpha_t \delta$, where δ is the peak width in the one-dimensional problem and the parameter α_t depends on the tolerance τ only. In other words, if we use a regular sampling plan then the desired accuracy (6.2.3) will be achieved for any number $N > N_t = 1 + 1/\alpha\delta$ of sample units. For a quadratic function with the peak width $\delta = 0.06$ and the tolerance $\tau = 0.25$ the threshold number providing the error below the given tolerance has been computed as $N_t = 21$ (i.e., the distance between sample units is $h = 0.05$). Any equidistant grid of sample units with the number $N > 21$ will then give us an accurate estimate of the pest abundance.

The discussion above can be summarised as follows. For the given value $\delta = 0.06$ of the peak width, on a grid of randomly distributed sample units the probability of getting an accurate estimate achieves its maximum $\tilde{p}(N) \approx 0.23$ for $N = 20$. For larger values of N the probability then becomes smaller. However, on a regular grid of about the same number $N = 21$ of sample

units, an accurate result is obtained with the probability being equal to one. Furthermore, the probability remains one for larger values of N. Therefore, for $N \sim 20$ (or larger), on a regular grid the estimate of the population size is not a stochastic variable anymore, while on a random grid of the same size it still essentially stochastic and the probability of obtaining an accurate result remains relatively low. Thus a grid of equidistant sample units is more favorable compared to a random sampling plan.

Clearly, the argument above is not complete, as for a random sampling plan there is the possibility that more than one sample unit falls inside the peak sub-domain. Generally speaking, a similar computation for the probability $\tilde{p}_2, \tilde{p}_3, \ldots, \tilde{p}_N$ would be needed in order to be able to conclude about the efficiency of a random distribution of sample units, Here $\tilde{p}_m$, $m = 2, 3, \ldots, N$, is the probability of having the error with the given range (6.2.4) when m sample units fall into the peak sub-domain. The total probability $\tilde{P}(N) = \sum_{m=1}^{N} \tilde{p}_m$ should then be computed to determine if $\tilde{P}(N) \approx 1$. We do not make this computation, however, let us note that in Section 6.1 it was shown that, in the case $\delta \ll 1$ (*i.e.* narrow peaks which is the main focus of this chapter), the probability of the event that more than one sample unit fall into the peak area is significantly less than 50% and gets smaller with a decrease in δ. Thus, having restricted our analysis to the case of high population aggregation (i.e. $\delta \ll 1$), corrections to the equation (6.5.1) are expected to be small.

Based on our present results, therefore, we believe that locating sample units at the nodes of an equidistant Cartesian grid is a better option than using a random sampling plan, when evaluating the pest abundance for a highly aggregated density distribution. It should be noted that a regular sampling plan is comparable in terms of effort to implement. A random sampling plan eliminates the bias error, but, as we already discussed in Section 6.1, the bias problem does not exist when a highly aggregated density distribution (a single peak) is considered. Finally, another important argument in favour of a grid of equidistant sample units, is that such a grid may be better suited for a multi-patch distribution. If we have a multi-patch density function where all patches have approximately the same width (i.e., a collection of several peaks scattered over the monitored area), then installing a grid with the number of sample units $N \geq N_t$ will

detect all the patches, while we cannot guarantee the same result when the sampling plan is random.

6.6 Chapter 6 Conclusions

We have revisited the problem discussed in the previous chapter of estimating pest abundance when the spatial pattern of the population is highly aggregated. Since the number N of sample units is limited, achieving a sufficiently accurate estimate is a matter of chance. In this chapter we have focused on the statistical method (2.1.4), which depends on the sample mean density (6.1.1). This method is space-implicit and thus permits the sample units (grid nodes) to be located randomly. As such, we have extended the theoretical approach developed earlier to handle a random sampling plan.

By using the assumption that a single sample unit falls within the patch of non-zero density, we have obtained theoretical predictions of the probability of achieving a sufficiently accurate estimate for both the 1D and 2D problems. It has also been shown that there is a critical number N^* of sample units for which the probability achieves its maximum value. Using a number $N > N^*$ of sample units may lead to a reduced chance of achieving a sufficiently accurate estimate. Comparisons with results from the previous chapter indicate that for the same number N^* of sample units, a regular sampling plan is more reliable than a random distribution when the population is highly aggregated.

Our analysis has been conducted under the assumption that the entire population is localised to a single patch. Whilst this kind of distribution has ecological significance as it corresponds to the early stage of biological invasion [90], it is somewhat of an extreme case. The technique we have presented, however, could be extended to handle the case where there are multiple patches of pests across the field which is often observed in reality [5]. Assuming the patches are on average the same size, it is then a matter of multiplying the probability p of achieving an accurate estimate for a single patch, by the total number of patches. When using a random sampling plan, this problem is complex as the number of patches is unknown. However, if a regular sampling plan is used instead, and the total number of sample units N is sufficient to detect one patch, then it will be sufficient to detect them all. Of course it may be that the

number N is too large to be practical. Further analysis is needed to determine whether in this situation a smaller number of randomly distributed sample units would in fact be more suitable.

Chapter 7

Evaluating Pest Abundance in the Presence of Noise

So far we have assumed that the population density data used to formulate an estimate of pest abundance is precise. In other words, we have considered the measurements of population density obtained by sampling to accurately reflect the true pest population density at each sample unit location. In practice this assumption often does not hold, for example when trapping is used as the sampling technique, and instead the density data are affected by a measurement error. This chapter consists of the work presented in [33] where we investigate the impact of the random component of this measurement error. Namely, we study how noise in the density data affects the accuracy of a pest abundance estimate. The study is restricted to the 1D case and the agricultural field is represented by the unit interval $D = [0, 1]$.

In Section 7.1 it is discussed how noise in the density data arises and how it propagates to uncertainty in the abundance estimate and its accuracy. We present a means of quantifying this uncertainty in Section 7.2 and give a credible interval for the relative approximation error when noise is present. We apply our theory to both standard mathematical and ecological significant test cases in Sections 7.3 and 7.4 respectively and compare the accuracy of estimates formed from noisy data with those formed from exact data. The chapter conclusions are presented in Section 7.5.

7.1 The Uncertainty of an Abundance Estimate from Noisy Data

We consider a trapping procedure conducted in an agricultural field to collect data on the pest population. Trapping is a sampling technique widely used for pest insect abundance evaluation [2, 12, 36, 45, 56]. Traps are installed in the field, exposed for a prescribed length of time, then the traps are emptied and the pests are counted. It was discussed in the introduction that it is a widespread situation in ecological monitoring that financial, ecological and other restrictions require the number of traps installed in an agricultural field to be relatively small [56, 60]. For example, the number of traps installed over an agricultural field in the United Kingdom very rarely exceed a few dozen [12, 36, 45], where a linear size of the field is typically of the order of a few hundred meters.

Since the traps are exposed for a certain time period, and the spatial pattern of the pest population changes over time, the question arises of whether the population density distribution can really be recovered from such a sampling procedure. We make a key assumption that the exposure time of the traps is selected such that within this time frame, the population density distribution does not change significantly. That is, we assume that the area over which the population has spread is small in comparison to the size of the characteristic spatial heterogeneity, *i.e.* cluster size. The exposure time therefore must vary depending on the target species. For instance, in a study of ground beetles (*Carabidae*), traps were emptied every 2 days [45]. The dispersal of one particular species *P.melanarius* has been estimated to be less than 55m over 30 days [102]. Thus, during the time the traps are exposed, an insect will be less than 3.7m from the position it held at the beginning of the traps exposure. This distance is an order of magnitude smaller than the estimated typical cluster size of greater than 30m [45]. Our assumption that the population density distribution has not altered significantly would therefore be reasonable.

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 [19, 84]. The conversion of trap counts to density is, however, by no means straightforward. Let us

assume the traps to be passive traps, *i.e.* no attractant such as bait or a pheromone is used to draw the pest insects towards the trap. The movement of the insects can be considered to be random, thus, in the finite time for which the trap is exposed the trap count will only be a fraction of the true number of insects local to the trap. Consequently, rather than providing an absolute count, the traps provide a count that is relative to the activity and the density of the insects. Converting relative abundance estimates into the true population density local to the trap has been the focus of much research [47, 73, 84]. The result of this conversion is of course not the precise pest population density local to the trap but rather a measurement, and measurements are subject to a *measurement error*.

Let us denote the measured pest density at the trap location x_i by $\tilde{f}_i$, and f_i represents the corresponding exact pest density. The relationship between the measured pest density $\tilde{f}_i$ and the the true pest density f_i is then $f_i = \tilde{f}_i + \varepsilon_{m_i}$ where ε_{m_i} is the measurement error. A measurement error is considered to consist of two components: a random component, and a systematic component [10]. In other words, the measurement error ε_{m_i} can be expressed as $\varepsilon_{m_i} = \varepsilon_{r_i} + \varepsilon_{s_i}$ where ε_{r_i} and ε_{s_i} represent the random and systematic error respectively. The random error is the result of noise in the data and thus any $\varepsilon_{r_i}, i = 1, \ldots N$ can be either positive or negative with equal probability. The systematic error on the other hand is caused by some source of bias and therefore every $\varepsilon_{s_i}, i = 1, \ldots N$ is consistently either positive or negative. We focus on the impact of noise in the data and as such we ignore the systematic contribution to the measurement error. That is to say we redefine the relationship between the measured quantity $\tilde{f}_i$ and the true value f_i as

$$f_i = \tilde{f}_i + \varepsilon_{r_i}.$$

The random error component ε_{r_i} of a measured pest density $\tilde{f}_i$ given in the above equation is in essence a realisation of a random variable. We consider the true pest density f_i to be some unknown constant. Since $\tilde{f}_i$ is the sum of an unknown constant f_i and a realisation ε_{r_i} of a random variable, it can in turn also be considered a realisation of another random variable. There is thus an *uncertainty* associated with a measured pest density $\tilde{f}_i$.

We have previously ignored any error in the density data and instead focused solely on the

error imparted by the numerical integration procedure itself. That is, we have considered the following abundance estimate I_a constructed from the numerical integration of exact data:

$$I_a = \sum_{i=1}^{N} w_i f_i, \tag{7.1.1}$$

and studied the relative error E_{rel} of this estimate which we recall is defined as

$$E_{rel} = \frac{|I - I_a|}{I}. \tag{7.1.2}$$

The exact pest abundance $I > 0$ would be given by

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

if the pest population density function $f(x)$ were known almost everywhere across the domain of the field which is here taken to be the unit interval. We again require that the estimate to be sufficiently accurate, namely that the relative error satisfies the condition (2.5.1)

$$E_{rel} \leq \tau,$$

where an accuracy tolerance $\tau \in (0.2, 0.5)$ is considered acceptable.

Now, we take the random error present in the density data into account. Applying a method of numerical integration (7.1.1) to the measured pest densities $\tilde{f}_i, i = 1, \ldots, N$ gives the following estimate of the pest abundance:

$$\tilde{I} = \sum_{i=1}^{N} w_i \tilde{f}_i. \tag{7.1.3}$$

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

$$\tilde{E}_{rel} = \frac{|I - \tilde{I}|}{I}. \tag{7.1.4}$$

The focus of our investigation is to establish how the introduction of noise to the data set $\{f_i\}$

affects the accuracy of the abundance estimation, that is, to determine how the error $\tilde{E}_{rel}$ differs from E_{rel}. We return to considering a regular sampling plan (2.2.2) across the unit interval thus the trap locations are defined as

$$x_i = \frac{i-1}{N-1}, \quad i = 1, \ldots, N. \tag{7.1.5}$$

Our attention is restricted to the composite trapezium rule (2.2.6) in this chapter, and additional numerical integration methods are then studied in the next chapter.

In our work we simulate the uncertainty in the density data 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 [40])

$$p(\tilde{f}_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{f}_i - \mu_i}{\sigma_i}\right)^2\right\}, \tag{7.1.6}$$

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,

$$u(\tilde{f}_i) = \sigma_i. \tag{7.1.7}$$

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

$$\tilde{f}_i \in [f_i - z\sigma_i, f_i + z\sigma_i] \tag{7.1.8}$$

with probability

$$P(z) = \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right), \tag{7.1.9}$$

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

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

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,

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

where we refer to $\nu \in (0, 1)$ as the *measurement tolerance*. Equating the interval above to that given by (7.1.8) gives the following relation between the standard deviation σ_i and the measurement tolerance ν:

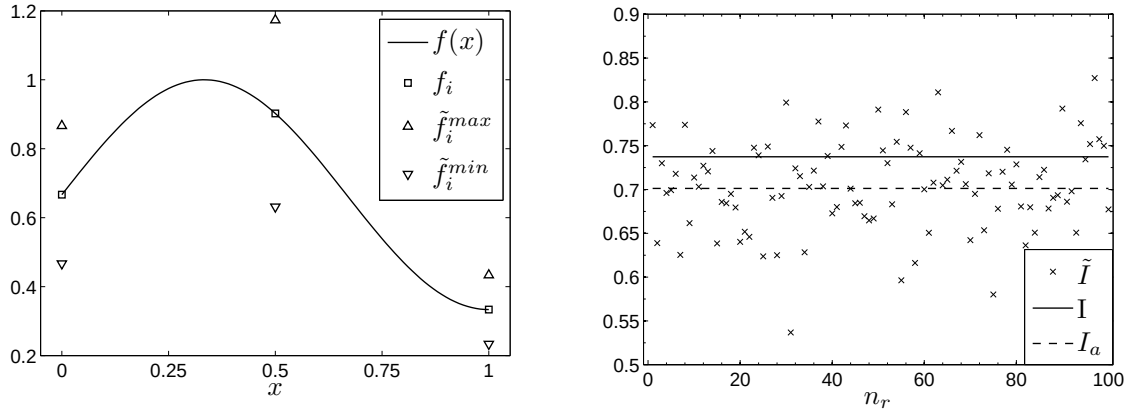
$$\sigma_i = \frac{\nu f_i}{z}. \tag{7.1.10}$$

It is worth noting here that our means of introducing noise into the data 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 larger 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) [98]. However, the measurement tolerance ν 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 Figure 7.1a. The ecologically relevant (*i.e.* non-negative) function $f(x)$ has been defined as

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

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



(a) (b)

Figure 7.1: 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 (7.1.8) with probability $P(z)$ as defined by (7.1.9). 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 $\nu = 0.3$ 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).

on the interval $[0, 1]$. In the example shown in Figure 7.1a 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 composite trapezium rule (2.2.6) is $I_a = 0.701184$, while the error is $E_{rel} = 0.049115$ which is much lower than the required tolerance τ.

We then consider the perturbed data as shown in Figure 7.1a. 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

$$\tilde{f}_i = f_i + \gamma\sigma_i, \tag{7.1.11}$$

where γ is a random variable taken from the standard normal distribution, and σ_i is defined according to (7.1.10). The measurement tolerance is set as $\nu = 0.3$. We also fix $z = 3$, therefore, the probability that each realisation $\tilde{f}_i$ lies within the range (7.1.8) 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 composite trapezium rule (2.2.6) 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 Figure 7.1b. 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 $\tilde{I}$. In the next section we quantify the uncertainty in the accuracy $\tilde{E}_{rel}$ of the approximated pest abundance induced by noise in the density data.

7.2 Quantifying the Evaluation Accuracy in the Presence of Noise

In this section we establish a credible interval for $\tilde{E}_{rel}$. Consider for each $i = 1, \ldots, N$ the density measurement $\tilde{f}_i$ to be a realisation of the normally distributed random variable F_i with density distribution (7.1.6). It can be seen from (7.1.3) 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

$$\tilde{I}_F = \sum_{i=1}^{N} w_i F_i. \tag{7.2.1}$$

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, assuming there is no correlation between trap data, the standard deviation $\sigma_{\tilde{I}}$ which quantifies the uncertainty $u(\tilde{I})$ associated with an estimate is

$$u(\tilde{I}) = \sigma_{\tilde{I}} = \sqrt{\sum_{i=1}^{N} w_i^2 u^2(\tilde{f}_i)}, \tag{7.2.2}$$

(e.g., see [23]).

We now determine the probability density function of the random variable $\tilde{E}_{rel}$. For the sake

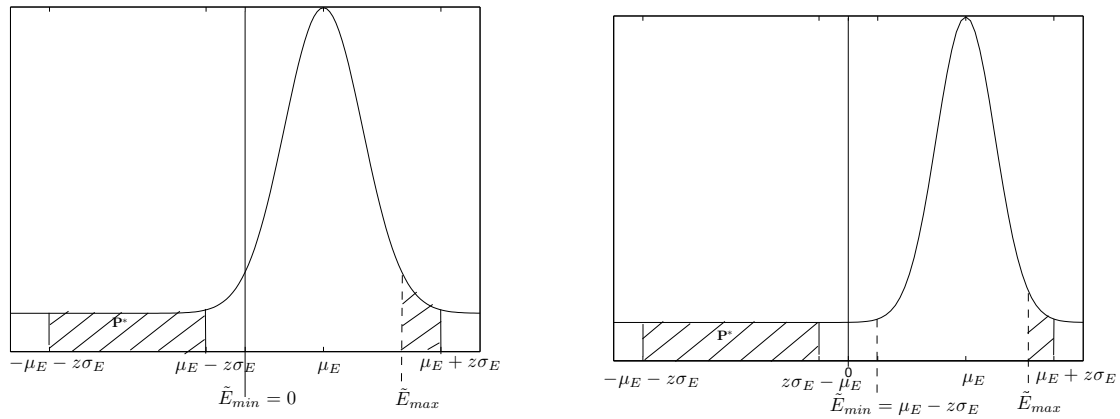


Figure 7.2: The probability density function of the quantity E as described by (7.2.4). Reflecting the negative contributions in the y-axis yields the folded normal distribution of $\tilde{E}_{rel}$. The upper and lower limits of the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ to which $\tilde{E}_{rel}$ 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 $|I - I_a| \leq z\sigma_{\tilde{I}}$ and (b) when $|I - I_a| > z\sigma_{\tilde{I}}$. See Appendix D for the details of how $\tilde{E}_{min}$ and $\tilde{E}_{max}$ are calculated.

of convenience let us first consider the following auxiliary quantity

$$E = \frac{I - \tilde{I}}{I}. \tag{7.2.3}$$

Since E is a linear function of $\tilde{I}$ which is a realisation of a normally distributed random variable, E 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

$$p(E) = \frac{1}{\sigma_E\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{E - \mu_E}{\sigma_E}\right)^2\right\}, \tag{7.2.4}$$

and the quantity E belongs to the range

$$E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E] \tag{7.2.5}$$

with probability $P(z)$ given by (7.1.9). Examples of the probability density function of E are

shown in Figure 7.2.

We have

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

and $\tilde{E}_{rel}$ becomes a realisation of a random variable with a folded normal distribution (*e.g.*, see [52]). The probability density function of $\tilde{E}_{rel}$ is then formed from that of E by reflecting the the negative contributions in the y-axis and is given by the following expression

$$\begin{aligned} p(\tilde{E}_{rel}) &= \frac{1}{\sigma_E\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{\tilde{E}_{rel}-\mu_E}{\sigma_E}\right)^2\right\}+\exp\left\{-\frac{1}{2}\left(\frac{\tilde{E}_{rel}+\mu_E}{\sigma_E}\right)^2\right\}\right] \qquad (7.2.6)\\ &= \frac{I}{\sigma_{\tilde{I}}\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{I(1-\tilde{E}_{rel})-I_a}{\sigma_{\tilde{I}}}\right)^2\right\}+\exp\left\{-\frac{1}{2}\left(\frac{I(1+\tilde{E}_{rel})-I_a}{\sigma_{\tilde{I}}}\right)^2\right\}\right], \end{aligned}$$

where the mean value is

$$\mu_{\tilde{E}_{rel}} = \left(1-\frac{I_a}{I}\right)\left[1-2\Phi\left(\frac{I_a-I}{\sigma_{\tilde{I}}}\right)\right]+\frac{\sigma_{\tilde{I}}}{I}\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{1}{2}\left(\frac{I_a-I}{\sigma_{\tilde{I}}}\right)^2\right\}, \qquad (7.2.7)$$

and the standard deviation is

$$\sigma_{\tilde{E}_{rel}} = \sqrt{\mu_E^2+\sigma_E^2-\mu_{\tilde{E}_{rel}}^2}. \qquad (7.2.8)$$

We now seek a range $[\tilde{E}_{min}, \tilde{E}_{max}]$ to which $\tilde{E}_{rel}$ belongs with probability $P(z)$. It can be seen from (7.2.4) (see also Figure 7.2) that the range of the error $\tilde{E}_{rel}$ 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 (7.2.4) 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 (7.2.6) contains part but not all of the range (7.2.5). That occurs when the distance between the true pest abundance I and the estimate I_a formed from exact data satisfies the condition $|I-I_a| \leq z\sigma_{\tilde{I}}$ (see Figure 7.2a). 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 (7.2.5) in the case that μ_E is positive (see Figure 7.2b) or, when μ_E is negative, is entirely inclusive.

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

$$\tilde{E}_{min} = \begin{cases} 0 & \text{for } E_{rel} \leq z\frac{\sigma_{\tilde{I}}}{I}, \\ \\ E_{rel} - \dfrac{z\sigma_{\tilde{I}}}{I} & \text{for } E_{rel} > z\frac{\sigma_{\tilde{I}}}{I}, \end{cases} \tag{7.2.9}$$

and the upper limit is given by

$$\tilde{E}_{max} = \begin{cases} E_{rel} + \dfrac{\sigma_{\tilde{I}}}{I}\Phi^{-1}\left[2\Phi(z) - \Phi\left(z + \dfrac{2IE_{rel}}{\sigma_{\tilde{I}}}\right)\right], \text{ for } E_{rel} \leq z\dfrac{\sigma_{\tilde{I}}}{I}, \\ \\ E_{rel} + \dfrac{\sigma_{\tilde{I}}}{I}\Phi^{-1}\left[\Phi(z) - \Phi\left(z - \dfrac{2IE_{rel}}{\sigma_{\tilde{I}}}\right) - \Phi\left(z + \dfrac{2IE_{rel}}{\sigma_{\tilde{I}}}\right) + 1\right], \text{ for } E_{rel} > z\dfrac{\sigma_{\tilde{I}}}{I}, \end{cases} \tag{7.2.10}$$

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 [14]), where $\alpha = 100P(z)$, for the error $\tilde{E}_{rel}$ 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 (7.2.9) and (7.2.10) 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.

7.3 Assessing the Impact of Noise: Standard Test Cases

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 previous chapters 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 we take the unit interval $[0, 1]$ to represent the agricultural field. Therefore, the traps are located according to (7.1.5). The exact pest population densities are then given by $f_i \equiv f(x_i), i = 1, \ldots, N$.

We have already discussed how the accuracy of an abundance estimate depends on the spatial heterogeneity of the population density function. Let us begin with a test case which is easy to handle whereby the population is spread over the entire domain. We define the function $f(x)$ as:

$$f(x) = \exp(x) \sin(3\pi x)^2 + \cos(\pi x)^2. \tag{7.3.1}$$

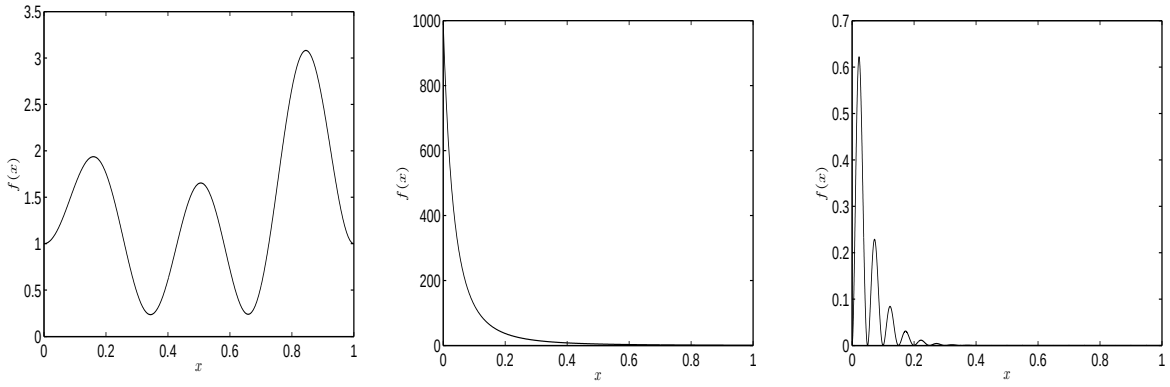
The density function consists of three wide peaks as can be seen in Figure 7.3a. We fix the number N of traps and generate measured values of the pest density by perturbing each exact pest density f_i a total of $n_r = 100,000$ times according to the transformation (7.1.11). We therefore have n_r sets of measured values $\{\tilde{f}_i\}$. For each set of data an estimate of the pest abundance is obtained by implementing the compound trapezium rule (2.2.6) and the relative error is then calculated. To confirm that these $n_r = 100,000$ estimates of $\tilde{E}_{rel}$ are indeed realisations of a random variable with a folded normal distribution with mean $\mu_{\tilde{E}_{rel}}$ and standard deviation $\sigma_{\tilde{E}_{rel}}$ we calculate the sample mean

$$\bar{\mu}_{\tilde{E}_{rel}} = \frac{1}{n_r} \sum_{i=1}^{n_r} \tilde{E}_{rel_i}, \tag{7.3.2}$$

and the sample standard deviation

$$s_{\tilde{E}_{rel}} = \sqrt{\frac{1}{n_r - 1} \sum_{i=1}^{n_r} \left(\tilde{E}_{rel_i} - \bar{\mu}_{\tilde{E}_{rel}} \right)^2}, \tag{7.3.3}$$

and make a comparison with the theoretical quantities given by (7.2.7) and (7.2.8) respectively.



(a) (b) (c)

Figure 7.3: The test cases to validate the evaluation error $\tilde{E}_{rel}$. (a), (b), and (c) are defined over the unit interval $[0, 1]$ by the functions given in (7.3.1), (7.3.5), and (7.3.6) respectively.

N	$\mu_{\tilde{E}_{rel}}$	$\bar{\mu}_{\tilde{E}_{rel}}$	$\frac{\lvert\mu_{\tilde{E}_{rel}} - \bar{\mu}_{\tilde{E}_{rel}}\rvert}{\lvert\mu_{\tilde{E}}\rvert}$	$\sigma_{\tilde{E}_{rel}}$	$s_{\tilde{E}_{rel}}$	$\frac{\lvert\sigma_{\tilde{E}_{rel}} - s_{\tilde{E}_{rel}}\rvert}{\lvert\sigma_{\tilde{E}_{rel}}\rvert}$
3	5.61487e-02	5.60751e-02	1.30966e-03	4.22788e-02	4.23984e-02	2.82837e-03
5	4.04341e-02	4.03461e-02	2.17619e-03	3.05006e-02	3.04113e-02	2.82837e-03
9	3.20344e-02	3.20420e-02	2.37235e-04	2.42023e-02	2.41924e-02	4.07894e-04
17	2.27749e-02	2.28342e-02	2.60330e-03	1.72067e-02	1.72749e-02	3.96428e-03
33	1.61567e-02	1.61861e-02	1.82543e-03	1.22065e-02	1.22618e-02	4.53146e-03
65	1.14429e-02	1.14904e-02	4.14803e-03	8.64526e-03	8.67210e-03	3.10411e-03

Table 7.1: Comparison between the theoretical mean and standard deviation of the quantity $\tilde{E}_{rel}$ as defined by (7.2.7) and (7.2.8), and their numerical counterparts (7.3.2) and (7.3.3) over several grids of N traps. The theoretical means and standard deviations are shown in the columns labelled $\mu_{\tilde{E}_{rel}}$ and $\sigma_{\tilde{E}_{rel}}$ respectively and the sample mean and standard deviations are labelled $\bar{\mu}_{\tilde{E}_{rel}}$ and $s_{\tilde{E}_{rel}}$. 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.

We then establish the following proportion

$$P_{num} = \frac{\tilde{n}_r}{n_r}, \tag{7.3.4}$$

where $\tilde{n}_r$ is the number of the relative errors $\tilde{E}_{rel}$ which fall within the range $[\tilde{E}_{min}, \tilde{E}_{max}]$ as defined by (7.2.9) and (7.2.10) 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 (7.3.2)-(7.3.4) are recalculated.

N	P_{num}	$\frac{\lvert P(3)-P_{num}\rvert}{\lvert P(3)\rvert}$
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

Table 7.2: Comparison between the theoretical probability $P(z)$ as defined by (7.1.9) that $\tilde{E}_{rel}$ lies within the range $[\tilde{E}_{min}, \tilde{E}_{max}]$ and the numerical probability P_{num} computed according to (7.3.4) 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.

We apply the above procedure to the test case (7.3.1), where the number of traps is subsequently increased to be $N = 3, 5, \ldots, 65$. We select the measurement tolerance as $\nu = 0.3$. As can be seen in Table 7.1, for each value of N we have good agreement between the sample mean $\bar{\mu}_{\tilde{E}_{rel}}$ and the theoretical mean $\mu_{\tilde{E}_{rel}}$, and likewise between the sample and theoretical standard deviations $s_{\tilde{E}_{rel}}$ and $\sigma_{\tilde{E}_{rel}}$. We fix $z = 3$ therefore we have the theoretical probability that $\tilde{E}_{rel}$ lies within the range $[\tilde{E}_{min}, \tilde{E}_{max}]$ as $P(z) \approx 0.9973$. It can be seen from Table 7.2 that the corresponding numerical probability P_{num} as given by (7.3.4) is indeed approximately 0.9973. We are therefore satisfied that the range given by (7.2.9) and (7.2.10) can be used to make reliable conclusions about the error $\tilde{E}_{rel}$ of an estimated pest abundance based on measured data $\tilde{I}$.

We now directly compare the quantities E_{rel} and $\tilde{E}_{rel}$ in order to understand how using noisy data rather than exact pest population densities impacts the accuracy of a pest abundance estimate. Let us introduce further test cases to consider alongside that prescribed by the function (7.3.1). It is more difficult *i.e.* a higher number N of grid nodes are required to obtain an accurate abundance estimate when the population is located in a small sub-domain of the field. We therefore introduce test cases where the density is either concentrated in a narrow layer as defined by the following function (see Fig. 7.3b):

$$f(x) = (x + 0.1)^{-3}, \tag{7.3.5}$$

or is located within a small sub-domain of the unit interval and also exhibits oscillatory behaviour (see Fig. 7.3c):

$$f(x) = \exp(-20x)\sin(20\pi x)^2. \tag{7.3.6}$$

For an increasing number N of regularly spaced traps, the relative error $E_{rel}(N)$ of an approximation based on exact data is calculated. The mean value $\mu_{\tilde{E}_{rel}}$ 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 (7.2.7) and (7.2.9), (7.2.10) respectively for the same set of values of N. The measurement tolerance is fixed as $\nu = 0.3$ 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 Figure 7.4. An estimate of the integral I is considered to be accurate if it satisfies the condition (2.5.1). We select the tolerance $\tau = 0.25$ which has been recommended for routine monitoring [86] and lies within the acceptable range for ecological applications given in Section 7.1. The line $\tau = 0.25$ is therefore also plotted so as to determine when the estimates become sufficiently accurate.

It can be seen in Figure 7.4a that for the test case (7.3.1) where the population is spread across the entire domain, the estimates based on exact data are sufficiently accurate for the whole 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 $\mu_{\tilde{E}_{rel}}$ has a less steep gradient than its E_{rel} 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 $\mu_{\tilde{E}_{rel}}$ 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 error decreases. Meanwhile, it is important to note the mean error $\mu_{\tilde{E}_{rel}}$ does converge to zero in the theoretical limit of an infinite number of traps (e.g., see [23]).

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

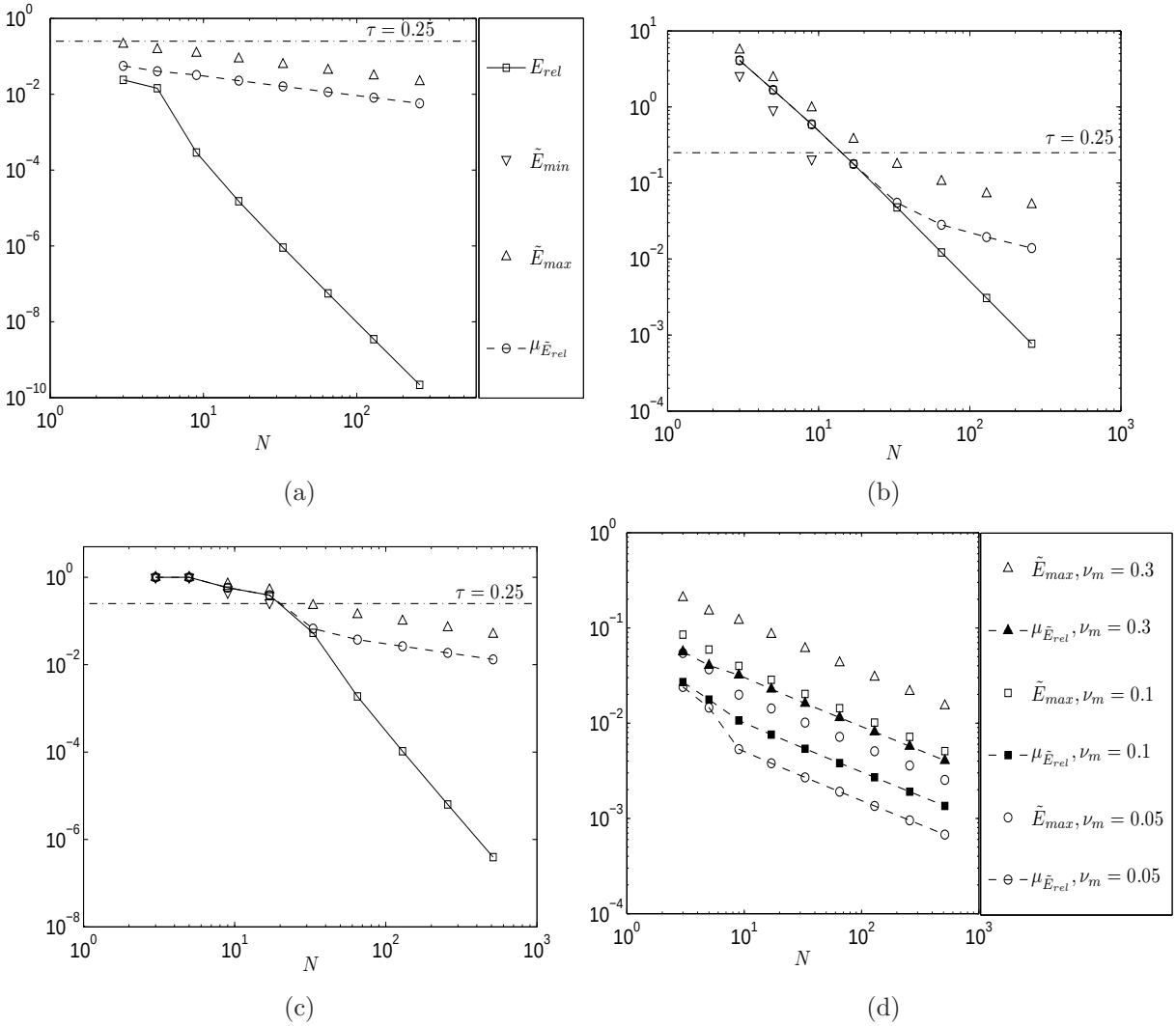


Figure 7.4: (a)-(c) The error for the approximation based on exact data E_{rel} is compared with the mean error $\mu_{\tilde{E}_{rel}}$ of an approximation based on noisy data alongside the limits of the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ for the test cases (7.3.1), (7.3.5) and (7.3.6) respectively as shown in Figure 7.3a-7.3c. The measurement tolerance is fixed as $\nu = 0.3$ and $z = 3$ in each case. The legend for each figure is as shown in (a). (d) Mean error $\mu_{\tilde{E}_{rel}}$ of an approximation based on noisy data and the upper limit of the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ for the test case (7.3.1) as shown in Figure 7.3a where values $\nu = 0.05, 0.1, 0.3$ of the measurement tolerance have been selected. We fix $z = 3$ as before.

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 testing density distributions (7.3.5) and (7.3.6), the number of traps N has to be sufficiently increased before the desired level of accuracy $E \leq \tau = 0.25$ is obtained (see Figure 7.4b and Figure 7.4c). 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 $\mu_{\tilde{E}_{rel}}$ 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 $\tilde{E}_{rel}$ 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 Figure 7.4c 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. We now investigate the impact of noise on an estimate's accuracy as the quantity ν is varied. Let us again consider the simpler test case (7.3.1) as shown in Figure 7.3a. Figure 7.4d shows the convergence curves for different values of the measurement tolerance: $\nu = 0.05, 0.1$ 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.

7.4 Assessing the Impact of Noise: Ecological Test Cases

Although informative, the test cases in the previous section 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 use data simulated from the spatially explicit form of the 1D Rosenzweig-MacArthur model (3.2.7–3.2.8) supplied by the authors of [70]. This system of equations is solved numerically at a fixed time $t = \hat{t} > 0$ for different parameter values to generate four ecologically meaningful test cases which are shown in Figure 7.5. The test cases of Figures 7.5a–7.5d have an increasing number of peaks and we refer to them as the monotone, single peak, three peak, and the multi-peak test case respectively.

The exact pest abundance I is computed using the composite trapezium rule (2.2.6) from the exact data f_i obtained on a very fine grid of $N_f = 2^{15} + 1$ traps. Once we have found the values of the pest density function $f(x)$ at the trap locations $x_i, i = 1, ..., N_f$, 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 (2.2.6).

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 instead computed using the composite trapezium rule (2.2.6) from the exact data fi obtained on a very fine grid of Nf = 215 + 1 traps.

Let us fix the number of traps as $N = N_1$ and calculate an estimate I_a based on exact data. The relative error E_{rel} is then calculated from (7.1.2). The mean error $\mu_{\tilde{E}_{rel}}$ of an estimate formed from noisy data is found from (7.1.2) and the limits of the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ are

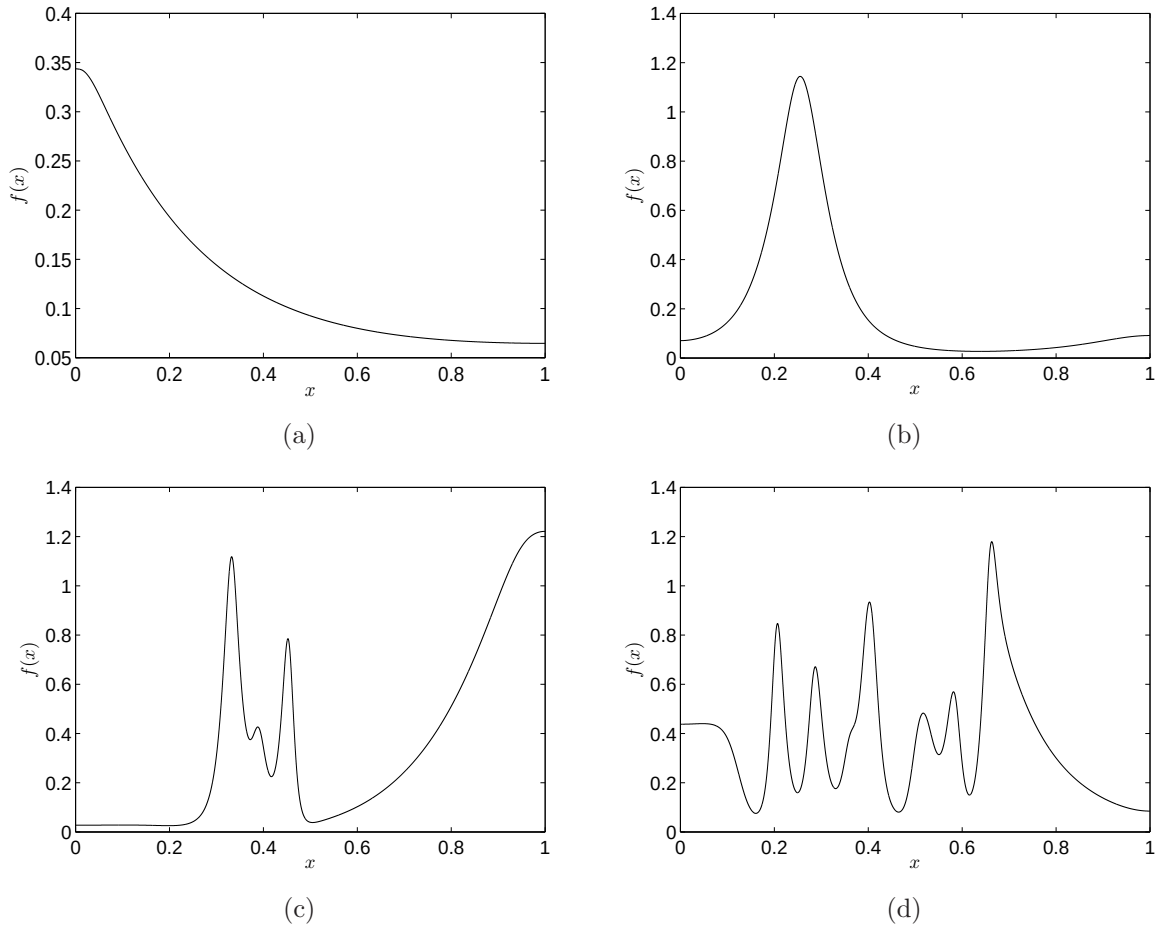


Figure 7.5: Ecologically meaningful test cases as considered in [70], generated by the model (3.2.7–3.2.8) at different times t and for various choices of the dimensionless diffusion coefficient d: (a) $t = 5, d = 10^{-4}$ (b) $t = 50, d = 10^{-4}$ (c) $t = 100, d = 10^{-5}$ (d) $t = 400, d = 10^{-5}$. In each case the other parameter values as well as the initial and boundary conditions are as given in the caption of Figure 3.3.

calculated from (7.2.9) and (7.2.10). 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 Figure 7.6. The measurement tolerance is fixed as $\nu_{=}0.3$ 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 level of approximation accuracy, 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 $\mu_{\tilde{E}_{rel}}$ of the estimate based on perturbed data. This can been seen in Figure 7.6a where the results for the monotone density distribution of Figure 7.5a are presented. For the monotone function a good level of accuracy is already achieved on coarse grids (e.g., see $N = 5$ in the figure), and the error E_{rel} obtained from exact data is several orders of magnitude smaller than the mean error $\mu_{\tilde{E}_{rel}}$ when N increases. However, it is important to emphasise that the mean error is already below the required tolerance even on very coarse grids. Furthermore 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 Figures 7.6b-7.6d where the more 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 $\mu_{\tilde{E}_{rel}}$ lie close to each other.

It should be mentioned that, as shown in Figures 7.6c and 7.6d 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 non-zero 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 (7.3.6) in the previous section. 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.

As discussed in previous chapters, a generic behaviour of the approximation error is that the accuracy of the approximation I_a worsens when the spatial complexity of the density function increases. 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 Figure 7.6d that for a multi-peak density function (*i.e.* the function that presents an ecologically important case of a patchy population density)

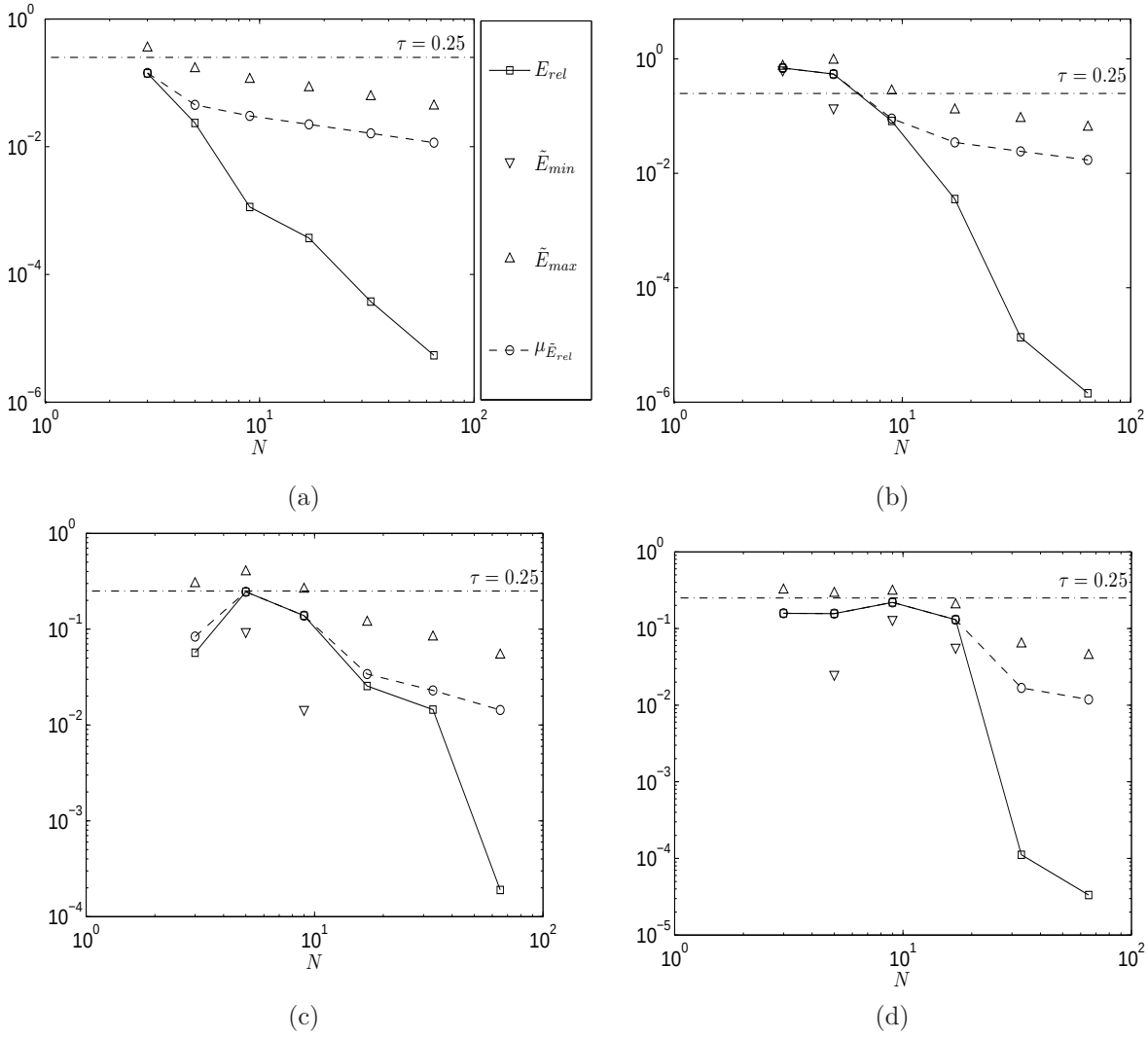


Figure 7.6: The error for the approximation based on exact data E_{rel} is compared with the mean error $\mu_{\tilde{E}_{rel}}$ of an approximation based on noisy data and the limits of the range $[\tilde{E}_{min}, \tilde{E}_{max}]$ for the ecologically meaningful (a) Monotone, (b) Single peak, (c) Three peak and (d) Multi-peak test cases as shown in Figures 7.5a - 7.5d respectively. The measurement tolerance is fixed as $\nu = 0.3$ and we set $z = 3$ in each case. The legend for all figures is as shown in (a).

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 the accuracy of evaluation on coarse grids of traps.

7.5 Chapter 7 Conclusions

In this chapter, we have considered a trapping procedure where noise (random error) can be expected in the resulting measurements of pest population density. The impact of noise on the accuracy of a pest abundance estimate obtained by the composite trapezium rule (2.2.6) has been investigated.

By considering each measured pest population density to be normally distributed about the true density, the error $\tilde{E}_{rel}$ of an estimate constructed from noisy data can be quantified. Assuming the density data are uncorrelated, an expression for the mean error $\mu_{\tilde{E}_{rel}}$ has been given, and we have found a range to which a realisation of the error $\tilde{E}_{rel}$ belongs with probability $P(z)$. In other words, we have constructed an α percent credible interval for the error $\tilde{E}_{rel}$ of an estimate based on measured data, where $\alpha = 100P(z)$. The theoretical results have been verified for various 1D density distributions, including those with ecological significance.

We have demonstrated that the approximation error when noise is present 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 $\tilde{E}_{rel}$ 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.

The accuracy of the estimate has also been studied for the ecologically significant scenario where the number N of traps is small *i.e.* on coarse computational grids. 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 may 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.

Numerical experiments 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.

We have only considered the composite trapezium rule (2.2.6) of numerical integration in this chapter. It has been shown that the accuracy of approximation on exact data is crucial when the number N of traps is small. Applying other methods of evaluation may give an estimate of pest abundance that is more accurate on coarse grids of traps in the presence of noise and we investigate this in the next chapter.

Chapter 8

Comparing Methods of Numerical Integration in the Presence of Noise

We continue our consideration of uncorrelated density data obtained from a trapping procedure which is thus subject to an inherent random error (noise). We extend the study conducted in the previous chapter in order to compare the impact of noise in the population density data on an estimate of pest abundance for various evaluation methods. In particular, we compare the performance of the statistical rule (2.1.4), the composite trapezium rule (2.2.6), and the composite Simpson's (2.2.7) rule in the presence of noise. We continue to work in 1D, however the approach can readily be extended to the 2D problem. This chapter consists of work given in [34].

The work we have carried out so far into the impact of noise in the population density data on the abundance estimation accuracy has shown that the effects when the number N of nodes where density data are available is large (*i.e.* on fine computational grids) differ to when the number N is small (on coarse grids). Thus we consider the cases separately: the performance of the different evaluation methods using noisy data is compared on fine computational grids in Section 8.1, and on coarse grids in 8.2. The theoretical conclusions drawn in these sections are then verified for ecologically relevant test cases in Section 8.3. Concluding remarks are provided

in 8.4.

8.1 The Accuracy of Integrating Noisy Data: Fine Grids

We wish to compare the performance of different means of evaluating pest abundance when noise is present in the population density data. As in the previous chapter, we consider a regular distribution of traps (7.1.5) across the unit interval which is taken to represent the agricultural field. We compare three means of forming an estimate namely: the statistical rule (2.1.4), the composite trapezium rule (2.2.6), and the composite Simpson's (2.2.7) rule.

We again take the relationship between the true pest density f_i and the measured (noisy) pest density $\tilde{f}_i$ at a trap location $x_i \in [0,1]$ to be

$$f_i = \tilde{f}_i + \varepsilon_{r_i}$$

where ε_{r_i} is the random component of the measurement error (see Section 7.1) and is a realisation of a normally distributed random variable. We assume ε_{r_i} to have a mean of zero and further take the true pest density f_i to be some unknown constant. Hence, the measured pest density $\tilde{f}_i$ is a realisation of a normal distributed random variable with mean equal to the corresponding true density f_i.

We recall that estimates I_a and $\tilde{I}$ of abundance formed from exact and noisy density data respectively are given by the following formulae

$$I_a = \sum_{i=1}^{N} w_i f_i \qquad \tilde{I} = \sum_{i=1}^{N} w_i \tilde{f}_i,$$

and for convenience we restate the weight coefficients $w_i, i = 1, \ldots, N$ for the evaluation methods under consideration below. On the unit interval, they are:

$$w_i = \frac{1}{N}, \quad i = 1, \ldots, N$$

for the statistical rule (2.1.4),

$$w_i = h, \quad i = 2, \ldots, N-1, \quad w_i = \frac{h}{2}, \ i = 1, \text{ or } i = N.$$

for the composite trapezium rule, and

$$w_i = \frac{4h}{3}, \ i = 2, 4, \ldots, N-1, \quad w_i = \frac{2h}{3}, \ i = 3, 5, \ldots, N-2, \quad w_i = \frac{h}{3}, \ i = 1, \text{ or } i = N.$$

for the composite Simpson's rule, where $h = 1/(N-1)$. The relative errors E_{rel} and $\tilde{E}_{rel}$ for estimates constructed from exact and noisy data are given by (7.1.2) and (7.1.4), namely,

$$E_{rel} = \frac{|I - I_a|}{I} \qquad \tilde{E}_{rel} = \frac{|I - \tilde{I}|}{I}.$$

It was demonstrated in Section 7.2 that the quantity $\tilde{E}_{rel}$ is a realisation of a random variable with a folded normal distribution, and an expression for the mean error was given (7.2.7). An α percent credible interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ was constructed for $\alpha = 100P(z)$, with the limits (7.2.9–7.2.10) and where $P(z)$ is described by (7.1.9). We use the quantities (7.2.7) and (7.2.9–7.2.10) to assess the relative integration error of an estimate formulated from noisy density data.

In this section we compare the performance in the presence of noise of the statistical rule (2.1.4), the composite trapezium rule (2.2.6) and Simpson's rule (2.2.7) on a fine computational grid. We thus analyse the formulae for the mean value of $\tilde{E}$ as well as the limits of its credible interval for the case when the number N of traps is large. It can readily be seen from (7.2.7) and (7.2.9–7.2.10) that the behaviour of these quantities are dictated by the accuracy E_{rel} of the estimate obtained when exact density data values $f_i, i = 1 \ldots, N$ are considered, as well as $\sigma_{\tilde{I}}$, namely the uncertainty (7.2.2) associated with the estimate $\tilde{I}$ of abundance formed from noisy data.

For large N, the fixed distance h between the traps is small. We recall from Chapter 2 (see Sections 2.2 and 2.5) that for small h the relative error E_{rel} of an estimate formed from exact

data behaves according to the asymptotic error estimates

$$E_{rel} \leq Ch^q \tag{8.1.1}$$

where the constant C and the asymptotic convergence rate q depend on the method of evaluation. When the statistical method (2.1.4) is applied on a regular grid of nodes $q = 1$, and for the composite trapezium rule (2.2.6) and composite Simpson's rule (2.2.7) we have $q = 2$ and $q = 4$ respectively. Thus on fine grids of precise data, the statistical rule and the composite Simpson's rule yield the least and most accurate estimates respectively.

We now look at the behaviour of the uncertainty term $u(\tilde{I})$ for large N (*i.e.* for small h). Let us introduce the term u_{max} such that

$$u_{max} = \max_{i \in \{1,\ldots,N\}} u(f_i).$$

From the above and the equation (7.2.2) it follows that

$$u(\tilde{I}) = \sqrt{\sum_{i=1}^{N} w_i^2 u^2(\tilde{f}_i)} \leq u_{max} \sqrt{\sum_{i=1}^{N} w_i^2}.$$

Let us first consider the uncertainty associated with an estimate $\tilde{I}_{stat}$ formed by the statistical rule. Substituting the weights (2.1.4) into the above and recalling the distance between the traps is fixed as $h = 1/(N-1)$ we obtain

$$u(\tilde{I}_{stat}) \leq \frac{u_{max}}{\sqrt{N}} = \left(\frac{1}{h} + 1\right)^{-\frac{1}{2}} u_{max}.$$

For small h we have

$$u_{max} \left(\frac{1}{h} + 1\right)^{-\frac{1}{2}} \approx h^{\frac{1}{2}} u_{max},$$

thus the convergence of the uncertainty $u(\tilde{I}_{stat})$ is of the order $k = 1/2$. Similar expressions can be found for the uncertainty associated with an estimate $\tilde{I}_{trap}$ formed by implementation of the

composite trapezium rule (2.2.6)

$$u(\tilde{I}_{trap}) \leq h\left(\frac{1}{h}-\frac{1}{2}\right)^{\frac{1}{2}} u_{max} \approx h^{\frac{1}{2}} u_{max},$$

and likewise that associated with the estimate $\tilde{I}_{Simp}$ formulated from the composite Simpson's rule (2.2.7)

$$u(\tilde{I}_{Simp}) \leq \frac{h}{3}\left(\frac{10}{h}-2\right)^{\frac{1}{2}} u_{max} \approx h^{\frac{1}{2}} u_{max}.$$

Thus the orders of convergence of the uncertainty terms $u(\tilde{I}_{trap})$ and $u(\tilde{I}_{Simp})$ are, as for $u(\tilde{I}_{stat})$, also $k = 1/2$ provided the distance h between traps is small.

It can be seen from (7.2.9) that as N grows large, and the relative error E_{rel} of an estimate formed from exact data grows sufficiently small, the quantity $\tilde{E}_{min}$ becomes zero. In our study of the effect noise has on the accuracy of an estimate formed by the composite trapezium rule conducted in the previous chapter, we found that when the number of traps N is large, the uncertainty $u(\tilde{I})$ associated with the estimate $\tilde{I}$ is the dominant contribution to the error of an estimate $\tilde{I}$. Therefore, in this chapter's study of the three methods (2.1.4), (2.2.6) and (2.2.7) we anticipate that as N increases, the quantities $\mu_{\tilde{E}_{rel}}$ and $\tilde{E}_{max}$ will converge at a rate of $k = 1/2$ in accordance with the behaviour of the uncertainty term $u(\tilde{I})$ rather than the order q as described by the error estimate (8.1.1). This conclusion is validated in Section 8.3.

8.2 The Accuracy of Integrating Noisy Data: Coarse Grids

We now turn our attention to the ecologically relevant case of a coarse grid of traps where the number of population density data points is small. In many cases having a small number of traps installed in the field means that the computational grid of data used to formulate an estimate of abundance is coarse. We recall from Chapter 3 that for this class of grids, the asymptotic error estimates do not hold. It was discussed in Chapters 4 and 5 that the matter of grid coarseness is related to the degree of heterogeneity of the pest population density function highly aggregated density distributions being the most difficult case for pest abundance evaluation. An estimate of pest abundance can be very inaccurate when the total pest population size is evaluated from

a strongly heterogeneous density pattern, while the same grid of traps will provide very good accuracy for another, quasi-homogeneous, density distribution. Ecologists and farmers often have to deal with pest insect density distributions that have a considerable degree of aggregation [22, 55, 62], therefore the study of abundance estimation accuracy on coarse grids is important. Our previous investigation has been made under the assumption that the density data is exact, thus we now need to further investigate the accuracy on coarse grids in the case that noise is present in the data.

Once again, the mean error and the bounds of the credible interval are generally determined by the uncertainty $u(\tilde{I})$ associated with the estimate $\tilde{I}$ formulated from measured pest densities, and the error E_{rel} of an estimate I_a formulated from exact values of the pest population density. We begin our study of coarse grids by discussing the uncertainty quantity $u(\tilde{I})$. Below we compare the uncertainty in the estimate obtained when the composite trapezium rule (2.2.6) is employed with the uncertainty for the statistical rule (2.1.4) and composite Simpson's rule (2.2.7) estimates on coarse grids.

From (7.2.2) it is clear that the uncertainty $u(\tilde{I})$ associated with an estimate formed from measured data $\tilde{I}$ will increase in magnitude as the magnitude of the weights of the numerical integration method increase. On comparing the weights of the statistical rule estimate (2.1.4), the composite trapezium rule (2.2.6), and composite Simpson's rule (2.2.7) on the unit interval which are given in Section 8.1, it can be seen that none of these methods has uniformly larger weights than another. For example, the weights of the composite trapezium rule are $w_1 = w_N = h/2$ and $w_i = h, i = 2, \ldots N-1$ whereas for the statistical rule they are uniformly $w_i \equiv 1/N$. Thus, whilst the weights corresponding to the interior nodes are larger for the composite trapezium rule than the statistical rule estimate, the converse is true for those at the exterior nodes. Consequently, employing a method which by the asymptotic error estimate (8.1.1) is ordinarily considered more accurate, could in fact lead to a larger associated uncertainty.

For instance the use of composite Simpson's rule (2.2.7) may result in a larger uncertainty in the estimate $\tilde{I}$ than that yielded by the composite trapezium rule (2.2.6). This occurs when

the following condition is satisfied

$$\frac{h^2}{9}\left(u_1^2+u_N^2\right)+\frac{16h^2}{9}\sum_{i=1}^{\frac{N-1}{2}}u_{2i}^2+\frac{4h^2}{9}\sum_{i=2}^{\frac{N-1}{2}}u_{2i-1}^2>\frac{h^2}{4}\left(u_1^2+u_N^2\right)+h^2\sum_{i=2}^{N-1}u_i^2,$$

where $u_i \equiv u(f_i)$. The above can be expressed as

$$u_1^2+u_N^2<C_1\sum_{i=1}^{\frac{N-1}{2}}u_{2i}^2-C_2\sum_{i=2}^{\frac{N-1}{2}}u_{2i-1}^2, \tag{8.2.1}$$

where the coefficients are $C_1 = 28/5$ and $C_2 = 4$.

Likewise, composite Simpson's rule could lead to a greater uncertainty associated with the estimate $\tilde{I}$ than that associated with statistical rule (2.1.4). This will happen when we have

$$\frac{h^2}{9}(u_1^2+u_N^2)+\frac{16h^2}{9}\sum_{i=1}^{\frac{N-1}{2}}u_{2i}^2+\frac{4h^2}{9}\sum_{i=2}^{\frac{N-1}{2}}u_{2i-1}^2>\frac{1}{N^2}\sum_{i=1}^{N}u_i^2.$$

Using the fact that $h = 1/(N-1)$ and rearranging gives

$$u_1^2+u_N^2<C_3\sum_{i=1}^{\frac{N-1}{2}}u_{2i}^2-C_4\sum_{i=2}^{\frac{N-1}{2}}u_{2i-1}^2, \tag{8.2.2}$$

where we have $C_3=\dfrac{7N^2+18N-9}{8N^2-18N+9}$ and $C_4=\dfrac{5N^2-18N+9}{8N^2-18N+9}$.

Finally, implementing the composite trapezium rule will give rise to a larger uncertainty than the statistical rule when the following condition is satisfied:

$$u_1^2+u_N^2<C_5\sum_{i=2}^{N-1}u_i^2, \tag{8.2.3}$$

where $C_5 = (8N-4)/(3N^2-8N+4)$.

The conditions (8.2.1), (8.2.2) and (8.2.3) can be used to decide which method is best to use on a coarse grid of traps in order to reduce the uncertainty of evaluation. Consider, for example, the condition (8.2.3) and let a very coarse grid of $N = 3$ traps be installed. The

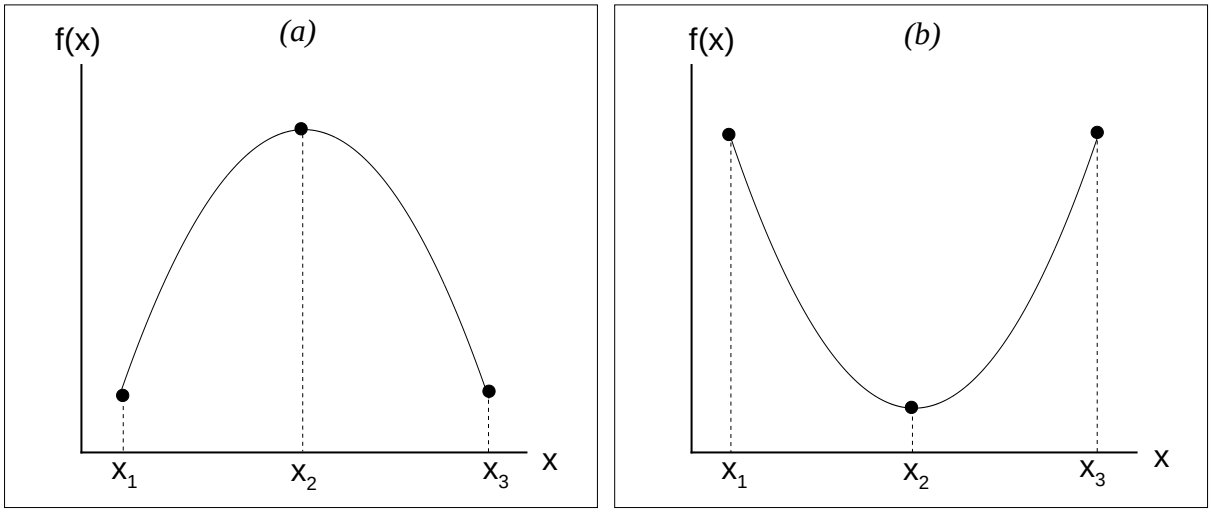


Figure 8.1: A sketch of the density distribution $f(x)$ where evaluation of the pest abundance is done on a coarse grid of $N = 3$ traps. The measured data are available at the points x_1, x_2 and x_3 of a regular grid of traps. (a) The density is localised close to the centre of the domain. (b) The density is localised close to the boundaries.

inequality (8.2.3) is then written as $u_1^2 + u_3^2 < C_5 u_2^2$, where $C_5 = 20/7 \approx 3$. Hence, if the spatial pattern of a density distribution is such that the density is concentrated close to the domain centre (see a sketch of the density function shown in Figure 8.1a), the inequality (8.2.3) holds. The uncertainty generated by the composite trapezium rule (2.2.6) is in this case larger than the uncertainty generated by the statistical method (2.1.4). On the contrary, if the density is localised close to the boundaries (see Figure 8.1b) then the composite trapezium rule yields a smaller uncertainty of evaluation. Similar analysis can be done for conditions (8.2.1) and (8.2.2).

It follows from the conditions (8.2.1), (8.2.2) and (8.2.3) that on coarse grids the error in the pest abundance estimate depends on the spatial pattern of the density function when noisy data are used for the evaluation. This is the same conclusion made in previous chapters when exact density data was considered. We thus reiterate that the knowledge of spatial pattern of the pest insect density distribution is crucial when pest abundance is evaluated on coarse grids and any information about spatial pattern must be used to its fullest extent.

We now discuss the role of the error E_{rel} of an estimate formed from exact density data in the behaviour of the error $\tilde{E}_{rel}$ of an estimate constructed from noisy data. It was shown in the

previous chapter that if for small N the error E_{rel} is already significant, then the contribution made by the uncertainty term $u(\tilde{I})$ discussed above is negligible. We remember that the error E_{rel} is (unavoidably) incurred by approximating the pest abundance by numerical integration as opposed to analytically integrating the continuous population density distribution. Meanwhile, the uncertainty term $u(\tilde{I})$ is the result of the noise in the density data. In other words, it was found in the previous chapter that on coarse grids the impact of noise in the data is insignificant in comparison to the error imparted by the method of numerical integration used to form an abundance estimate. This conclusion was drawn from a study of the composite trapezium rule (2.2.6) only, thus in the next section we investigate whether it hold true for the statistical rule (2.1.4) and the composite Simpson's rule (2.2.7) by conducting numerical experiments on several ecologically meaningful test cases. The accuracy on fine grids will also be investigated.

8.3 Numerical Study of the Impact of Noise

In this section we test the conclusions of the previous two sections for a variety of ecologically meaningful data. Estimates of the pest abundance are obtained by employing the methods (2.1.4), (2.2.6) and (2.2.7) over a series of increasingly refined grids of traps, *i.e.* for increasing values of the number N of traps. The mean error of estimates formed from measured data $\mu_{\tilde{E}_{rel}}$ is calculated for each value of N, as are the lower and upper limits of the credible interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ using (7.2.7) (7.2.9) and (7.2.10) respectively. To assess the impact of noise on the accuracy of an estimate, a comparison is made with the relative errors of the estimates based on exact data E_{rel}. Plots of the convergence curves of all error quantities are given and the results are discussed.

Since we intend to investigate the accuracy of numerical integration for a broad variety of density patterns, we generate six ecologically significant test cases from the model (3.2.7–3.2.8) by inputting different parameter values. Plots of the resulting pest density functions are shown in Figures 8.2a - 8.2f. The test cases are chosen such that the level of difficulty in obtaining an accurate estimate of pest abundance increases as we move from test case 1 through to test case 6. A test case is considered more difficult the higher the number N of traps needed to obtain a sufficiently accurate estimate. Test case 1, as shown in Figure 8.2a, is a smooth, monotonous

function. The structure of the density function can therefore be detected from a small amount of data *i.e.* a small number N of installed traps and an accurate estimate of abundance can be readily produced. The number of peaks present in the density function increases in the subsequent test cases until we reach test case 5 as shown in Figure 8.2 which has a complicated multi-peak structure. More information about the pest density function, which corresponds to a higher number N of installed traps, will be required to detect the more complex peak structure and thus obtain an accurate estimate. Meanwhile, test case 6 provides an example of the most difficult case whereby the pest population is located within a small sub-domain of the field. The difficulties of handling such distributions, also known as *peak functions* have been discussed in detail in Chapters 5 and 6. If we consider a fixed number N of installed traps, we expect the estimate of abundance to be most accurate for test case 1 and least accurate for test case 6.

It should be noted that the quantities $\mu_{\tilde{E}_{rel}}$, $\tilde{E}_{min}$ and $\tilde{E}_{max}$ were derived using the assumption that any measured pest density $\tilde{f}_i$ is normally distributed about the true pest density f_i and belongs to the range (7.1.8) with a prescribed probability (7.1.9). The counterpart to this assumption is that there is a chance that any $\tilde{f}_i$ can lie outside of this range. In particular the theory does not discount a measured pest density $\tilde{f}_i$ being negative. Of course a negative pest density is senseless, therefore instead each measured pest density should be considered to belong to a truncated normal distribution. The effects of such a truncation are investigated in Appendix E and are shown to be small when the parameter $z \geq 1$. We thus ensure that z satisfies this condition and consider $\mu_{\tilde{E}_{rel}}$, $\tilde{E}_{min}$ and $\tilde{E}_{max}$ to be reliable indicators of accuracy behaviour of an abundance estimate in the presence of noise.

The quantities we use to assess the impact of noise on an estimate of pest abundance, namely E_{rel}, $\mu_{\tilde{E}_{rel}}$, $\tilde{E}_{min}$, and $\tilde{E}_{max}$, all depend on the true value of the pest abundance I. In order to obtain I we use the same approach used throughout previous chapters and solve the system of equations (3.2.7–3.2.8) on a very fine regular grid. For all test cases shown in Figure 8.2 the number of nodes on the fine grid was fixed as $N_f = 4097$. The data $f_i, i = 1, \ldots, N_f$ obtained on the fine grid was then integrated using the composite trapezium rule (2.2.6) and the result was taken to be the exact pest abundance I.

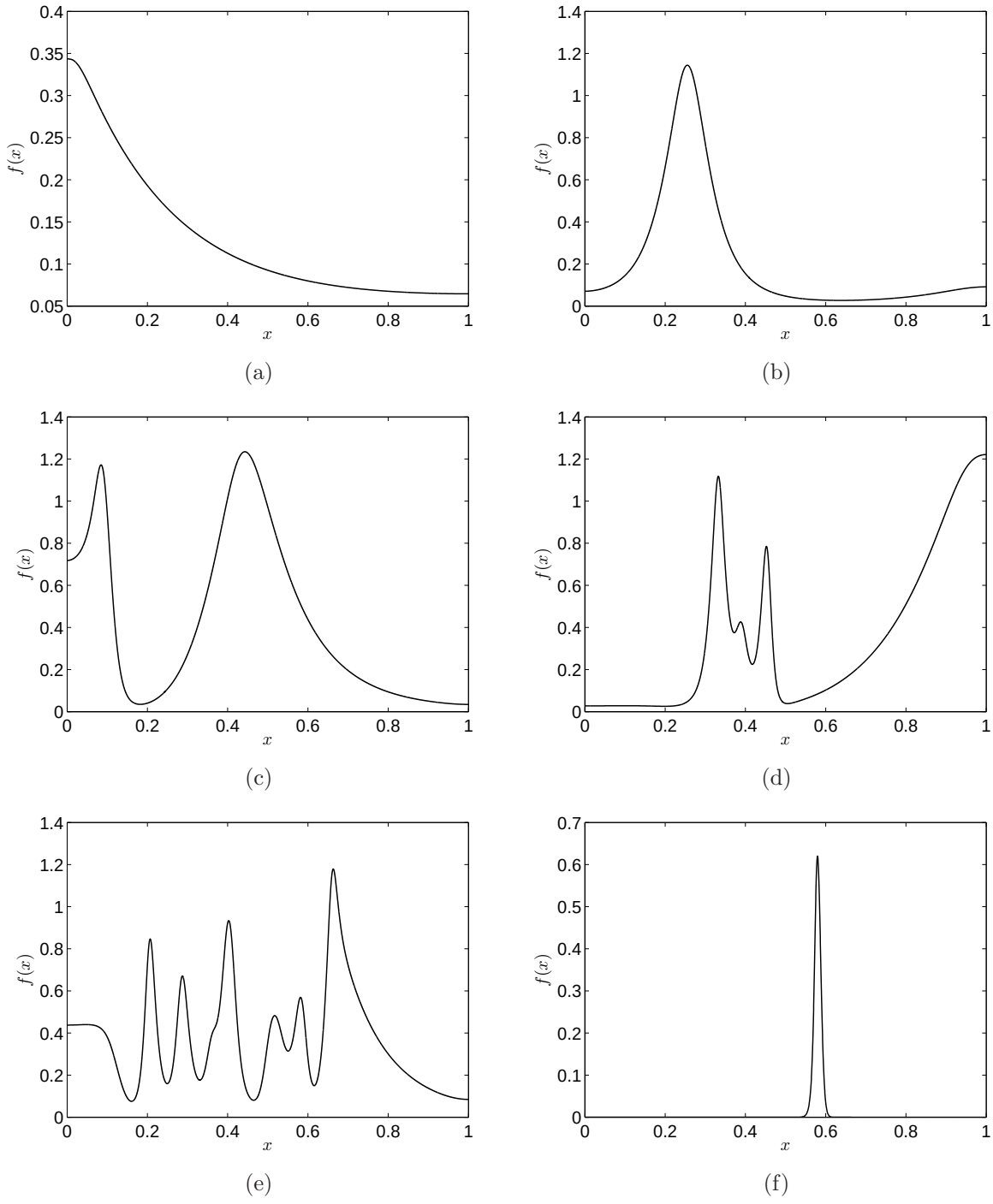


Figure 8.2: Ecologically significant test cases. A spatial distribution $f(x)$ of the pest population density function is obtained from the model (3.2.7–3.2.8) at different times t and for different values of the diffusivity d. (a) Test case 1, $d = 10^{-4}$, $t = 5$, (b) Test case 2, $d = 10^{-4}$, $t = 50$, (c) Test case 3, $d = 10^{-5}$, $t = 50$, (d) Test case 4, $d = 10^{-5}$, $t = 100$. (e) Test case 5, $d = 10^{-5}$, $t = 400$. (f) Test case 6, $d = 3 \times 10^{-6}$, $t = 10$. The other parameter values as well as initial and boundary conditions are as given in the caption of Figure 3.3.

An estimate of pest abundance based on exact data I_a is obtained using each of the weight choices (2.1.4), (2.2.6), and (2.2.7), on a series of regular grids of traps and the error E_{rel} is calculated according to (2.1.7) as usual. The initial grid has the number of traps fixed as $N = N_1$, where N_1 is odd. The number of traps in subsequent grids is then calculated as $N_s = 2N_{s-1} - 1$ for $s \geq 2$. This process is repeated as many times as necessary to fully show the behaviour of the convergence. The quantities $\mu_{\tilde{E}_{rel}}, \tilde{E}_{min}$ and $\tilde{E}_{max}$ are then evaluated for each value of N from (7.2.7), (7.2.9) and (7.2.10) where the measurement tolerance (see Section 7.1) has been set as $\nu = 0.3$. We have fixed $z = 3$ so the probability that a single realisation of the error $\tilde{E}_{rel}$ lies within the range $[\tilde{E}_{min}, \tilde{E}_{max}]$ is $P(z = 3) \approx 0.9973$.

Figures 8.3a–8.3f compare the error E_{rel} of an estimate formed from exact data with the mean error $\mu_{\tilde{E}_{rel}}$ of an estimate formed from noisy data. Once the grid of traps becomes sufficiently refined, the convergence rate of the error quantity E_{rel} behaves according to the asymptotic error estimates (8.1.1) and the composite Simpson's rule (2.2.7) yields a more accurate estimate than the composite trapezium rule (2.2.6), which in turn is superior than the estimate provided by the statistical rule (2.1.4).

Meanwhile, it can also be seen from Figures 8.3a–8.3f that for each method of numerical integration the mean error of an estimate formed from noisy data, $\mu_{\tilde{E}_{rel}}$, converges at the slower rate of $k = 1/2$ as explained in Section 8.1. The difference in the convergence rates of E_{rel} and $\mu_{\tilde{E}_{rel}}$ demonstrates that when the number N of traps is large, the accuracy of an estimate may be severely hampered by the presence of noise. In the presence of noise, the composite Simpson's rule (2.2.7) is not superior to the methods (2.1.4) and (2.2.6), as is the case when precise data are used on fine grids. It should be noted, however, that as N grows large the estimate of pest abundance based on exact data I_a tends to the true pest abundance I and hence the error of an estimate based on exact data E_{rel} tends to zero. Consequently, the probability mass function of the quantity $\tilde{E}_{rel}$ transitions to a special case of the folded normal distribution, namely the half normal distribution [27]. In other words, the probability mass function skews towards zero, thus it becomes more probable that a smaller rather than a larger error will be obtained.

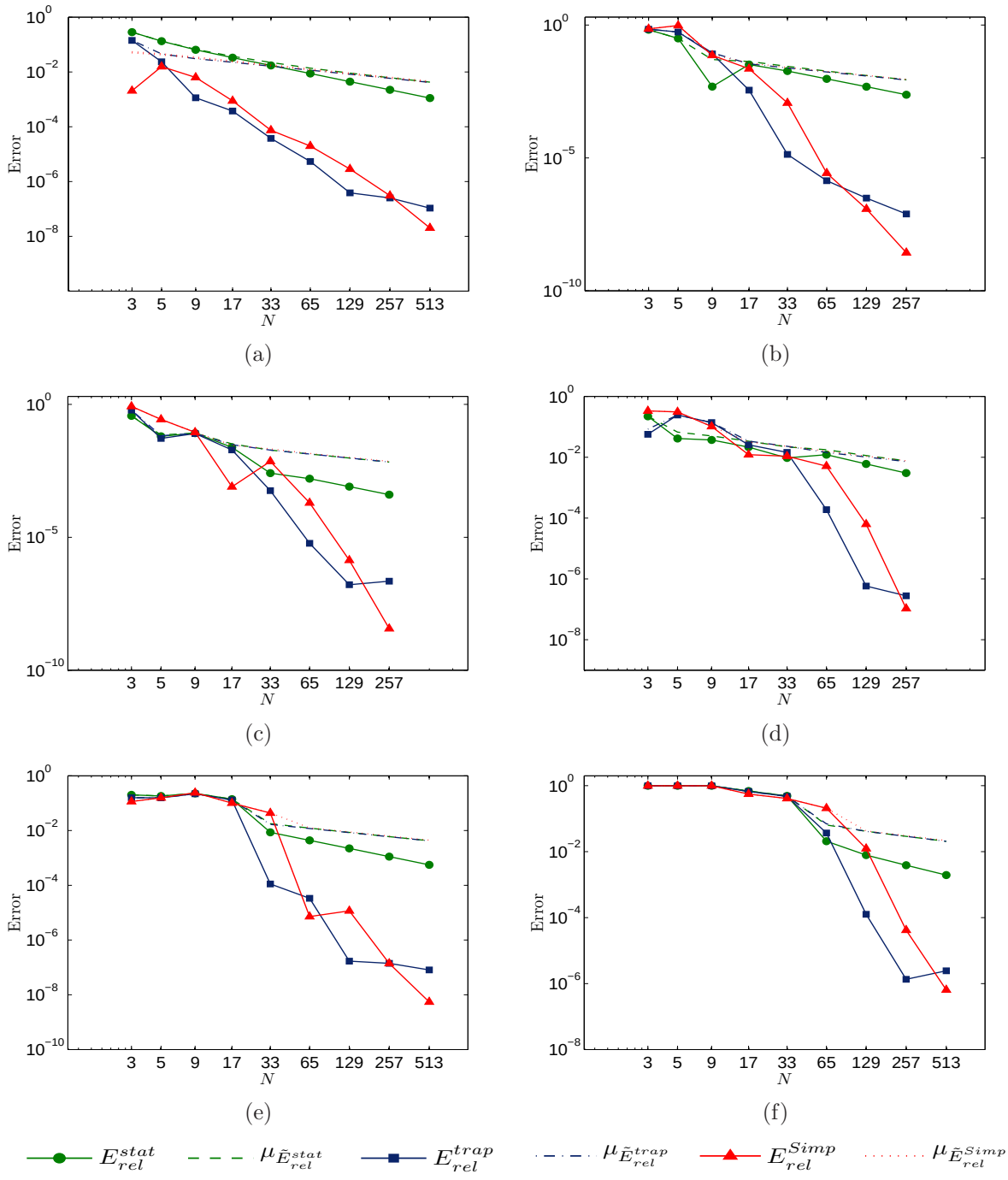


Figure 8.3: Convergence curves for the density distributions depicted in Figure 8.2a – 8.2f respectively. The mean error $\mu_{\tilde{E}_{rel}}$ of an estimate formed from noisy data (dashed lines) is compared with the error E_{rel} of an estimate constructed from exact data (solid lines). The caption in all figures is as given above. The superscripts 'stat', 'trap' and 'Simp' indicate that the estimate of the pest abundance was calculated either by the statistical rule (2.1.4), or was formed using the composite trapezium rule (2.2.6) or Simpson's rule (2.2.7).

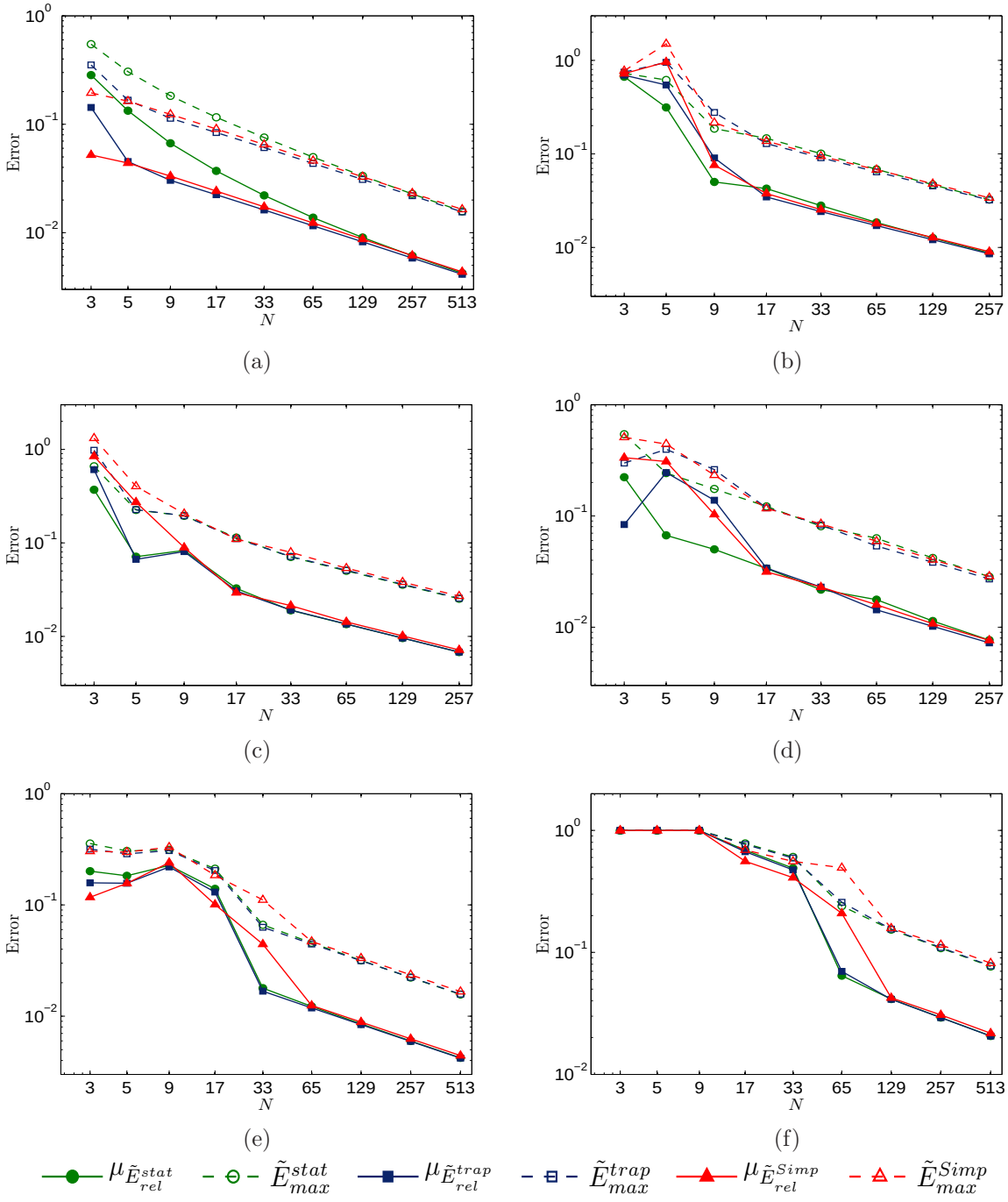


Figure 8.4: Convergence curves for the density distributions $f(x)$ depicted in Figure 8.2a – 8.2f respectively. Plots of the mean error quantity $\mu_{\tilde{E}_{rel}}$ are shown (solid lines) alongside the quantity $\tilde{E}_{max}$ (dashed lines). The caption in all figures is as given above and the same superscript notation is used as in Figure 8.3.

Figures 8.4a–8.4f show the mean error $\mu_{\tilde{E}_{rel}}$ more clearly, as well as the quantity $\tilde{E}_{max}$. The curves of $\tilde{E}_{max}$ are shown to be parallel to that of $\mu_{\tilde{E}_{rel}}$ for larger N, therefore confirming that the convergence rate of $\tilde{E}_{max}$ is also $k = 1/2$ as expected. Thus for large N, the uncertainty associated with the estimate of pest abundance $\tilde{I}$ caused by noise in the data $\tilde{f}_i, i = 1, \ldots, N$ is the dominant factor affecting the accuracy of an estimate. One interesting feature shown in Figures 8.4a–8.4f is that when the number N of traps is large, the difference between the values of the quantity $\mu_{\tilde{E}_{rel}}$ for each of evaluation methods is very small, as is the difference between the values of $\tilde{E}_{max}$. This confirms our previous conclusion that whilst the more sophisticated composite Simpson's method (2.2.7) outperforms the composite trapezium rule (2.2.6) and the statistical method (2.1.4) when the data on the pest population density are precise, there is little difference between the methods when the data is noisy.

Tables 8.1a–8.1f give the values of $\tilde{E}_{min}$ for the test cases 1–6. As can be seen from (7.2.9), the definition of this quantity depends on the relative error E_{rel} of an estimate constructed from exact data. A sufficiently accurate estimate is needed for the quantity $\tilde{E}_{min}$ to be zero, thus the grid of N traps needs to be sufficiently refined to resolve the heterogeneity of the pest population density. Tables 8.1a–8.1f confirms that the point at which $\tilde{E}_{min}$ becomes consistently zero varies depending on the spatial pattern of the pest population density function of the corresponding test case (compare with Figures 8.2a-8.2f). For the easier to handle spatial density distributions *e.g* the monotone function of test case 1 (see Figure 8.2a), the quantity $\tilde{E}_{min}$ is non-zero only for the estimate formed by the statistical rule (2.1.4) on the grid of $N = 3$ traps (see Table 8.1a). Test cases 5 and 6 as shown in Figure 8.2e and Figure 8.2f on the other hand require further grid refinement before $\tilde{E}_{min}$ becomes consistently zero. This happens for all numerical integration methods considered in the paper after the grid has been refined to $N = 33$ and $N = 65$ traps for test case 5 and test case 6 respectively.

Now let us consider the behaviour of the error quantities when the number N of traps is small. Figures 8.3a–8.3f confirm the findings of the previous chapter. It can be seen from Figure 8.3 that for small N the accuracy of an estimate of pest abundance formed from noisy data is determined by the accuracy of an estimate based on exact data. That is, for small N

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0.0057	0	0
5	0	0	0
9	0	0	0

(a)

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0.6088	0.6415	0.6686
5	0	0.1391	0.4180
9	0	0	0

(b)

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0.0819	0.2363	0.3724
5	0	0	0.1423
9	0	0	0
17	0	0	0

(c)

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0	0	0.1590
5	0	0	0.1742
9	0	0.1742	0
17	0	0	0

(d)

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0.0452	0	0
5	0.0600	0.0252	0.0118
9	0.1407	0.1308	0.1499
17	0.0679	0.0568	0.0170
33	0	0	0
65	0	0	0

(e)

N	$\tilde{E}_{min}^{stat}$	$\tilde{E}_{min}^{trap}$	$\tilde{E}_{min}^{Simp}$
3	0.9982	0.9980	0.9977
5	0.9984	0.9984	0.9985
9	0.9985	0.9985	0.9985
17	0.5949	0.5697	0.4267
33	0.3757	0.3562	0.2627
65	0	0	0

(f)

Table 8.1: The quantity $\tilde{E}_{min}$ for the test cases shown in Figure 8.2a – 8.2f respectively. The same superscript notation is used as in Figure 8.3. For the larger values of N where the values of $\tilde{E}_{min}$ are not displayed, they are uniformly zero.

the quantity $\tilde{E}_{rel}$ is strongly dependent on E_{rel} and is just slightly affected by the uncertainty $u(\tilde{I})$ caused by noise being present in the data. This is evident from the fact that in general the curves representing the mean error $\mu_{\tilde{E}_{rel}}$ of the estimates formed from noisy data lie close to their corresponding curves E_{rel} when N is small. In some cases the estimate of the pest abundance I_a based on exact data already achieves good levels of accuracy even when N is small. For example, for Test case 1 (see Figure 8.3a), this is evident for the estimates formed by implementing the composite Simpson's rule. Here, there is a clear difference between the quantities E_{rel}^{Simp} and $\mu_{\tilde{E}_{rel}^{Simp}}$ even on very coarse grids of $N = 3$ and $N = 5$ traps (compare the solid red line with closed triangles with the dashed red line). Whereas considering the estimates formed by the statistical rule on the same coarse grids, it can be seen that there is little difference between E_{rel}^{stat} and $\mu_{\tilde{E}_{rel}^{stat}}$ since the accuracy remains poor until the grid of traps is further refined (compare the solid green line with closed circles with the dashed green line).

Figures 8.3f and 8.4f exhibit the behaviour on grids with a small number N of traps whereby the quantities $E_{rel}, \mu_{\tilde{E}_{rel}}, \tilde{E}_{max}$ and also $\tilde{E}_{min}$ (see Table 8.1f) lie very close to each other. As discussed in the previous chapter this is the result of how we consider the noisy data $\tilde{f}_i$ to be related to the true population density values f_i. At the nodes of these grids the values of f_i are very small. Since we essentially consider the $\tilde{f}_i$ to be a percentage of the corresponding f_i, in this instance the noisy data will lie close to the true data.

We continue to consider the ecologically relevant scenario where the number of grid nodes N is small. Tables 8.2a - 8.2f provide further evidence to support the assertion that for each numerical integration method, the magnitude of $\mu_{\tilde{E}_{rel}}$ is mainly defined by E_{rel} on coarse grids of traps. In other words, the impact of the uncertainty $u(\tilde{I})$ in an estimate caused by noise in the density data is dominated by the error E_{rel} which is imparted by the means of obtaining an estimate *i.e.* the method of numerical integration. Tables 8.2a - 8.2f gives $u(\tilde{I})$, alongside the quantities E_{rel} and $\mu_{\tilde{E}_{rel}}$ for all test cases on the grids of $N = 3, 5$ and 9 nodes. For each fixed value of N, the uncertainties $u(\tilde{I})$ associated with an estimate are compared for each numerical integration method. The maximum of these uncertainties is given in bold, and the minimum is given in grey text. The same comparison is made for the relative errors E_{rel} of an estimate based on exact data, and the mean errors $\mu_{\tilde{E}_{rel}}$ of an estimate formulated from noisy data.

It is shown in Tables 8.2a–8.2f that the numerical integration method which yields the maximum or minimum value of $\mu_{\tilde{E}_{rel}}$ for a fixed number N of traps is the same as that which generates the maximum or minimum value of E_{rel} for all test cases. Therefore, the accuracy of an estimate based on exact data should be used to assess which method is superior when N is small. The tables also demonstrate the point made in Section 8.2, that there are instances when the uncertainty term $u(\tilde{I})$ is greater when the composite Simpson's rule is employed than when the composite trapezium rule and/or the statistical rule is used to estimate pest abundance. For example, this occurs for test case 2 when $N = 5$ as shown in Table 8.2b as well as for test case 3 when $N = 3$. Other examples of this happening can be seen in the remaining tables, as can examples of when the uncertainty associated with an estimate formed by the composite trapezium rule exceeds that associated with the corresponding statistical method estimate.

N	3	5	9
$u(\tilde{I}^{stat})$	**0.0121**	**0.0081**	**0.0055**
$u(\tilde{I}^{trap})$	0.0099	**0.0067**	**0.0049**
$u(\tilde{I}^{Simp})$	**0.0085**	0.0068	0.0054
E_{rel}^{stat}	**0.2838**	**0.1327**	**0.0643**
E_{rel}^{trap}	0.1409	0.0235	**0.0011**
E_{rel}^{Simp}	**0.0021**	**0.0156**	0.0063
$\mu_{\tilde{E}_{rel}^{stat}}$	**0.2839**	**0.1335**	**0.0668**
$\mu_{\tilde{E}_{rel}^{trap}}$	0.1428	0.0452	**0.0304**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.0521**	**0.0439**	0.0332

(a)

N	3	5	9
$u(\tilde{I}^{stat})$	**0.0042**	**0.0229**	0.0132
$u(\tilde{I}^{trap})$	0.0037	0.0285	**0.0148**
$u(\tilde{I}^{Simp})$	**0.0036**	**0.0379**	**0.0108**
E_{rel}^{stat}	**0.6681**	**0.3138**	**0.0048**
E_{rel}^{trap}	0.6948	0.5459	**0.0823**
E_{rel}^{Simp}	**0.7214**	**0.9595**	0.0723
$\mu_{\tilde{E}_{rel}^{stat}}$	**0.6681**	**0.3139**	**0.0502**
$\mu_{\tilde{E}_{rel}^{trap}}$	0.6948	0.5459	**0.0907**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.7214**	**0.9595**	0.0760

(b)

N	3	5	9
$u(\tilde{I}^{stat})$	**0.0403**	0.0244	**0.0170**
$u(\tilde{I}^{trap})$	0.0518	**0.0263**	**0.0262**
$u(\tilde{I}^{Simp})$	**0.0659**	**0.0182**	0.0176
E_{rel}^{stat}	**0.3701**	0.0628	0.0824
E_{rel}^{trap}	0.6069	**0.0526**	**0.0798**
E_{rel}^{Simp}	**0.8438**	**0.2725**	**0.0888**
$\mu_{\tilde{E}_{rel}^{stat}}$	**0.3701**	0.0711	0.0830
$\mu_{\tilde{E}_{rel}^{trap}}$	0.6069	**0.0666**	**0.0806**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.8438**	**0.2725**	**0.0893**

(c)

N	3	5	9
$u(\tilde{I}^{stat})$	**0.0407**	**0.0255**	**0.0175**
$u(\tilde{I}^{trap})$	0.0306	0.0177	**0.0145**
$u(\tilde{I}^{Simp})$	**0.0205**	**0.0157**	0.0164
E_{rel}^{stat}	0.2205	**0.0414**	**0.0369**
E_{rel}^{trap}	**0.0568**	0.2455	**0.1384**
E_{rel}^{Simp}	**0.3341**	**0.3084**	0.1027
$\mu_{\tilde{E}_{rel}^{stat}}$	0.2230	**0.0670**	**0.0501**
$\mu_{\tilde{E}_{rel}^{trap}}$	**0.0837**	0.2455	**0.1384**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.3341**	**0.3084**	0.1031

(d)

N	3	5	9
$u(\tilde{I}^{stat})$	**0.0190**	**0.0150**	**0.0105**
$u(\tilde{I}^{trap})$	0.0210	0.0160	0.0108
$u(\tilde{I}^{Simp})$	**0.0248**	**0.0176**	**0.0110**
E_{rel}^{stat}	**0.2009**	**0.1828**	0.2268
E_{rel}^{trap}	0.1579	0.1567	**0.2193**
E_{rel}^{Simp}	**0.1148**	**0.1563**	**0.2401**
$\mu_{\tilde{E}_{rel}^{stat}}$	**0.2009**	**0.1828**	0.2268
$\mu_{\tilde{E}_{rel}^{trap}}$	0.1580	0.1567	**0.2193**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.1173**	**0.1563**	**0.2401**

(e)

N	3	5	9
$u(\tilde{I}^{stat})$	**1.30e-06**	**9.53e-07**	**6.87e-06**
$u(\tilde{I}^{trap})$	1.54e-06	1.03e-06	**7.13e-07**
$u(\tilde{I}^{Simp})$	**1.89e-06**	**1.05e-06**	7.41e-07
E_{rel}^{stat}	**0.99852**	0.99867	0.998711
E_{rel}^{trap}	0.99833	**0.99861**	**0.99869**
E_{rel}^{Simp}	**0.99815**	**0.99871**	**0.998714**
$\mu_{\tilde{E}_{rel}^{stat}}$	**0.99852**	0.99867	0.998711
$\mu_{\tilde{E}_{rel}^{trap}}$	0.99833	**0.99861**	**0.99869**
$\mu_{\tilde{E}_{rel}^{Simp}}$	**0.99815**	**0.99871**	**0.998714**

(f)

Table 8.2: Contributions to the error of an estimate calculated from noisy data. For grids of a small number of traps and for each of the numerical integration methods (2.1.4)-(2.2.7), the quantities $u(\tilde{I})$, E_{rel}, and $\mu_{\tilde{E}_{rel}}$ are compared. The same superscript notation is used as in Figure 8.3. For a fixed value of N, the greatest of each quantity is highlighted in bold and the lowest is given in grey text. The position of the bold/grey text in the $\mu_{\tilde{E}_{rel}}$ row matches that of the E_{rel} row.

Accuracy control on coarse grids remains, perhaps, the most difficult issue in the general problem of pest abundance evaluation since the asymptotic error estimates do not hold. As explained in Chapters 5 and 6, on coarse grids the approximation error can be considered a random variable and we have thus recommended that a method of numerical integration should instead be assessed probabilistically. In other words, the probability of obtaining a desired level of accuracy should be calculated rather than the error of an estimate. The initial methodology we presented for such an assessment was for exact data only, however the findings of this paper indicate that the results would also apply for noisy data as the effects of noise can be ignored on coarse grids.

8.4 Chapter 8 Conclusions

The study of pest abundance evaluation in the presence of noise conducted in the previous chapter has been extended to consider other methods of numerical integration. The performance of the composite trapezium rule (2.2.6) has been compared with that of the statistical rule (2.1.4) and the Simpson's rule (2.2.7). For each method we have studied the behaviour of the mean error $\mu_{\tilde{E}_{rel}}$ arising when pest abundance is evaluated from noisy data. We have also investigated the credible interval $[\tilde{E}_{max}, \tilde{E}_{max}]$ to which the error of evaluation belongs with a given probability $P(z)$.

We have considered the cases of a large number of traps and a small number of traps separately as different accuracy criteria should be applied in the former and latter case. If the number of traps is large enough such that the grid of traps is fine, the methods of numerical integration can be compared based on their asymptotic convergence rate. It has been demonstrated that for each of the evaluation methods considered, the mean error of an estimate formed from noisy data converges to zero at the same rate. This is despite the fact that the methods have different convergence rates when applied to exact data. This confirms the findings of the previous chapter, that noise becomes a dominant feature of the approximation when the number N of traps is large. This conclusion, however, does not immediately result in the recommendation to dismiss more advanced (and therefore more accurate on exact data) methods of pest abundance evaluation for the sake of methods less accurate yet easy to implement. As noted in Section

8.3, it becomes more probable that a smaller rather than a larger error will be obtained as the error E_{rel} of an estimate formed from exact data tends to zero. Hence a method of numerical integration with a higher convergence rate when applied to exact data may still be superior when perturbed data are considered on fine grids of traps. Further study is required to verify this assertion.

When the number N of traps is small and the grid of traps is coarse, the asymptotic error estimates do not hold. Generally, the mean error and the bounds of the credible interval are determined by the uncertainty $u(\tilde{I})$ associated with the estimate formulated from measured pest densities as well as the error E_{rel} of an estimate formulated from exact values of the pest population density. We have shown that on coarse grids the uncertainty depends on the spatial pattern of the density function. Hence, any a priori knowledge about the density distribution could be helpful in deciding which integration method should be used to reduce the uncertainty of the evaluation.

Meanwhile, it was found that on coarse grids the most significant contribution to the error of an estimate $\tilde{I}$ formed from measured data is the quantity E_{rel} *i.e.* the error imparted by the numerical integration method. Our numerical experiments confirmed that the impact of the uncertainty $u(\tilde{I})$ was negligible in many ecologically meaningful test cases where E_{rel} was large. Thus finding a method of evaluating pest abundance which is accurate on coarse grids when exact density data are considered is critical to the pest monitoring problem.

Chapter 9

Concluding Remarks

9.1 Discussion and Conclusions

We have considered the ecological monitoring problem of estimating pest insect abundance in an agricultural field. Such an estimate can be used to decide whether or not it is necessary to intervene and implement a control action to reduce the pest population size. Means of pest control, the most commonly used being pesticides, are costly, time consuming to administer, and can have damaging effects on the environment and the ecosystem. At the same time, the damage caused to crops by pests is significant and needs to be prevented as much as possible. The importance of making a correct pest management decision is clear and a more accurate pest abundance estimate means greater confidence can be placed in the resulting pest management decision.

Typically, a statistical approach is used to obtain an estimate of pest abundance. In this thesis, we have instead built upon the work conducted in [70, 71, 72] and applied numerical integration to the abundance evaluation problem. An estimate is formed by numerically integrating the discrete pest population density function obtained by sampling the population. The motivation behind applying numerical integration methods in the problem is that their weight coefficients can be chosen such that the asymptotic convergence rate exceeds that of the statistical method. Such methods thus have the potential to yield more accurate results.

Restrictions imposed by the nature of the ecological scenario, however, mean that this differs

greatly from a standard numerical integration problem. Whereas conventionally it is assumed that the number of data points can be made large, in pest monitoring there is a practical limit to the number N of sampling locations which can be used within a single field, and this number can be particularly small in the case of routine monitoring. Thus we are forced to seek an accurate approximation from a very limited amount of data.

We reviewed the results of [70, 71, 72]. It was explained that when the number of data points where the integrand is available is small, the asymptotic error estimates cannot be relied upon. Computational grids for which the error does not behave according to these estimates are described as coarse grids, meanwhile grids where the asymptotic error estimates hold are considered to be fine grids. These definitions depend on the nature of the integrand, rather than explicitly the number of data points. A grid that is considered as fine for one function may be coarse for another. The transition point from coarse to fine grids depends on the spatial heterogeneity of the integrand *i.e.* the pest population density function. The asymptotic error estimates were shown to hold even for a small number of data points when the population is spread homogeneously over the entire domain. For such a spatial distribution of pests, numerical integration methods with a higher convergence rate will produce a more accurate estimate of abundance. Meanwhile a higher number of grid nodes is required for the asymptotic error estimates to hold when a more heterogeneous population density distribution is considered. This number of nodes may be too large to be used in practice, and thus the coarse grid problem arises. A consequence of the asymptotic error estimates not holding on coarse grids is that the conventional approach of using a method with a higher convergence rate to improve accuracy may not be effective.

Some other typical techniques used in standard numerical integration problems to improve accuracy were not available to us. With some prior knowledge of the spatial pattern of the pest population, the accuracy of an estimate could be improved by adapting the sampling plan accordingly. For instance more data could be collected in areas that required a higher level of resolution. However, such information is not usually available thus we made no assumptions about the spatial heterogeneity of the pest population density function. Furthermore, our focus

was on the accurate estimation of the population size at a fixed point in time. As such we could not use grid refinement or adaptive grid movement to improve the accuracy. In pest monitoring, both of these techniques would require repeating the sampling procedure either with an increased number of sample units in the former case, or with a different spatial configuration of sample units (*i.e.* a different sampling plan) in the latter. Clearly the initial conditions cannot be recreated; pest insects are living creatures and the population density function changes over time. The time parameter would have to be taken into account to permit the use of the aforementioned techniques but this was outside the scope of our study. Instead we considered the time independent problem. It is worth noting, however, that grid refinement and adaptive grid movement could be applied whilst still considering the problem as time independent if the target pest species is sufficiently slow moving. For example in the case of weed species, the population density distribution will not change significantly between rounds of sampling. Grid refinement or grid adaptation techniques, widely used in conventional mathematical integration problems, could then be exploited to improve the accuracy of the abundance estimate.

The accuracy that can be achieved for a fixed, small number of data points also depends on the spatial heterogeneity of the integrand *i.e.* the pest population density function. Accurate estimates can be obtained when the density is spread over the entire domain, even if the density function is heterogeneous. More inaccurate estimates are expected when the density is patchy (distinct areas of zero and non-zero density) with the situation worsening the higher the severity of the aggregation as the patches of non-zero density may be missed by the sampling procedure. We aimed to investigate how the accuracy on coarse grids could be assessed, since the asymptotic error estimates conventionally used to conclude about performance do not hold. We also embarked on a search for means of improving the accuracy of estimates on coarse grids under the conditions outlined above.

Though the precise spatial pattern of the pest population itself is not usually known, estimates of the rate at which the population diffuses exist for several species. It was explained how the diffusion rate is a defining parameter of the spatial pattern of the population density function and how it can be used to estimate the characteristic width of spatial heterogeneity. We then

investigated how knowledge of the diffusion rate could be exploited to control the integration error so that it is within a prescribed tolerance.

For highly aggregated population density distributions we found that on coarse grids not only do the asymptotic error estimates not hold, but the integration error behaves as a random variable. An abundance estimate may still achieve a prescribed accuracy level, but it is a matter of chance. We introduced a new class of computational grid, namely an ultra-coarse grid, where the probability of achieving an error within a certain tolerance is $p < 1$. By modelling a component of spatial heterogeneity, *i.e.* a single peak, as a quadratic function and considering the integration error as a random variable, we were able to find an estimate of the number N_t of regularly spaced data points required to ensure sufficient accuracy in terms of the species diffusion rate. We were also able to describe the probability $p < 1$ of obtaining a desired accuracy level when the number of data points is $N < N_t$. We have proposed that such a probabilistic assessment of accuracy should be used to compare methods of numerical integration on ultra-coarse grids where the method with the highest probability of achieving the desired accuracy would be recommended.

We then extended our probabilistic study of ultra-coarse grids to handle random sampling plans. The statistical means of estimating pest abundance is spatially implicit and permits such a distribution of sample units. Furthermore, whilst a regular sampling plan is often used in ecological monitoring, a random sampling plan is sometimes favoured instead. The concern is that a non-random sampling plan will correlate with the spatial pattern of the population distribution and thus the abundance estimate will be biased. Although sampling randomly works well when the population is spread over the entire field, we showed that a regular sampling plan may be better suited to handling highly aggregated populations.

We considered the scenario whereby the entire population is concentrated within a small patch of the field. Under the assumption that a single sample unit lies within the patch of non-zero density, we constructed a theoretical prediction of the probability of achieving an accurate abundance estimate for a given number N of randomly located sample units, and found the critical number N^* of sample units for which this probability achieves its maximum value.

Increasing N beyond this point leads to a decrease in the chance of estimating the abundance accurately. This is contrary to the behaviour when a regular sampling plan is used, where an increase in N ultimately leads to sufficient accuracy being guaranteed. Furthermore, on comparison of our theoretical predictions, for a fixed number N the probability of obtaining an estimate within a prescribed tolerance of the true abundance was higher for a regular distribution than a random distribution of sample units.

The above conclusions were reached under the assumption that the pest population density data at the sample unit locations are precise. As discussed in the introduction, there are sampling techniques which produce an absolute estimates of abundance at the sample unit location thus a true reflection of the local population density can be recovered. Such methods, however, are labour intensive and time consuming. More commonly, more cost-effective sampling methods which produce a relative abundance estimate at the sample unit location are used instead. For example trapping is an often used sampling technique, however, the trap counts are relative to trapping technique. We considered a passive trapping process *i.e.* no attractant was used to draw the target species into the trap. Consequently, the trap counts are relative to the activity of the target pest as well as the density. Means of converting such counts to the population density exist, however the resulting density data are of course not exact. The data are instead measurements which include a measurement error. A key component of this measurement error is the random error (noise). We thus conducted a study of the impact of this noise on an estimate of pest abundance and compared it to the error imparted by the numerical integration process.

Our assumption that the resulting measurement of pest population density is a realisation of a normal random variable with the mean equal to the true density at the trap location means that the relative approximation error in the presence of noise is a realisation of a folded normal distribution. We constructed a confidence interval for the error of an estimate formulated from noisy data and then assessed the impact of noise on the accuracy of an abundance estimate by investigating the behaviour of the lower and upper limits of this confidence interval as well as the mean of the error. It was found that the convergence rate of a numerical integration method is significantly slowed down in the presence of noise. As such, when the number of traps is large,

noise is the dominant feature of the evaluation error. Meanwhile when the number of traps is small and the error attributed to the numerical integration of the data is large, noise in the data makes little impact and the matter of reducing the integration error on coarse grids remains the main concern.

It is evident from our research that any a priori information regarding the spatial distribution of the pest population should be used to its fullest extent. Based on the findings presented in this thesis, in the case of estimating the pest abundance from precise density data, we recommend the use of higher order numerical integration methods if the number of sample units is sufficient to resolve the spatial heterogeneity of the population distribution as the conventional asymptotic error estimates hold. For the small number of sample units used in pest monitoring programmes, this corresponds to when the pest population is known to be spread over the entire agricultural field. As indicated in Section (8.4), more research needs to be done before a recommendation can be made regarding which numerical integration method should be used when the population is spread over the entire domain but the density data is noisy. Likewise, further research is required to determine the best means of evaluating pest abundance when the population is known to be aggregated. In this case it is a matter of chance as to whether the data from the small number of sample units yields a sufficiently accurate abundance estimate. We suggest our probabilistic means of assessing a numerical integration method on such ultra-coarse grids should be extended so that more methods can be compared and a recommendation as to which one to implement can be made. Our study suggests that the same hierarchy could then be applied when the density data is noisy, as the integration error is dominant when the number of sample units is small and the population density distribution is patchy, however this needs to be confirmed through further research. Suggested extensions to the work conducted in this thesis are provided in the next section.

9.2 Directions for Future Work

We now discuss the limitations of the work conducted in this thesis and the avenues of inquiry left open. Furthermore, we outline the next steps which could be taken en route to providing more concrete recommendations to be used in the practice of pest monitoring.

We have demonstrated how to yield an estimate of pest abundance from real-world multi-dimensional pest population density data, and results for the 2D problem have been presented. For parts of our study we have reduced the problem to 1D for simplicity, however, these results can readily be extended to two dimensions. Numerical tests presented in this thesis have been performed using simulated data. The numerical integration of field data has been discussed in [68, 71], thus further testing could henceforth be conducted using field data. We have considered a simplified representation of an agricultural field, namely a rectangular domain. The study could be extended to consider a curvilinear boundary.

The probabilistic assessment of the error induced by numerical integration in Chapters 4 and 5 was founded on treating a peak (local maximum) of the population density function as a quadratic. As such, we were able to evaluate the probability of achieving a sufficiently accurate estimate, and also estimate the minimum number of sample units needed to guarantee a prescribed level of accuracy, but only for numerical integration methods with a degree of precision r less than the order of a quadratic *i.e.* for $r < 2$. To consider other numerical integration methods with a higher degree of precision, a higher order replacement of the peak should be used in the analysis. A comparison between the performance of the often used statistical method 2.1.4 and an array of higher order numerical integration methods should then be made in order to provide recommendations of which method to use on ultra-coarse computational grids.

We made an initial investigation of how locally shifting the sample unit locations away from a regular grid affects the accuracy of numerical integration and the minimum number of sample units needed to guarantee the error is within a chosen tolerance (see Chapter 3, Section 3.4 and Chapter 4, Section 4.5). The mean error produced by an irregular sampling plan was compared to the error of an estimate obtained on a regular grid of sample units. The variance of the error should also be studied to fully understand the impact of perturbations on the prescribed sampling plan. This investigation should be extended to consider how the probability of achieving an accurate estimate on ultra-coarse regular grids is affected.

As mentioned in the introduction, other estimates exist of the minimum number of sample units required for sufficient accuracy. Extending the probabilistic study of Chapters 4 and 5

to handle the 2D problem would mean that a comparison could then be made between these estimates.

In Chapter 6 we evaluated the probability of achieving a sufficiently accurate estimate on implementing the statistical method (2.1.4) when a random sampling plan is used. We arrived at our theoretical predictions by considering a highly aggregated density distribution where the entire population is confined within a small patch of the field. We also simplified the analysis by considering the event that a single sample unit is located randomly within the patch of non-zero density. For completeness, the cases when more than one sample units fall into this patch should be incorporated into the probability evaluation, however as was explained in Section 6.5, it is expected this additional contribution will be small in comparison. The study should, of course, also be extended to handle multi-patch distributions.

The impact of noise in the population density data obtained via passive trapping on the accuracy of abundance estimation was studied in Chapters 7 and 8. The impact was assessed by considering the mean error, and by constructing a confidence interval. The results on coarse grids were clear; the impact of noise in the density data is negligible in comparison to the error imparted by the numerical integration method. As mentioned in Section 8.4, however, further investigation is required in order to make a recommendation of which numerical integration method to use on fine grids when noise is present. A means of making such a recommendation would be to utilise equation (7.2.8) to calculate the standard deviation of the error when noise is present, alongside the mean, and then subsequently calculate the skewness of the probability density distribution. The method of numerical integration with the stronger positive skew, i.e. the mass is skewed more towards zero error would then be the recommended method.

We also assumed that the trap counts were uncorrelated however, in reality there may be some correlation between trap counts if the traps are installed sufficiently close to each other and this should be investigated. Furthermore, we conducted this study for the 1D counterpart of the pest monitoring problem; similar analysis should be carried out in 2D. The systematic component of the density measurement error was ignored, as mentioned in Section 7.1. Its contribution to the abundance estimation error should be evaluated and compared to the error

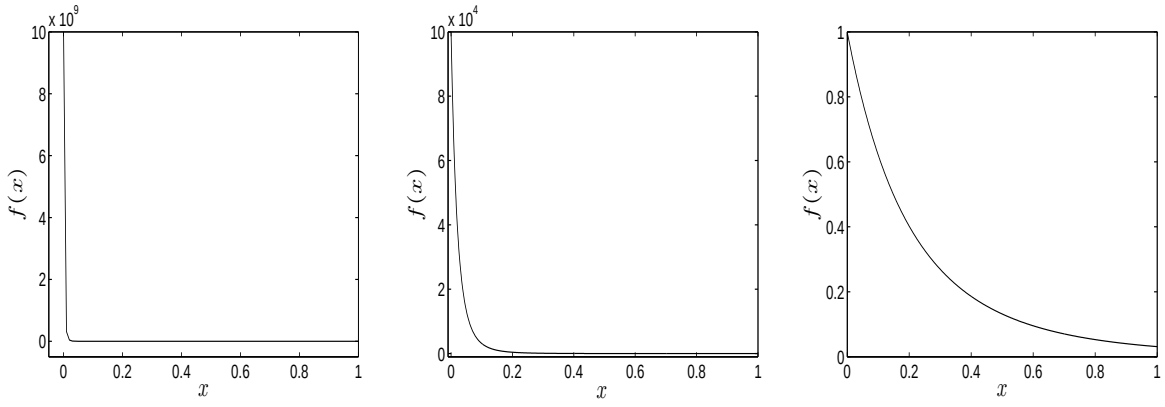
imparted by the numerical integration procedure.

Our work has focused on assessing the accuracy of a pest abundance estimate at a fixed point in time. Once the outstanding issues outlined above have been resolved, the time dependent problem should be considered.

We have considered numerical integration as a means of formulating an abundance estimate. Numerical integration is based on (piecewise) polynomial interpolation of the integrand. An alternative interpolating technique used in abundance estimation is kriging. Kriging approximates the data statistically (predominantly using least-squares). While it is widely employed in ecological application, its potential in terms of accuracy is still unclear when data used for kriging is sparse. Results of our study reveal that the accuracy on coarse grids should be formulated in probabilistic rather than deterministic terms, thus the use of the kriging technique on sparse data should be revisited. This may constitute another topic of future research.

We have demonstrated that numerical integration techniques can be used to evaluate pest abundance however we do not proclaim to have presented a method ready to use in pest monitoring. As indicated above, much further research needs to be conducted. Instead, we have made some important first steps towards achieving this ultimate objective.

Appendix A



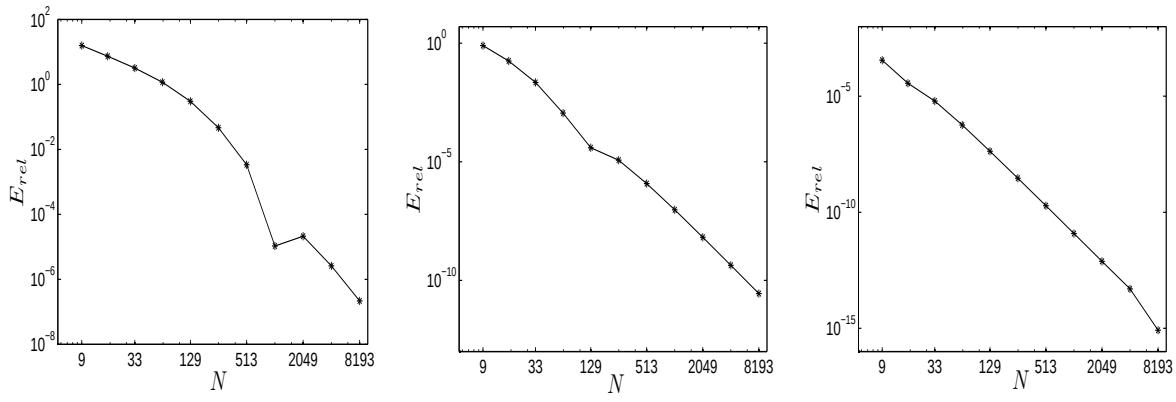
(a) (b) (c)

Figure A.1: The pest population density is described by the function A.0.1 where (a) $\delta = 0.01$, (b) $\delta = 0.1$ and (c) $\delta = 1$.

In Chapter 3, Section 3.1 several numerical integration methods were applied to test cases with mathematically interesting characteristics over a series of regular grids with an increasing number N of nodes. On plotting the convergence curves for the test case described by function (3.1.3), namely

$$f(x) = \frac{1}{(x + 0.01)^5}, \qquad x \in [0, 1]$$

a jump in accuracy was evident for LII method approximation when the interpolating polynomial degree was set as $k = 3$. For convenience both the function and relevant convergence curve are redrawn in figures A.1a and A.2a respectively. We now investigate the jump in accuracy further.



(a) (b) (c)

Figure A.2

Let us consider a more generalised version of the above function, namely,

$$f(x) = \frac{1}{(x+\delta)^5}, \tag{A.0.1}$$

where on setting $\delta = 0.01$ we obtain the original function of interest (3.1.3) restated above. Plots of this function for different values of δ are shown in Figure A.1b where we have chosen $\delta = 0.1$ and in Figure A.2b where $\delta = 1$. Corresponding convergence curves of estimates generated by the LII method for $k = 3$ are shown in Figure A.2b and A.2c. Comparing the graphs of Figure A.2 it can be seen that as δ is increased the jump becomes less severe. Its position is also related to the order of δ. When $\delta = 0.01$ the jump occurs when n is of the order 10^3. When δ is ten times larger, $\delta = 0.1$ the number of sub-intervals n corresponding to the jump decreases tenfold to become of order 10^2. On increasing the value of δ by a factor of ten once again, that is we now have $\delta = 1$, there is no longer a jump visible. We are thus satisfied that the jump in accuracy is a result of cancellation caused by the geometry of the interpolation stencil, (see Section 2.3).

Appendix B

The analysis conducted in chapters 4 and 5 uses the assumption that we can treat a peak (a local maximum) as a quadratic function. In making this assumption, however, we introduce an error. Here we investigate the significance of this error. Suppose that the population density function $f(x)$ has a local maximum at the point x_i which lies somewhere on the unit interval $[0, 1]$. Let the population density function value at the point x_i be available, as well as at the points $x_{i-1} = x_i - h$ and $x_{i+1} = x_i + h$ where $h > 0$ is an arbitrary parameter. Knowing the function value at three nodes allows us to construct a quadratic interpolating polynomial to approximate the peak. Such interpolation was discussed in Section 2.2. The coefficients of the quadratic are usually found by the condition that at three nodes $f(x) = Q(x)$. We require that the maximum of the quadratic coincides with the maximum of the peak, thus we define the quadratic to be of the following form

$$Q(x) = B - A(x - x_i)^2$$

and find the coefficients A and B using only two collocation conditions, namely $f(x_{i-1}) = Q(x_{i-1})$ and $f(x_i) = Q(x_i)$. As such, we have

$$A = \frac{f(x_i) - f(x_{i-1})}{h^2}, \qquad B = f(x_i).$$

Let us investigate the error we introduce to our analysis by treating the peak as a quadratic

h	0.125	0.0625	0.0312	0.0156
E_{int}^{max}	0.3325	0.0922	0.0123	0.0011

Table B.1: Maximum interpolation error (B.0.1) for the quadratic approximation the single peak of $f_1(x)$ for various values of h.

function. To do this we consider the maximum interpolation error, that is the maximum distance between the actual density function $f(x)$ and the quadratic replacement $Q(x)$ within the vicinity of the peak $[x_{i-1}, x_{i+1}]$

$$E_{int}^{max} = \max_{x \in [x_{i-1}, x_{i+1}]} |f(x) - Q(x)|. \tag{B.0.1}$$

Table B.1 shows the maximum interpolation error introduced when the single peak of the function $f_1(x)$ shown in Figure 4.1a is approximated by a quadratic for various values of h. The maximum of the peak is located at $x = 0.255249$ thus this is taken as the value of the node x_i. We recall that the function $f_1(x)$ is discrete and available at the nodes of a very fine regular grid of $N_f = 32,769$ nodes on the unit interval $[0, 1]$. Thus the maximum interpolation error (B.0.1) was calculated over the set of points belonging to the peak vicinity namely the sub-interval $[x_{i-1}, x_{i+1}]$. The table shows that a decrease in the parameter h results in a decrease in the interpolation error (B.0.1). This is also evident from Figure 4.2.

We now look at the difference in the integration error when the actual density function $f(x)$ and the quadratic replacement $Q(x)$ are integrated in the vicinity of the peak $[x_{i-1}, x_{i+1}]$. The chosen means of numerical integration is the compound trapezium rule (2.2.6), *i.e.* the peak is interpolated by linear polynomials on the local grid of three nodes $x_{i-1} = x_i - h, x_i$ and $x_{i+1} = x_i + h$ for a fixed value of h as detailed in Section 4.2. The relative integration error (2.1.7) is calculated in the vicinity of the peak for the function $f(x)$ and the quadratic replacement $Q(x)$ and are denoted E_{rel}^{f} and E_{rel}^{Q} respectively. The exact integral of the peak used in the computation of these error terms is obtained by applying the composite trapezium rule (2.2.6) to the subset of the $N_f = 32,769$ data points which lie in the peak domain $[x_{i-1}, x_{i+1}]$.

Table B.2 shows the integration error (2.1.7) when the integral is calculated in the vicinity of the peak for both the density distribution $f_1(x)$ shown in Figure (4.1a), and the quadratic

h	0.125	0.0625	0.0312	0.0156
E^f_{rel}	0.0641	0.0464	0.0279	0.0091
E^Q_{rel}	0.1839	0.0961	0.0341	0.0096

Table B.2: The integration error (2.1.7) when the approximate integral I_a is computed on a grid of three nodes local to the peak using the compound trapezium rule (2.2.6). The integration errors E^f_{rel} and E^Q_{rel} computed for the peak of the density distribution $f_1(x)$ and its quadratic replacement $Q(x)$ are given for several values of h.

h	0.0312	0.0156	0.0078	0.0039
E^f_{rel}	0.0545	0.0532	0.0267	0.0080
E^Q_{rel}	0.1986	0.1024	0.0325	0.0086

Table B.3: The integration error (2.1.7) when the approximate integral I_a is computed on a grid of three nodes local to the peak using the compound trapezium rule (2.2.6). The integration errors E^f_{rel} and E^Q_{rel} computed for the first peak of the density distribution $f_2(x)$ and its quadratic replacement $Q(x)$ are given for several values of h.

approximation of the peak $Q(x)$. Table B.3 shows corresponding integration errors local to the first peak of the multi-peak population density function $f_2(x)$ shown in Figure 4.1b. In both cases we consider that E^Q_{rel} lies sufficiently close to E^f_{rel}. Thus we proceed with the approach of treating a peak as a quadratic in our analysis.

Appendix C

Here, we give the details of the probability analysis conducted in Chapter 6 for the 2D case. The procedure is similar to that used for the 1D case shown in Section 6.2. Consider N sample units installed over the domain $D = [0, 1] \times [0, 1]$, where we assume that only one sample unit is located within the peak sub-domain D_u, and any other units fall outside D_u, where the density distribution is zero. The location of this single sample unit is denoted $\mathbf{r}_0 = (x_0, y_0)$, and is parameterised as

$$x_0 = r\cos\theta + x^*, \quad y_0 = r\sin\theta + y^*, \tag{C.0.1}$$

where $r \in [0, R]$ and $\theta \in [0, 2\pi]$. The location of $\mathbf{r}_0$ is randomised by considering r and θ as uniformly distributed random variables. The population density at this location, written as $u(x_0, y_0) \equiv u_0$, is then calculated as

$$f_0 \approx Q(x_0, y_0) = B - A\left((x_0 - x^*)^2 + (y_0 - y^*)^2\right) = A(R^2 - r^2), \quad r \in [0, R], \tag{C.0.2}$$

where we have used the fact that $B = AR^2$. The mean density $M(N)$ is then

$$M(N) = \frac{f_0}{N} = \frac{A(R^2 - r^2)}{N}.$$

The true mean density $\bar{M}$ is computed as

$$\bar{M} = \int_0^1 \int_0^1 f(x,y)\,dx\,dy = \frac{1}{2}A\pi R^4.$$

Once again we require that the error (6.2.3) is sufficiently small, therefore we impose condition (6.2.4). From the above values of $M(N)$ and $\bar{M}$ we obtain

$$\frac{(1-\tau)A\pi R^4}{2} \leq \frac{A(R^2-r^2)}{N} \leq \frac{(1+\tau)A\pi R^4}{2}. \tag{C.0.3}$$

Let us first consider the upper limit in (C.0.3), namely

$$\frac{A(R^2-r^2)}{N} \leq \frac{(1+\tau)A\pi R^4}{2}. \tag{C.0.4}$$

Solving for r we obtain

$$r \geq r_I = R\sqrt{1-\frac{N(1+\tau)\pi R^2}{2}}, \tag{C.0.5}$$

where r_I exists for

$$N \leq N_{2D}^* = \frac{2}{(1+\tau)\pi R^2}. \tag{C.0.6}$$

We now consider the inequality

$$\frac{(1-\tau)A\pi R^4}{2} \leq \frac{A(R^2-r^2)}{N}. \tag{C.0.7}$$

After some rearrangement we arrive at

$$r \leq r_{II} = R\sqrt{1-\frac{N(1-\tau)\pi R^2}{2}}. \tag{C.0.8}$$

The limit r_{II} exists when

$$N \leq N_{2D}^{**} = \frac{2}{(1-\tau)\pi R^2}. \tag{C.0.9}$$

As $\tau \in (0,1)$, the number $N_{2D}^* < N_{2D}^{**}$. We consequently have three cases to consider when evaluating the probability $p(N)$ that the error (6.2.3) is within the prescribed tolerance τ.

Case 1: $N \leqslant N_{2D}^{*}$.

For this range of N, both r_I and r_{II} exist. Therefore the admissible range of the parameter r is

$$r_I \leq r \leq r_{II}. \tag{C.0.10}$$

Since r is a uniformly distributed random variable the probability $p(N)$ of $M(N)$ being sufficiently close to the true mean density $\bar{M}$ can be computed as

$$p_I(N) = (r_{II} - r_I)/(r_{max} - r_{min}) = (r_{II} - r_I)/R,$$

where $r_{min} = 0$ and $r_{max} = R$. From (C.0.5) and (C.0.8) we thus have

$$p(N)_I = \sqrt{1 - \frac{N(1-\tau)\pi R^2}{2}} - \sqrt{1 - \frac{N(1+\tau)\pi R^2}{2}}.$$

Case 2: $N_{2D}^{*} < N \leqslant N_{2D}^{**}$.

In this instance, r_I no longer exists, but the inequality (C.0.4) always holds. Therefore the lower limit in (C.0.10) should be replaced by $r_{min} = 0$. The admissible range now becomes $0 \leq r \leq r_{II}$, therefore the probability $p(N)$ is described by

$$p_{II}(N) = (r_{II} - 0)/(r_{max} - r_{min}) = r_{II}/R.$$

Substituting in the values for r_I and r_{II} we arrive at

$$p_{II}(N) = \sqrt{1 - \frac{N(1-\tau)\pi R^2}{2}}.$$

Case 3: $N > N_{2D}^{**}$.

When the number of sample units N exceeds the limit N_{2D}^{**}, neither r_I nor r_{II} exist, and the inequalities (C.0.4) and (C.0.7) never hold. There is thus no admissible range of r for this range of N. The probability that the error (6.2.3) is sufficiently small is then $p_{III}(N) = 0$.

Appendix D

We seek the upper and lower limit of the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ to which the quantity $\tilde{E}_{rel}$ belongs with probability $P(z)$ given by (7.1.9) as discussed in Section 7.2 of Chapter 7. 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 (7.2.2). 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 $\tilde{E}_{rel}$. 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. D.1(a), 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:

$$\tilde{E}_{min} = 0. \tag{D.0.1}$$

Case 2: $|I - I_a| > z\sigma_{\tilde{I}}$

In this instance, from Fig. D.1(b) we can see that the range $[I_a - z\sigma_{\tilde{I}}, I_a + z\sigma_{\tilde{I}}]$ does not include

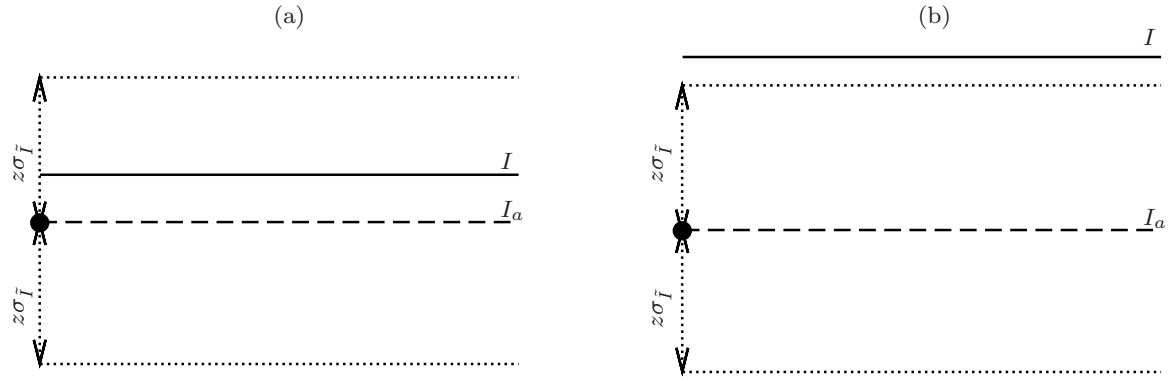


Figure D.1: Finding the interval $[\tilde{E}_{rel}, \tilde{E}_{max}]$ to which $\tilde{E}_{rel}$ belongs with probability $P(z)$. (a) Case 1: $|I - I_a| \leq z\sigma_{\tilde{I}}$. In this case, the exact value of the integral I lies within the credible interval for $\tilde{I}$ thus the lower limit of the credible interval for $\tilde{E}_{rel}$ 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.

the exact value of the integral I. Either we have $I_a \leq I$ 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

$$\tilde{E}_{min} = E_{rel} - \frac{z\sigma_{\tilde{I}}}{I}, \tag{D.0.2}$$

which is a strictly positive quantity as the condition $|I - I_a| > z\sigma_{\tilde{I}}$ of course means that $E_{rel} > z\sigma_{\tilde{I}}/I$, 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 $\tilde{E}_{rel}$ as

$$\tilde{E}_{min} = \begin{cases} min\left\{E \geq 0 : E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E]\right\}, \text{for } \mu_E \geq 0, \\ \\ \left|max\left\{E \leq 0 : E \in [\mu_E - z\sigma_E, \mu_E + z\sigma_E]\right\}\right|, \text{for } \mu_E < 0 \end{cases}$$

where E is defined by (7.2.3). In other words we find the value of the quantity E closest to zero which lies within the range (7.2.5) and then take the absolute value as $\tilde{E}_{min}$ (see Figure 7.2).

Let us now consider the upper limit $\tilde{E}_{max}$ of the credible interval of $\tilde{E}_{rel}$. 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 (7.1.9). As mentioned above, $\tilde{E}_{rel}$ is a realisation of a random variable with a folded normal distribution. This distribution is formed by reflecting the negative quantities of the distribution (7.2.4) 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 $\tilde{E}_{rel}$ 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 Figure 7.2a the probability P^* is given by

$$P^* = \int_{-\mu_E - z\sigma_E}^{\mu_E - z\sigma_E} p(E)\, dE. \tag{D.0.3}$$

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

$$\int_{\tilde{E}_{max}}^{\mu_E + z\sigma_E} p(E)\, dE = P^*. \tag{D.0.4}$$

Using the transformation

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

from (D.0.3)and (D.0.4) 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

$$\tilde{E}_{max} = \mu_E + \sigma_E \Phi^{-1}\left[2\Phi(z) - \Phi\left(z + 2\frac{\mu_E}{\sigma_E}\right)\right]. \tag{D.0.5}$$

Case 2: $|I - I_a| > z\sigma_{\tilde{I}}$

Similar calculations for this case as illustrated in Figure 7.2b yield

$$\tilde{E}_{max} = \mu_E + \sigma_E \Phi^{-1}\left[\Phi(z) - \Phi\left(z - \frac{2\mu_E}{\sigma_E}\right) - \Phi\left(z + \frac{2\mu_E}{\sigma_E}\right) + 1\right], \tag{D.0.6}$$

Earlier we assumed $\mu_E \geq 0$. Since the probability density function (7.2.6) 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 equations (D.0.5) and (D.0.6) so that they hold for arbitrary μ_E.

Appendix E

The means of assessing the impact of noise on a pest abundance estimate discussed in section Chapters 7 and 8 were founded on the assumption that with probability $P(z) = \text{erf}(z/\sqrt{2})$ a measured pest density $\tilde{f}_i$ is within a fixed percentage ν of the true pest density. That is, we assume that with probability $P(z)$, each $\tilde{f}_i$ satisfies the condition

$$\tilde{f}_i \in [(1-\nu)f_i, (1+\nu)f_i]$$

where $\nu \in (0,1)$. The counterpart of this assumption of course is that there is a chance that each $\tilde{f}_i$ could lie outside of this range. The theory therefore allows for $\tilde{f}_i, i = 1, \ldots, N$ to be negative. The probability of a single realisation $\tilde{f}_i$ being negative is

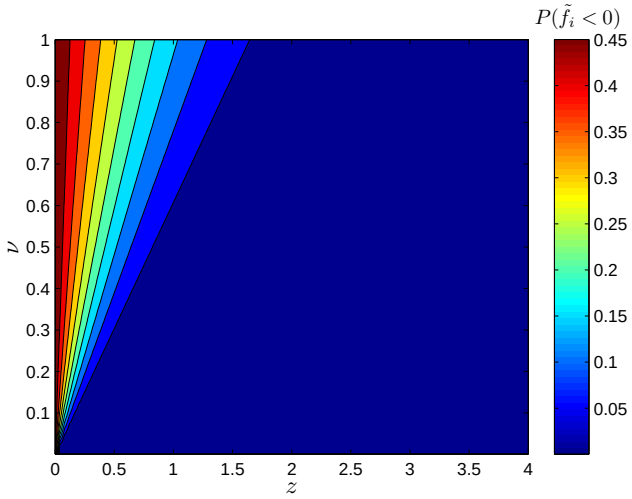
$$P(\tilde{f}_i < 0) = \Phi\left(\frac{-z}{v}\right)$$

and is shown in Figure E.1 for varying measurement tolerance ν and parameter z.
Whilst in the practical context a negative pest density is senseless, we now provide justification that this theory can still be applied for a range of values of ν and z.

Noise is introduced to the function values f_i to generate the values $\tilde{f}_i$ using the transformation

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

Figure E.1: Probability of a single realisation $\tilde{f}_i$ being negative for various values of the measurement tolerance ν and the parameter z.



where γ is a random variable taken from the standard normal distribution and we recall that $\sigma_i = \nu f_i/z$. For a fixed value of N, the above transformation is applied $n_r = 100,000$ times to each value of f_i thus generating $n_r = 100,000$ sets of perturbed data $\tilde{f}_i, i = 1, \ldots, N$. Each set of noisy data is then integrated and the relative error is calculated, thus a total of n_r realisations of the error quantity $\tilde{E}_{rel}$ are generated. We have chosen the sample mean density as the method of evaluation. The number of realisations $\tilde{n}_r$ of $\tilde{E}_{rel}$ which lie within the interval $[\tilde{E}_{min}, \tilde{E}_{max}]$ is then calculated and the following proportion is established

$$P_{num} = \frac{\tilde{n}_r}{n_r}. \tag{E.0.1}$$

for each value of N. The same sets of noisy data $\tilde{f}_i$ generated above are then considered again and any negative values are replaced with zero. As before, the relative error $\tilde{E}_{rel}$ of the estimate of the pest density obtained are calculated for each data set. We again estimate the abundance by the sample mean density. The proportion (E.0.1) is again evaluated for each value of N. We evaluate the proportion (E.0.1) for the data which includes negative values of $\tilde{f}_i$, which we denote P_{raw}. We then replace the negative values of $\tilde{f}_i$ and evaluate (E.0.1) to find the proportion P^+. The relative difference between the two proportions is calculated

$$P_{diff} = \frac{|P_{raw} - P^+|}{P_{raw}}$$

and is shown for varying ν and z in Figure E.2 where the density distributions of Figure 8.2b and 8.2f have been considered for both a small number $N = 3$ and large number $N = 257$ number of grid nodes. Visual inspection of the figures indicates that for $z \geq 1$ the relative difference is very small. Thus for this range of z values the theoretical credible interval of $\tilde{E}_{rel}$ can be considered to be reliable.

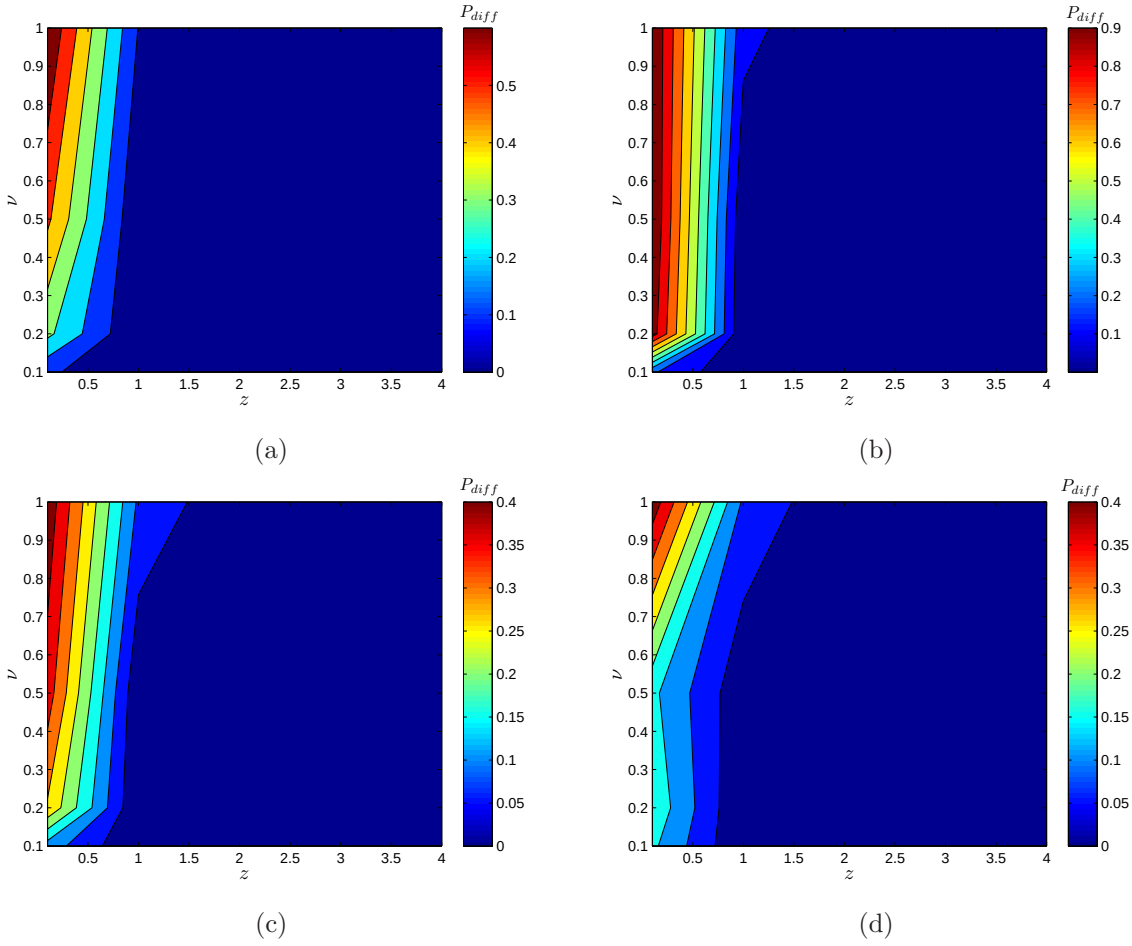


Figure E.2: Difference to theoretical proportion when negative values replaced with zero. Means of estimate evaluation is the sample mean density. (a) Single peak test case (see Figure 8.2b), number of grid nodes fixed as $N = 3$. (b) Single peak test case (see Figure 8.2b), number of grid nodes fixed as $N = 257$. (c) Narrow peak test case (see Figure 8.2f), number of grid nodes fixed as $N = 3$. (d) Narrow peak test case (see Figure 8.2f), number of grid nodes fixed as $N = 257$.

List of References

[1] M. C. R. Alavanja, M. K Ross, and M. R. Bonner. Increased cancer burden among pesticide applicators and others due to pesticide exposure. *CA: A Cancer Journal for Clinicians*, 63(2):120–142, 2013.

[2] C. J. Alexander, J. M. Holland, L. Winder, C. Woolley, and J. N. Perry. Performance of sampling strategies in the presence of known spatial patterns. *Annals of Applied Biology*, 146(3):361–370, 2005.

[3] A. Alyokhin, M. Baker, D. Mota-Sanchez, G. Dively, and E. Grafius. Colorado potato beetle resistance to insecticides. *American Journal of Potato Research*, 85(6):395–413, 2008.

[4] M. Ausden. Invertebrates. In: *Ecological Census Techniques: a Handbook*, (Ed. Sutherland, W. J.), pp. 139–177, Cambridge University Press, 1996.

[5] H. J. Barclay. Modelling the effects of population aggregation on the efficiency of insect pest control. *Researches on Population Ecology*, 34(1):131–141, 1992.

[6] G. I. Barenblatt. *Scaling, Self-similarity, and Intermediate Asymptotics*, volume 14. Cambridge Univ Pr, 1996.

[7] S. L. Bates, J.-Z. Zhao, R. T. Roush, and A. M. Shelton. Insect resistance management in gm crops: past, present and future. *Nature Biotechnology*, 23(1):57–62, 2005.

[8] D. Belkić and K. Belkić. *Signal Processing in Magnetic Resonance Spectroscopy with Biomedical Applications*. CRC Press, 2010.

[9] M. R. Binns, J. P Nyrop, and Van Der Werf W. *Sampling and Monitoring in Crop Protection: the Theoretical Basis for Designing Practical Decision Guides*. CABI Publishing, Wallingford, 2000.

[10] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of Measurement Data Guide to the Expression of Uncertainty in Measurement, JCGM 100: 2008 GUM 1995 with Minor Corrections, first ed.* Joint Committee for Guides in Metrology, Paris, 2008.

[11] A. L. Birmingham, E. Kovacs, J. P. Lafontaine, N Avelino, J. H. Borden, I. S. Andreller, and G. Gries. A new trap and lure for drosophila melanogaster (diptera: Drosophilidae). *Journal of Economic Entomology*, 104(3):1018–1023, 2011.

[12] R. P. Blackshaw. The annual leatherjacket survey in Northern Ireland, 1965–82, and some factors affecting populations. *Plant Pathology*, 32(3):345–349, 1983.

[13] C. I. Bliss. Statistical problems in estimating populations of Japanese beetle larvae. *Journal of Economic Entomology*, 34(2):221–232, 1941.

[14] W. M Bolstad. *Introduction to Bayesian statistics (2nd ed.)*. Wiley, 2007.

[15] J. A. Browde, L. P. Pedigo, T. A. Degooyer, L. G. Higley, W. K. Wintersteen, and M. R. Zeiss. Comparison of sampling techniques for grasshoppers (orthoptera: Acrididae) in soybean. *Journal of Economic Entomology*, 85(6):2270–2274, 1992.

[16] C. J. Budd, W. Huang, and R. D. Russell. Adaptivity with moving grids. *Acta Numerica*, 18(111-241):79–83, 2009.

[17] R. L. Burden and J. D. Faires. *Numerical Analysis*. 6th ed. Brooks/Cole Publishing Company, Pacific Grove, USA, 1997.

[18] A. J Burn, T. H. Coaker, and P. C. Jepson. *Integrated Pest Management*. Academic Press, New York, 1987.

[19] J. A. Byers, O. Anderbrant, and J. Löfqvist. Effective attraction radius: A method for comparing species attractants and determining densities of flying insects. *Journal of Chemical Ecology*, 15(2):749–765, 1989.

[20] P. Christou, T. Capell, A. Kohli, J. A. Gatehouse, and A. M. R. Gatehouse. Recent developments and future prospects in insect pest control in transgenic crops. *Trends in plant science*, 11(6):302–308, 2006.

[21] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2(1):197–205, 1960.

[22] H. N Comins, M. P Hassell, and R. M May. The spatial dynamics of host–parasitoid systems. *Journal of Animal Ecology*, pages 735–748, 1992.

[23] M. G. Cox. The area under a curve specified by measured values. *Metrologia*, 44(5):365, 2007.

[24] J. Crank. *The Mathematics of Diffusion*, volume 1. Clarendon press Oxford, 1979.

[25] Pimentel D. and A. Greiner. Environmental and socio-economic costs of pesticide use. In: *Techniques for reducing pesticide use: economic and environmental benefits.*, (Eds. Pimentel, D.), pp. 51–78, John Wiley & Sons, 1997.

[26] G. Dahlquist and Å. Björck. *Numerical Methods in Scientific Computing*, volume 1. Society for Industrial and Applied Mathematics, 2008.

[27] C. Daniel. Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, 1(4):311–341, 1959.

[28] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, New York, 1975.

[29] P. M. Davis. Statistics for describing populations. In: *Handbook of Sampling Methods for Arthropods in Agriculture*, (Eds. L. P. Pedigo and G. D. Buntin), pp. 33–54, CRC Press, London, 1994.

[30] M. Dehghan, M. Masjed-Jamei, and M. R. Eslahchi. The semi-open Newton–Cotes quadrature rule and its numerical improvement. *Applied Mathematics and Computation*, 171(2):1129–1140, 2005.

[31] D. Dent. *Insect pest management*. CABI Publishing, Wallingford, 2000.

[32] N. L. Embleton and N. B. Petrovskaya. On numerical uncertainty in evaluation of pest population size. *Ecological Complexity*, 14:117–131, 2013.

[33] N. L. Embleton and N. B. Petrovskaya. A novel approach to evaluation of pest insect abundance in the presence of noise. *Bulletin of mathematical biology*, 76(3):718–743, 2014.

[34] N. L. Embleton and N. B. Petrovskaya. On the accuracy of estimating pest insect abundance from data with random error. *Ecological Complexity*, 2014.

[35] E. W. Evans, R. A. Rogers, and D. J. Opfermann. Sampling grasshoppers (orthoptera: Acrididae) on burned and unburned tallgrass prairie: night trapping vs. sweeping. *Environmental Entomology*, 12(5):1449–1454, 1983.

[36] A. W. Ferguson, Z. Klukowski, B. Walczak, J. N. Perry, M. A. Mugglestone, S. J. Clark, and I. H. Williams. 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. *Entomologia Experimentalis et Applicata*, 95(2):161–171, 2000.

[37] R. A. Fisher. The wave of advance of advantageous genes. *Annals of Human Genetics*, 7(4):355–369, 1937.

[38] A. M. R. Gatehouse, N. Ferry, M. G. Edwards, and H. A. Bell. Insect-resistant biotech crops and their impacts on beneficial arthropods. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1569):1438–1452, 2011.

[39] H. H. Goldstine. *A History of Numerical Analysis from the 16th through the 19th Century*. Springer-Verlag, New York, 1977.

[40] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. 3rd ed. Oxford University Press, Oxford, 2001.

[41] J. Gwinner, R. Harnisch, and O. Mück. *Manual on the prevention of post-harvest grain losses*. GTZ, Eschborn, 1996.

[42] J. R. Hagler and C. G. Jackson. Methods for marking insects: current techniques and future prospects. *Annual Review of Entomology*, 46(1):511–543, 2001.

[43] L. G. Higley and L. P. Pedigo. *Economic Thresholds for Integrated Pest Management*, volume 9. University of Nebraska Press, Lincoln, 1996.

[44] H. M. T. Hokkanen. Trap cropping in pest management. *Annual Review of Entomology*, 36(1):119–138, 1991.

[45] J. M. Holland, J. N Perry, and L. Winder. The within-field spatial and temporal distribution of arthropods in winter wheat. *Bulletin of Entomological Research*, 89(6):499–513, 1999.

[46] S. H. Hutchins. Techniques for sampling arthropods in integrated pest management. In: *Handbook of Sampling Methods for Arthropods in Agriculture*, (Eds. Pedigo, L. P. and Buntin, G. D.), 1994.

[47] M. J. W. Jansen and J. A. J. Metz. How many victims will a pitfall make? *Acta Biotheoretica*, 28(2):98–122, 1979.

[48] P. C. Jepson and J. R. M. Thacker. Analysis of the spatial component of pesticide side-effects on non-target invertebrate populations and its relevance to hazard analysis. *Functional Ecology*, pages 349–355, 1990.

[49] P. M Kareiva. Local movement in herbivorous insects: applying a passive diffusion model to mark-recapture field experiments. *Oecologia*, 57(3):322–327, 1983.

[50] M. Kogan. Integrated pest management: historical perspectives and contemporary developments. *Annual review of entomology*, 43(1):243–270, 1998.

[51] D. E. Legg and R. D. Moon. Bias and Variability in Statistical Estimates. In: *Handbook of Sampling Methods for Arthropods in Agriculture*, (eds. Pedigo, L. P. and Buntin, G. D.), 1994.

[52] F. C. Leone, L. S. Nelson, and R. B. Nottingham. The folded normal distribution. *Technometrics*, 3(4):543–550, 1961.

[53] M. Liebman and E. Dyck. Crop rotation and intercropping strategies for weed management. *Ecological applications*, pages 92–122, 1993.

[54] F. J. Louws, C. L. Rivard, and C. Kubota. Grafting fruiting vegetables to manage soilborne pathogens, foliar pathogens, arthropods and weeds. *Scientia Horticulturae*, 127(2):127–146, 2010.

[55] H. Malchow, S. V. Petrovskii, and E. Venturino. *Spatiotemporal Patterns in Ecology and Epidemiology–Theory, Models, and Simulation. Mathematical and Computational Biology Series*. Chapman & Hall/CRC, Boca Raton, 2008.

[56] J. G Mayor and M. H. Davies. A Survey of Leatherjacket Populations in South-west England, 1963–74. *Plant Pathology*, 25(3):121–128, 1976.

[57] R. L. Metcalf and W. H. Luckmann. *Introduction to Insect Pest Management*, volume 101. Second edition. John Wiley & Sons, London, 1992.

[58] A. Milne. The centric systematic area-sample treated as a random sample. *Biometrics*, 13(2):270–297, 1959.

[59] J. D. Murray. *Mathematical Biology: an Introduction*, volume 1. Springer-Verlag, Berlin, 1989.

[60] P. Northing. Extensive field based aphid monitoring as an information tool for the UK seed potato industry. In *Potatoes: Viruses and their Vectors, Edinburgh, UK, 16 September 2009, (ed. Dale, F.)*, number 94 in Aspects of Applied Biology, pages 31–34. Association of Applied Biologists, 2009.

[61] E.-C. Oerke. Crop losses to pests. *The Journal of Agricultural Science*, 144(01):31–43, 2006.

[62] A. Okubo. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, 22:1–94, 1986.

[63] A. Okubo and S. A. Levin. *Diffusion and Ecological Problems: Modern Perspectives*, volume 14. Springer Verlag, 2001.

[64] L. Ott and M. Longnecker. *An Introduction to Statistical Methods and Data Analysis*. Duxbury Pr, 2008.

[65] M. A. Pascual and P. Kareiva. Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology*, 77(2):337–349, 1996.

[66] L.P. Pedigo and M.E. Rice. *Entomology and Pest Management*. Pearson Prentice Hall, New Jersey, 2009.

[67] N. B. Petrovskaya and N. L. Embleton. Evaluation of peak functions on ultra-coarse grids. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 469(2153):20120665, 2013.

[68] N. B. Petrovskaya and N. L. Embleton. Computational Methods for Accurate Evaluation of Pest Insect Population Size. (In Press). In: *Ecological Modelling Applied to Entomology*, (eds. W. A. C. Godoy and C. P. Ferreira), 2014.

[69] N. B. Petrovskaya, N. L. Embleton, and S. V. Petrovskii. Numerical Study of Pest Population Size at Various Diffusion Rates. In: *Dispersal, Individual Movement and Spatial Ecology, Lecture Notes in Mathematics, vol.2071*, (eds. Lewis, M. A., Maini, P. K. and Petrovskii, S. V.), 2013.

[70] N. B. Petrovskaya and S. V. Petrovskii. The coarse-grid problem in ecological monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 466(2122):2933, 2010.

[71] N. B. Petrovskaya, S. V. Petrovskii, and A. K. Murchie. Challenges of ecological monitoring: estimating population abundance from sparse trap counts. *Journal of The Royal Society Interface*, 9(68):420–435, 2012.

[72] N. B. Petrovskaya and E. Venturino. Numerical integration of sparsely sampled data. *Simulation Modelling Practice and Theory*, 19(9):1860–1872, 2011.

[73] S. V. Petrovskii, D. Bearup, D. A. Ahmed, and R. Blackshaw. Estimating insect population density from trap counts. *Ecological Complexity*, 10:69–82, 2012.

[74] S. V. Petrovskii and B. L. Li. *Exactly Solvable Models of Biological Invasion.* CRC Press, 2006.

[75] S. V Petrovskii, B. L. Li, and H. Malchow. Quantification of the spatial aspect of chaotic dynamics in biological and chemical systems. *Bulletin of mathematical biology*, 65(3):425–446, 2003.

[76] S. V. Petrovskii and H. Malchow. Spatio-temporal chaos in an ecological community as a response to unfavourable environmental changes. *Advances in Complex Systems*, 4(2):227–249, 2001.

[77] S. V Petrovskii, H. Malchow, F. M. Hilker, and E. Venturino. Patterns of patchy spread in deterministic and stochastic models of biological invasion and biological control. *Biological Invasions*, 7(5):771–793, 2005.

[78] S. V. Petrovskii, A. Y. Morozov, and E. Venturino. Allee effect makes possible patchy invasion in a predator–prey system. *Ecology Letters*, 5(3):345–352, 2002.

[79] D. Pimentel. Amounts of pesticides reaching target pests: environmental impacts and ethics. *Journal of Agricultural and environmental Ethics*, 8(1):17–29, 1995.

[80] D. Pimentel. Pesticides and pest control. In: *Integrated Pest Management: Innovation-Development Process, Volume 1*, (Eds. Peshin, R. & Dhawan, A. K.), pp. 83–88, Springer, 2009.

[81] D. Pimentel and M. H. Pimentel. *Food, Energy, and Society.* CRC press, 2008.

[82] D. Pimentel (ed.). *Techniques for Reducing Pesticide Use: Economic and Environmental Benefits.* John Wiley & Sons, 1997.

[83] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics.* Springer Verlag, Berlin, 2007.

[84] D. A. Raworth and M.-Y. Choi. Determining numbers of active carabid beetles per unit area from pitfall-trap data. *Entomologia Experimentalis et Applicata*, 98(1):95–108, 2001.

[85] W. K Reisen and H. D Lothrop. Effects of sampling design on the estimation of adult mosquito abundance. *Journal of the American Mosquito Control Association*, 15(2):105–114, 1999.

[86] D. S. Robson and H. A. Regier. Sample size in Petersen mark-recapture experiments. *Transactions of the American Fisheries Society*, 93(3):215–226, 1964.

[87] M. L. Rosenzweig and R. H. MacArthur. Graphical representation and stability conditions of predator-prey interactions. *American Naturalist*, 97(895):209–223, 1963.

[88] J. R. Ruberson (ed.). *Handbook of Pest Management.* Marcel Dekker Inc., New York, 1999.

[89] J. A. Sherratt and M. J. Smith. Periodic travelling waves in cyclic populations: field studies and reaction–diffusion models. *Journal of the Royal Society Interface*, 5(22):483, 2008.

[90] N. Shigesada and K. Kawasaki. *Biological Invasions: Theory and Practice*. Oxford University Press, USA, 1997.

[91] A. V. Shoffner and J. F. Tooker. The potential of genotypically diverse cultivar mixtures to moderate aphid populations in wheat (triticum aestivum l.). *Arthropod-Plant Interactions*, 7(1):33–43, 2013.

[92] J. B. Silver. *Mosquito Ecology: Field Sampling Methods*. 3rd ed. Springer, New York, 2008.

[93] A. C. Smigocki, S. Ivic-Haymes, H. Li, and J. Savić. Pest protection conferred by a beta vulgaris serine proteinase inhibitor gene. *PloS one*, 8(2):e57303, 2013.

[94] G. W. Snedecor and W. G. Cochran. *Statistical Methods*. The Iowa State University Press, Ames, 1980.

[95] F. Sohrabi, P. Shishehbor, M. Saber, and M. S. Mosaddegh. Lethal and sublethal effects of imidacloprid and buprofezin on the sweetpotato whitefly parasitoid eretmocerus mundus (hymenoptera: Aphelinidae). *Crop Protection*, 45:98–103, 2013.

[96] T. R. E. Southwood and P. A. Henderson. *Ecological Methods*. Third edition, Blackwell Science Ltd., 2000.

[97] M. R. Spiegel. *Theory and Problems of Probability and Statistics*. McGraw-Hill, New York, 1992.

[98] R. G. D Steel, J. H Torrie, and D. A. Dickey. *Principles and Procedures of Statistics*. McGraw-Hill Book Company, Inc., New York, 1960.

[99] V. M. Stern, R. F. Smith, R. Van den Bosch, and K. S. Hagen. The integration of chemical and biological control of the spotted alfalfa aphid. part i. the integrated control concept. *Hilgardia*, 29(2):81–101, 1959.

[100] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, New York, 1980.

[101] A. Taboada, C. Pérez-Aguirre, and T. Assmann. A new method for collecting agile tiger beetles by live pitfall trapping. *Entomologia Experimentalis et Applicata*, 145(1):82–87, 2012.

[102] C. F. G. Thomas, L. Parkinson, and E. J. P. Marshall. Isolating the components of activity-density for the carabid beetle pterostichus melanarius in farmland. *Oecologia*, 116(1-2):103–112, 1998.

[103] L. N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review*, 50(1):67–87, 2008.

[104] P. Turchin. *Complex Population Dynamics: a Theoretical/Empirical Synthesis*, volume 35. Princeton Univ Press, Princeton, 2003.