Methods for Accurate Evaluation of Population Abundance from Ecological Data

by

MANAL ALQHTANI

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
College of Engineering and Physical Sciences
The University of Birmingham
August 2018

Abstract

An accurate evaluation of total population density is required in many ecological and biological field. To protect crops from pest attacks, the population density of pests must be evaluated adequately. Accurate information obtained as a result of trapping in ecological monitoring is beneficial for decision-making purposes when implementing a control action. In pest monitoring, a classic technique of evaluating density based on a statistical method may result in poor accuracy. Accuracy can be optimised by applying alternative numerical integration methods to the problem. We explain how insufficient information regarding population density negatively affects the accuracy of estimation. Consequently, a coarse grid problem arises where the numerical integration methods are no longer valid. The evaluation of integration error is now a random variable and the probabilistic approach is used, due to the uncertainty in sampling data. In this thesis several population models have been considered to explain that the value of correlation coefficient on a coarse sampling grid is lost even if the true value is close to one. Phenomenon of ghost synchronisation has been observed when the value of correlation coefficient on a coarse sampling grid is close to one but in reality the dynamics are not correlated.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Natalia Petrovskaya, for her help and support throughout my studies and for her excellent guidance. I would also like to thank my friend Nina Embleton for her unlimited support and patience.

I would like to take this chance to thank Najran University for enabling me to continue my PhD studies at the University of Birmingham. It was a great opportunity to extend my knowledge and develop my skills. Thank you to the School of Mathematics at the University of Birmingham for everything and for making my studies unforgettable experience. I am extremely grateful to the administrator at the Graduate School of Mathematics, Janette Lowe, for her collaboration throughout my research. I would like to extend my thanks to the staff and students I have met during my study. My gratitude goes to those who take the time to read my thesis.

My gratitude goes to my father and mother and to the rest of my family, specially to my sisters (Mona and Norah), who have been so supportive and helpful during my studies. This thesis is dedicated to my children (Abdulelah, Joanna, and Haitham), without their love and support, I would not have been able to complete this work.

Contents

Nomenclature

The following notation will be used throughout this thesis, unless otherwise stated.

List of Acronyms

Acronym	Description
$1d$	One-dimensional
$2d$	Two-dimensional
IPM	Integrated Pest Management

List of Greek Symbols

Symbol	Description
δ	Width of peak (local maximum)
Δx	Characteristic length of spatial heterogeneity
τ	Accuracy tolerance of an estimate of pest abundance
μ_e	Mean of the relative integration error
σ_e	Standard deviation of the relative integration error

List of Latin Symbols

(**Symbol Description**

D	Domain representing the agricultural field
f	Pest population density
E_{abs}	Absolute integration error formed from exact data
E_{rel}	Relative integration error form from exact data
h	Grid step size represents distance between sample units
I	Exact population density abundance
I_a	Population density approximation formed from exact data
L	Length of the agricultural field in a one-dimensional problem
k	Degree of interpolating polynomial
S	Area of agricultural field
p	Probability of obtaining an accurate evaluation of the problem
r	Degree of accuracy of a numerical integration method
M	Sample mean of pest population density
N	Number of grid nodes
q	Order of convergence of a numerical integration formula
$Pk(x)$	An interpolating polynomial of degree k
w	Weight of a method of numerical integration

Chapter 1

Introduction

1.1 The Concept of Integrated Pest Management

Pests are probably the factor that most threatens the production of food worldwide [124, 137]. According to [43], the term 'pest' involves any organism that in some way causes damage to humans or the ecosystem. Food crops are very weak and are prone to attacks from pests, both while growing and after harvesting. Annual estimates of the losses due to pests in worldwide food production, in the period before harvesting, lie between 35% and 42% [124, 137]. In this period of growth, pests mainly consist of weeds, arthropods and plant pathogens [101, 162]. Furthermore, around $14 - 15\%$ of this percentage is documented to be as a result of pest insects [144, 146]. Pest invasions of stored crops, whether insects, birds, or micro-organisms, lead to crop losses of between $10 - 25\%$ after harvesting [64]. These losses have to be minimised professionally, by managing the abundance of pest populations using some means of control.

To prevent increases in pest population densities from becoming a problem in the first place, methods of preventative pest management must be used. The regulation and management of pest control is therefore crucial, since it has detrimental effects on human health, the environment, and the global economy [190]. Ecological monitoring methods

are used to present accurate estimates of pest insect density, which are used to make effective decisions about the means of control required. Preventative pest management has several key aims, one of which is to prevent, or at least to minimise, the harm caused by pests that target agricultural crops. Many approaches have been utilised to successfully achieve this goal. For instance, crop rotation and intercropping are considered to be some of the oldest methods of pest management. The aims are to introduce heterogeneity, and to destabilise the life cycle of the pest, which in turn reflects positively on the preventative pest management program. The conventional process of planting assumes that the same kind of crop is grown in a single location. In crop rotation, different types of crop are grown in a single location that hosts different kinds of pest. Intercropping involves planting various types of crop at the same time, in the same location [87, 169]. This protects the crop, and may reduce the area of agricultural land that must be targeted for further management by applying the intercropping tactic [77]. In this strategy, crops are spread that are specifically for use as decoys, being particularly attractive to pests but with no other function, pests are then located in localised domains which require less management. The cultivation of crops that have been proven to be resistant to pest attacks is another precautionary measure. According to [101], grafting has been employed for centuries to manage pathogens and is considered beneficial for the control of arthropod pests and weeds. Another preventative measure is to plant crops that have been genetically modified, to be resistant to pest attacks. However, this precautionary management process is not yet fully understood. For this reason, it may give rise to potential risks, and much research is required before this strategy can be adopted [7, 32, 58, 171].

Another approach to pest management is to implement a control action as a preventative method, in order to kill pests. In recent years, this approach has been used to control the diffusion of agricultural pests in an agricultural field across the world by the application of pesticides. The use of pesticides is widespread in worldwide agriculture, and around

3×10^9 kg of pesticides are used yearly around the world [143]. It has been argued that this increased use of pesticides may lead to serious negative effects upon mankind and the environment at the same time. For instance, according to [3], a new generation of pests that have become resistant to this strategy will result from the indiscriminate and regular use of control actions, which may make the future task of management increasingly difficult. Furthermore, the application of pesticides has been considered to be a main cause of the increase in ecological pollution [145, 179, 190]. According to [145], just 0.1% of pesticides used are absorbed by the targeted insects, whereas around 99.9% are absorbed into the environment in some other manner. In terms of public health, a strong link between illness and exposure to regular doses of pesticide, has been found [19].

It is obvious from the above that no method of preventative pest management is completely effective, and there is therefore a growing demand to introduce the concept of integrated pest management (IPM). IPM can be defined as the incorporation of several different methods of pest management, which work together in a more sustainable way to protect food production from pest attacks [89]. IPM consists of three successive stages. In the first stage, the preventative measures of pest management are implemented. In the second stage, pest density is monitored. Finally, a decision on whether to implement a control action is made by comparing abundance to threshold values [26, 112].

1.2 Estimating Pest Population Density

In order for an IPM programme to achieve its aims, it is very important to understand that different types of pest species have different behaviour. We restrict our attention to insect pests, and the generic term 'pest' will be used refer to insect pests henceforth unless otherwise stated. The monitoring methodologies in an IPM programme will vary, according to the varieties of pest behaviour. The IPM programme used will also depend on the specific area of the environment that is being targeted by pest attacks. Our scope

will be concentrated specifically on the area of pest management for crops immediately prior to harvesting. In an agricultural field performing a complete census of pests is a complex process, therefore, the population density will be estimated using alternative methods. To collect data related to the pest population in a specific agricultural field, many approaches can be used to estimate pest populations by taking a sample of the whereabouts of pests [6, 18, 82, 108, 174].

A direct counting of the number of pests in-situ can be done from determining a sample unit of plant or a specific area of habitat. A 'knockdown' approach can be used to count species that are more unobtrusive. 'Knockdown' is a process that depends on dislodging the pests from plants carefully, and in some cases they can then be taken to a laboratory where a count can be made.

In order to obtain the arithmetic mean number M of pests per sample unit, the following equation is widely used:

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

where f_i denotes the individual sample counts and N is the number of sample counts in an agricultural field [37]. After determining the arithmetic mean number of pests per unit area, the number of pests in the entire agricultural field can be calculated by multiplying this mean number by the entire area of the agricultural field [172]. Then, this estimation will be used as the 'absolute' abundance of pests in this area, where the direct counts of samples will reflect the number of pests in the sample unit. This process is known as the statistical method, and has a history of successful application [172, 177].

Although the sampling process is useful, its performance over a large area requires a long time and is expensive to implement. In such cases, the alternative method of direct counting in-situ, using netting, is used to calculate the number of pests. Nets can be used for a defined period of time, then the number of pests can be determined by counting those caught inside the nets [131, 174]. It is beneficial to use netting in order to count the pests

in a large agricultural field, because it is faster and more cost-effective than the previous method.

Another widely used method of sampling is trapping. After traps have been installed in an agricultural field for a specific period of time (for example one week), the pests are counted after emptying the traps. There are two different types of trap. Active traps attract pests with the help of some biological, chemical or physical means. Passive traps rely on the movement of pests. Traps can be installed in arbitrary places. For instance, some ecologists prefer to install traps in random locations, whereas, others prefer to follow a specific sampling pattern [2, 49]. In many ecological test cases, traps are placed at the nodes of a rectangular grid [50, 78]. In the trapping process, the sample mean density can be calculated by applying equation (1.2.1) to the data obtained from the trapping process [29, 156], which then provides a discrete data distribution of the pest population density. It should be noted here that, in the trapping and netting techniques presented, the relative estimate of the arithmetic mean number for pest abundance can be obtained, instead of the absolute value of the estimation. This is because the arithmetic mean equation measures the efficiency of the netting and trapping procedures and the conditions whilst sampling, rather than introducing a direct estimate of pest density. Therefore, the comparison will be made between the relative estimates that have been acquired under the same conditions and by the same technique.

Mark-Release-Recapture technique can be used successfully to estimate the population pest density. The Mark-Release-Recapture technique involves marking a number of individuals in a natural population, returning them to that population, and subsequently recapturing some of them as a basis for estimating the size of the population at the time of marking and release [65]. It is based on the principle that if the proportion of population was marked in some way , returned to the original population and then, after complete mixing, a second sample was taken, the proportion of marked individuals in the second

sample would be the same as was marked initially in the total population. This statement can be translated into the following formula:

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

where $\hat{I}$ denotes the number of pests initially caught, marked, and released in the first sample, I represents the size of whole population, C is the total number of pests caught in the second sample, and $\hat{C}$ is the number of marked recaptured pests in the second sample. Equation (1.2.2) can be written as follows:

$$I = \frac{\hat{I} \times C}{\hat{C}}. \tag{1.2.3}$$

In scientific studies, this process works very well but it is not suited for routine monitoring due to the intensive laboratory work required. Finally, it is obvious from the above that the correct choice of monitoring methodology is essential for accurate results from an IPM programme.

1.3 The Importance of Obtaining Accurate Estimations

The most important aim of an IPM programme is to provide an accurate estimation of the pest population size at the typical agricultural area. This importance of obtaining an accurate estimation of pest population density has been discussed in [136]. After estimating the pest population abundance in an agricultural field, a decision of implementing a control action will be taken depending on the comparison of the estimation to some threshold value(s). For instance, if this estimate falls below the value of threshold, the decision in this case would be that no action is needed. However, according to [15], if

the estimation exceeds this threshold, then the decision is to implement a control action immediately. Many studies have proven that the application of pesticides is a widely used in order to reduce damage from pest attacks [49, 179]. Thus, it can be seen from the above that precision is of utmost importance in order to arrive at the correct decision with regards to whether to implement a control action. If the true value of population density is known, then the decision can be made easily, however, this cannot be known, therefore reliable information with which to make an evaluation of population density is required. Therefore, some tolerance τ of the absolute value of population size I will be considered to define the accuracy of estimation I_a to be within the range $[I - \tau, I + \tau]$. Then, the risk can be occurred if the threshold value is within this range. If the estimation can be quantified, the risk can be quantified as well. In addition, when values of tolerance τ are small enough a more accurate evaluation of population will be acquired which in turn will reduce the risk of making incorrect decisions about pest management.

Different examples of estimation of pest population density will be considered in order to illustrate the importance of producing an accurate evaluation of pest abundance. The underestimation of the pest population size may lead to the non application of pesticides, whereas in reality there may be a strong need to implement a control action. The loss of crops will increase as a consequence of this decision. According to [144], the value of crops lost as a direct result of pest attacks with the application of pesticides has been estimated to be two billion dollars yearly. On the other hand, an overestimation of pest densities in an agricultural field may lead to the implementation of a control action that may be unnecessary. In addition, the application of pesticides has significant negative effects upon the environment, costing huge amounts of money yearly [87]. Therefore, in order to prevent the unwanted effects of the application of pesticides, there is an urgent need to apply reliable methods that produce an accurate evaluation of pest population abundance. There are several techniques that have been used to improve the measure-

ment of accuracy. One method is to increase the number of samples in an agricultural field to calculate the value of pest abundance with a sufficiently large number N in equation (1.2.1), and then to consider it as the exact value of the pest population. In ecological research, the number of traps can be increased to hundreds per given agricultural area. However, in real-life pest monitoring programmes, the number of traps does not exceed twenty [37], and in some cases it ranges between one and a few per typical agricultural area [38]. There are several reasons to prevent farmers from using a large number of traps in an agricultural field. For instance, a large number of traps requires an increase in the amount of labour required, which in turn increases the cost of this process. Another problem is that installing a large number of traps may damage the agricultural field, because traps are known to be a disturbance to the agricultural field. Therefore, in view of these drawbacks, the number of traps is normally taken to be small [123] and equation (1.2.1) does not suit accurate evaluation well. Another drawback of equation (1.2.1) is that it does not take into consideration the distance between traps, or trap installation patterns. A consequence result of this neglection is that significant information about the pest population size is missed. So, the need to obtain accurate alternative methods has become urgent. It is worth noting here that the required degree of accuracy will differ, according to the aim of the monitoring process. The error range of estimation may lie between $20 - 100\%$ in routine ecological monitoring [128, 164]. However, this percentage error range becomes 10% for monitoring research purposes [131].

It may be argued that the main concern in an IPM programme is to try ensuring the accuracy of the estimation of the pest population in a given agricultural field. Therefore, intensive research is still under way into methods that may present more accurate estimations of pest abundance than existing statistical methods based on (1.2.1). However, these alternative techniques need to undergo sufficient studies to ensure a balance between their advantages and disadvantages. The decision as to whether these methods can or

should be adopted can only be taken once the degree of accuracy of each method has been critically scrutinised. It has been noticed that in current ecological literature, the most significant concern raised was about the data collection method [15, 16, 43, 123, 131, 184], whereas in the study of evaluation accuracy, the prime focus should be on the method of data processing.

Numerical integration methods present an alternative to statistical methods of evaluation. In recent years, many studies have been carried out on these methods in order to present an overall view of their application to ecological problems [46, 134, 135, 136, 137]. These studies have shown that numerical integration methods are more reliable than standard statistical methods (1.2.1), which can be considered as a simple form of numerical integration. It has been demonstrated that advanced numerical integration methods tend to provide a more effective estimation of pest density. Generally, from a mathematical viewpoint, the problem of estimating the pest population size is considered as a problem of calculation of the following integral

$$I = \int_D \int f(x, y) dx dy,$$

where the integrand $f(x, y)$ is the pest population density.

Clearly, in ecological problems the function $f(x, y)$ is only known at discrete points $f_{ij} = f(x_i, y_j)$, and in this situation it is impossible to find the analytical solution. In order to evaluate this integral, numerical integration methods must be used to approximate the solution [44, 96, 97, 103, 199].

It is commonly known to mathematicians that numerical integration methods can be used to successfully approximate the solution numerically when dealing with a set of discrete data. The evaluation accuracy will improve depending on the amount of available data. In other words, in order to achieve a more accurate estimation of the pest population size, the amount of data must increase accordingly [139]. In this case, the approximated solution will converge to the analytical solution. This means that accurate evaluation of

Figure 1.1: A typical research study area with traps installed at the nodes of a rectangular 12×12 grid with regular spacing.

the approximated solution requires a vast array of data to be available. However, in pest monitoring, a large amount of data requires more equipment and labour to perform the physical measurements. Furthermore, in terms of trying to generate simulated data, this will require a high computational effort. Therefore, the amount of data may be limited in various types of application. The problem of integrating sparse data is completely different from the conventional integration problem. Unfortunately, in the previous literature this significant problem has been disregarded. However, in recent years, it has been focused on dealing particularly with sparse data in ecology. It has been illustrated in the works [136, 137, 138] that the problem of pest monitoring in an agricultural field has been considered. In this study, the evaluation of pest population size in an agricultural field has been treated as a problem of the numerical integration of sparse data.

A sampling plan is an essential and clearly affects the accuracy of population abundance estimates. The samples must be installed in professional way to capture adequate information about true pest abundance. According to [50, 78], pests may be randomly

distributed within a field, while the field's conditions are homogeneous. However, pests usually prefer to distribute as an aggregated spatial pattern. Thus, the sampling plan must be investigated carefully in order to ensure accurate estimation of pest population abundance. Traps are used as a method of monitoring in an IPM programme, where traps are positioned as the nodes of a computational grid. There are different approaches to installing traps in a desired agricultural field. As presented in [196], a common approach is to use a regular scheme of installation. Another approach is to install traps randomly to avoid selection bias [15]. However, the second approach has several drawbacks. For instance, the process of installing and collecting takes a long time, as a lot of effort may be required to find the positions of the traps again [196]. Therefore, according to [51, 78, 130], the best way to install traps is for them to be installed at the uniform grid nodes of the agricultural field, as a rectangle for simplicity, as can be seen in Figure 1.1 [137].

It has been shown in [134, 139] that the standard evaluation technique does not work when the available data is sparse because of insufficient information in data collected on coarse grids. A probabilistic approach is used to evaluate the total population size on coarse sampling grids. Recent cross-disciplinary research has accelerated in the field of computational and numerical algorithms, including methods for solving real-world ordinary differential equations, numerical integrals, and optimization, which can be expressed as estimation algorithms. They are utilised to approximate the value of a variable, numerical, or stochastic integral, the solution of a differential equation, or the location of a maximum, as well as to optimize a multi-objective algorithm. Such techniques utilise principles of statistical inference and generalised probability theory. As such, they have been applied to these concepts in computer science, most notably artificial intelligence and modern machine learning.

Understanding and developing numerical techniques such as learning algorithms can be powerful in a wide range of multi-disciplinary research areas and industries. Combining

such platforms encourages cross-fertilisation of research in computation, probability and numerical methods. The interface of research between probabilistic and numerical methods has been applied to diverse and wide-ranging areas such as the Internet of Things, the development of algorithms for neural networks used in self-driving vehicles, and intelligent analysis systems, such as big data, stochastic methods in finance, and Monte-Carlo simulation [72, 73, 90, 111].

1.4 The Objective of the Study

As mentioned previously, the problem of obtaining an accurate estimation of pest population abundance has been discussed in the ecological literature, where the focus has been essentially on the method of collecting data. The consideration in this thesis will be on the method of processing the data. Numerical integration methods will be discussed as alternative techniques to the existing classical approach (statistical methods).

The theory of numerical integration and the method for applying these approaches in ecological problems will be demonstrated. It will be shown that numerical integration methods may provide more accurate estimations than the classical approach depending on the sample mean. However, the ecological problem differs from common problems of numerical integration [136]. Firstly, there is a limit to the number of traps that can be used. As a result, the problem now is to obtain an accurate estimation of the pest population density from the data available from a very limited number of traps. Secondly, a refinement of the computational grid is not possible. Grid refinement is the process of obtaining a more accurate approximation by increasing the number of grid points. However, the application of this process to ecological data means the number of installed traps must to be increased, which is impossible [25]. Finally, it is very difficult to predict the location of the pest population within a patchy environment in advance. Thus the process of distributing more traps will be expensive and may require a higher degree of

accuracy. In order to deal with the above restrictions, it is reasonable to adopt studies to investigate which factors may impact the degree of accuracy within the framework of the problem. Consequently, the accuracy of pest population abundance evaluations may be controlled, after understanding the factors that affect the accuracy of numerical integration methods.

This thesis has the following structure. The basic population dynamics model will be introduced in Chapter 2. An overview of ecological models and their proprieties will be presented, with spacial attention paid to the Rosenzweig- MacArthur model. A numerical solution to the Rosenzweig -MacArthur model and its applications will be introduced. In Chapter 3, we will introduce the alternative approaches to the widely used statistical method, with an overview of the mathematical theory behind numerical integration methods. The numerical integration methods and their application to a one-dimensional $1d$ case will be explained first. The work will extend to examine two-dimensional $2d$ problems. Regular computational grids will mainly be considered, due to unexpected ecological reasons the application will extend to involve random patterns of sampling. The Richardson extrapolation technique will be illustrated in Chapter 3 as an alternative technique to numerical integration methods. In Chapter 3, we will demonstrate the impact of numerical integration approaches when dealing with ecological problems, due to such restrictions imposed. Furthermore, this chapter will explain that the accuracy of population density distribution evaluation depends on the spatial pattern of distribution. In addition, Chapter 3 will prove that accuracy depends on the amount of data available, with sparse data (coarse grid nodes), this accuracy is quite low and further refinement of our computational grid nodes is highly required. In Chapter 4, our attention will turn to a highly aggregated distribution, which is a special type of distribution. It will be shown that this pattern of distribution is difficult to handle, due to nature of the problem and the limited amount of data that can be obtained. In this chapter, a new

classification of computational grid nodes will be considered, ultra-coarse grids. It will be explained in Chapter 4 that the accuracy of an estimate from ultra-coarse grids will be probabilistic rather than deterministic, and achieving an acceptable degree of accuracy will become a matter of chance. Different test cases that reflect some realistic patterns will be considered, and the probabilistic approach will be applied to treat the properties of pattern on ultra-coarse grid nodes. In Chapter 4, application of the probabilistic approach will be extended to involve $2d$ problems. In Chapter 5, there is an overview of spatial synchronisation, and means of measurement will be introduced. After this, synchronisation in a mathematical framework will be discussed. It will be investigated how the coarseness strongly affects the presence and absence of synchronisation. The ecological reasons behind the phenomenon of ghost synchronisation will be introduced, with an instructive example. Finally, Chapter 5 will investigate factors that enhance or deteriorate synchronisation.

Chapter 2

Population Dynamics Model: Basic Theory

According to [103, 119], any ecosystem involves a variety of biological communities. Within these biological communities. different ecological interactions exist between species and their environments. Several studies have been carried out to illustrate the dynamic processes of these vital ecological interactions, for instance, predator-prey interaction, competition interaction, resource production, control actions to manage pests.

One of the most significant ecological interactions are consumer - resource interactions [188]. Consumer- resource interactions are considered a kind of population cycle that directly result from oscillations at trophic interactions. In the light of the diversity of models that relate to trophic interactions, however, not all models present cycles of trophic interactions. Then, in order to classify consumer-resource interactions, two different ecological factors must be considered [103, 119]. Firstly, intimacy represents the relationship between each consumer and the organism that feeds on it. Secondly, lethality represents changes in the trophic interactions as a result of the death of an organism due to consumption by predators. Therefore, using these factors, it can be said that predators record a high percentage on lethality and low on intimacy, whereas, parasitoids record

high percentages on both scales, and grazers are low on both scales. Finally, compared to predators, parasites are low in lethality, but high in intimacy. Then, as a result of the diversity in functional classification, mathematical models which are interested to investigate the trophic interactions will be affected[188].

These ecological interactions require interpretation, alongside theoretical studies, by applying appropriate mathematical approaches. In order to choose an appropriate approach to describe our ecological problem mathematically, the relationship between the species found in our agricultural field and their environment has to be adequately understood [103, 119, 188]. The mathematical approach is very useful for understanding the temporal and spatial dynamics of species in an ecological domain, which can include predator-prey interactions, competition for food, the process of reproduction, and so on. Another fundamental issue is to determine whether the population densities must be thought of as continuous- discrete, or a hybrid of continuous and discrete population densities [103]. Depending on the ecological properties and requirements of the system, continuous or discrete approaches can be used accordingly. Furthermore, in some ecological cases, due to the potential for overlapping between spatial and temporal scales, a mixed approach can be applied to describe some continuous processes and some as discrete processes. Therefore, in order to choose a successful model, we have to determine the ecological requirements of our ecological system. Dealing with continuous population models for a single species differs from coping with continuous models for an interacting population [119]. Continuous models for an interacting population will be considered in our research. Species are known to be distributed heterogeneously. Where the important processes related to these phenomena is to understand formation of ecological patterns, which poorly understood [103, 119]. A predator and its prey is an optimal example of ecological species interacting.

In order to understand ecological pattern formation, the historical perspective must be

considered [12]. Today, the habitation ranges of different species may be completely different from what they were a few thousand years ago, due to continuing changes in climatic conditions, and recently the profound impact of human activities, consequently, the location of ecological borders has also changed. The historical dimension of species distribution has to be considered carefully, together with the recent process of species spread. Whether due to invasion or changing ecological borders, the redistribution of species is correlated with a travelling boundary. The most important question that need to be answered is this: can the spatiotemporal patterns that are obtained by the spread of the travelling density front, be considered to be the main process by which ecological invasions occur? According to [125, 189], in the wake of invasion, spatiotemporal density is observed to oscillate in ecological data. Many theoretical studies have been carried out to prove that the oscillation of formation patterns decays promptly as a result of a travelling propagation front [45, 166, 167, 198]. A traveling population front represents the simplest example of a theoretical spatiotemporal pattern. A system that has a propagation front usually separates the agricultural area into two different regions, for instance one area is typified by high population density, while the other area is low with different properties. The population in each region is known to be distributed homogeneously. Different patterns of spatial population interactions will be investigated in the upcoming sections, after presentation of a brief explanation of predator responses to fluctuations in prey density.

The predator-prey system is a spatiotemporal model where the population densities are functions of time and space. From a mathematical viewpoint, the predator-prey model is a system of coupled diffusion-reaction equations that describe the spatiotemporal dynamics of pest population density [119]. The predator-prey model involves a couple of equations; one describes the change over time U' of prey population U and the second describes the change over time V' of predator population V, where in this model, the pest

is the predator. These changes in the densities of prey and predator can be interpreted as follows:

$$U' = \text{diffusion term} + \text{reaction terms.}$$

This represents changes in prey density, and it can be rewritten as follows:

$$U' = \text{diffusion term} + \text{birth rate of prey} - \text{ mortality of prey not due to predator} - \text{mortality due to predator.}$$

Changes in predator density are presented as follows:

$$V' = \text{diffusion terms} + \text{reaction terms.}$$

This can be written as follows:

$$V' = \text{diffusion terms} - \text{mortality rate of predator} + \text{reproduction rate of predator.}$$

It can be seen that in the absence of predators, a logistic rate of the prey birth and death cycle occurs. Furthermore, if the density of predators is not very high, then they will interfere with each other while searching for prey, which leads one to assumption a linear rate of death for predators. In addition, it is common to assume that the predator reproduction rate is proportional to the predator mortality rate [188].
This can be generalised to a system of n species U_1, U_2,...,U_n, whose temporal dynamics (in the absence of space) can be described by an ordinary differential equation as follows:

$$\frac{dU_i(T)}{dT} = f_i(U_1, U_2, U_3, ..., U_n);\ i = 1, 2, ..., n, \tag{2.0.1}$$

where U_i denote the population densities of the ith species at a specific time T, and n is the number of species in the agricultural field. The functions f_i are the nonlinear effects of the birth and death of species. In order to deal with the problems caused by spatial

aspects, diffusion terms will be added to equation (2.0.1) so that it becomes:

$$\frac{\partial U_i(R,T)}{\partial T} = D_i \nabla^2 U_i(R,T) + f_i(U_1, U_2, U_3, ..., U_n); i = 1, 2, ..., n, \tag{2.0.2}$$

where $R = (X, Y)$, D_i refer to the diffusion coefficient of the ith species, and ∇^2 is the Laplace operator which can be presented as: $\nabla^2 = \frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2}$. The system (2.0.2) is a system of nonlinear partial differential equations. The analytical solution is not generally available; therefore, there is an urgent need to employ numerical methods to solve this system. In order to do this, the system has to be transformed to a dimensionless form [103, 119]. This process of system transformation to a dimensionless form is usually called scaling. The reaction term in equation (2.0.2), which includes the functions f_i, not only depends on the population densities of species, but a number of parameters that affect the values of the function f_i as well. Consider a two-component system, with prey U and predator V. Then, according to the previous consideration of the functions f_i, system (2.0.2) can be rewritten as follows:

$$\frac{\partial U(R,T)}{\partial T} = D_1 \nabla^2 U(R,T) + P(U) - E(U,V) \tag{2.0.3}$$

$$\frac{\partial V(R,T)}{\partial T} = D_2 \nabla^2 V(R,T) + \kappa E(U,V) - M(V), \tag{2.0.4}$$

where U and V denote the population densities of prey and predator respectively. P denotes the linear growth rate of prey numbers, E refers to the rate of predation, $M(V)$ is the predator mortality rate, and the predation efficiency is represented by the coefficient κ. The term $E(U, V)$ in the system of equations (2.0.3), (2.0.4) represents the functional response of the predator to prey density, which is defined as the temporal rate of prey killed per individual predator [188]. According to [79], there are three different classifications of functional responses. Then, the system of equations (2.0.3), (2.0.4) represents the

general form of predator-prey model, which can take different forms depending on the type of predator response to prey density [103, 119]. Recently, the most common forms of predator response to prey population are types I, II of the Holling types [103, 119] as will be explained later.

2.1 Responses of Predators to Fluctuations in Prey Density

This section will focus on different kinds of predator responses to changes in prey density. Population levels and individual predators will be affected by changes in prey density. There are three main kinds of predator response to changes in prey population size [79, 80]. A functional response represents changes in prey density as a result of predation [188]. According to [157], an aggregation response represents predators' responses to spatial changes in prey population size. An aggregation response can be defined as predator response to changes in prey density as a result of moving in specific ways to have finally prey aggregated in such areas. Functional and aggregation responses can be considered as direct results of the individual behaviour of prey, and they occur over a short timescale [188]. Variations in prey abundance will affect the predator population. For instance, let us consider a numerical response which occurring on a slower timescale. To illustrate the concept of numerical response, let us consider that predator reproduction occurs only once a year. In this case, if prey population increases in a straightforward fashion after a period of reproduction, the predator may take another year to increase their population density. Predators will take time to increase their population size to the point that they affect the abundance cycle of their prey.

Generally, functional and aggregation responses, which occur on a fast timescale, will not present any delay into the dynamic of predator-prey interactions and population densities. A numerical response will introduce time lags into the dynamic of population density, as

it occurs on a slower predator-prey interaction timescale, related to behaviour. However, in some cases this process can be broken, when the aggregation responses become very slow as a result of slow rate of predator movement, compared to the irregular invasion rate of prey. These types of predators response will be discussed greater detail in the following subsections.

2.1.1 Functional Response

The term 'functional response' was introduced by [173]. This term has become widely used due to the work conducted by Holling [79, 80]. According to [79], functional responses have been classified into three common types I, II, and III and these nomenclatures has been widely used in the literature. For a given unit of time, the average number of prey killed by an individual predator is known as the functional response. In the case of plant-herbivore models, the quantity of prey can be measured as biomass instead of individuals. Note that the functional response gives the rate of prey consumption per predator biomass. The reason for this, is that the functional response is considered the main part of the prey equation. However, the term 'consumption' which represents increases in the number of predators and predator reproduction, are the main part of the predator equation. Holling modified the functional forms of types I, II, and III by considering two periods of an individual predator's time to be 'searching for' and 'handling of' prey. Then the total time t that a predator spends in the predation process is the total time spent on searching and handling as follows $T_t = T_s + T_h$ where T_t is the total time, T_s denotes time spent searching for prey, and T_h represents time spent handling prey.

A functional response of type I is the simplest functional response pattern, which represents a linear relationship between number of prey killed and prey density; a function increases linearly up to the point that it forms a flat, horizontal line, as shown in Figure 2.1 [79]. A functional response of type I is considered firstly in the Lotka-Volterra

predator-prey model. A functional response of type I is often observed in passive predators, such as spiders, where the number of flies caught is proportional to the density of flies. A functional response of type I has the following form:

$$f(u) = aT_s u, \tag{2.1.1}$$

where a is a constant Holling called the 'discovery rate' or the slope of the line, T_s is time available for searching, and u represents prey population density in general. If the handling time for prey is not required, then the total time is considered to be time spent searching for prey as follows: $T_s = T_t$. Consequently, if predator population density equals P, prey population density will drop in line with the following equation aT_sPu. However, if each predator needs time to handle each individual prey animal, then the searching time for prey will reduce to $T_s = T_t - T_h f(u)$. A functional response of type I has different effects on the dynamics of the predator-prey system, depending on the value of the slope line [12]. It has been illustrated in [8] that, at a value of slope equals one, all population densities result in the prey species having the same probability of being killed by the predator species. If the value of slope is greater than one, the population dynamics tend to stabilise, where the probability of being eaten increases as a result of increasing prey density. For a slope with a value of less than one, as prey density increases, the risk of being eaten decreases, which reflects the situation at the maximum consumption rates in a functional response of type I.

Combining the time available for searching $T_s = T_t - T_h f(u)$, with equation (2.1.1), leads to $f(u) = aT_t u - aT_h u f(u)$, which implies:

$$f(u) = \frac{aT_t u}{(1 + aT_h u)}, \tag{2.1.2}$$

which is the functional response of type II. The second type of functional response is

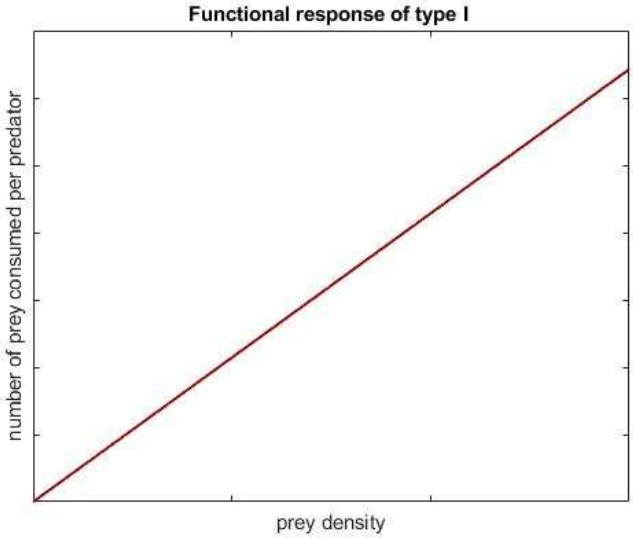

Figure 2.1: Functional response of type I

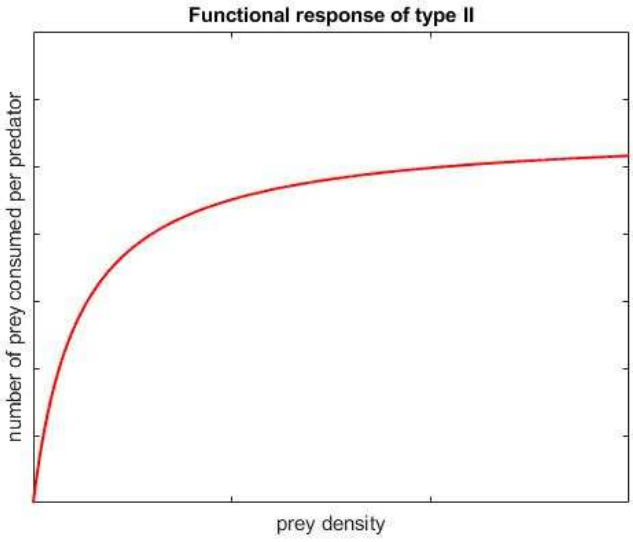

Figure 2.2: Functional response of type II

known as a hyperbolic response, and depends on strategies at the individual level. A hyperbolic response appears as a monotonically decreasing gradient curve, with increasing prey population density, to eventually reach saturation at a constant rate of prey consumption, as shown in Figure 2.2. It can be noticed from equation (2.1.2) that, in the total time interval T_t the prey population density drops according to the quantity $\frac{aT_tPu}{(1+aT_hu)}$. From equation (2.1.2), the term aT_h is dimensionless. If the handling time becomes much longer than discovery time, then aT_h is given a large value. On the other hand, the value of aT_h is small with a large value for discovery time, which converts a functional responses of type II to a functional response of type I, where the total time is considered to be $T_t = T_s$: according to the definitions presented by Holling [79], there is no difference between functional response of type I and II. Therefore, it can be seen that a functional response of type I does the same job, and has the same propriety, as type II, but is not as effective [188]. The predator-prey system presented by wolves and caribou is an ideal example of a type II functional response. When the number of wolves is constant, the caribou population increases, as a result the proportion of caribou killed per wolf drops as caribou population density increases. In other words, with a high density of caribou, the search time wolves spend finding caribou is very little, whereas almost of time is spent on handling prey.

A type III functional response often appears as a general form of the type II functional response, but at low prey population densities [79]. A functional response of type III, is a general form of the functional response of type II which can be derived as:

$$f(u) = \frac{aT_tu^k}{(1 + aT_hu^k)}. \tag{2.1.3}$$

A functional response of type III occurs when the gradient of the curve first increases. Then with an increasing prey population density it decreases, as can be seen in Figure

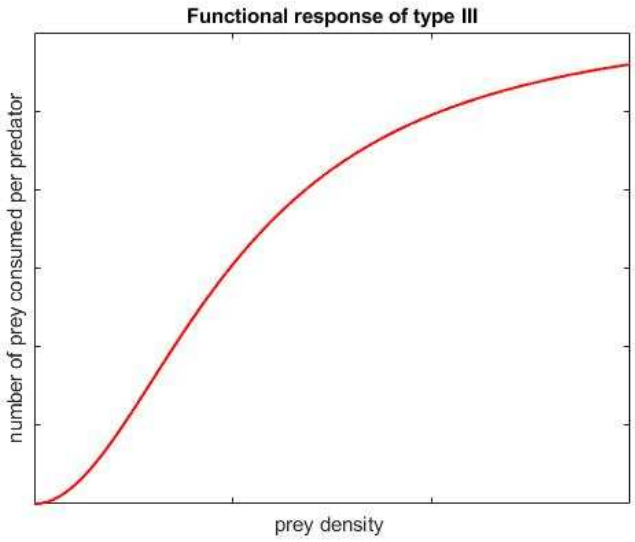

Figure 2.3: Functional response of type III

2.3. At high prey population densities a functional response of type III occurs when predators' search activities increase. For instance, many species of predator respond to kairomones and increase their activity accordingly. This ecological behaviour is known as 'sigmoidal' behaviour, and is attributed in the predator population to the existence of learning time [158]. According to [79, 188], learning time is defined as improvement in the ability of predators to handle prey effectively as prey density increases. It has been detected in [158], different distinct behaviours that can change from a functional response of type II to type III. Such predators will learn better or specialist techniques in order to develop their ability to efficiently search and handle prey within a particular area. If the prey population density is low, then the possibility of a predator finding prey becomes extremely low. The predator will find prey infrequently due to their low density; as a result, they will not have enough experience to develop their skills in capturing and subduing available prey species. An example to illustrate this mechanism was introduced in [79], concerning shrews and deer mice that feed on sawflies.

2.1.2 Aggregative Response

As mentioned previously, an aggregative response is a direct result of explicit spatial dynamics. Aggregative responses are not given enough importance, and many text books do not refer to aggregative responses; and they all in confuse between aggregative and numerical responses in some cases. For instance, aggregative have been considered to be a kind of numerical response by some authors. However, it is known that aggregative responses result in a local increase in predator numbers. Changes in predator birth and mortality rates represent a numerical response. A numerical response occurs over a longer timescale than aggregative and functional responses, as mentioned above. Therefore, there is a high demand to specify an explicitly spatial process to deal with aggregative responses. Let us consider that simple spatially implicit functional forms can be used to approximate aggregative responses. With very variable prey populations in many different spatial locations, let us suppose that movement of predators is free over the whole area. An optimal free distribution of predator aggregative will provide the simplest pattern of predator aggregation. In order to generate the ideal aggregative free distribution pattern of predator, predators must forage in an area of prey density that exceeds the threshold level of prey density. When the prey density is below this threshold, the number of predators will be zero. Thus, the number of predators in areas of high prey density will be equal V_{high}; the highest density of predators equals the total number of predators in the landscape divided by the number of high-density patches of prey. At high-prey density locations, the rate of predation by each predator will be considered as a constant C. According to [157], The functional response of a predator to the high-prey density locations will be saturated, and according to [157] the total killing rate will be presented as follows:

$$\text{total killing rate} = \begin{cases} 0 & \text{if} \quad u < f, \\ g & \text{if} \quad u > f, \end{cases} \tag{2.1.4}$$

where $g = CV_{high}$ is constant. However, this is not a common situation because at discontinuous steplike fashion when prey density equal to threshold level f ($i.e\ u = f$), then the killing rate will be changed consequently.

2.1.3 Numerical Response

According to [188], changes in predator density that are represented by a function of prey and predator population are known as numerical response which were first introduced in [173]. A numerical response consists of three main components: growth rate of individual predators, ratio of predator reproduction, and percentage of predator deaths [11]. The processes of individual growth and predator reproduction are considered to be a single process that increases predator biomass. The ratio of predator deaths is used to estimate the amount of energy required to maintain a stable population. According to [60], the relationship between the energy extracted from prey consuming and the energy available is a linear function of consumed food. Thus, the growth rate in the predator population per capita can be presented as follows:

$$\frac{dV}{Vdt} = \chi(I - \mu). \tag{2.1.5}$$

In order to understand the meaning of equation (2.1.5), the equation has been formulated in terms of energy use. According to energy use , let us consider the amount of energy in predator biomass as V. I represents the ratio of energy in prey biomass that is consumed by an individual predator. The assimilation efficiency is presented by χ. Then, χI is the rate of assimilated prey energy per predator, and $\chi\mu$ represents the maintenance term. The rate of growth and reproduction is introduced by $\chi(I - \mu)$. Then, when the rate of consumption is zero, individual predator biomass will decay as $\chi\mu$ consequently. The next step is to rewrite equation (2.1.5) in terms of population level. Let P denotes predator density, μ is zero population growth, which represents consumption rate when individual

predator works to satisfy its requirements of maintenance. The χ parameter presents the rate of increases in the predator population. Finally, in the absence of prey, predator mortality rates are represented by $\chi\mu$.

A beneficial starting point for predator-prey interactions can be provided by equation (2.1.5). However, in order to apply this to real-life situations, equation (2.1.5) must be modified. For instance, according to [60], in the absence of prey, the assumption that consider predator mortality rates are constant is unrealistic. Then, to link equation (2.1.5) to the prey equation, let us consider the consumption rate equals the predation rate. Equation (2.1.5) can then be rewritten as follows:

$$\frac{dP}{Pdt} = \chi(\frac{CN}{d+N} - \mu). \tag{2.1.6}$$

Here, predator responses are represented as a hyperbolic functional response.

2.2 The Rosenzweig - MacArthur Model

The spatially explicit predator-prey system from the Rosenzweig- MacArthur model will be used in this chapter to generate series of ecological data. A computer simulation will be used to obtain ecological data, instead of depending on data acquired from experiments. In order to apply numerical integration methods directly to ecological data obtained from experiments, the performance of these methods needs to be assessed by computing the integration error. Hence, the quantity of ecological data must increase in order to plot the convergence curve $E(N)$ of the integration error as a function of N grid nodes. Then, the original condition of the experiments must be reproduced as many times as required to collect ecological data on refined grids, which is indeed restrictive to our applications. In other words, it is extremely difficult to ensure that the same initial conditions will be available when repeating experiments, thus, according to [139], ecological data obtained

from computer simulations will be used instead. Let us assume that our ecological system consists of two species. In interacting between the species, the population density dynamics of each species will be affected axiomatically. There are three different kinds of effective ecological interaction [103, 119]. The predator-prey interaction represents the first type of ecological interaction. This kind of interaction can be defined as the change in a growth rate of species, where one of the populations will increase and the second population will consequently decrease. In this kind of interaction, it assumed that a simple interaction may lead to oscillation in population densities. For instance, an increase in the prey population will increase the growth rate of the predator population. As a result, more predators will consume more prey, which leads to a decrease in the growth rate of prey. Consequently, the growth rate of the predator species will decline due to a smaller quantity of available food. When the predator population consequently drops, the growth rate of the prey will increase to allow the trophic cycle to start again. Therefore, as can be seen from the previous explanation of predator-prey interaction, such oscillations impact the population density of species. The second type of ecological interaction shows a decrease in the growth rate of each population, which is known as ecological competition. Finally, if the growth rate of each population increases, this reflects a relationship variously known as coexistence, mutualism, or symbiosis. According to [103, 119], there are several models that can be employed to generate ecological data. The most common framework for presenting predator-prey interactions is the following:

$$\frac{du}{dt} = \mathrm{A} - \mathrm{B} \tag{2.2.1}$$

$$\frac{dv}{dt} = \mathrm{C} - \mathrm{D}, \tag{2.2.2}$$

where u, v are densities of prey and predator respectively; A is the rate of prey growth in the absence of predators; B is the rate which prey are killed by predators; C represents

the decline in the predator population in the absence of prey; and D is the growth rate of the predator population due to prey consumption. By considering the rate of consumed prey to introduce new predators is proportional to the killing rate, we have :

$$\frac{du}{dt} = r(u)u - f(u,v)v \tag{2.2.3}$$

$$\frac{dv}{dt} = \chi f(u,v) - \delta(v)v, \tag{2.2.4}$$

where in the absence of predators, $r(u)$ is the rate of prey growth per capita, whereas, in the absence of prey, $\delta(v)$ is the rate of predator decline per capita. $f(u,v)$ is the functional response of the predator. Finally, χ is the increase in the predator growth rate as a result of prey consumption. Equations (2.2.3), (2.2.4) represent the simplest form of Lotka-Volterra model. In order to generalise Lotka-Volterra models, is necessary to add more sophisticated assumptions into equations (2.2.3), (2.2.4). According to [193], in the absence of predators, prey grows logistically in the following manner:

$$\frac{du}{dt} = r_0(u(1 - \frac{u}{k})) - auv \tag{2.2.5}$$

$$\frac{dv}{dt} = \chi auv - \delta_0 v. \tag{2.2.6}$$

The Volterra model is distinguished by a point of stable equilibrium. Therefore, the Volterra model should be more stable than the Lotka-Volterra model. In addition, we can conclude that any density dependence will contribute to the stability of the system.
The Rosenzweig- MacArthur model [161, 188] is a general form of the Lotka-Volterra model [100, 193]. The latter is a system of equations that has a functional response incorporated into it. The functional response can be defined as the temporal rate at which each predator kills prey [79], and it does not have any terms of self-limitation (i.e. in both resource and consumer equations, there is no direct dependence on density feed-

back). The Lotka-Volterra model depends on assumptions that the densities of resource (prey) and consumer (predator) will increase or decrease exponentially, in the absence of effects from other species. Although it is very widespread because of its simplicity, the Lotka-Volterra model has not been successfully applied to pest population density issues in a real-life agricultural environment. Hence, it has not been considered to be a realistic model [100, 193]. Therefore, in order to be applicable to real-life applications, the Lotka-Volterra model must be modified. The functional response will therefore be replaced by a hyperbolic response (non-linear response); and by adding self-limitation to the resource (prey) in a logistic way (*i.e* direct population dependence), the result is the Rosenzweig-MacArthur model [161, 188]. The Rosenzweig-MacArthur model is a spatiotemporal model, where the population densities are functions of time and space. From a mathematical viewpoint, the Rosenzweig-MacArthur model is a system of coupled diffusion-reaction equations that describe the spatiotemporal dynamics of pest population density [119]. Let us consider the general form of the predator-prey model given by equations (2.0.3), (2.0.4), where, U and V denote the population densities of prey and predator respectively, P denotes the linear growth rate of prey, E refers to the rate of predation, $M(V)$ is the predator mortality rate, and predation efficiency is represented by the coefficient κ. The term $E(U,V)$ in the system of equations (2.0.3), (2.0.4), representing the functional response of the predator to the prey density, which is defined as the temporal rate of prey killed per individual predator [188]. According to [79], there are three different classifications of functional response. However, the Rosenzweig-MacArthur Model depends only on the functional response of type II:

$$E(U,V) = A\frac{UV}{U+H}, \tag{2.2.7}$$

where A is the rate of predation and H denotes the half-saturation prey density.

The assumption of considering that when the population density of prey becomes roughly equals zero the growth rate of predator per capita will reach its maximum, where this value decreases monotonically as a result of increasing population density. In reality, ecological populations do not always follow this assumption. Several studies have been conducted to prove that, per capita the population growth rate arrives its maximum value at nonzero value of population density. This ecological phenomenon is known as the Allee effect, and the associated population dynamics is often called the Allee effect. At low population density, the Allee effect is defined as a reduction in individual fitness [103, 115, 119]. Another definition of the Allee effect represents the positive synchronisation between population density and individual fitness; in other words, the Allee effect reflects positive population density dependence. The Allee effect can be created by a variety of mechanisms involving biological interactions, predation, environmental conditions, and mating systems. At low population densities, the Allee effect can be classified in different types, depending on the nature of population density. A strong Allee effect appears as a result of population shrinkage at low densities [127, 194], which has been paid special attention in this case. A weak Allee effect is noticeable when the rate of proliferation is positive and increasing. If the proliferation rates are positive, but decreasing when the densities are low, the null hypothesis of the Allee effect has occurred. The Allee effect is considered affect virtually all aspects of interactions between species in space and time [5, 14].

In addition, attention will be restricted to the one-dimensional case for the purposes of simplicity; then, equations (2.0.3), (2.0.4) can be presented as follows:

$$\frac{\partial U(X,T)}{\partial T} = D_1 \frac{\partial^2 U}{\partial X^2} + \alpha U(1 - \frac{U}{K}) - A\frac{UV}{U+H} \tag{2.2.8}$$

$$\frac{\partial V(X,T)}{\partial T} = D_2 \frac{\partial^2 V}{\partial X^2} + \kappa A \frac{UV}{U+H} - MV. \tag{2.2.9}$$

The system of equations (2.2.8), (2.2.9) depends on eight meaningful ecological parameters. These parameters are defined as follows: D_1 and D_2 are the diffusion rate of prey and predator densities respectively, α is the rate of linear growth of the maximum prey per capita, K denotes the carrying capacity of the prey, A is the rate of attacks from a predator, κ is the coefficient of food assimilation efficiency, and M is the parameter that represents the mortality rate of the predator.

The above system should be presented in a dimensionless form for simplicity. Thus, we consider the following: $D_1 = D_2 = D$, $x = X/L$, where L is the length of our field. Time in a dimensionless form is $t = \alpha T$. The scaling of the prey population density is $u = U/K$. In the same way, the dimensionless form of the predator population size is $v = VA/\alpha K$. Consequently, the system of equations (2.2.8), (2.2.9) can be written in the dimensionless form as follows:

$$\frac{\partial u(x,t)}{\partial t} = d\frac{\partial^2 u}{\partial x^2} + u(1-u) - \frac{uv}{u+a} \tag{2.2.10}$$

$$\frac{\partial v(x,t)}{\partial t} = d\frac{\partial^2 v}{\partial x^2} + k\frac{uv}{u+a} - mv, \tag{2.2.11}$$

where the dimensionless parameters are introduced as $d = D/\alpha l^2$, $k = \kappa A/\alpha$, $m = M/\alpha$, and $a = H/L$. Therefore $u(x,t)$ and $v(x,t)$ now depend on the combinations of dimensionless parameters k, m, a and d.

The system of equations (2.2.10), (2.2.11) is known as the spatial one- dimensional Rosenzweig-MacArthur model [188].

2.3 Stability of the Solutions

Generally, the properties of spatial systems differ from the ones of nonspatial systems, as the dynamics of spatial patterns are much richer than nonspatial ones. Let us now

consider a nonspatial system from the system (2.2.10), (2.2.11) as:

$$f(u,v) = u_t = u(1-u) - \frac{uv}{u+v}; \qquad g(u,v) = v_t = \kappa\frac{uv}{u+a} - mv. \tag{2.3.1}$$

In order to investigate the stability of nonlinear system of equations (2.3.1), standard linear stability analysis can be used [4, 61]. It is difficult to solve a system of nonlinear equations (2.3.1), therefore, the system must be linearised. Linearisation is a method used to assess the local stability of an equilibrium point of a system of nonlinear differential equations [114]. Only three stationary points (u^*, v^*) can be obtained from the phase plane (u, v). We can have the extinction state $(0, 0)$, a state of predator absence $(1, 0)$, and the coexistence of species $(\bar{u}, \bar{v})$, where:

$$\bar{u} = \frac{pa}{1-p}, \qquad \bar{v} = (1-\bar{u})(a+\bar{u}), \tag{2.3.2}$$

where $p = m/\kappa$.

Eigenvalues of the Jacobian matrix can be used to determine the type of stationary points [75, 192]. The Jacobian matrix of the system (2.3.1) at the fixed point (u^*, v^*) is given as follows:

$$J = \begin{bmatrix} \frac{\partial f}{\partial u}(u^*, v^*) & \frac{\partial f}{\partial v}(u^*, v^*) \\ \frac{\partial g}{\partial u}(u^*, v^*) & \frac{\partial g}{\partial v}(u^*, v^*) \end{bmatrix} \tag{2.3.3}$$

Eigenvalues of J are values of λ that satisfy the following equation:

$$det(J - \lambda I) = 0, \tag{2.3.4}$$

where I is the 2×2 identity matrix. Equation (2.3.4) is as follows:

$$\lambda^2 - \lambda tr(J) + det(J) = 0, \tag{2.3.5}$$

where equations (2.3.4) and (2.3.5) are called the characteristic equation of matrix J. As mentioned above, depending on the values of the eigenvalues, the type of fixed point can be determined. If the eigenvalues are real positive numbers, then the equilibrium point is an unstable point. The equilibrium point is a stable point if both eigenvalues are real negative numbers. The equilibrium point is a saddle unstable point if the eigenvalues are real numbers with different signs [33, 102].

Changes in the stability of steady states can be revealed by tracing any changes of sign in $tr(J)$, $det(J)$ and $\Delta = (trJ)^2 - 4det(J)$. For any values of the parameters m, a, and k in the system of equations (2.3.1), the fixed point $(0, 0)$ is a saddle point, whereas point $(1, 0)$ becomes a saddle point at $a < \frac{1-p}{p}$; otherwise, $(1, 0)$ is a stable point [103] . For the stationary point $(\bar{u}, \bar{v})$, we have:

$$tr(\bar{J}) = \frac{p}{1-p}[(1-a) - p(1+a)], \quad det(\bar{J}) = \kappa p[1 - p(1+a)], \tag{2.3.6}$$

where $\bar{J}$ is Jacobean matrix at equilibrium point $(\bar{u}, \bar{v})$. The stability of coexistence state is changed to $a = \frac{1-p}{1+p}$. The system of equations (2.2.10), (2.2.11) has three parameters a, m, and k. However, $tr(J)$ relies on the ratio $p = m/k$ but not on each parameter individually. Therefore, variations in k or m lead to changes in the structure of the phase plane (p, a) only. Knowing the dynamics of nonspatial patterns helps to predict the dynamics of a spatial system.

2.4 The Numerical Solution to the Spatially Explicit Form of the Rosenzweig-MacArthur System

The spatially explicit form of the Rosenzweig-MacArthur predator-prey model is a presentation of an ecological system given by a mathematical model using partial differential equations. This system of equations can be solved in a closed form by applying the an-

alytical approach when dealing with a simple system of equations. However, for a more complex system of equations, the analytical scheme becomes much more complicated to employ. Thus, numerical methods are used to find approximated solutions that are sufficiently close to the exact solution. Finite difference methods can be successfully utilised as a numerical approach to evaluate the solution for the partial differential equations. The finite difference methods are numerical approach that discretise of differential equations to convert them into a system of algebraic equations. In this section, we first consider the parabolic partial differential equation, which is sometimes called the diffusion equation or heat equation. The parabolic equation can be written as follows [27, 150]:

$$\frac{\partial u}{\partial t} = d\nabla^2 u, \tag{2.4.1}$$

where $\nabla^2 = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})$ and d denotes the diffusion rate. In order to simplify our explanation, we will consider function u as a function of two variables, x and t. Then, equation (2.4.1) can be rewritten as follows:

$$\frac{\partial u}{\partial t} = d\frac{\partial^2 u}{\partial x^2}, \tag{2.4.2}$$

for any $0 < x < 1$, $t > 0$. In order to approximate the solution to the one-dimensional parabolic partial differential equation, the forward difference approximation of the first and the second derivatives is used [117]. In addition, to fulfil the requirements for the uniqueness of the solution, the initial and boundary conditions have to be placed. Therefore, let us consider, for instance, the initial condition of the equation (2.4.2) as $u(x, 0) = u_0, \ \forall x \in [0, 1]$, and the boundary conditions will be $u(0, t) = U_0, \quad u(1, t) = U_1$, $\forall\ t > 0$. The discrete equation for equation (2.4.2) can be written as follows:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = d[\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2}], \tag{2.4.3}$$

where Δt is the time-step size, h represents the grid-step size, and u_i^n approximates the values of $u(x_i, t^n)$ of the exact solution for any (x_i, t^n), where $x_i = ih;\ \ i = 0, 1, ..., N$ and $t^n = n\Delta t$. Once we have determined discrete forms of the initial and boundary conditions as follows: $u_i^1 = u_0,\ \ u_1^{n+1} = U_0,\ \ u_N^{n+1} = U_1$, the parabolic initial boundary value problem can be approximated as follows:

$$u_i^{n+1} = u_i^n + \frac{d\Delta t}{h^2}[u_{i+1}^n - 2u_i^n + u_{i-1}^n]. \tag{2.4.4}$$

For the purpose of simplicity, the term $\frac{d\Delta t}{h^2}$ can be considered in equation (2.4.4) as follows:

$$\lambda = \frac{d\Delta t}{h^2}. \tag{2.4.5}$$

Then the equation (2.4.4) can be written as follows:

$$u_i^{n+1} = \lambda u_{i+1}^n + (1 - 2\lambda)u_i^n + \lambda u_{i-1}^n. \tag{2.4.6}$$

Equation (2.4.6) represents the explicit method [150]. The notion of an explicit method depends on calculating the state of equation (2.4.6), at a later time, from the state of equation (2.4.6) at the current time. Then, the partial differential equation can be solved numerically without the need to solve the system of algebraic equations.

The finite difference scheme is consistent if the operator reduces to the original differential equation as the increments in the independent variables vanish. The finite difference scheme shown in (2.4.6) will reproduce the original differential equation (2.4.2) as $h \to 0$ and $\Delta t \to 0$; therefore, the proposed scheme for the numerical solution of this equation is consistent. It must be mentioned here that any finite difference scheme based on an approximation of the derivatives should be consistent. Another important characteristic of a numerical scheme in the solution of partial differential equation is stability. Stability

means that the error caused by a small perturbation in the numerical solution remains bound; otherwise, the numerical solution will be unstable. In order to have a stable solution for any explicit system, the Courant- Friedrichs -Lewy (CFL) condition has to be held [93, 150]. The CFL condition represents the most important condition for any explicit scheme to be stable, and is:

$$|\lambda| = \left|\frac{d\Delta t}{h^2}\right| \leq \frac{1}{2}, \quad \text{or } |\Delta t| \leq \left|\frac{h^2}{2d}\right|, \tag{2.4.7}$$

where λ is referred to as the CFL number. The issues of stability and accuracy that are raised when the explicit scheme is used are fully understood. However, the choice is justified because of the simplicity of computation. We are interested in the solution at small time t. In order to hold the CFL condition, the time-step size must be very small, and on fine grids with $h << 1$. However, it does not increase the cost of computations in small time steps. Therefore, the explicit method is conditionally stable. Finally, as $\Delta t \to 0$ and $h \to 0$, the numerical solution will converge to the exact solution, which represents the definition of convergence [27, 150]. From previous discussion, we can say that consistency plus stability gives convergence. It is worth nothing here that the topic of stability of finite differences schemes is beyond the scope of our study and we do not discuss it in details. The interested reader can find more detailed information in [27, 61, 150].

Therefore, the discrete equations for the system of equations (2.2.10), (2.2.11) are given as follows:

$$u_i^{n+1} = u_i^n + \lambda[u_{i+1}^n - 2u_i^n + u_{i-1}^n] + \Delta t[u_i^n(1 - u_i^n) - (\frac{u_i^n v_i^n}{u_i^n + h})] \tag{2.4.8}$$

$$v_i^{n+1} = v_i^n + \lambda[v_{i+1}^n - 2v_i^n + v_{i-1}^n] + \Delta t[\kappa(\frac{u_i^n v_i^n}{u_i^n + h}) - m v_i^n]. \tag{2.4.9}$$

The discrete initial and boundary conditions for the system of equations (2.4.8), (2.4.9) can be written as follows:

$$u_i^1 = u_0,\ v_i^1 = v_0,\ u_1^{n+1} = u_2^{n+1},\ v_1^{n+1} = v_2^{n+1},\ u_N^{n+1} = u_{N-1}^{n+1},\ v_N^{n+1} = v_{N-1}^{n+1}.$$

It is clear from the previous explanation of the explicit method that in order to achieve a stable solution for the system of equation (2.4.6), the time-step size Δt must be very small. As a result, the CFL condition is considered to be a restriction of the implementation of the explicit scheme. In order to deal with a partial differential equation that contains some terms which may lead to an unstable solution, an alternative method should be applied instead of the explicit method, which requires the time-step size to be extremely small. Therefore, the implicit method may be applied in this situation to achieve a stable solution to the system of partial differential equations, without any restriction on the time-step size [27, 74] where the implicit method is unconditionally stable. Therefore, for the simplicity of computation, and since we are interested in solutions at small time steps in order to hold the stability condition, the explicit forward Euler method will be used in the following sections in the numerical simulation.

2.5 Results of $1d$ Numerical Simulation

Heterogeneity is a common feature in the spatial distribution of ecological population density [104, 105]. The heterogeneity of population density distributions have degrees of complexity and variability in the structure of habitats (plant/animal groupings). In an ecosystem, a wide range of practical and theoretical problems are effected by heterogeneity [104]. Areas of high population density usually alternate with uninhabited area, or with areas of low population density. This distribution within an ecosystem is known as patchiness or aggregation as well as by other names [103, 119]. Populations densities are often distributed irregularly. According to [94, 149], patchiness is considered to be a separate phenomenon, which often emerges as a result of biological interaction rather

than due to the environmental properties of the different locations. Therefore, the characteristics of spatial structure are often uncorrelated or weak correlated with the patchiness of population heterogeneity which is an important phenomenon.

In patchy population densities, self-organised biological patterns may be a result of different aspects of species kinetics. Patchy patterns that appear as a result of flock or swarm formation are not considered here. Therefore, spatial scale must be large enough to prevent the patchiness that occurs as a result of dealing with the small scale emergence of flock or swarm formations. In some ecological cases the mechanism of self-organised ecological diagrams mainly depends on properties of the landscape especially on the degree of heterogeneity. A species areal consists of groups of habitats in a very fragmented agricultural field. The population distribution inside each habitat will be considered homogeneous, if the size of habitat is quite small due to small fluctuations. In order to describe this phenomenon mathematically, coupled systems of ordinary differential equations will be employed. Each system is a space-discrete model that describe the kinetics of population density inside each habitat. It has shown in [84, 85] that when the inter-habitat becomes weak, the dynamics of predator-prey population density oscillation become uncorrelated or desynchronised. The dynamics of the subpopulations even at periodic kinetics of local populations become chaotic due to weak coupling.

The choice of initial conditions is very important to demonstrate the dynamics of a spatiotemporal system (2.2.10), (2.2.11). When dealing with chaotic patterns both species are distributed over the whole domain. The simplest pattern of the initial distribution is to be spatially homogenous. When the initial condition is homogenous, the distribution will remain homogenous over time, and we will not be able to generate a spatial distribution from the initial distribution. Therefore, in order for spatiotemporal patterns to emerge, the homogenous distribution must be perturbed [103, 119]. In the hypothetical

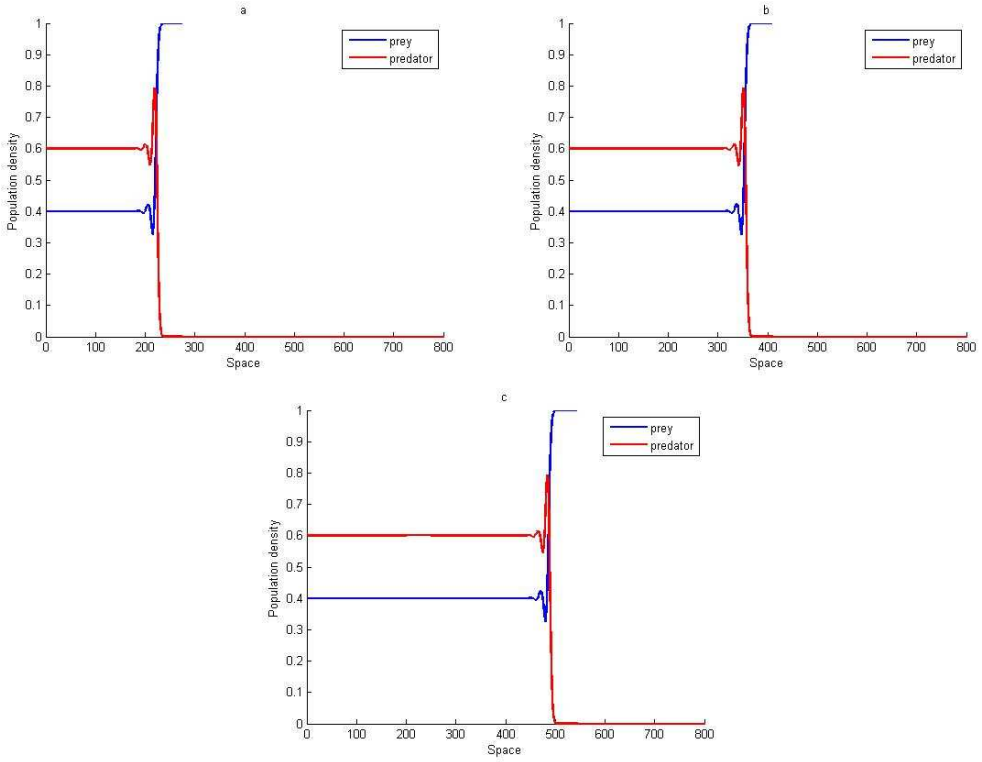

Figure 2.4: Prey and predator density versus space obtained from system (2.2.10), (2.2.11) for parameters $k = 2.0,\ p = 0.4,$ and $a = 0.6$, where the initial distribution is given by (2.5.2), (2.5.3) with $d_v = 200,\ v_0 = 0.5$. (a) At $t = 100$. (b) At $t = 200$. (c) At $t = 300$.

distribution, we have the following initial conditions:

$$u_i^1 = \bar{u}; \quad v_i^1 = \bar{v} + \epsilon x + \delta, \tag{2.5.1}$$

where ϵ, and δ are parameters. The system of equations (2.2.10), (2.2.11) relies mainly on the values of $(\epsilon, \delta\ )$.

A smooth monotonic distribution will be generated in the case of $\delta = 10^{-2}$, and $\epsilon = 10^{-5}$; this biological behaviour is expected from system (2.2.10), (2.2.11). Indeed, in the theoretical ecology diffusion reaction, equations involving only two species are considered to be too simple to provide any complex distribution patterns over time. Therefore, in order to generate more complicated distributions, different values of parameters in the

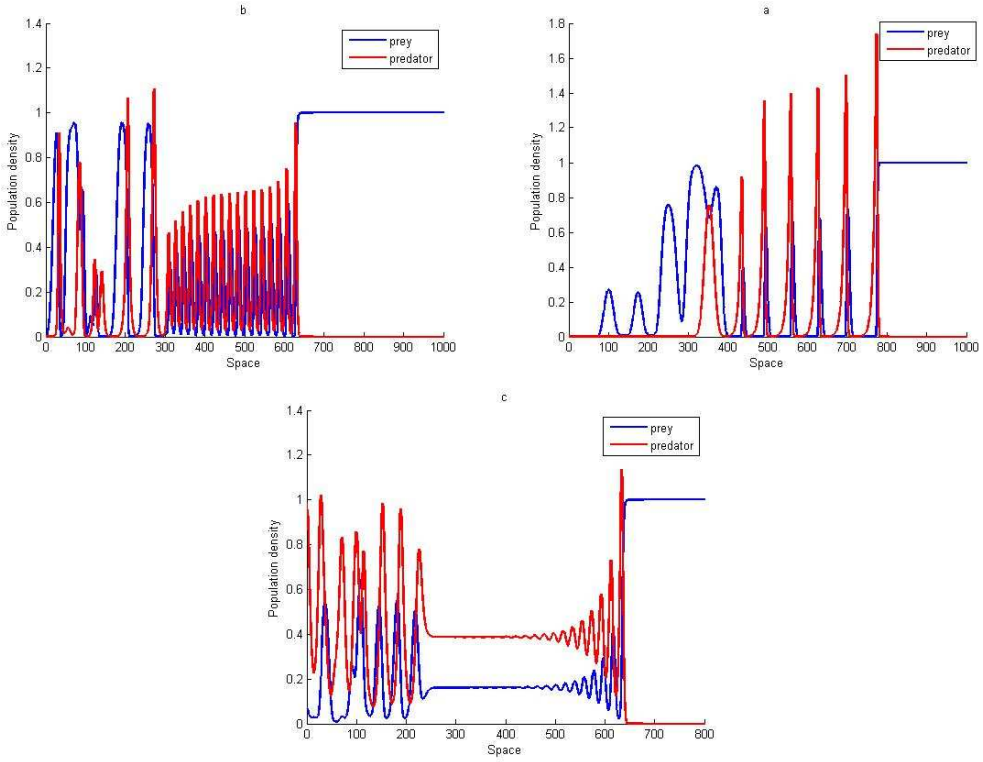

Figure 2.5: Prey and predator density versus space obtained from system (2.2.10), (2.2.11) for parameters $k = 2.0$, and $a = 0.1$ where the initial distribution is given by (2.5.2), (2.5.3) with $d_v = 200$, $v_0 = 0.5$. (a) At $t = 290$ and $p = 0.2$. (b) At $t = 300$ and $p = 0.5$. (c) At $t = 400$, $a = 0.35$, and $p = 0.3$.

system of equations (2.2.10), (2.2.11) will be chosen. An irregular spatiotemporal pattern can be generated over the entire domain by slightly altering the value of $\delta = -1.5. \times 10^{-2}$. In this case, the initial distribution will introduce very irregular dynamics into a sub-domain, which will, with time become distributed over the whole domain.

For different values of parameters in system of equations (2.2.10), (2.2.11) several patterns of oscillations in population densities were investigated in the work of [103]. Figure 2.4 shows patterns of population front distribution, where the oscillation is prominent at the front but flattens out promptly behind the front to have monotone population density. Over time, the length of the unstable plateau may grow up to the maximum value, then, approximately over the whole domain, the population density distribution becomes

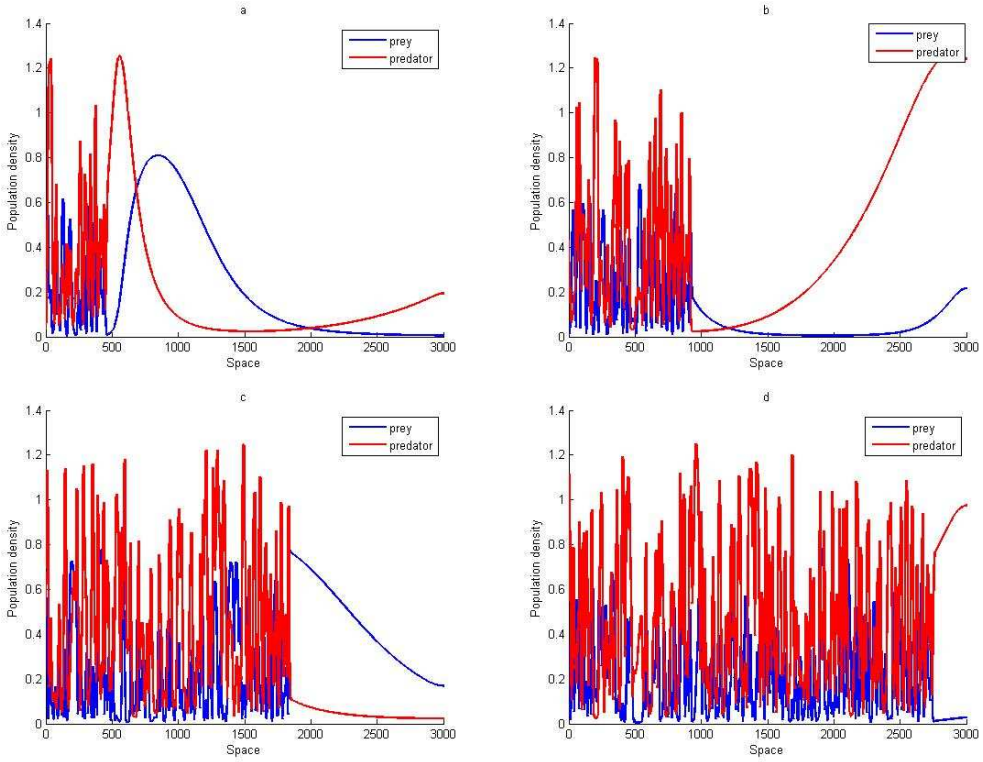

Figure 2.6: Prey and predator density versus space obtained from system (2.2.10), (2.2.11) for parameters $k = 2.0,\ p = 0,35,$ and $a = 0.3$, where the initial distribution is given by (2.5.1) with $\epsilon = 10^{-5}$, and $\delta = -1.5.(10^{-2})$. (a) At $t = 1000$ (b) At $t = 2000$. (c) At $t = 4000$. (d) At $t = 6000$.

homogeneous. By choosing different values for the parameters given in the system of equations (2.2.10), (2.2.11), with initial conditions (2.5.2), (2.5.3), several population density distributions have been generated, as depicted in Figure 2.5.

$$u_i^1 = 1 \quad \text{for any x}, \tag{2.5.2}$$

$$v_i^1 = v_0 \quad \text{if } |x| < \frac{d_v}{2} \quad \text{and} \quad v_i^1 = 0 \quad \text{if} \quad |x| > \frac{d_v}{2}. \tag{2.5.3}$$

It can be noticed from Figure 2.5a, and Figure 2.5b that the homogenous distribution of the population density, which was the common feature of the pattern behind the travelling front in Figure 2.4, is now more varied, with much more distinct population oscillations.

The domain that is occupied by oscillations, introduces a travelling wave pattern that spreads over entire field in the same shape and the diffusion speed is constant. In another pattern of travelling front, depicted in Figure 2.5c, a plateau appears in the weak of the travelling population wave, behind the strong oscillation front, and inside this area, the population density is a quasi-homogenous. With time, the emerged plateau may grows gradually to finally reach its maximum length as the overall length of our domain.

This phenomenon of irregularity araises a question as to whether this irregular pattern is chaotic. To answer this, different measures will be used to measure the sensitivity of the solutions to small oscillations. In fact the difference between perturbed solutions and unperturbed ones at the initial dynamic stage remains small; however, this difference steadily grows over time [103]. Hence, the chaos in the system (2.2.10), (2.2.11) must be in the spatiotemporal system, as this phenomenon cannot exist in the nonspatial system, which only produces simple dynamics.

The chaotic system appears as a result of a wave of chaos moving between the two regions. Furthermore, different chaos diagrams can be obtained by choosing different sets of initial conditions. By considering the system (2.2.10), (2.2.11) with initial conditions given by (2.5.1), two different patterns of spatiotemporal population densities have been generated, as shown in Figure 2.6. It can be seen from Figure 2.6a that for different values of time t, the dynamics of the system change, at $t = 1000$, a strongly oscillatory dynamic pattern inside the sub-domain of the system is generated, and this irregular pattern spreads throughout the domain over time, as shown in Figure 2.6b, and Figure 2.6c. This strongly irregular pattern grows steadily over time to encompass the whole domain, as depicted in Figure 2.6d.

2.6 Results of $2d$ Numerical Simulation

The system of equations used to generate the $1d$ ecological data has been discussed above. We will now consider a two-dimensional $2d$ counterpart to the system (2.2.8), (2.2.9). As mentioned above, the data obtained from the experiments was not sufficient to achieve the desired degree of accuracy. As a result of this, the number of grid nodes needed to be increased in order to approximate the solution effectively. However, this has practical problems because it is not possible to re-create the same initial conditions of the experiment, which must be replicated in order to increase the number of points in the grid. Therefore, data obtained by applying the computer simulation will be used instead. According to [119, 188], the system of equations that represents the spatially explicit predator-prey model in the 2d case is as follows:

$$\frac{\partial U(X,Y,T)}{\partial T} = D_1\left(\frac{\partial^2 U}{\partial X^2} + \frac{\partial^2 U}{\partial Y^2}\right) + \left(\frac{4\nu}{(K-U_0)^2}\right) \times U(U-U_0)(K-U) - \frac{AUV}{U+B} \tag{2.6.1}$$

$$\frac{\partial V(X,Y,T)}{\partial T} = D_2\left(\frac{\partial^2 V}{\partial X^2} + \frac{\partial^2 V}{\partial Y^2}\right) + \kappa\frac{AUV}{U+B} - MV, \tag{2.6.2}$$

where U and V are the densities of prey and predator at time $T > O$, the position of traps is (X, Y) ; $0 < X < 1, 0 < Y < 1$, and other parameters have evident ecological meaning as discussed for a $1d$ case [119]. The domain is the unit square $[0, 1] \times [0, 1]$. The system above must be rewritten in a dimensionless form as follows:

$$\frac{\partial u(x,y,t)}{\partial t} = d\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \beta u(u-b)(1-u) - \frac{uv}{1+Au} \tag{2.6.3}$$

$$\frac{\partial v(x,y,t)}{\partial t} = d\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \kappa\frac{uv}{1+Au} - mv. \tag{2.6.4}$$

In the system of equations (2.6.3), (2.6.4) the dimensionless parameters are represented as follows: $A = K/B$, $b = U_0/K$, $\beta = 4\nu BK/A\kappa(K-U_0)^2$, $m = M/a$, $d_1 = D_1/aL^2$

and $d_2 = D_2/aL^2$, where L is the size of the given agricultural field, D_1 and D_2 denote the diffusion rates, K represents the parameter of prey carrying capacity, U_0 is the Allee threshold density, ν is the maximum value of the growth rate of prey per capita, A represents the rate of predator attacks, coefficient B describes the half-saturation prey density, parameter κ presents food assimilation efficiency, and M is the predator mortality rate. Let us notice that the system (2.6.3), (2.6.4) is different from the system (2.2.10),

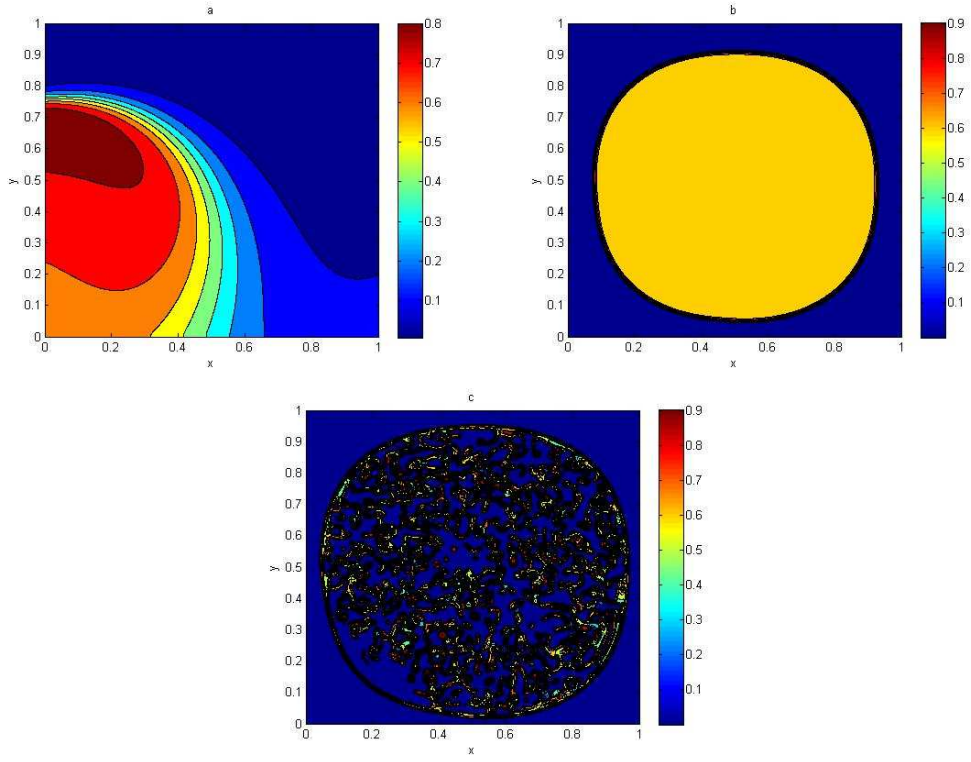

Figure 2.7: Examples of the pest population density function generated by the system of equations (2.6.3), (2.6.4). (a) Ecological test case 1 (ETC1). (b) Ecological test case 2 (ETC2). (c) Ecological test case 3 (ETC3).

(2.2.11). In this chapter we follow the discussion in [137], where a similar system describing the dynamics of prey-predator interaction has been introduced. Our choice of model (2.6.3), (2.6.4) is based on the fact that it is well known that this model can provide the wealth of spatial patterns that we intend to study when our technique of numerical integration is introduced.

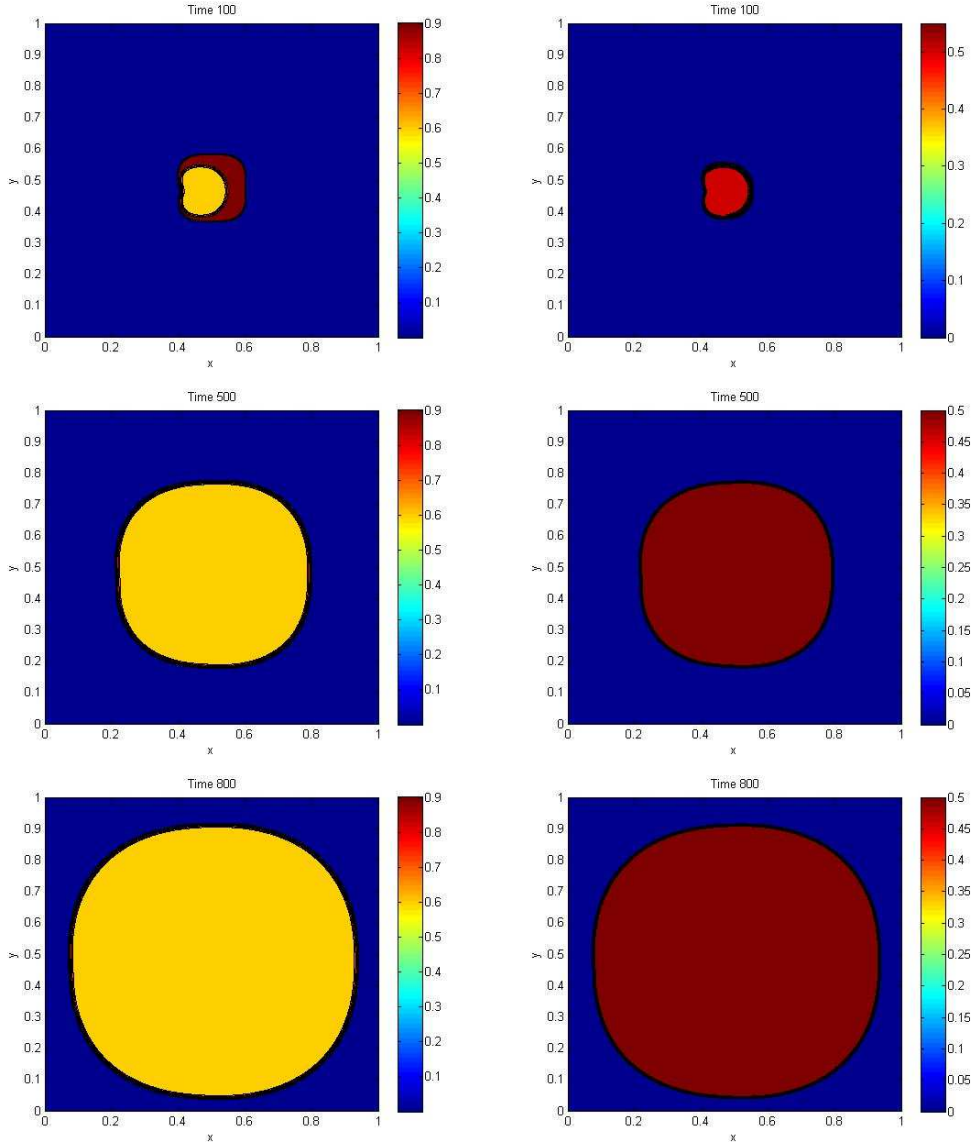

Figure 2.8: The population density distribution of prey and predator: (top) at time $t = 100$, (middle) at time $t = 500$, (bottom) at maximum time $t = 800$, where the left column is for prey and the right column is for predator.

In order to solve the system of equations (2.6.3), (2.6.4) numerically, the discrete equations for (2.6.3), (2.6.4) must be obtained as follows:

$$u_{i,j}^{n+1} = u_{i,j}^{n} + \lambda[u_{i+1,j}^{n} + u_{i-1,j}^{n} + u_{i,j+1}^{n} + u_{i,j-1}^{n} - 4u_{i,j}^{n}] + \Delta t[\beta u_{i,j}^{n}(u_{i,j}^{n} - b)(1 - u_{i,j}^{n}) - \frac{u_{i,j}^{n} v_{i,j}^{n}}{1 + A u_{i,j}^{n}}], \tag{2.6.5}$$

$$v_{i,j}^{n+1} = v_{i,j}^{n} + \lambda[v_{i+1,j}^{n} + v_{i-1,j}^{n} + v_{i,j+1}^{n} + v_{i,j-1}^{n} - 4v_{i,j}^{n}] + \Delta t[\kappa \frac{u_{i,j}^{n} v_{i,j}^{n}}{1 + A u_{i,j}^{n}} - m v_{i,j}^{n}], \tag{2.6.6}$$

where Δt is the time-step size, and $u_{i,j}^{n}$, $v_{i,j}^{n}$ approximate the values of $u(x_i, y_j, t^n)$, and $v(x_i, y_j, t^n)$ of the exact solution for any (x_i, y_j, t^n), where $x_i = ih$; $i = 0, 1, ..., N_x$, $y_j = jh$; $j = 0, 1, ..., N_y$, h is the grid step size and $t^n = n\Delta t$. The System of equations (2.6.5), (2.6.6) is conditionally stable if the stability condition $|\lambda| = \left| \frac{d\Delta t}{h^2} \right| \leq \frac{1}{4}$ holds, where $h = h_x = h_y$.

Once the system has been rewritten in a more convenient dimensionless form, it can then be used to generate a variety of spatiotemporal patterns of ecological data [103]. Firstly, we generate a continuous front of prey density over the domain $[0, 1] \times [0, 1]$, where the total number of grid nodes is $N_f = 1025 \times 1025$, where in x-axis and y-axis the number of grid nodes is $N_x = N_y = 1025$. As shown in Figure 2.7a, the population density is obtained at $t = 50$, $\beta = 3$, $b = 0.28$, $A = 0.5$, $m = 0.48$, $D = 10^{-6}$, and under the following discrete initial conditions:

$$u_{i,j}^{1} = u_* + \epsilon_{ux} x + \epsilon_{uy} y,$$

$$v_{i,j}^{1} = v_* + \epsilon_{vx} x + \epsilon_{vy} y,$$

where (u_*, v_*) is the equilibrium state $\epsilon_{ux} = 0.007$, $\epsilon_{uy} = 0.008$, $\epsilon_{vx} = 0.008$ and $\epsilon_{vy} = -0.007$. It can be seen from Figure 2.7a, that the maximum value of prey population density is located close to the x-axis, while it declines to zero in the area that represents the right-hand side of the graph. Figure 2.7b shows the second spatial distribution of population density when the population density is constant in the centre of the domain;

inside the round area the values for the population densities are constant. Also, it can be seen that the almost identical prey distribution is split from the rest of the domain which indicates that our population density may be established from the centre of the domain. This distribution was obtained at $t = 800$, $m = 0.5$ and the remaining parameters are as presented in the previous test case. Finally, the more complicated pattern than above, is shown in Figure 2.7c. The population density in the figure represents a patchy invasion [165]. This complex structure of population density was acquired at $t = 1200$, $m = 0.45$, and the values of the other parameters are the same as in test case 1. The initial conditions in both the second and third test cases are as follows:

$$u^1_{i,j} = u_0 \text{ if } x_{11} < x < x_{12} \text{ and } \ y_{11} < y < y_{12} \text{ , otherwise } u^1_{i,j} = 0,$$

$$v^1_{i,j} = v_0 \ \text{ if } x_{21} < x < x_{22} \text{ and } \ y_{21} < y < y_{22}, \text{ otherwise } v^1_{i,j} = 0,$$

where $x_{11} = 0.42$, $x_{12} = 0.53$, $y_{11} = 0.45$, $y_{12} = 0.55$, $x_{21} = 0.42$, $x_{22} = 0.48$, $y_{21} = 0.45$, $y_{22} = 0.51$, and u_0, v_0 are the initial densities of prey and predator respectively. Changing the values for the initial densities of prey and predator leads to generation of a different type of spatial population density distribution each time, as will be illustrated.

Initial conditions play an important role in the properties of our generated spatial population density distribution. In order to investigate this statement, let us consider the second two dimensional ecological test case which is denoted as ETC2, with different values of initial conditions. The rest of the parameters in the system (2.6.3), (2.6.4) are the same as considered in ETC2.

Let us first consider that the initial populations of prey and predator overlap (*i.e* the patch of predators is located within the patch of prey) by considering the following initial condition:

$$u^1_{i,j} = u_0 \text{ if } x_{11} < x < x_{12} \text{ and } \ y_{11} < y < y_{12} \text{ , otherwise } u^1_{i,j} = 0,$$

$$v^1_{i,j} = v_0 \ \text{ if } x_{21} < x < x_{22} \text{ and } \ y_{21} < y < y_{22}, \text{ otherwise } v^1_{i,j} = 0,$$

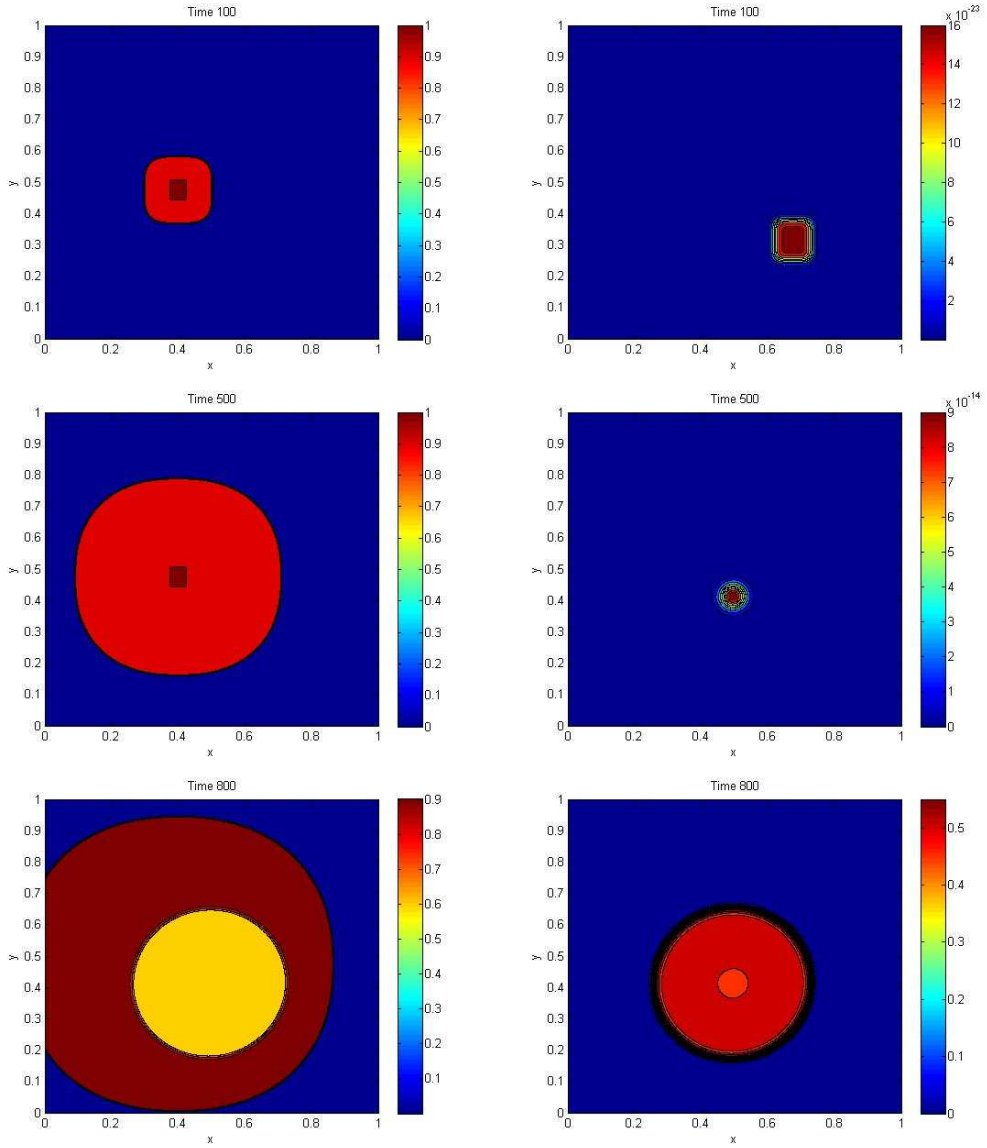

Figure 2.9: The population density distribution of prey and predator: (top) at time $t = 100$, (middle) at time $t = 500$, (bottom) at maximum time $t = 800$, where the left column is for prey and the right column is for predator.

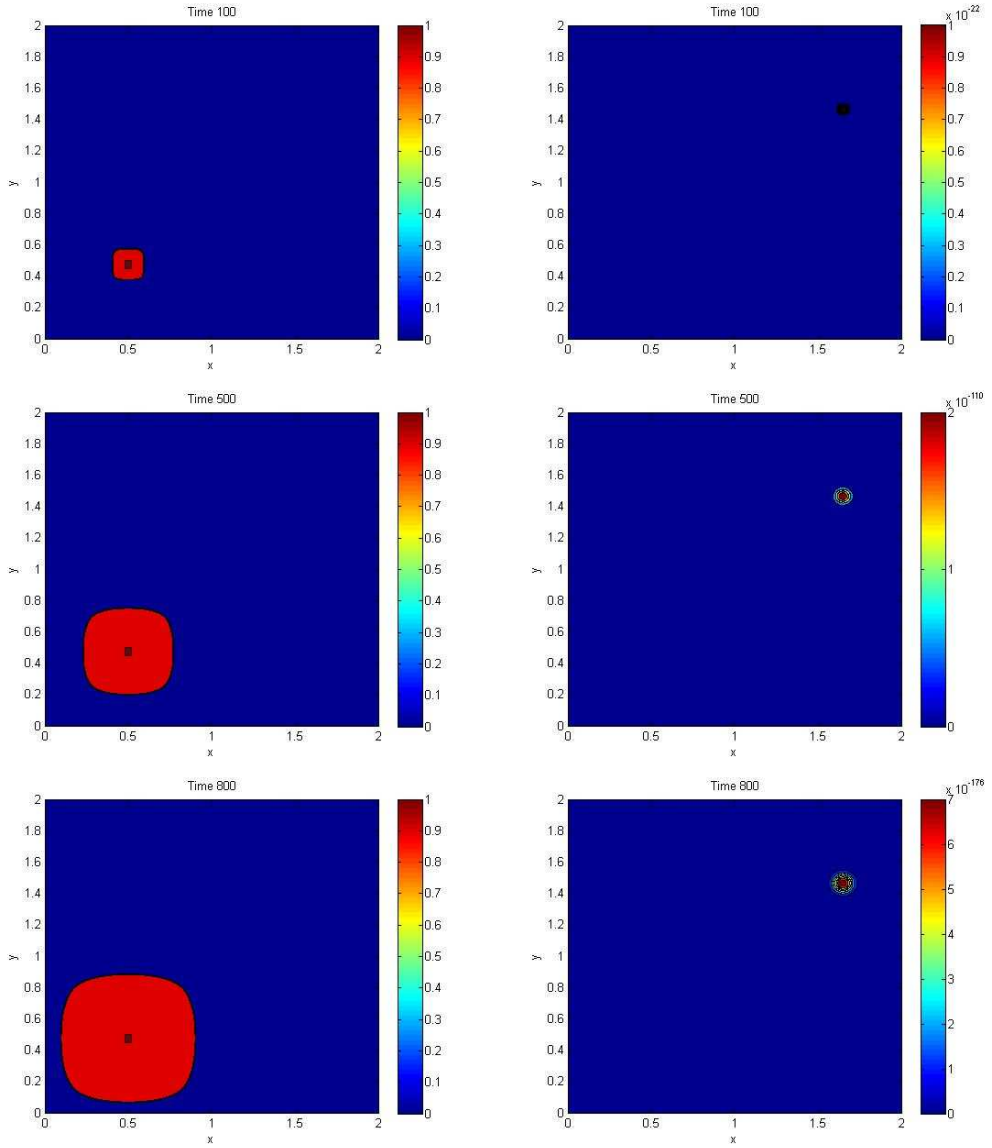

Figure 2.10: The population density distribution of prey and predator: (top) at time $t = 100$, (middle) at time $t = 500$, (bottom) at maximum time $t = 800$, where the left column is for prey and the right column is for predator.

where $x_{11} = 0.42$, $x_{12} = 0.53$, $y_{11} = 0.45$, $y_{12} = 0.55$, $x_{21} = 0.45$, $x_{22} = 0.48$, $y_{21} = 0.45$, $y_{22} = 0.48$, and u_0, v_0 are the initial densities of prey and predator respectively . We expect as the prey patch grows and spreads throughout the domain over time, the predator patch will grow at the same time, as the predators will be in the domain of the prey, and therefore its source of food will be available and its population will increase over time, as depicted in Figure 2.8. It can be seen from Figure 2.8 that the populations of prey and predator increase at the same time, because they are overlapping.

Now let us consider an initial distribution of prey separate from the initial distribution of predator, (*i.e* there is no overlapping between them). The initial populations in this case will be as follows:

$$u^1_{i,j} = u_0 \text{ if } x_{11} < x < x_{12} \text{ and } \ y_{11} < y < y_{12} \text{ , otherwise } u^1_{i,j} = 0,$$

$$v^1_{i,j} = v_0 \ \text{ if } x_{21} < x < x_{22} \text{ and } \ y_{21} < y < y_{22}, \text{ otherwise } v^1_{i,j} = 0,$$

where $x_{11} = 0.42$, $x_{12} = 0.53$, $y_{11} = 0.35$, $y_{12} = 0.45$, $x_{21} = 0.25$, $x_{22} = 0.38$, $y_{21} = 0.35$, $y_{22} = 0.45$, and u_0, v_0 are the initial densities of prey and predator respectively. The predator will strive to stay alive, but lacks the source of food (prey) due to the small distance between the initial population patches of populations leading to a reduction in the size of the predator patch over time. Meanwhile, the population of prey spreads throughout the domain over time, until it reaches the predator patch. At that point, the predator starts to consume the available prey and its population increase gradually over time, as shown in Figure 2.9. The predator population tends to be zero when considering large distances between initial patches of prey and predator. The initial population distribution in this case is as follows:

$$u^1_{i,j} = u_0 \text{ if } x_{11} < x < x_{12} \text{ and } \ y_{11} < y < y_{12} \text{ , otherwise } u^1_{i,j} = 0,$$

$$v^1_{i,j} = v_0 \ \text{ if } x_{21} < x < x_{22} \text{ and } \ y_{21} < y < y_{22}, \text{ otherwise } v^1_{i,j} = 0,$$

where $x_{11} = 0.42$, $x_{12} = 0.53$, $y_{11} = 0.45$, $y_{12} = 0.55$, $x_{21} = 1.45$, $x_{22} = 1.48$, $y_{21} = 1.63$, $y_{22} = 1.67$, and u_0, v_0 are the initial densities of prey and predator respectively and the population density distributions are depicted in the Figure 2.10. The domain in this case is $[0, 2]$, instead of $[0, 1]$, to increase the distance between initial patches of prey and predator without exceeding the boundaries of our domain.

2.7 Chapter 2 Conclusion

The Rosenzweig-MacArthur model has been derived in this chapter from a general form of predator-prey model. The general properties of this system were discussed. Numerical solutions for the Rosenzweig-MacArthur system were illustrated. Ecological data can be generated using computer simulations as mentioned previously. For this purpose, different ecological models were created. Population density can be distributed throughout the domain in different patterns, depending on the availability of the prey. For instance, if the prey population is concentrated in a narrow area of the stochastic domain or is patchy, the predator-prey system will be generated according to the aggregated source. Different shapes for ecological population density can be obtained by choosing a variety of values for the parameters in each model, and by using a combination of initial and boundary conditions. This brief introduction to the ecological systems, along with this useful information regarding the properties and numerical solutions of the spatially explicit predator-prey system from the Rosenzweig- MacArthur model, is used to generate $1d$ and $2d$ ecological data with different values for the initial conditions.

Chapter 3

Evaluation of the Total Population Size From Ecological Data

The main purpose of this chapter is to illustrate how to apply numerical integration methods to standard mathematical test cases. The discussion will be extended to involve an evaluation of population density in the framework of ecological monitoring. The basic concepts of numerical integration will be investigated in Section 3.1. In addition, the accuracy of these methods will be discussed. Moreover, it is very important to draw attention to their application. This application will first be focused on one-dimensional cases, in order to simplify the explanation, and then it will be extended to involve two-dimensional cases. Therefore, we will first deal with a straight line of traps, installed at uniform intervals.

3.1 Basic Notions of Numerical Integration Methods

Information about the pest population has been obtained by applying sample means in a specific agricultural field, This information will be combined in order to evaluate the

pest population abundance. According to [37], the classical approach used to estimate abundance relies on sample means, as demonstrated in Chapter 1. It has been assumed in [29, 156] that the count of samples can be manipulated for each sample unit location to estimate the pest population abundance. The exact value of population abundance I can be approximated by using the following equation:

$$I \approx I_a = AM(N) = \frac{A}{N}\sum_{i=1}^{N} u_i, \tag{3.1.1}$$

where A denote the area of the entire agricultural field, $M(N)$ represents the sample mean of pest density given in the equation (1.2.1), N denotes the total number of samples, and u_i gives the pest population abundance at each location of sample units $x_i, \quad i = 1, 2, ..., N$. According to [63], the sample mean density $M(N)$ tends to reach the exact mean density $\bar{M}$ as a result of increasing the total number of sample units N. Hence, the estimate of the pest population density I_a tends approximately to reach the exact value of population abundance I when dealing with a large value of N.

Equation (3.1.1) provides an evaluation of the pest population abundance as a sum of the available values of pest population density. Numerical integration approaches can be deduced readily from general form the equation (3.1.1). The solution to problem of estimating the pest population density is considered to be the value of the following integral I over domain of agricultural field D:

$$I = \int\int_{D} u(x, y)dxdy, \tag{3.1.2}$$

where the integrand function $u(x, y)$ is the pest population density. However, in ecological monitoring problems, the function $u(x, y)$ is known only for a finite number of sample units N. As a result, the function of pest population density is discrete points $u(x, y) \equiv$

$u(x_i, y_j) = u_{ij}$. For discrete population density, it is impossible to find the analytical solution. In order to evaluate this integral, numerical integration methods must be used to approximate the solution [44, 96, 97, 103, 199], as demonstrated in Chapter 1.
In the $1d$ cases, integral (3.1.2) becomes:

$$I = \int_a^b u(x)dx, \tag{3.1.3}$$

where $u(x) \equiv u_i,\ \ i = 1, 2, .., N$. According to [38], the general form of numerical integration can be obtained by the weighted sum as follows:

$$I \approx I_a = \sum_{i=1}^{N} w_i u_i, \tag{3.1.4}$$

where w_i, $i = 1, 2, ..., N$ rely on numerical integration methods. Formula (3.1.1) is considered as a simple form of numerical integration formulas, where:

$$w_i = \frac{A}{N}, \quad i = 1, 2, ..., N. \tag{3.1.5}$$

As N tends to reach infinity, the approximate integral I_a tends towards the true integral I; this condition is introduced as:

$$I_a(N) \to I, \text{ as } N \to \infty. \tag{3.1.6}$$

The accuracy of estimation is associated with the concept of absolute approximation error. If the value of the integral I in (3.1.1) is known, then the integration error $E_{abs}(N)$ can be defined as follows [150]:

$$E_{abs}(N) = |I - I_a(N)|\,. \tag{3.1.7}$$

The absolute integration error measures the magnitude of the distance between the exact value of the integral and an estimate of the integral I_a. Another definition of integration error is represented by the relative approximation error, where the accuracy of estimate I_a is computed as a proportion of the true solution I as follows:

$$E_{rel}(N) = \frac{E_{abs}}{|I|} = \frac{|I - I_a(N)|}{|I|}. \tag{3.1.8}$$

From condition (3.1.6) if we consider the absolute error (3.1.7), we obtain the following condition:

$$E_{abs}(N) \to 0, \text{ as } N \to \infty. \tag{3.1.9}$$

It can be seen from the above that formula (3.1.7) gives an accurate definition of the integration error only if the value of the integral I is known. However, in real-life applications, the exact value of the integral I is not available, and so formula (3.1.7) can not be applied. Therefore, according to [38], in order to overcome this obstacle it is beneficial to present the asymptotic error estimate of the numerical integration method. The asymptotic error estimate depends on the assumption that the distance between the sample units is small, which means that the number N of sample units must be large. The asymptotic error estimates can be expressed in the form:

$$E_{abs} \leq Ch^q sup_{x\in[a,b]} \left|u^{r+1}(x)\right|, \tag{3.1.10}$$

where C, q and r are constants and depend on the choice of weight coefficient in formula (3.1.4). Clearly, q represents the order of convergence of the methods used(*i.e.* the rate at which E_{abs} converges to zero and $h \to 0$). r denotes the degree of precision, *i.e* the maximum degree of the bivariate polynomial for which the method produces an exact result. A more accurate estimation of pest population density will be acquired by

choosing methods with a higher order of convergence.

Let us now consider a one-dimensional computational regular grid nodes generated as follows:

$$x_1 = a, \quad x_i = x_{i-1} + h, \quad x_N = b, \quad i = 2, ...N-1. \tag{3.1.11}$$

where the grid step size h in this case equals $h = \frac{b-a}{N-1}$. According to [28], equation (3.1.4) represents the Newton-Cotes formula of degree (N). The procedure for replacing the function u(x), by a fitting polynomial $p^n(x)$ to approximate the value of integral I at a set of district equidistant points, is known as the Newton-Cotes formulae. According to [182], the polynomial $p^n(x)$ is defined by the Lagrange interpolation formula as follows:

$$p^n(x) \equiv \sum_{i=1}^{n} u_i L_i(x),$$

where L_i is the Lagrange polynomial. Then, the exact value of integral I can be approximated by integrating the interpolating polynomial $p^n(x)$ as follows:

$$I = \int_a^b u(x)dx \approx I_a = \int_a^b p^n(x)dx,$$

The simplest evaluation of integral (3.1.3) can be achieved when the function $u(x)$ is replaced by a constant at each sub-interval $[x_i, x_{i+1}]$. Then, the composite midpoint rule has the following form [27, 38]:

$$I = \int_a^b u(x)dx \approx 2h \sum_{i=0}^{N/2} u(x_{2i}), \tag{3.1.12}$$

where the weight coefficients on (3.1.4) are known as $w_i = 2h$. The degree of the interpolating polynomial $p(k)$ is increased as $k = 1$, to replace function $u(x)$ on (3.1.4) by a linear polynomial. Then points x_i and x_{i+1} are linked by a straight line at each sub-interval $[x_i, x_{i+1}]$ to derive the formula known as the trapezoidal rule, which has the following

form [27, 38]:

$$I = \int_a^b u(x)dx \approx \frac{h}{2}[u_1 + 2\sum_{i=2}^{N-1} u_i + u_N], \tag{3.1.13}$$

where the weight coefficients on (3.1.4) are known as $w_1 = w_N = \frac{h}{2}$ and $w_i = h, i = 2, ...N-1$. Creation of a straight line requires two points; therefore, the total number of grid nodes required is $N \geq 2$. The composite trapezoidal rule has the same degree of precision and the same convergence rate as the composite midpoint rule.

By considering the degree of the interpolation polynomial as $k = 2$, when the function $u(x)$ on (3.1.4) is replaced by a quadratic interpolating polynomial, the composite Simpson's rule will be derived as follows [27, 38]:

$$I = \int_a^b u(x)dx \approx \frac{h}{3}[u_1 + 2\sum_{i=1}^{\frac{N-3}{2}} u_{2i+1} + 4\sum_{i=1}^{\frac{N-1}{2}} u_{2i} + u_N], \tag{3.1.14}$$

where the weight coefficients on (3.1.4) are given by $w_1 = w_N = \frac{h}{3}, w_i = \frac{2}{3}h$ for $i = 3, 5, .., N-2$, and $w_i = \frac{4}{3}h$, $i = 2, ..., N-1$. The composite Simpson's rule provides an approximation of the exact integral I, with a degree of precision $r = 3$ and a convergence rate equal $q = 4$.

3.2 Applications of Numerical Integration Methods

In this section, we provide a detailed study of asymptotic convergence based on condition (3.1.10). In order to validate this statement, we have applied the composite Newton-Cotes rules to the twenty-four mathematical test cases. However, this section will introduce just three of these, that have mathematically interesting characteristics. The table below presents these twenty-four mathematical test cases, along with their domains and exact values for the integrals. These mathematical test cases have been investigated in [47], where the polynomial approximations of one, three and five degrees have been applied to the integrand functions in the table. In addition, the composite Simpson's rule has been

employed as a method of numerical integration for these problems.

Test case (TC)	[a,b]	u(x)	I
1	[0,1]	x^3	$1/4$
2	[0,1]	x^6	$1/7$
3	[0,1]	x^{10}	$1/11$
4	[0,1]	$\sqrt{x^7}$	$2/9$
5	[0,π]	$sin(x)$	2
6	[1,10]	$1/x$	$ln(10)$
7	[0,1]	e^x	(e-1)
8	[0,1]	$x/(e^x-1)$	$7.7750463411e-01$
9	[-4,4]	$1/(1+x^2)$	$2tan^{-1}(4)$
10	[-1,1]	$1/(x^4+x^2+0.9)$	$1.5822329637e+00$
11	[0,$\pi/2$]	$1/(1+sin^2(x))$	$\pi\sqrt{2}/4$
12	[0,5]	$1/120(x-1)(x-2)....(x-5)$	$-47.50/144.0$
13	[-1,1]	$23/25cosh(x)-cos(x)$	$46/25sinh(x)-2sin(x)$
14	[0,1]	$(x+0.01)^{-5}$	$\frac{((0.01)^{-4}-(1.01)^{-4})}{4}$
15	[0,1]	$1/\sqrt{x+0.0001}$	$2.0\sqrt{1.0001}-\frac{1}{50}$
16	[0,1]	$1/(x+0.0001)$	$ln(1.0001)$
17	[0,1]	$1/((230x-30)^2+1)$	$1/230(tan^{-1}(200)+tan^{-1}(30))$
18	[0,1]	$1/(x+0.01)$	$ln(101)$
19	[0,10]	$50/(\pi(1+2500x^2))$	$tan^{-1}(500)/\pi$
20	[0,1]	$2/(2+sin(10\pi x))$	1.1547006690
21	[0.1,1]	$\frac{\sin(100\pi x)}{(\pi x)}$	$9.098637539e-03$
22	[0,2π]	$xsin(30x)cos(x)$	$(-60/899)\pi$
23	[0,1]	$4\pi^2xsin(20\pi x)cos(2\pi x)$	$-20x/99$
24	[0,2π]	$e^{-x}sin(10x)$	$\frac{10-e^{-2x}}{10.1}$

Table 3.1: Mathematical test cases

As can be seen from Table 2.1, test case five denoted as TC5 represents a standard trigonometric function, which has a simple behaviour pattern as shown in Figure 3.1a. From Figure 3.1a, it can be seen that the statistical method has recorded the lowest degree of accuracy, whereas, as can be seen from Figure 3.1b both the trapezoidal rule and the midpoint rule behaved as we expected, they achieved a satisfactory degree of accuracy. Furthermore, the convergence curves of the trapezoidal rule and the midpoint

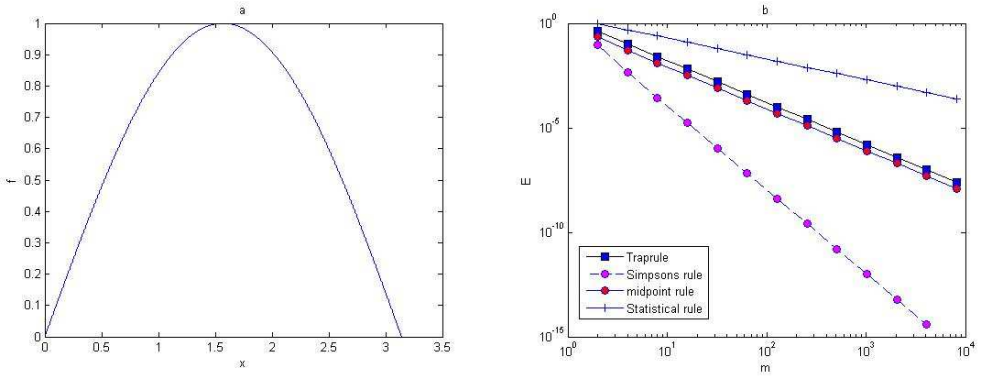

Figure 3.1: The results of the numerical integration methods of the function TC5. (a) TC5 plotted over its domain. (b) Comparison of the convergence rates for the trapezoidal rule, midpoint rule, statistical method, and the Simpson's rule.

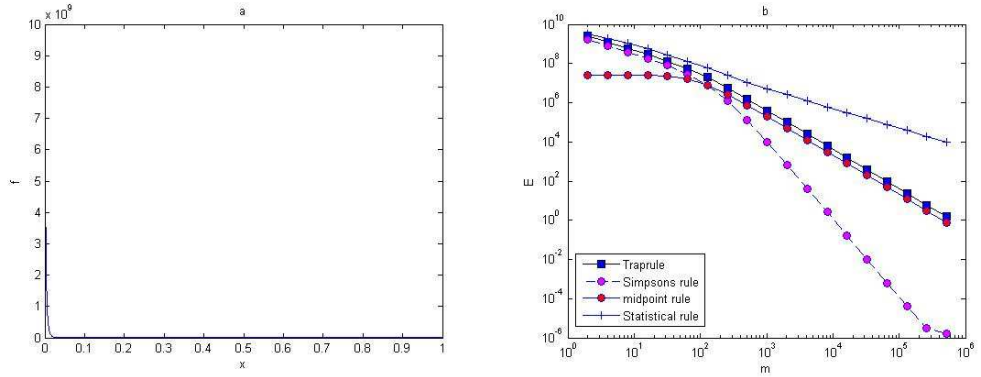

Figure 3.2: The results of the numerical integration methods of the function TC14. (a) TC14 plotted over its domain. (b) Comparison of the convergence rates for the trapezoidal rule, the midpoint rule, statistical method and the Simpson's rule.

rule are parallel to each other, as shown in Figure 3.1b. However, a more accurate approximation was acquired by using the composite Simpson's rule. The convergence curve of the composite Simpson's rule was much faster than other rules, which confirms that the numerical integration methods used are valid. TC14 shows a function that has a sharp boundary layer. This function has a more complicated behaviour pattern, as shown in Figure 3.2a. TC14 was chosen to illustrate that the error estimate does not always hold. Figure 3.2b shows the results of the integration error after applying the statistical method, trapezoidal rule, midpoint rule, and Simpson's rule to TC14. When inspecting

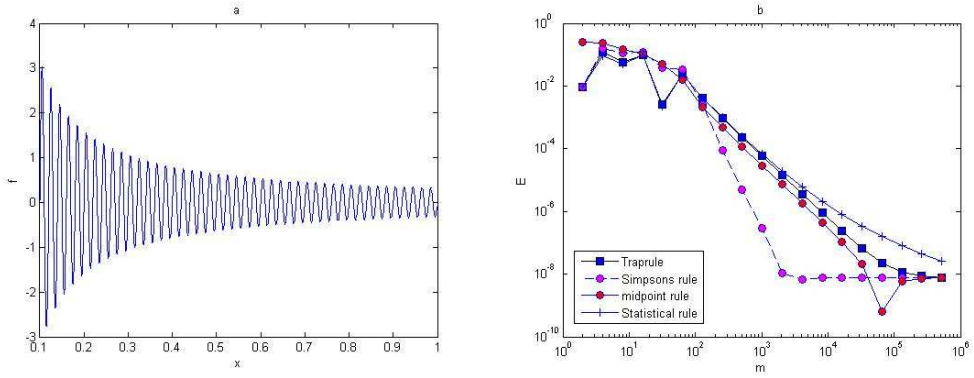

Figure 3.3: The results of the numerical integration methods of the function TC21. (a) TC21 plotted over its domain. (b) Comparison of the convergence rates for the trapezoidal rule, the midpoint rule, statistical method and the Simpson's rule.

this figure, it can be seen that from the outset, all methods have achieved a high value for the integration error. Therefore, in order to achieve an acceptable evaluation of the integration error, the grid has to be refined. The composite Simpson's rule started to converge faster when the number of sub-intervals N becomes $N = 256$, but prior to that, the grid can be considered a coarse grid. Furthermore, at $N = 10^5$, which is a huge number of grid nodes, it can be argued that the composite Simpson's rule has achieved a desirable value for the integration error. However, to achieve the same accuracy by applying the trapezoidal and midpoint rules, the number of nodes must be doubled many times. Also, the statistical method did not achieve a satisfactory degree of accuracy, even when using a huge number of grid nodes. Thus, it can again be concluded that the composite Simpson's rule is much faster than other compound methods.

A quickly oscillating integrand function has been studied in TC21. This integrand has complex behaviour, as shown in Figure 3.3a. The convergence curves for all methods oscillated from the outset, as presented in Figure 3.3b. It is essential to refine the grid sufficiently, in order to provide an accurate estimation of the integration error. Then, for any $N \geq 512$, a sufficient degree of accuracy for the integration error has been achieved at $E \leq \tau = 10^{-5}$. Hence, another criterion of accuracy can be considered to be that the

following condition must hold:

$$E \leq \tau, \tag{3.2.1}$$

where τ is a fixed tolerance number.

As can be seen from the previous discussion, several numerical integration methods can be applied to the ecological problem of the evaluation of pest population abundance, and the method that has the fastest convergence rate will be chosen, depending on formula (3.1.10). It is worth noting here that the method that converges fastest usually has some restrictions on its applications. It is noticeable from previous applications that the composite Simpson's rule may provide the most accurate approximation. However, it is impossible to apply it to an arbitrary number of grid nodes, as the total number N of grid nodes should always be an odd number when this rule is applied. In contrast, there are no restrictions when applying the trapezoidal rule and the midpoint rule, which have slower convergence rates, but can be applied for an arbitrary number N.

3.3 Two-Dimensional Newton-Cotes Formulae

As well as applying the numerical integration methods in the one-dimensional case, they can be easily extended to the two-dimensional case. Let us consider a set of points x_i, $i = 1, 2, ..., N_x$, on the interval $[a, b]$ where $x_1 = a$, $x_{i+1} = x_i + h_x$, $i = 1, ..., N_x - 1$, $x_{N_x} = b$ and h_x is the grid step size defined as $h_x = b - a/N_x - 1$. In order to generate a one-dimensional grid in the y-direction, let us consider points y_j, $j = 1, 2, ..., N_y$ on the interval $[c, d]$ where $y_1 = c$, $y_{j+1} = y_j + h_y$, $j = 1, 2, ..., N_y - 1$, $y_{N_y} = d$ and h_y is the grid step size defined as $h_y = d - c/N_y - 1$. Then the position of any grid node on a computational grid in the unit square $D = [a, b] \times [c, d]$ will be given by (x_i, y_j).

The approach that employed composite numerical integration rules in the 1d case will be

the same in the 2d case. Let us consider the integral I given by

$$I = \int_c^d \left(\int_a^b f(x,y)dx \right) dy = \int_c^d F(y)dy, \tag{3.3.1}$$

where the population density function is discrete, ($i.e$ $f(x,y) \equiv f(x_i, y_j)$ for $i = 1, ..., N_x$ and $j = 1, ..., N_y$. The integral $F(y)$ can be presented as:

$$F(y) = \int_a^b f(x,y)dx \approx \sum_{i=1}^{N_x} w_i f(x_i, y), \tag{3.3.2}$$

Integral (3.3.1) can be presented as

$$I \approx I_{ij} = \sum_{j=1}^{N_y} w_j F(y_j). \tag{3.3.3}$$

where the approximation I_ij is given as :

$$I_{ij} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} w_i w_j f(x_i, y_j). \tag{3.3.4}$$

In order to compute the value of function $F(y)$ in sub-domain $c_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$, we apply the $1d$ Newton-Cotes formulae as discussed previously. After evaluating the value of $F(y)$, the same integration rule will be employed to compute integral (3.3.1). According to [36], the local approximation of integrand $f(x,y)$ will differ, depending on the different integration rules on a grid cell c_{ij}.

The simplest way to evaluate integral (3.3.2) is to replace the function $f(x,y)$ with a constant function at each sub-domain c_{ij}. Then, the composite midpoint rule will be derived as follows:

$$I_{ij} \approx w_{ij} f(x_{i+1/2}, y_{j+1/2}), \tag{3.3.5}$$

where $w_{ij} = h_x h_y$ at each midpoint node $(x_{i+1/2}, y_{j+1/2}) = (x_{i+h/2}, y_{j+h/2})$ of the grid cell $c_i j$. In the line of derivation the $2d$ Newton-Cotes formulae, the second formula is the composite trapezoidal rule. In order to employ the composite trapezoidal rule, a linear function on each sub-domain c_{ij} is used to approximate the integrand $f(x, y)$. The integral I_a is then approximated as follows [38, 150]:

$$I_{ij} \approx \frac{h_x h_y}{4} \{f(x_i, y_j) + f(x_{i+1}, y_j) + f(x_i, y_{j+1}) + f(x_{i+1}, y_{j+1}).\} \tag{3.3.6}$$

The exact value of integral (3.3.1) is approximated as follows [41, 42]

$$I = \sum_{i,j} I_{ij} + O(h_x^2) + O(h_y^2), \tag{3.3.7}$$

where I_{ij} is the approximate value of integral (3.3.1) after employing the composite trapezoidal rule, h_x is the grid step size in the x-direction, h_y is the grid step size in the y-direction. The terms $O(h_x^2)$, and $O(h_y^2)$ in equation (3.3.7) refer to the error term when applying the composite trapezoidal rule. Because we are dealing with mathematical test cases where the distance between grid nodes is equidistant, then we will consider $h_x \equiv h_y = h$; therefore, we can rewrite equation (3.3.7) as follows [41, 42]:

$$I = \sum_{i,j} I_{ij} + O(h^2). \tag{3.3.8}$$

It follows from equation (3.3.8) that the two-dimensional composite trapezoidal rule has an error of the second order. The next form of the $2d$ Newton-Cotes formulae is the compound Simpson's formula. It can be derived by using a quadratic polynomial to approximate the value of the integrand $f(x, y)$ in the square cell c_{ij}. The two-dimensional

Simpson's rule has the following formula [38, 150]:

$$\begin{aligned} I_{ij} \approx \tfrac{h_x h_y}{36}(f(x_i, y_j) + f(x_i, y_{j+2}) + f(x_{i+2}, y_j) + f(x_{i+2}, y_{j+2}) \\ +4[f(x_i, y_{j+1}) + f(x_{i+1}, y_j) + f(x_{i+2}, y_{j+1}) + f(x_{i+1}, y_{j+2})] + 16f(x_{i+1}, y_{j+1})). \end{aligned} \tag{3.3.9}$$

After applying the two-dimensional composite Simpson's rule, the exact value of integral (3.3.1) can be rewritten as follows [41, 42]:

$$I = \sum_{i,j} I_{ij} + O(h_x^4) + O(h_y^4), \tag{3.3.10}$$

where I_{ij} is the approximate value of the integral (3.3.1), h_x is the grid step size in the x-direction, h_y is the grid step size in the y-direction. The terms $O(h_x^4), O(h_y^4)$ in the equation (3.3.10) denotes the error term when applying the composite Simpson rule. We will consider $h_x \equiv h_y = h$ and rewrite equation (3.3.10) as follows [41, 42]

$$I = \sum_{i,j} I_{ij} + O(h^4). \tag{3.3.11}$$

It follows from equation (3.3.11) that the $2d$ composite Simpson's rule has the fourth order of approximation.

3.4 Numerical Integration Methods for the $1d$ Ecological Data

Once the Rosenzweig-MacArthur model has been deduced, and the numerical solution of the Rosenzweig-MacArthur system has been obtained, equations (2.2.10), (2.2.11) are used to generate different sets of $1d$ ecological data. In order to generate various ecological test cases, the values of time t and diffusion rate d will be arbitrary, and we will fix other parameters such as $m = 0.7$, $h = 0.3$, $K = 2$ in the system of equations (2.2.10),

(2.2.11). The simplest ecological test case denoted as 1ETC1, representing a one-peak spatial distribution at $t = 50$ and $d = 10^{-4}$, is shown in Figure 3.4a. A more complex spatial distribution is presented by a three-peak function in ecological test case 1ETC2, shown in Figure 3.4b. The solution for this test case has been generated at $t = 100$, and $d = 10^{-5}$. Finally, according to Figure 3.4c, a more sophisticated ecological test case 1ETC3 represents a multi-peak function at $t = 400$ and $d = 10^{-5}$. Let us note that

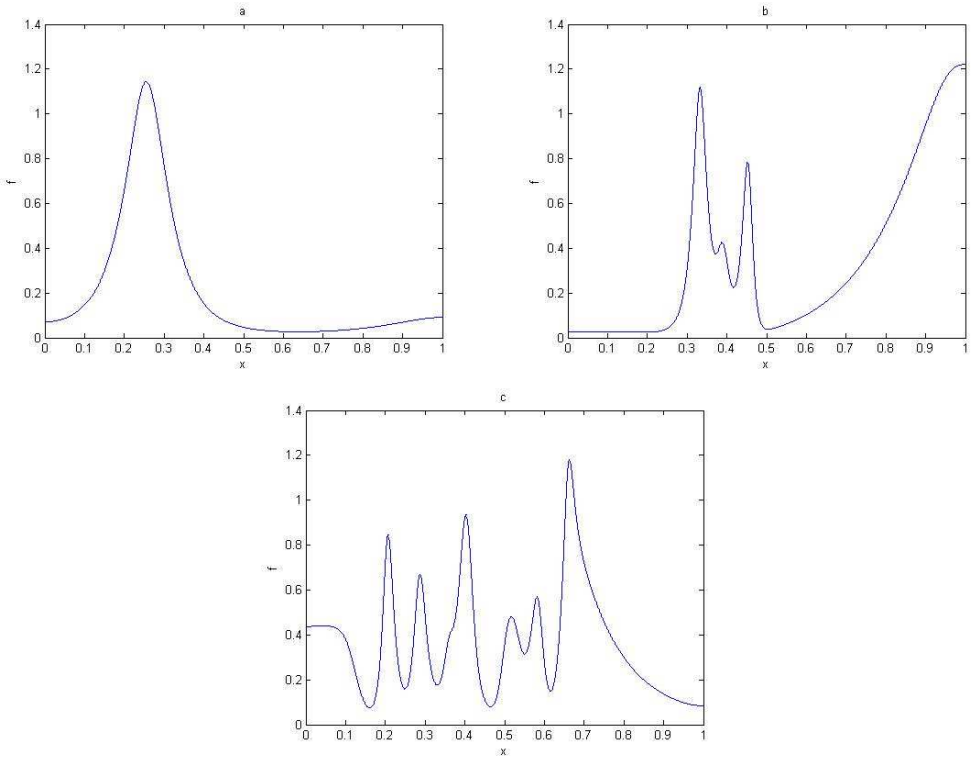

Figure 3.4: Ecological test cases at fixed parameters $m = 0.7$, $h = 0.3$, $K = 2$: (a) 1ETC1 one-peak function at $t = 50$ and $d = 10^{-4}$, (b) 1ETC2 three-peak function at $t = 100$, and $d = 10^{-5}$, and (c) 1ETC3 multi-peak function at $t = 400$ and $d = 10^{-5}$.

the exact value of the integral is not available to us, as the population density is not analytically defined. Conversely, we have a set of values for the distributed population densities at selected trap positions. Then, the exact value of the integral in this case will be computed by applying the trapezoidal rule (3.1.13) to this ecological data, with numerical solution on a very fine uniform grid to be considered as the exact value of the integral.

The system of equations (2.2.10), (2.2.11) is solved to find the population densities $u(x)$ for a very fine regular mesh with the total number $N_f + 1$ of points over the interval $[0, 1]$, where $N_f = 2^{15}$ is the number of sub-intervals. Once the value of the population densities has been obtained and the integral is evaluated on the fine grid, several approximations of these populations will be extracted from the data generated on the grid of N_f intervals. The values of the density function are not recomputed; instead, they will be extracted from the finest population density function, which has already been computed. In order

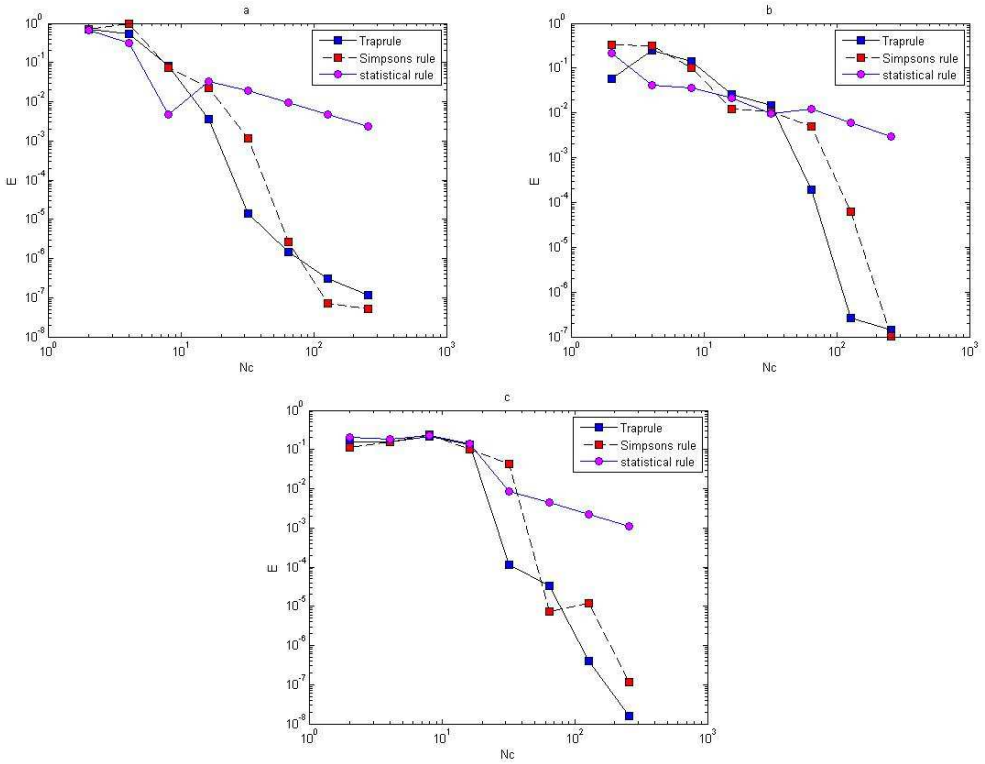

Figure 3.5: Convergence curves for $1d$ ecological test cases on a regular grid. (a) 1ETC1, (b) 1ETC2, and (c) 1ETC3.

to illustrate the procedure for extracting the coarse grid from the fine grid, let us first denote the number of sub-intervals in the fine grid and in the coarse grid as N_f and N_c respectively. It should also be considered that x^f and x^c denote the positions of traps in the finest grid and in the coarse grid respectively. It must be pointed out that the

endpoints in the fine grid will be the same as in the coarse mesh. Then, the procedure for extracting the coarse grid nodes is determined by the ratio $s = N_f - 1/N_c - 1$, as follows[152]:

$$x_j^c = x_i^f;\ \ i = 1, s+1, 2s+1, ...N_f - s, N_f;\ \ j = 1, 2, ..., N_c.$$

The same technique will be applied to extract the coarse density function f^c from the fine density function f^f. For instance, if we have a grid of five grid nodes $N_f = 4$, and it is considered the fine grid, then the grid elements will be $x^f = \left\{x_1^f, x_2^f, x_3^f, x_4^f, x_5^f\right\}$. Depending on the previous ratio s, the process of extracting the coarse grid of three sub - intervals $N_c = 2$ from this fine grid will be as follows: $x^c = \left\{x_1^f, x_3^f, x_5^f\right\}$.

After the population density of the pest insects in the fine grid has been obtained, and the different population densities have been extracted from the fine grid, then different numerical integration methods will be applied to these coarse functions to approximate the value of integral I [182]. According to [38], the statistical rule (1.2.1), which has the lowest convergence rate $E = Ch$, the trapezoidal rule (3.1.13) with its convergence rate $E = Ch^2$ and Simpson's rule (3.1.14), which achieved the highest convergence rate for the integration error $E = ch^4$, will be applied to these ecological test cases. Once the exact value I_{Nf} of the integral I has been obtained, and our approximation integrals have been computed, then the relative integration error is presented as follows:

$$E = \frac{|I_{Nf} - I_{Nc}|}{|I_{Nf}|}, \tag{3.4.1}$$

where I_{Nf}, and I_{Nc} are the exact value and the approximate value of the integral I respectively.

The convergence curves for 1ETC1 on coarse grids do not behave as expected from asymptotic error (3.1.10) as shown in Figure 3.5a. Once the grid has been refined sufficiently, the improvement in accuracy can be seen from the convergence curves of the trapezoidal

and Simpson's rules at $N_c = 8$. Meanwhile, the statistical method, which has lower order of convergence $q = 1$, starts to behave as expected at $N_c = 16$, as shown in Figure 3.5a. Simpson's rule with an order of convergence of $q = 4$, evaluated the problem less accurately than the trapezoidal rule, which has an order of convergence of $q = 2$. Also, it can be noted that when the number of sub-intervals becomes $N_c = 64$, the compound Simpson's rule will converge faster and will present an accurate estimation of the integration error, as can be seen from Figure 3.5a. In fact, when our applications are extended to involve more sophisticated test cases, i.e. 1ETC2 and 1ETC3, it is readily seen that the asymptotic error estimate does not hold until the grid has been refined sufficiently. As shown in Figure 3.5b, the errors for ecological test case 1ETC2 have achieved the desirable limit when the number of sub-intervals becomes $N_c = 32$. The accuracy achieved for the integration error by the estimate produced using the trapezoidal rule is higher than the degree of accuracy achieved by applying the statistical method. However, for 1ETC3 it can be observed that all convergence curves from the coarsest grid oscillated, as shown in Figure 3.5c. According to Figure 3.5c after increasing the number of sub-intervals to $N_c = 128$, the convergence curves have behaved as expected and the trapezoidal rule has presented an accurate estimate of pest abundance.

3.5 Numerical Integration Methods for the $2d$ Ecological Data

The $2d$ ecological test cases, which are depicted in Figure 2.7, will be examined in this chapter to investigate the performance of numerical integration methods for estimating $2d$ ecological population density. Once our $2d$ ecological test cases have been generated, then numerical integration methods can be applied to these test cases, as has been done with the $1d$ ecological problems. As discussed previously, the exact value of the integral is required in order to compute the integration error. Since the exact value of the integral is

not known, because we cannot solve the system (2.6.3), (2.6.4) analytically, the composite trapezoidal rule (3.3.6) will be applied to this system at very fine number $N_x = N_y = N_f = 2^{15} + 1$ of grid nodes in each direction, in order to provide an accurate evaluation of the integral. The integral evaluation made on the fine grid will be considered to be the exact value of the integral. Then, the coarse grids will be extracted from the finest number of grid nodes, as has been done for the $1d$ ecological test cases. Once we have extracted the population densities $u(x, y)$ at the nodes of a coarse grid, then the composite trapezoidal rule (3.3.6) and the compound Simpson's rule (3.3.9) will be applied to this coarse data. Following this, the integration error can be computed as follows:

$$E = \frac{|I_{Nf} - I_N|}{|I_{Nf}|}, \tag{3.5.1}$$

where I_{Nf} is the exact value of the integral I , I_N is the approximate values of the integral I. According to [128], in ecological test cases a relative error of 100% can be considered an acceptable degree of accuracy. It can be seen from Figure 3.6a that the values of the relative error are small for the first ecological test case, even with a small number of traps $N = 3$ in each direction. However, the convergence curves of both numerical integration methods oscillate on coarse grids. Once the grid had been refined sufficiently (i.e. the grid has $N = 33$ nodes in each direction), the convergence curves behave as expected. The convergence rate of the composite Simpson's rule is much faster than the convergence rate of the compound trapezoidal method, and it presents the most accurate estimate of the integration error.

It can be readily seen from Figure 3.6b that the composite trapezoidal rule provides the most accurate estimation of the integration error, even in a small number of traps $N = 3$ in each direction, while the convergence curve of the composite Simpson's rule oscillates, even when the number of grid nodes has been refined, as shown in Figure 3.6b.

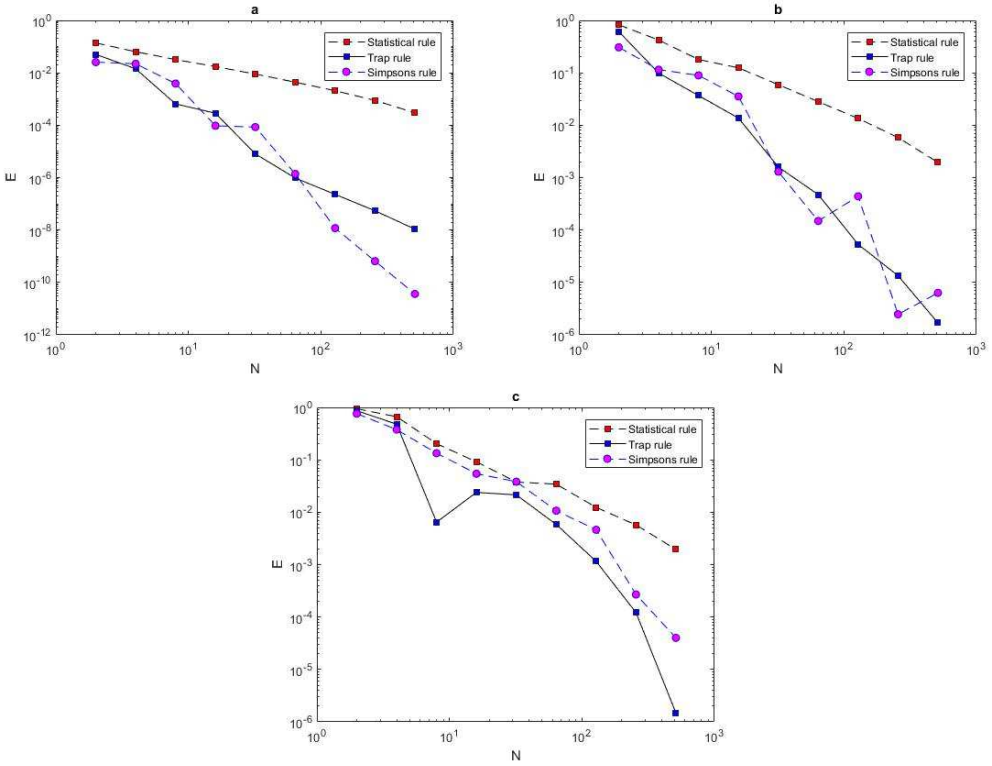

Figure 3.6: The convergence curves of the numerical integration methods for the $2d$ ecological test cases: (a) 2d ecological test case 1 (ETC1), (b) 2d ecological test case 2 (ETC2), and (c) 2d ecological test case 3 (ETC3).

The same poor accuracy on coarse grids has been observed in ecological test case ETC3. It can be seen from Figure 3.6c that the convergence curves of both integration methods oscillate on coarse grids, and the grid has to be refined to achieve a desirable degree of accuracy. As shown in Figure 3.6c, once the grid has been refined adequately (i.e. the grid has $N = 65$ number of grid nodes in each direction), the convergence curve of the trapezoidal rule starts to behave as expected. However, according to Figure 3.6c, the convergence curve of the Simpson's rule is still oscillating, even when the number of grid nodes increases significantly. It must be noticed from Figure 3.6 that the lowest degree of accuracy was achieved by applying the statistical rule to the $2d$ ecological test cases.

3.6 The Numerical Integration Methods on Irregular Ecological Data

We have previously discussed numerical integration methods on a regular grid. We started with $1d$ standard mathematical test cases.After validating our numerical integration methods with standard mathematical test cases, we extended the application of numerical integration methods to estimate pest population abundance in an agricultural field.

For some ecological reasons, we must investigate numerical integration methods on an irregular computational grid. In some ecological cases, we may think of some obstacles, for instance a bush or a tree, which force us to deal with the irregular landscape of agricultural fields. In other words, even if we consider our sampling plan to be a regular computational grid, we may need to move some traps from ideal positions in order to account for such obstacles. As a result, we will have samples on an irregular grid.

There are different types of irregular grid nodes. The main factor affecting this classification is the degree of irregularity. We have three types of irregular grid node: a slightly irregular grid, a quasi-random grid, and a random grid. The term 'slightly irregular grid' refers to the perturbation of a single grid node being moved from its position to produce a new sampling plan for an irregular grid pattern. To generate a slightly irregular grid node, we first generate a set of regular grid nodes $x_i, \ \ i = 1, 2, .., N$. Then, for some selected i from $i = 2, ..., N-1$, a single grid node x_i will be perturbed as follows:

$$x_i^{irreg} = x_i + h(\gamma - \frac{1}{2}), \tag{3.6.1}$$

where x_i represents the position of nodes on computationally regular grids, and $\gamma \in (0, 1)$ is a random variable with a uniform distribution.

In order to increase the level of irregularity while preserving the same structure, a quasi-

random gird can be used. The same generation strategy as with a slightly irregular grid will be applied to generate a quasi-random grid, with some modifications. Instead of perturbing a single interior grid node, all interior grid nodes will be perturbed. According to [113], when our agricultural field is divided into sections, and within each section a sample is taken from a random position, then we have a random sampling plan called centric systematic sampling. This system is related to our generation of quasi-random grids. However, the endpoints in our quasi-random system are fixed at the endpoints of the original interval $[a, b]$. In order to avoid counter bias in our evaluation of pest population density, the random sampling plan will be considered [20, 91, 159, 170]. The use of distribution samples as a systematic strategy will be beneficial for dealing with pest population density distribution. Then, the quasi-random distribution will be as follows:

$$x_i = a + \gamma_i(b - a);\ i = 1, .., N, \tag{3.6.2}$$

where $\gamma_i \in (0, 1)$ is a random variable taken from the uniform distribution. For $i = 1, .., N$, x_i represents the irregular grid nodes that are listed in ascending order, with fixed end points as follows:

$$x_1 = a, \quad x_N = b.$$

In order to estimate the pest population density on an irregular computational grid, numerical integration methods will be applied. The statistical rule and forms of the trapezoidal rule and Simpson's rule will be used. As demonstrated previously, the statistical rule (1.2.1) does not has any spatial dependence; therefore, the same form can be applied on an irregular grid. Meanwhile, another form of the trapezoidal rule (3.1.13) and Simpson's rule (3.1.14) will be deduced to cope with irregular computational grid nodes. We will replace the integrand function, with an interpolation polynomial to be integrated instead. When the function $u(x)$ is replaced by a linear polynomial, the derived formula

will be the trapezoidal rule which has the following form:

$$I \approx \widetilde{I} = \sum_{i=1}^{N} h_i \frac{(u_i + u_{i+1})}{2}, \tag{3.6.3}$$

where the grid step size $h_i = x_{i+1} - x_i$ is now a variable. Another form of numerical integration on irregular grid nodes can be derived by replacing the function $u(x)$ with a quadratic polynomial, then the derived polynomial will represent Simpson's rule as:

$$I \approx \widetilde{I} = \sum_{i=1}^{(N-1)/2} \frac{h_{2i-1} + h_{2i}}{6} \left(u_{2i-1} + 4u_{2i} + u_{2i+1}\right). \tag{3.6.4}$$

There is one restriction on applying this formula to any irregular grid nodes. This is that the number of grid nodes must always be an odd number.

In order to illustrate the convergence rate for numerical integration methods on irregular grid nodes, a sequence of grid nodes will be used. Then, the relative integration error (3.1.8) will be employed. However, when dealing with irregular computational grids, it is necessary to compute the mean error at each number N of grid nodes N. Our applications will be restricted to the quasi-random grid nodes. When dealing with quasi-random grid nodes, each grid generation will be repeated n_r times, then we will have n_r values of the relative integration error (3.1.8) for the total number N of grid nodes. As demonstrated above, the mean error can be calculated as follows:

$$\mu(E) = \frac{1}{n_r} \sum_{i=1}^{n_r} E_i. \tag{3.6.5}$$

Since in the ecological data there is no available function, then simulated ecological pop-

ulation density functions will be obtained from solving the system of equations (2.2.10), (2.2.11) on a very fine regular grid of $N_f = 2^{15} + 1$ nodes over the interval $[a, b] = [0, 1]$ numerically. Now let us consider we have fine grids x_i^f, $i = 1, .., N_f$ available, where:

$$x_1^f = a = 0,\ x_{N_f}^f = b = 1,$$

$$x_i^f = x_{i-1} + \tfrac{b-a}{N_f - 1};\ i = 2, .., N_f - 1.$$

In order to generate a slightly irregular grid of N, a regular grid nodes of N will first be extracted from the available fine grid as follows:

$$x_i = x_j^f;\ j = 1 + (i-1)(\frac{N_f - 1}{N - 1}),\ i = 1, ..., N. \tag{3.6.6}$$

A single interior node will be perturbed to generate a slightly irregular grid node, according to the following formula:

$$x_i = x_{j+\gamma}^f, \gamma \in [-\frac{N_f - 1}{2(N-1)}, \frac{N_f - 1}{2(N-1)}]. \tag{3.6.7}$$

For some selected $i = 2, .., N-1$, $\quad j$ is given in (3.6.6), and γ is a random integer selected from the uniform distribution.

Meanwhile, to generate quasi-random simulated ecological data, the endpoints will be fixed as follows:

$$x_1 = x_1^f,\ x_N = x_{Nf}^f.$$

The interior points will be perturbed as follows:

$$x_i = x_{j+\gamma}^f;\ \gamma \in [-\frac{N_f - 1}{2(N-1)}, \frac{N_f - 1}{2(N-1)} - 1]. \tag{3.6.8}$$

This will be done for some $i = 2, ..., N-1$. It can be seen from (3.6.8) that the upper limit of the interval that γ belongs to is one less than that in (3.6.7) in order to prevent

any coinciding nodes. The procedure that is applied to generate slightly irregular grid nodes will be applied to generate a quasi-random grid nodes with each interior grid node perturbed, instead of perturbing a single interior grid node. Then, the generated quasi-random grid nodes $x_i, i = 1, ..N$ will be sorted into ascending order, where the endpoints are fixed as follows:

$$x_1 = a = 0, \quad x_N = b = 1.$$

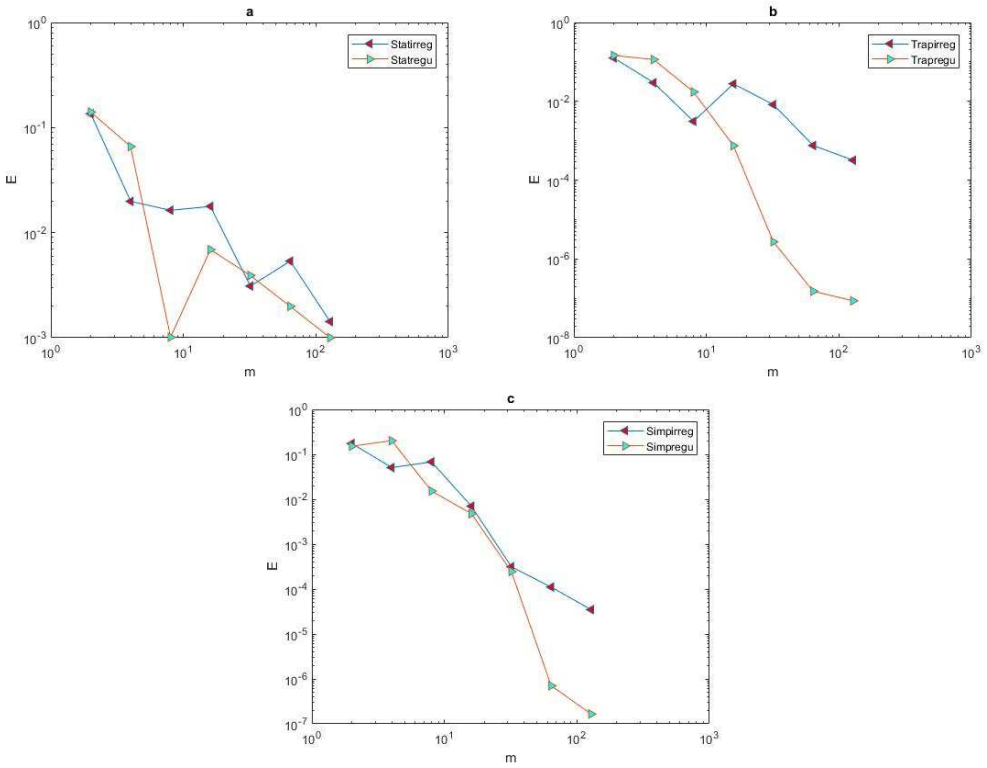

Figure 3.7: Comparison of the convergence curves for the one-peak function on quasi-random and regular grids: (a) the statistical rule, (b) the trapezoidal rule and (c) Simpson's rule.

The application of numerical integration methods on simulated ecological data will be restricted to quasi-random grid nodes. Now let us consider the simplest ecological test case, a function of one-peak, and more sophisticated ecological test cases, functions with three-peak and multiple-peak. These simulated ecological test cases are shown in Figure

3.4. Since there is no function available for integration in order to compute the exact value of the set of simulated discrete data, the trapezoidal rule (3.1.13) will be applied to approximate the extremely fine regular grid nodes of $N_f = 2^{15}+1$, to be then the exact value of the population pest abundance I. For each number of coarse grid nodes extracted from the very fine number N_f of regular grid nodes, the relative error is computed according to (3.1.8). For quasi-random grid generation, each generated grid will be repeated n_r times; then we will have n_r values of the relative integration error (3.1.8) for the total number N of grid nodes. As demonstrated above, the mean error can be calculated from formula (3.6.5). Convergence curves for the one-peak function over quasi-random grid nodes are

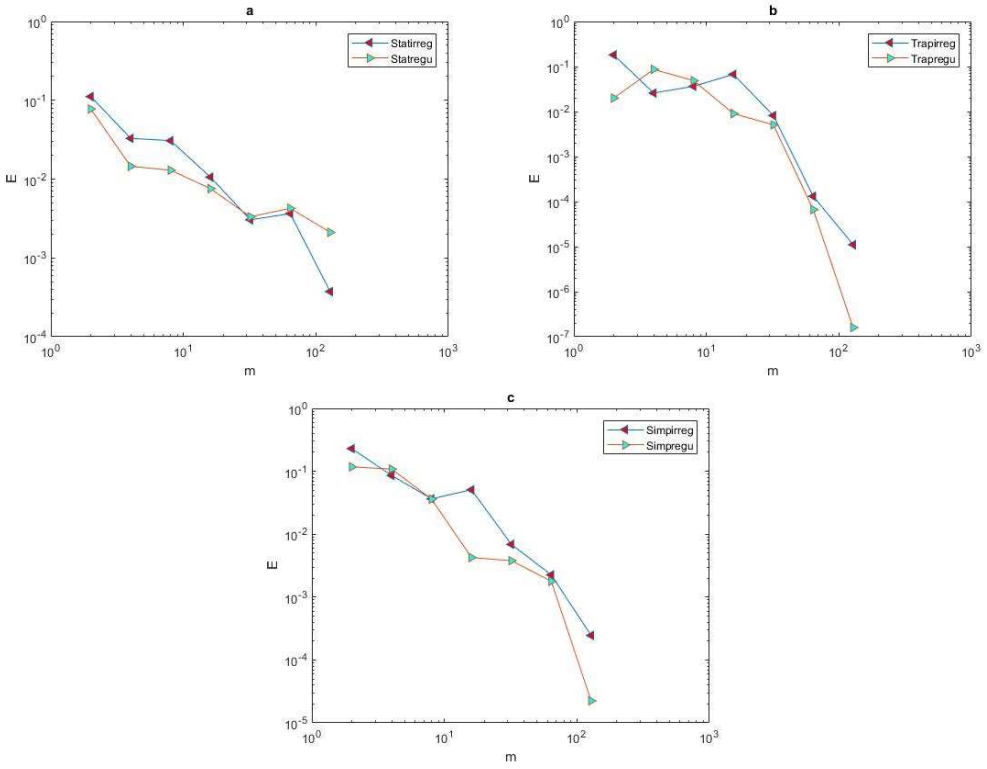

Figure 3.8: Comparison of the convergence curves for the three-peaks function on quasi-random and regular grids: (a) the statistical rule, (b) the trapezoidal rule and (c) Simpson's rule.

shown in Figure 3.7, for different numerical integration methods. The convergence curve of the mean error (3.6.5) employed for statistical rule (1.2.1) is presented in Figure 3.7a,

along with the convergence curve on regular grid nodes for comparison. The convergence curves for trapezoidal rule (3.1.13) and Simpson's rule (3.1.14) are presented in Figure 3.7b, and Figure 3.7c. It can be noted from Figure 3.7 that the integration error increases when each interior grid node is perturbed, no matter which numerical integration method is used.

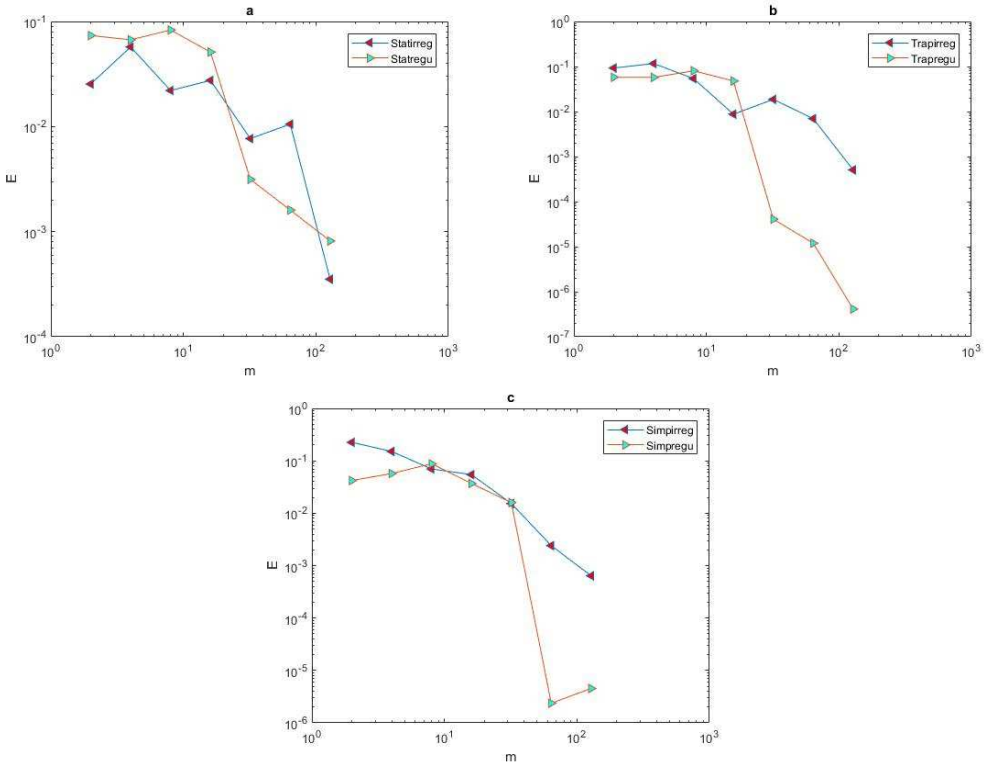

Figure 3.9: Comparison of the convergence curves for the multi-peak function on quasi-random and regular grids: (a) the statistical rule, (b) the trapezoidal rule and (c) Simpson's rule.

In order to investigate more complex simulated ecological data three-peak and multi-peak functions will be presented. The convergence curves for the three-peak function and multi-peak functions on a sequence of quasi-random and equidistant grid nodes are shown in Figure 3.8 and Figure 3.9 respectively. When dealing with more sophisticated ecological data, the convergence curves for different numerical integration methods on a sequence of quasi-random grid nodes offers a much poorer degree of accuracy. Furthermore, the

integration error on a sequence of quasi-random grid nodes becomes bigger than the integration error on a sequence of regular grid nodes. This observation is prominent when dealing with multi-peak functions. This is true even with the statistical rule which does not have any spatial dependence.

3.7 The Richardson Extrapolation

We have discussed in the previous sections that the composite Simpson's rule can converge much faster than the trapezoidal rule, and it therefore can present a more accurate evaluation of pest abundance. However, there are some restrictions that may prevent us from applying the composite Simpson rule. In real-life applications, we cannot ensure that the number N of grid nodes will always be an odd number in order to apply the composite Simpson's rule. Therefore, we require an alternative, more flexible method will potentially deliverer the same accuracy as the composite Simpson's rule. The alternative approach used to improve the degree of accuracy of our approximation, and which does not require an odd number of grid nodes, is called the Richardson extrapolation [24].

The Richardson extrapolation can be defined as a method that amalgamates several sequential approximations of the quantity of interest I, by applying a certain mathematical technique to attain more precise approximations $\tilde{I}$ [150]. The Richardson extrapolation can be used with any numerical integration method to increase the convergence rate of the approximations. If there is a method that converges as $O(h^m)$, the Richardson extrapolation can readily produce convergence as $O(h^{2m})$, where h is the grid step size and m refers to the degree of integration error of the numerical integration methods [38, 150]. This procedure will reduce the error by a factor of h. For instance, the convergence rate of the trapezoidal rule $O(h^2)$ can evolve into the accelerated convergence rate $O(h^4)$ by applying the Richardson extrapolation to the trapezoidal rule where $m = 2$. Whereas the convergence rate of Simpson's rule $O(h^4)$ can evolve into the accelerated convergence rate

$O(h^8)$ by applying the Richardson extrapolation to Simpson's rule where $m = 4$.

In order to derive the general formula of the Richardson extrapolation, let us introduce $\tilde{I}(h)$, which denotes the value of the quantity I, that is obtained from a grid with grid step size h. For any numerical integration method, the following computational formula holds:

$$I = \tilde{I}(h) + Ch^p, \tag{3.7.1}$$

where I is the analytic quantity of interest, p is the degree of the numerical integration methods used and C is a constant. Note that the work required to compute $\tilde{I}(h)$ increases sharply as $h \to 0$.

The Richardson extrapolation can be used asymptotically on two sequential grids to derive more accurate computational formulae. Let us consider two grids G_1 and G_2, where the grid step sizes are h_1 and h_2 respectively, and $h_2 = \gamma h_1$, $0 < \gamma < 1$. From equation (3.7.1) the computational formulae for grids G_1 and G_2 are:

$$I \approx \tilde{I}(h_1) + C_1 h_1^p = \tilde{I}_1 + C_1 h_1^p, \tag{3.7.2}$$

$$I \approx \tilde{I}(h_2) + C_2 h_2^p = \tilde{I}_2 + C_2 h_2^p. \tag{3.7.3}$$

Since $h_2 = \gamma h_1$, then $h_2^p = \gamma^p h_1^p$. Substituting h_2 into (3.7.1), we obtain:

$$I = \tilde{I}_2 + C_2 \gamma^p h_1^p. \tag{3.7.4}$$

Then, dividing (3.7.2) by $\gamma^p(C_2/C_1)$ and subtracting it from (3.7.4), we arrive at:

$$I(1 - \gamma^p(C_2/C_1)) \approx \tilde{I}_2 - \gamma^p(C_2/C_1)\tilde{I}_1, \tag{3.7.5}$$

so that

$$I \approx w\tilde{I}_1 + (1-w)\tilde{I}_2, \tag{3.7.6}$$

where $w = \frac{-\gamma^p(C_2/C_1)}{1-\gamma^p(C_2/C_1)}$, γ represents the ratio of the grid step size in neighbouring grids (G_1, G_2), i.e. $\gamma = h_2/h_1$ and constants C_1, C_2 are usually not known. Let us assume that $C_1 \approx C_2$, then we have

$$w = \frac{-\gamma^p}{1-\gamma^p}. \tag{3.7.7}$$

Equation (3.7.6) provides the general formula of the Richardson extrapolation for numerical integration methods.

3.8 Applications of the Richardson Extrapolation to the Mathematical and Ecological Test Cases

It has been mentioned above that the main goal of applying the Richardson extrapolation is to overcome the obstacle that we face when applying the composite Simpson's rule. In other words, we hope that the accuracy that is obtained by employing the composite Simpson's rule can be achieved by applying the Richardson extrapolation to the trapezoidal rule. Therefore, the Richardson extrapolation will be applied to the three mathematical test cases presented in Table 2.1. In the application of the Richardson extrapolation, the value of p in the formula (3.7.7) will differ depending on the numerical integration method used. According to [36], the convergence rate of the composite trapezoidal rule is $E = Ch^2$; therefore, the value of p in the formula (3.7.7) equals $p = 2$. Also the value of p is $p = 4$ when applying the composite Simpson's rule because the convergence rate of the Simpson rule is $E = Ch^4$. As can be seen from Figure 3.10, the convergence curves obtained by applying the composite Simpson's rule are parallel to the convergence curves obtained by employing the Richardson extrapolation of the composite trapezoidal rule in all three test cases. This means we have achieved the same accuracy,

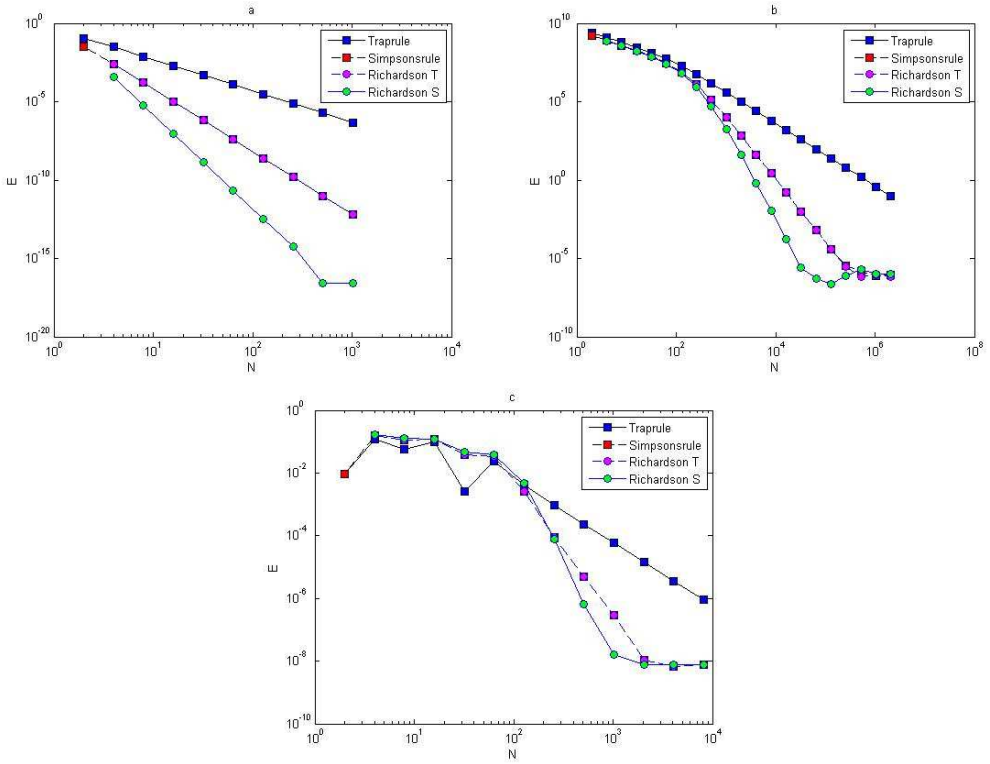

Figure 3.10: (a) The convergence curves for TC5, (b) the convergence curves for TC14 and (c) the convergence curves for TC21. Comparison of the convergence curves for the trapezoidal rule, Simpson's rule, the Richardson extrapolation of the trapezoidal rule and the Richardson extrapolation of Simpson's rule.

but with an even number of grid nodes. Furthermore, Figure 3.10 suggests that more accurate results have been achieved when applying the Richardson extrapolation to the composite Simpson's rule. Now we shall consider ecological test cases. It can be seen from previous applications of numerical integration methods to the $1d$ ecological test cases that these methods do not provide a satisfactory degree of accuracy for our approximations. Thus, an alternative method must be applied to improve the degree of accuracy of our approximations. As discussed previously, the Richardson extrapolation can be employed to the ecological test cases in the same way as we have done with the mathematical test cases. In standard mathematical test cases, the Richardson extrapolation (3.7.6) achieved a satisfactory degree of accuracy and was a reliable approach to use in order to overcome

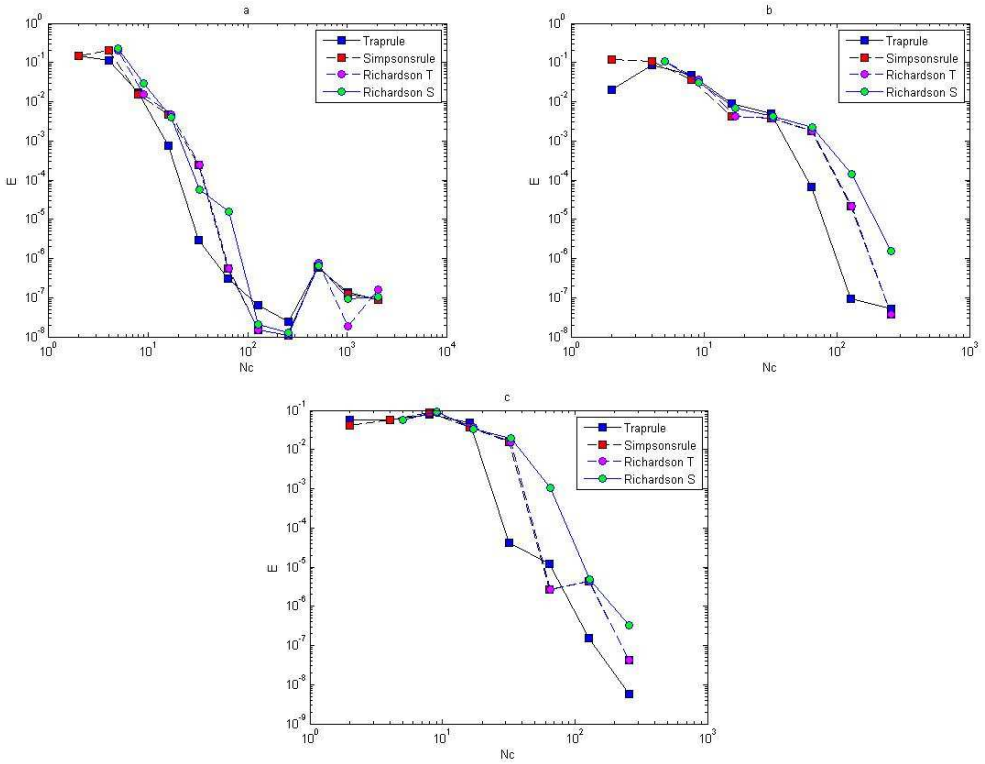

Figure 3.11: Convergence curves of the numerical integration methods and the Richardson extrapolation of the $1d$ ecological test cases on a regular grid: (a) 1ETC1, (b) 1ETC2, and (c) 1ETC3.

the restrictions that we found in our mathematical applications. However, this is not always the case when dealing with ecological problems.

It can be seen from Figure 3.11 that the Richardson extrapolation does not provide any improvement in the convergence rate of the integration error. Furthermore, the Richardson extrapolation applied to the composite Simpson's rule, has poor approximation accuracy as shown in Figure 3.11. Although the convergence curves of the Richardson extrapolation applied to the trapezoidal rule are parallel to the convergence curves of the composite Simpson's rule, the Richardson extrapolation of the composite trapezoidal rule is not a reliable approach to use in order to increase our degree of accuracy. The reason for the poor performance of the Richardson extrapolation in the ecological problem is that the convergence rate of the composite Simpson's rule is slower than the convergence rate of

the trapezoidal rule in complicated ecological test cases. In other words, due to the nature of ecological problems in regions where the available data is sparse, the trapezoidal rule provides more accurate estimates of the problem than Simpson's rule. Because this, the Richardson extrapolation failed to work well with ecological problems, as presented in Figure 3.5. Therefore, we cannot adopt the Richardson extrapolation method in ecological problems, and must seek an alternative approach to improve the convergence rate of $1d$ ecological test cases.

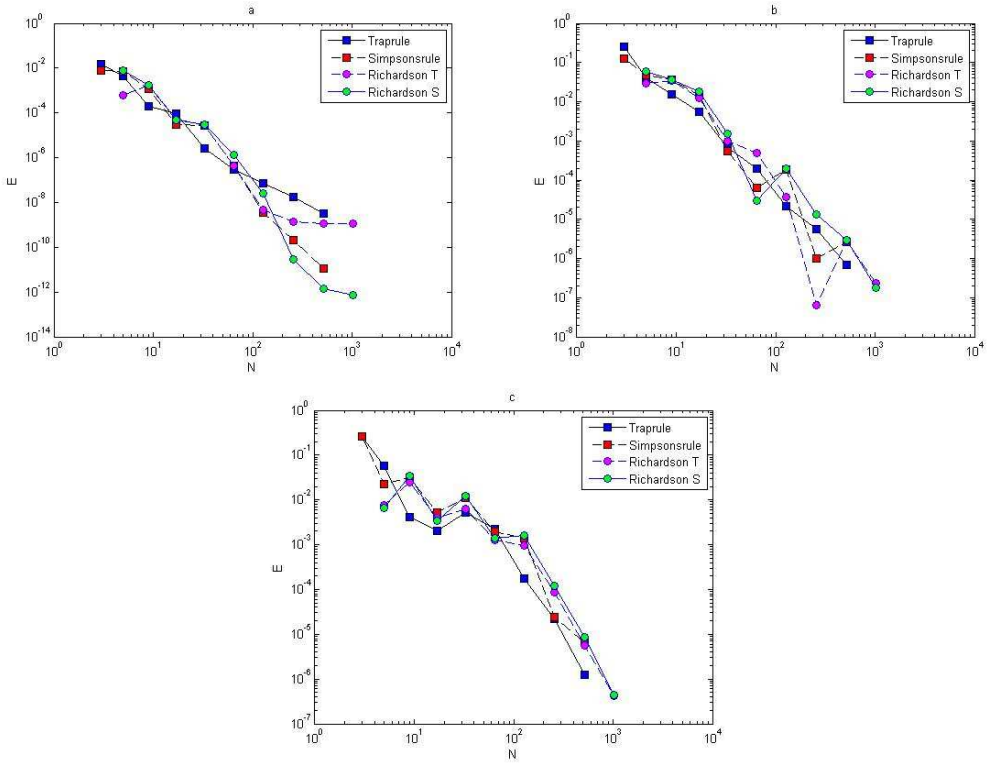

Figure 3.12: Convergence curves of the Richardson extrapolation of the numerical integration methods for the $2d$ ecological test cases: (a) ETC1, (b) ETC2, and (c) ETC3.

In terms of dealing with $2d$ ecological data, it can be argued from the discussion in the previous chapter that there is no specific method that be would described as a reliable approximation approach on coarse grids. In order to confirm this conclusion, the Richardson extrapolation for numerical integration methods will be applied to the previous $2d$

ecological test cases. The Richardson extrapolation (3.7.6) of the trapezoidal rule (3.3.6) and Simpson's rule (3.3.9) will be employed in these test cases with the purpose of checking whether a more accurate estimation of the population density can be achieved.
Figure 3.12 shows that, the convergence curves of Simpson's rule are parallel to the convergence curves of the Richardson extrapolation of the trapezoidal rule for all ecological test cases. The accuracy achieved by applying the composite Simpson's rule is achieved by employing the Richardson extrapolation of the trapezoidal rule. But this does not improve the degree of accuracy. That is because the composite Simpson's rule from the beginning failed to provide an acceptable degree of accuracy when dealing with more complicated ecological test cases, as shown previously in Figure 3.6. Moreover, the Richardson extrapolation of the composite Simpson's rule does not present any improvement in the convergence rates of the approximation. Therefore, according to these results, it can be argued that the Richardson extrapolation cannot be used as a reliable method to improve the degree of accuracy in ecological test cases.

3.9 Chapter 3 Conclusion

This chapter has discussed how to use numerical integration techniques (3.1.4) to estimate pest population density. The results of our study demonstrate that numerical integration methods are more accurate than statistical methods. In particular, Simpson's rule can be more accurate than the trapezoidal rule, and its convergence rate is much faster than the convergence rate of the trapezoidal rule in the mathematical test cases. The Richardson extrapolation may have a higher degree of accuracy, and it may provide a more accurate evaluation of the integration error when dealing with mathematical test cases. However, in ecological test cases the Richardson extrapolation failed to provide a satisfactory degree of accuracy to the problem under study, especially in coarse grids. The previous ecological results illustrated that the degree of accuracy depends on the spatial pattern

of the density distribution and on the degree of its spatial heterogeneity, as presented in [136, 138]. Therefore, an alternative approach must be applied to provide an accurate estimation for ecological problems when the data available is sparse.

We have also investigated numerical integration approaches applied to standard mathematical test cases, and to generated ecological data over a sequence of irregular computational grid nodes. It has been documented above that, depending on the numerical integration methods used on random computational grids, the degree of the convergence rate is affected. The results for irregular ecological data, can be summed up with the following remarks. Randomisation leads to larger values for the integration error, whether on coarse or fine grids. The convergence curves for numerical integration methods on regular computational grids always lie below those obtained from irregular computational grid nodes. Spatial dependence does not affect the results of accuracy, as illustrated when applying the statistical rule. Thus, more accurate numerical methods are required to deal with simulated irregular ecological data, as the numerical integration methods achieved much better results on equidistant ecological data. Finally, a prior knowledge of the spatial pattern of the pest population density will be beneficial to deal with an irregular distribution.

Chapter 4

Evaluation of the Total Population Density on Coarse Grids

Evaluation of the population density in many ecological and biological problems requires a satisfactory degree of accuracy. Insufficient information about the population density, obtained from sampling procedures negatively impacts on the accuracy of the estimate. When dealing with sparse ecological data, the asymptotic error estimate (3.1.10) fails to achieve a reliable degree of accuracy, as demonstrated previously. It is essential to investigate which factors affect the degree of accuracy of numerical integration methods. When the number of traps is less than the recommended threshold, the degree of accuracy will be negatively affected. Therefore, available methods cannot guarantee a satisfactory degree of accuracy, and in this sense the error will be probabilistic rather than deterministic. In other words, the probabilistic approach is used instead of the deterministic approach in this instance; by considering the error as a random variable, the chance of obtaining an accurate estimation can be quantified. In the probabilistic approach, we determine a threshold number of grid nodes required to guarantee a desirable level of accuracy with

the probability equal to one.

4.1 Characteristics of the Probabilistic Approach for Higher Aggregated Population Density

In this section, the probability approach will be used to illustrate that when the number of grid nodes is very small, the accuracy is probabilistic rather than deterministic (*i.e* achieving an accurate estimation becomes a matter of chance) [136]. Let us consider a peak function which represents an early stage of patchy invasion [136]. The population density function over interval $D = [a, b]$ will be presented as a single peak function located within sub-interval $D_u = [x_I, x_{II}]$, and the density will be zero elsewhere. Therefore, a highly aggregated density distribution $u(x)$ can be defined as:

$$u(x) = \begin{cases} f(x) > 0 & \text{for} \quad x \in [x_I, x_{II}] \\ f(x) = 0 & \text{otherwise} \end{cases} \tag{4.1.1}$$

The maximum of the peak is only obtained from the construction $x^* = \frac{(x_I + x_{II})}{2}$ and it is assumed to be located within this construction. An example of a $1d$ peak function is shown in Figure 4.1a. The location of the peak must be considered randomly within a truncated sub-interval $[a_I, b_I]$ taken from the original interval $[a, b]$ (*i.e* $a_I = a + \epsilon < x^* < b - \epsilon = b_I$). The restriction imposed on the location of the peak function is to prevent any important information on the population density being missed if the peak is located at the endpoints of the original interval $[a, b]$ or at the boundaries of domain.

A standard example of a peak function is given by substituting $f(x)$ in equation (4.1.1) as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x - x^*)^2}{2\sigma^2}), \tag{4.1.2}$$

where the location of the peak maximum x^* is a random variable within the truncated

sub-interval, and σ is the standard deviation. The definition of the peak width as a function of the standard deviation is represented as follows [126]:

$$\delta = 6\sigma. \tag{4.1.3}$$

Now let us generate peak function by using formula (4.1.2) over interval $[0, L]$ where $L = 300$, $\sigma = 8$, and the locations of peak will be changed randomly within $[a_I, b_I] = [40, 260]$ where $\epsilon = 40$ At a very small number of grid nodes, and when the function of highly aggregated density distributions is considered, the accuracy of the estimate cannot be determined and the integration error is dependent on the location of a random point x^*. Then, the error becomes a random variable as a result. The generated function will be integrated over the interval $[0, 300]$ and a series of an increasingly refined number of regular grid nodes $N = 3, 4, 5, ...$, by applying statistical rule (1.2.1). Then, the relative error (3.1.8) will be computed to assess the accuracy of the method used, by using the criterion (3.2.1). The condition $E \leq \tau$ must hold, where in ecological applications a good level of accuracy lies between $0.2 \leq \tau \leq 0.5$, in some cases, $\tau \approx 1$ may be considered acceptable level of accuracy, where τ is the accuracy tolerance [128, 164].

At each number of grid node, the computations will be repeated for $n_r = 10^4$ realisations of the random variable x^*. Then, the probability $p(h)$ of accurate numerical integration is computed as follows:

$$p(h) = \frac{\hat{n_r}}{n_r}, \tag{4.1.4}$$

where $p(h)$ is the probability obtained numerically, h is the grid step size on the number N of grid nodes $N = 3, 4, 5, 6, ...$, and $\hat{n_r}$ is the number of realisations for which the integration error satisfies the condition $e \leq \tau$ where $\tau = 0.25$. We repeat the computations by increasing the number of grid nodes until the grid step size becomes $h \leq \frac{\delta}{2}$ then,

the computations will be stopped. The peak function (4.1.2) is shown in Figure 4.1 for the peak width $\delta = 48$. The probability $p(h)$ when the condition (3.2.1) of the function (4.1.2) holds is presented in Figure 4.1b. The computation started from a grid of $N_1 = 3$

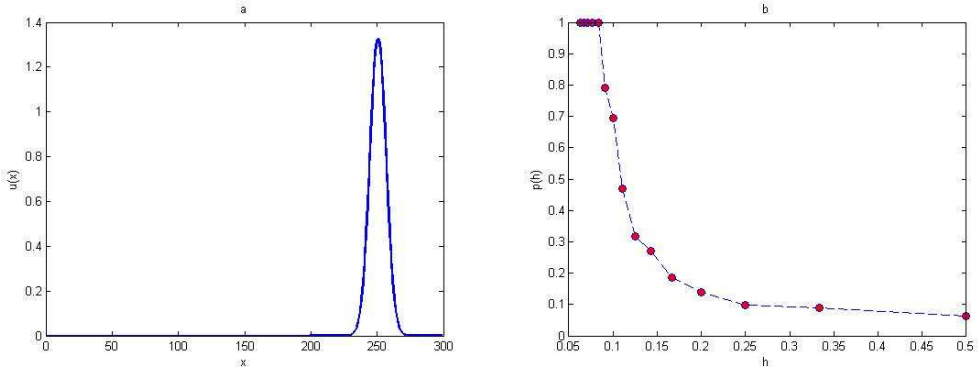

Figure 4.1: (a) A peak function (4.1.2). The peak width is $\delta = 6\sigma$ where $\sigma = 8$. (b) The probability $p(h)$ obtained by direct computation for the function (4.1.2).

nodes and ended on a grid of $N_{15} = 17$, nodes where the condition $h \leq \frac{\delta}{2}$ holds. It can be seen from Figure 4.1b that the integration error when integrating the function (4.1.2) on a very coarse grid of nodes will depend on the location of the peak, and the probability of achieving an accurate estimate is $p(h) < 1$. Meanwhile, on a grid of $N = 13$ nodes $h = \frac{1}{12}$ the error is deterministic and the error is $e \leq \tau = 0.25$ no matter where the peak is located. Hence, it follows from the results of integrating the function of highly aggregated density distributions that the integration error becomes a random variable, and the accuracy of numerical integration at a very small number of grid nodes is probabilistic rather than deterministic, even if an acceptable degree of accuracy is acquired.

The above discussion leads us to introduce a new type of grid to be added to the grid classification, which is defined as an ultra-coarse grid. Therefore, there are now three types of computational grid: fine grid, coarse grid, and ultra-coarse grid. With fine grids, the asymptotic error estimate (3.1.10) always holds. For coarse grid, the asymptotic error estimates do not hold, however, the error will be deterministic and it will satisfy the con-

dition (3.2.1) for a chosen τ. In the ultra-coarse grid , the asymptotic error estimate does not hold and the accuracy is not deterministic. The accuracy can only described in terms of the probability of achieving an error within a prescribed tolerance τ. The probability satisfies $p(h) < 1$ when the condition $e \leq \tau$ holds for a chosen tolerance τ. Therefore, the threshold number N^* of grid nodes must be determined, where at any number of grid node less than N^*, the probability is $p(h) < 1$ and the computational grid is defined as an ultra-coarse grid. At the number of grid nodes $N \geq N^*$, the transition from ultra-coarse to coarse grid takes place and the error becomes deterministic.

Now let us vary the value of standard deviation σ in the peak function (4.1.2) to investigate the impact of choosing different values of σ on the probability of achieving an accurate estimation. The peak function (4.1.2) will be considered and the value of σ will vary as follows $\sigma = 4, 6, 8, 10, 12, 15$. Then at each value of σ, the above computations will be repeated. All values of threshold numbers of grid nodes, alongside their value of σ, are presented in Table 4.1.

σ	4	6	8	10	12	15
N^*at $\tau = 0.25$	25	17	13	11	9	8
N^*at $\tau = 0.4$	22	15	11	9	8	6

Table 4.1: The number of grid nodes required to achieve an accurate estimate of the peak function (4.1.2) at fixed $\tau = 0.25, 0.4$ and several values of σ.

It can be seen from Table 4.1 that the number of grid nodes required for the condition $p(h) = 1$ to hold depends on the width of the peak function $\delta = 6\sigma$. By comparing the critical number of grid nodes required to achieve an accurate estimate at two different values of τ, it can be noted from Table 4.1 that there is an inverse relationship between peak width $\delta = 6\sigma$ and N^*. For a wide peak, the critical number of grid nodes required is small, whereas this number becomes bigger at narrow peaks, as shown in Table 4.1 and Figure 4.2b. Furthermore, the number of grid nodes to fulfill the accuracy requirements in the probabilistic approach depends on the value of tolerance τ, where from Table 4.1 it

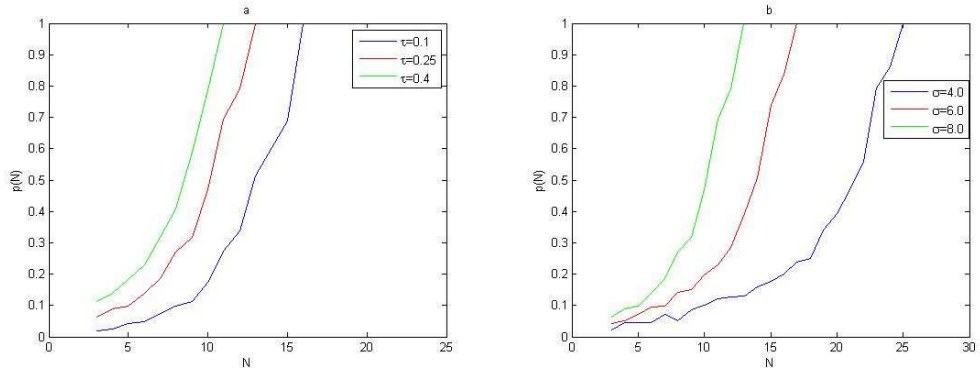

Figure 4.2: (a) Graphs of the probability function $p(N)$ for fixed $\sigma = 8$ in (4.1.2) and various values of tolerance τ. (b) Graphs of the probability function $p(N)$ for fixed $\tau = 0.25$ in (4.1.2) and various values of parameter σ .

can be seen that for a low level of accuracy $\tau = 0.4$, the required number of grid nodes is considerably less than the number required at $\tau = 0.25$ at the same values of peak width δ, as can be seen from Figure 4.3.

An instructive example is depicted in Figure 4.2a, where at fixed width $\sigma = 8$ in (4.1.2), the probability function $p(N)$ depends on the chosen tolerance τ: where at the high level of accuracy $\tau = 0.1$, the critical number of grid nodes N^* required for $p(N) = 1$ to hold is considerably bigger than the numbers required for $\tau = 0.25$ and $\tau = 0.4$ respectively.

σ	N^*at $P = 1$	N^*at $P = 4$	N^*at $P = 8$
4	25	24	23
6	17	16	15
8	13	12	10
10	11	10	9
12	9	8	8
15	8	7	7

Table 4.2: The number of grid nodes required to achieve an accurate estimate of the peak function $P = 1, 4, 8$ at fixed $\tau = 0.25$ and several values of σ.

We now consider the superposition of normal distributions given by the following equations:

$$u_j(x,t) = \frac{w(t)}{P} \sum_{m=1}^{P} f_j^m(x), \;\; j = A, B, \tag{4.1.5}$$

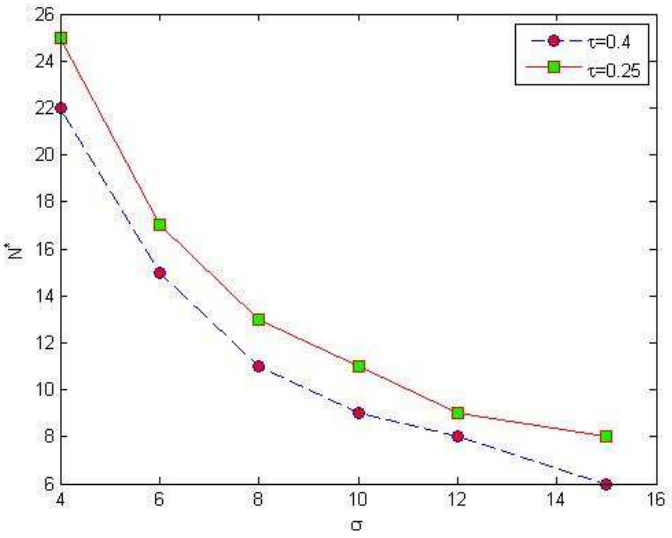

Figure 4.3: The critical number of grid nodes for the probabilistic approach at different values of σ and different values of τ.

where

$$f_j^m(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-(x^*)^m)^2}{2\sigma^2}). \tag{4.1.6}$$

This generates functions of four and eight peaks by substituting the value of P as $P = 4$, or $P = 8$, as shown in Figure 5.8a and Figure 5.8b respectively. The previous computations, which are applied to the peak function (4.1.2) at several values of standard deviation $\sigma = 4, 6, 8, 10, 12, 15$, will be repeated for functions with four and eight peaks respectively. The main aim of this variation in the number of peaks is to determine the effect of increasing the number of peaks on the probability of achieving an accurate estimate on a coarse grid nodes. For each value of σ at different numbers of peaks, $P = 1, 4, 8$, in the function, the number of grid nodes required for the probability condition $p(N) = 1$ to hold is recorded in Table 4.2. It can be noted from Table 4.2, by comparing the values of the number of grid nodes N^* with different numbers of peaks, that when the number of peaks becomes $P = 8$, the number of grid nodes N^* required to achieve an accurate estimate is slightly less than that number N^* required at number of peaks $P = 1, 4$ respectively as

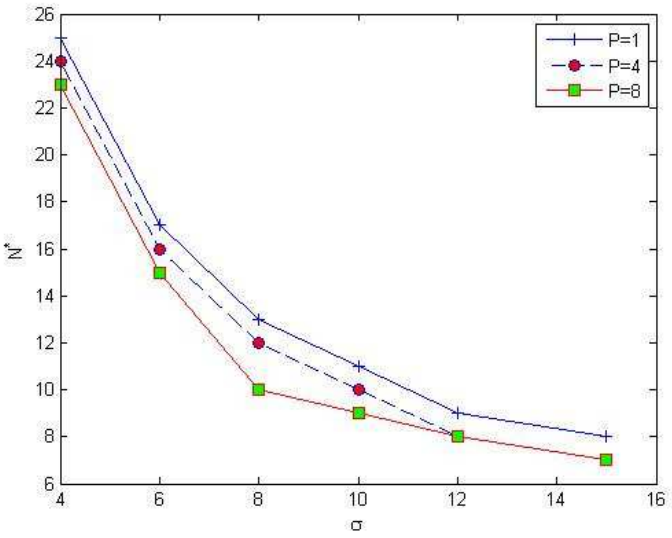

Figure 4.4: The number of grid nodes required to achieve an accurate estimate of the function with different peaks $P = 1, 4, 8$ at a fixed value of tolerance $\tau = 0.25$.

shown in Figure 4.4. Therefore, when the number of peaks is increased we have a higher probability of obtaining an accurate estimate.

Let us now expand the population density function given by (4.1.1) in the vicinity of the location of the maximum x^* as follows:

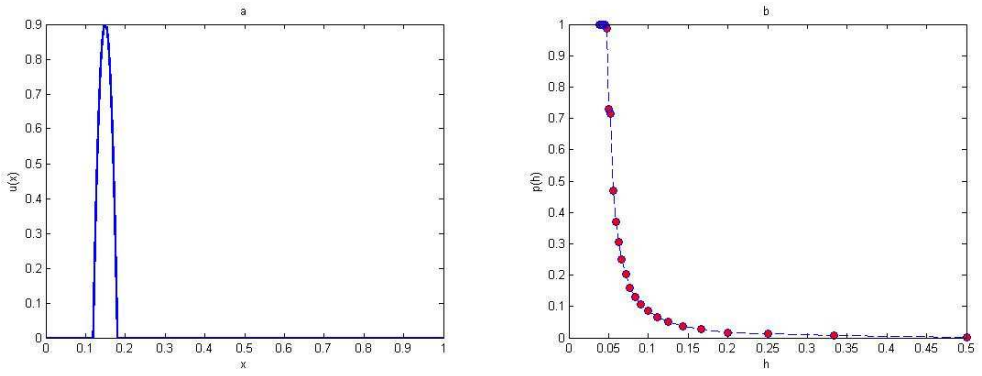

Figure 4.5: (a) A spatial population density given by quadratic function (4.1.8). The peak width is $\delta = 0.06$ and the parameters are $A = 1000$ and $B = 0.9$. (b) The probability $p(h)$ was obtained by direct computation for the function (4.1.8) at $\delta = 0.06$.

$$u(x) = u(x^*) + \frac{1}{2}\frac{d^2u(x^*)}{dx^2}(x - x^*)^2 + R(x). \tag{4.1.7}$$

In the vicinity of the peak, the remainder term $R(x)$ can be ignored, so that the density peak function is given by the quadratic function as follows:

$$u(x) \approx Q(x) = B - A(x - x^*)^2, \quad x \in [x_I, x_{II}], \tag{4.1.8}$$

where $A = -\frac{1}{2}(\frac{d^2u(x^*)}{dx^2}) > 0, \quad B = u(x^*) > 0$. The maximum value of the peak coincides with the maximum value of the quadratic function, which is symmetric around the location of x^*. The quadratic function $Q(x)$ must be always non-negative over sub-interval $[x_I, x_{II}]$, *i.e* $Q(x_I) = Q(x_{II}) = B - Ah^2 \geq 0$. Then, the following condition must hold:

$$h^2 \leq \frac{B}{A}. \tag{4.1.9}$$

Let us now use the quadratic function given in equation (4.1.8) to generate data over unit interval $[0, 1]$. A uniform grid of N nodes is generated in the domain $[0, 1]$, as $x_{i+1} = x_i + h$ where the grid step size $h = 1/N - 1, \quad i = 1, 2, ..., N - 1$. The population density in the vicinity of the peak will be considered, where for the interval $[x_{i-1}, x_i], \quad \delta = 2h$ is the width of the peak function. From equation (4.1.9), the width of the peak function becomes $\delta = 2\sqrt{\frac{B}{A}}$ and the roots are $x_I = x^* + \delta/2, \; x_{II} = x^* - \delta/2$. The grid step size h can be defined in terms of peak width as $h = \alpha\delta$, where the parameter $\alpha > 1$. The exact value of abundance I can be computed by treating the peak as a quadratic function, as follows:

$$I = \int_{x_I = x_{i-1}}^{x_{II} = x_{i+1}} Q(x)dx = 2Bh - \frac{2Ah^3}{3}. \tag{4.1.10}$$

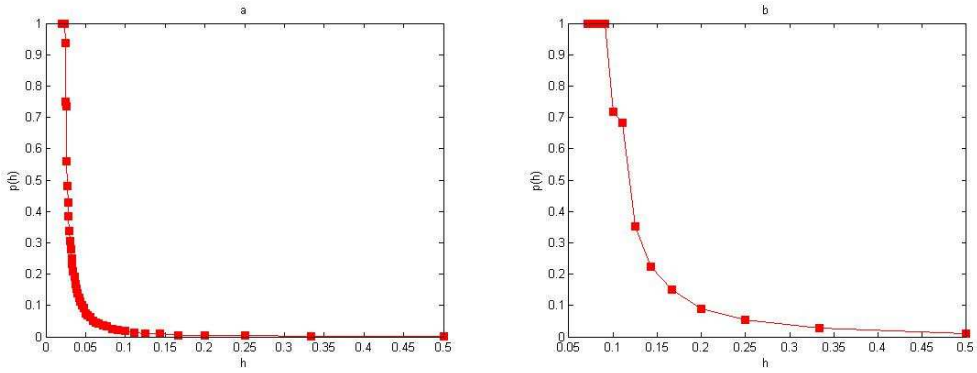

Figure 4.6: (a) The probability $p(h)$ obtained by direct computation of the function (4.1.8) at $\delta = 0.03$. (b) The probability $p(h)$ obtained by direct computation of the function (4.1.8) at $\delta = 0.12$.

In order to ensure that the relative error satisfies the condition $E_{rel} \leq \tau$, in the quadratic function (4.1.8), the chosen value for the peak is always positive *i.e* $A > 0$ and the grid step size h should satisfy the following condition $h \leq \sqrt{\frac{B}{A}}$.

Now let us consider the quadratic function given by (4.1.8). As demonstrated in the previous section, the probability of achieving a sufficiently accurate estimate when modelling a peak function with normal distribution will be computed. The width of the peak is fixed as $\delta = 0.03, 0.06, 0.12$, and the location of peak x^* is a random variable distributed uniformly over the interval $[\delta, 1-\delta]$, where the entire peak must be stationed within the unit interval $[0, 1]$. By fixing the tolerance $\tau = 0.25$, then the probability of achieving an accurate estimate at different values of peak widths can be calculated from equation (4.1.4). The density of quadratic function (4.1.8) at $\delta = 0.06$ is shown in Figure 4.5a.

Figure 4.5b shows the probability $p(h)$ of obtaining an accurate answer at peak width $\delta = 0.06$, from this figure, it can be seen that for a very small number of grid nodes, the probability is $p(h) < 1$ and the integration error itself is a random variable with a high magnitude. The probability becomes equal to one at the critical number of grid nodes $N^* = 23$. This required critical number of grid nodes N^* increase as the value of the peak width decreases (*i.e* a narrower peak) and it decreases as the peak width increases,

as can be seen from Table 4.3. Table 4.3 presents the number of grid nodes required to achieve an accurate estimate of the quadratic function at different values of peak widths δ. The probability curves for approximating the quadratic function with peak widths $\delta = 0.03, 0.12$ are displayed in Figure 4.6a, and Figure 4.6b. It can be deduced from Figure 4.5b, Figure4.6a and Figure4.6b that at different values of peak width, the required value of probability when the number of grid nodes N is very small has not been achieved, and the integration errors become random variables with high magnitudes, due to inadequate information about the integrand quadratic function. Furthermore, the relative error when the number of grid nodes is very small tends to be large ($i.e\ E_{rel} \to 1$), the accuracy in this case is probabilistic rather than deterministic, and achieving an acceptable level of accuracy will become a matter of chance in this case. By considering a peak as a quadratic function and the integration error as a random variable, the threshold number of grid nodes N^* required to achieve a prescribed level of accuracy has been determined, where at each regular grid with $N \geq N^*$, an adequate level of accuracy holds.

δ	N^*	h
0.03	43	$h = 0.0238 > \frac{\delta}{2}$
0.06	23	$h = 0.0455 > \frac{\delta}{2}$
0.12	12	$h = 0.0909 > \frac{\delta}{2}$

Table 4.3: The number of grid nodes required to achieve an accurate estimate of the population density given by peak function (4.1.8), at fixed $\tau = 0.25$ and several values of δ alongside values of grid step size h where $p(N) = 1$.

4.2 Evaluating the Arithmetic Mean and Probabilistic Mean Population Density on Coarse Grids

Let us now consider the standard approach to calculating the mean and standard deviation of random variables I_a given by (1.2.1), and the relative error e given by (3.1.8) on a grid of N nodes. The formulae (4.2.1), (4.2.2) give us the arithmetic mean and standard

deviation for uniform variables $I_a,\ e$ as follows [37]:

$$\hat{\mu_i} = \frac{1}{n_r}\sum_{i=1}^{n_r} y_i, \tag{4.2.1}$$

$$\hat{\sigma_i} = \sqrt{\frac{1}{n_r}\sum_{i=1}^{n_r}(y_i - \hat{\mu_i})^2}, \tag{4.2.2}$$

$$\mu_i = \sum_{i=1}^{n_r} p(y_i)y_i, \tag{4.2.3}$$

$$\sigma_i = \sqrt{\sum_{i=1}^{n_r} p(y_i)(y_i - \hat{\mu_i})^2}, \tag{4.2.4}$$

where n_r is the total number of realisations of the random variables, y_i is a random evaluation of the total population size I_a, or a random evaluation error e, $p(y_i)$ for a random y_i is the probability density function, and the subscript i can be I_a or e. According to [37], the following formulae are true:

$$|I - \mu_{Ia}(N)| \longrightarrow 0 \text{ as } N \longrightarrow \infty, \quad \mu_e \longrightarrow 0 \text{ as } N \longrightarrow \infty, \tag{4.2.5}$$

$$\sigma_{Ia} \longrightarrow 0 \text{ as } N \longrightarrow \infty, \quad \sigma_e \longrightarrow 0 \text{ as } N \longrightarrow \infty. \tag{4.2.6}$$

where μ_{Ia} is the arithmetic mean of a random variable of the total population size I_a, μ_e is the arithmetic mean of a random variable of the evaluation error e, σ_{Ia} is the standard deviation of a random variable of the total population size I_a, and σ_e is the standard deviation of a random variable of the evaluation error e. Let us now consider the quadratic function given in equation (4.1.8) to generate data over the unit interval $[0, 1]$. The width of the peak will be fixed as $\delta = 0.03, 0.06, 0.12$, and the location of peak x^* is a random variable distribute uniformly over the interval $[\delta, 1 - \delta]$, where the entire peak must be located within the unit interval $[0, 1]$. The above computations are then

repeated.

The density (4.1.8) at $\delta = 0.06$ is shown in Figure 4.5. The computation starts with a fixed number of grid nodes $N = 3$, and then the number of grid nodes increases as $x_{i+1} = x_i + h, \quad i = 1, 2, ...$ At each number of grid nodes, the computations will be repeated until the following condition holds:

$$\sigma_e \leq \tau = 0.25. \tag{4.2.7}$$

Once the data has been obtained, the arithmetic mean of approximate integrals I_a and

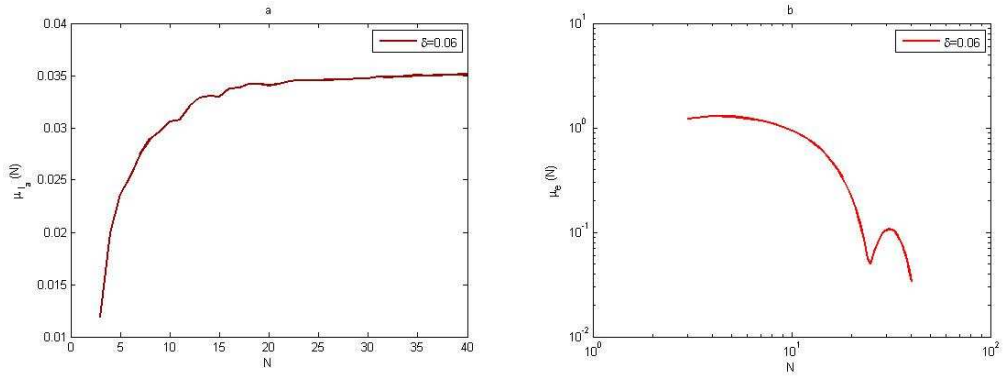

Figure 4.7: (a) The arithmetic mean for I_a at $\delta = 0.06$. (b) The arithmetic mean for e at $\delta = 0.06$.

.

numerical integration errors e will be computed by using (4.2.1), (4.2.2). It can be noted from Figure 4.7, Figure 4.8, and Figure 4.9, at each value of peak width δ, that the arithmetic mean of the approximated integral values tend to be very close to the exact value of the integral when the number of sample units is increased. Furthermore, at a narrow peak $\delta = 0.03$, the arithmetic mean converges slowly to the exact value of integral $I = 0.0045$. It also has a slightly oscillating feature, as shown in Figure 4.8a. At a wider peak width, the approximate value μ_{Ia} converges to the exact values of integrals $I = 0.0360$, and $I = 0.2880$ faster when the width of the peak increases to $\delta = 0.06$,

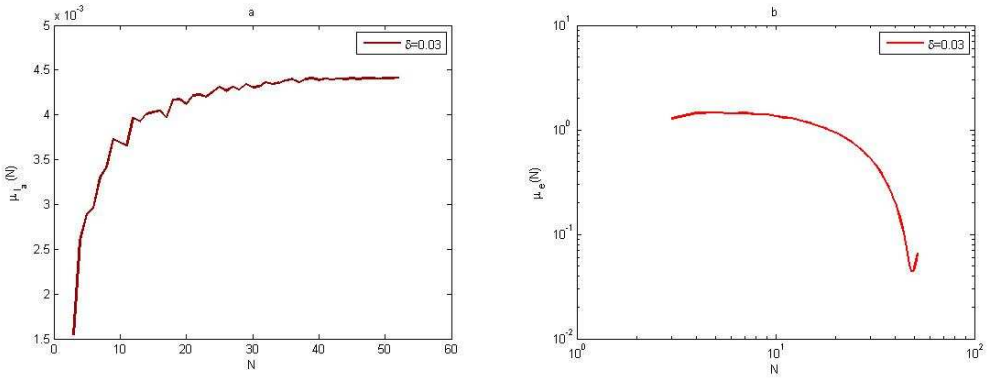

Figure 4.8: (a) The arithmetic mean for I_a at $\delta = 0.03$. (b) The arithmetic mean for e at $\delta = 0.03$.

.

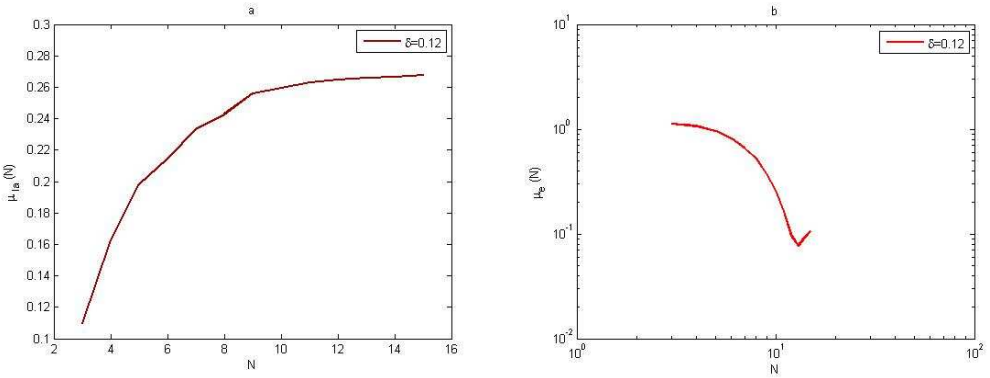

Figure 4.9: (a) The arithmetic mean for I_a at $\delta = 0.12$. (b) The arithmetic mean for e at $\delta = 0.12$.

.

and $\delta = 0.12$, as depicted in Figure 4.7a, and Figure 4.9a. It can be observed from Figure 4.7b, Figure 4.8b, and Figure 4.9b that the sample mean μ_e tends to be zero when the number of grid nodes increases, and the approximation converges faster to reach zero when the width of peak in the quadratic function becomes wider, as shown in Figure 4.7b, Figure 4.8b, and Figure 4.9b. Table 4.4 introduces the number of grid nodes required for the condition (4.2.7) to hold at different peak widths for quadratic function (4.1.8) and when the probability $p(N) = 1$. It appears from Table 4.4 that the number of grid nodes required for condition (4.2.7) to hold is slightly less than the number of grids required

for condition $p(N) = 1$ to hold. For the population density given by (4.1.8), Figure 4.10

δ	$p = 1$	$\mu_e \leq 0.25$
0.03	N^*= 43	N^* =39
0.06	N^*= 23	N^*=20
0.12	N^*= 12	N^*= 11

Table 4.4: The number of grid nodes required to achieve an accurate estimate of the peak function (4.1.8) at fixed $\tau = 0.25$ and several values of δ alongside the number of grid nodes require for the condition (4.2.7) to hold.

compares the arithmetic mean of error $\hat{\mu}_e$ and upper and lower bounds $\hat{\mu}_e \pm \hat{\sigma}_e$ with a tolerance of $\tau = 0.25$. When considering the peak function, the graphs of arithmetic mean error and bounds of error for different values of peak widths show the impact of peak width on the number of grid nodes needed to achieve the desirable accuracy $\hat{\mu}_e + \hat{\sigma}_e \leq \tau$. At the narrowest considered peak function $\delta = 0.03$, the number of grid nodes N required to fulfill the accuracy requirements is a significant $N = 41$ points, as shown in Figure 4.10b, whereas this number decays gradually while increasing the width of the peak as $\delta = 0.06$, and $\delta = 0.12$ to be $N = 21$, and $N = 11$, as shown in Figure 4.10a, and Figure 4.10c. Therefore, in order to achieve a desirable accuracy, the grid has to be refined when dealing with a quadratic function that has different peak widths, and the degree of refinement depends on the width of the peak, as demonstrated above. Since we are dealing with continuous random variables, then the probabilistic approach of computing the mean and standard deviation of variables I_a, or e given by (4.2.3), (4.2.4), will be used. Let us consider the range of variable y to be $y \in [y_i, y_{i+1}]$. The interval $[y_i, y_{i+1}]$ will be divided into M sub-intervals, where the size of each sub-interval is $h_y = (y_{i+1} - y_i)/M$. The following approximation will be considered $y_{i+1} = y_i + h_y$, for $i = 1, 2, ..., M$ and $\bar{y} = \frac{y_i + y_{i+1}}{2}$, which is the midpoint of the sub-interval $[y_i, y_{i+1}]$. Now let us consider the value of variable y that denotes the approximate value of integral I_a. The range of I_a is be $I_a \in [I_a(i), I_a(i+1)]$ and will be extracted to determine the frequencies $k_i,\ i = 1, ..., M$ of having the value I_a. Then, the probability of having I_a at each sub-interval will be

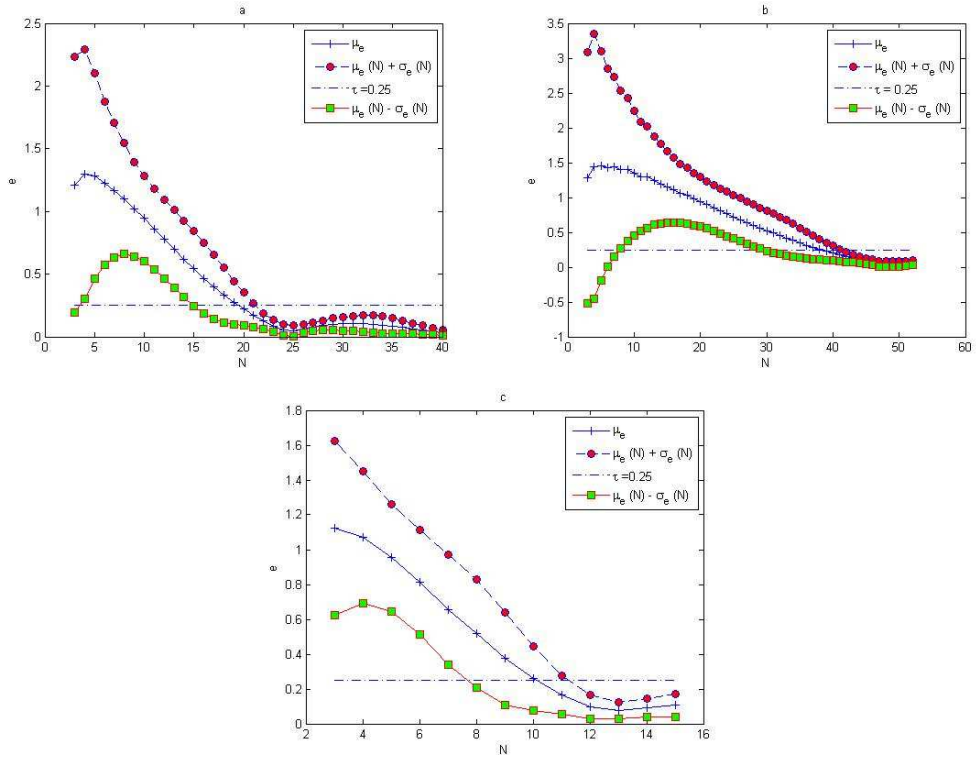

Figure 4.10: Analysis of the evaluation error for the quadratic function (4.1.8): (a) The function $\hat{\mu}_e(N)$ and $\hat{\mu}_e(N) \pm \hat{\sigma}_e(N)$ at $\delta = 0.06$. (b) The function $\hat{\mu}_e(N)$ and $\hat{\mu}_e(N) \pm \hat{\sigma}_e(N)$ at $\delta = 0.03$. (c) The function $\hat{\mu}_e(N)$ and $\hat{\mu}_e(N) \pm \hat{\sigma}_e(N)$ at $\delta = 0.12$.

computed using the following formula:

$$p_i = \frac{k_i}{n_r}, \quad i = 1, ..., M, \tag{4.2.8}$$

where $n_r = 10^4$ provides the number of random realisations. The probability of having I_a in each sub-interval with a coarse number of grid nodes $N = 3$ is very low, whereas the probability increases when dealing with fine number of grid points $N = 23$, as shown in Figure 4.11a, and Figure 4.11b. Table 4.5 shows combinations of statistical approaches used to investigate the difference between the arithmetic mean and standard deviation of I_a, and the probabilistic mean and standard deviation. From Table 4.5, it can be noted that the probability $p(N)$ at coarse grid nodes remains very small and the condition

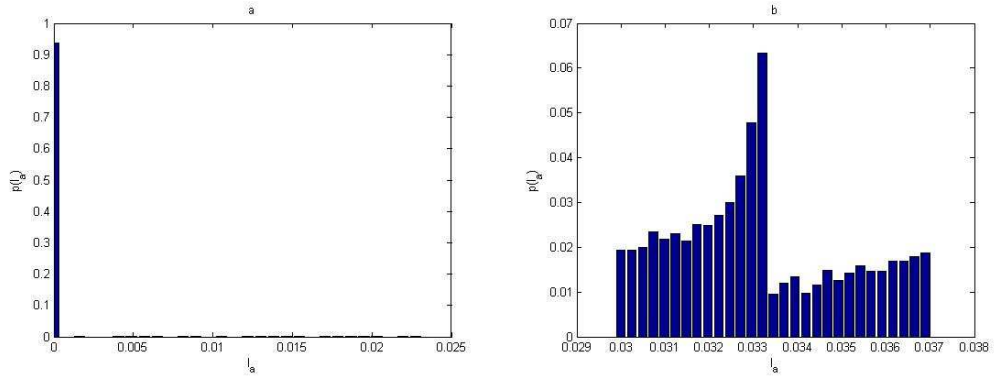

Figure 4.11: The probability of having the value I_a of sub-intervals $[I_a(i), I_a(i+1)]$ for the spatial population density distribution (4.1.8) at $\delta = 0.06$ when (a) a coarse grid of $N = 3$ points and (b) a fine grid of $N = 23$ points.

N	3	4	5	...	20	21	22	23	24
$p(N)$	0.0019	0.0067	0.0129	...	0.7151	0.7292	0.9877	1	1
$\hat{\mu}_{Ia}$	0.0133	0.0196	0.0239	...	0.0341	0.0342	0.0343	0.0344	0.0345
μ_{Ia}	0.0176	0.0225	0.0261	...	0.0342	0.0343	0.0344	0.0359	0.0360
$\lvert I - \mu_{Ia}\rvert$	0.0184	0.0135	0.0099	...	0.0018	0.0017	0.0016	0.0001	0
σ_{Ia}	0.2257	0.2249	0.2104	...	0.0376	0.0286	0.0211	0.0167	0.0102

Table 4.5: The probability $p(N)$, the arithmetic mean $\hat{\mu}_{Ia}$, the probabilistic mean μ_{Ia}, the difference $|I - \mu_{Ia}|$, and the probabilistic standard deviation σ_{Ia} computed on a sequence of regular computational grid nodes for quadratic function (4.1.8) at $\delta = 0.06$, $A = 1000$, and $I = 0.0360$.

$p(N) = 1$ is achieved when intensively refining the grid nodes. The threshold number of grid points required is $N^* = 23$, and at any number of grid nodes less than N^*, the condition $p(N) = 1$ does not hold. The arithmetic mean $\hat{\mu}_{Ia}$ converges to the exact size of population density I, as predicted in (4.2.5), but the probabilistic mean μ_{Ia} converges slightly quicker than $\hat{\mu}_{Ia}$ to the exact size I, as it accounts for random variables I_a, as shown in Table 4.5. It is known from (4.2.6) that the probabilistic standard deviation of I_a tends to be zero while $N \longrightarrow \infty$, as illustrated in Table 4.5. Furthermore, the difference between the exact value of the integral and the probabilistic mean of the integral tends to be zero when the number of grid nodes increases incrementally. Therefore, from these results the following conclusion is true: the probabilistic approach is more effective than

the deterministic approach when dealing with continuous random variables.

4.3 The Probability of Accurate Evaluation at Random Choice of the First Grid Node x_1 on a Sampling Grid

In previous sections, the probability of obtaining an accurate estimate was computed for regular computation grids and a random location of the maximum peak. The probability will now be computed for a quadratic peak function at a fixed location of the maximum peak, and the location x_1 on the uniform computational grid will be moved randomly within sub-domain $[a, h]$, where h is the grid step size on a grid of N points, and a is the left-hand endpoint of interval $[a, b]$. Let us consider the quadratic peak function given

δ	$p = 1$ at fixed x^*
0.03	N^*= 43
0.06	N^*= 22
0.12	N^*= 12
0.18	N^*= 9
0.24	N^*= 7

Table 4.6: The number of grid nodes required to achieve an accurate estimate of the peak function (4.1.8) at fixed $\tau = 0.25$ and several values of δ alongside the number of grid nodes required to hold the condition $p(N) = 1$.

by (4.1.8) as the first test case for computing the probability of achieving a sufficiently accurate estimate, as demonstrated in the previous section. The width of the peak is fixed as $\delta = 0.06$, and the location of peak x^* is fixed as $x^* = 0.672$ within the unit interval $[0, 1]$. Now the location of grid node x_1 is randomly moved over interval $[0, h]$, where $h = \frac{1}{N-1}$ is the grid step size. The grid nodes must be stationed within the unit interval $[0, 1]$; therefore, the last grid node x_N will be excluded. We provide $n_r = 10^4$ realisations of the random variable x_1 on a grid with a fixed location of the maximum peak at $x^* = 0.672$.

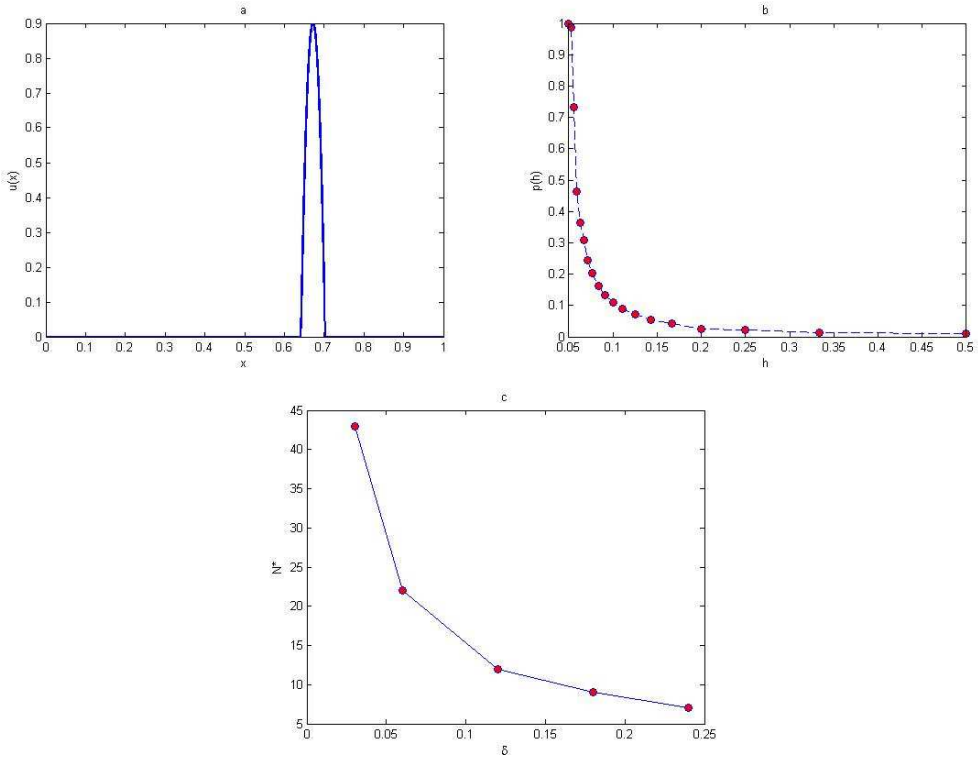

Figure 4.12: (a) The quadratic peak function (4.1.8) at $\delta = 0.06$ and $x^* = 0.672$ (b) The probability $p(h)$ obtained by direct computation for the function (4.1.8) at $\delta = 0.06$ and $x^* = 0.672$ (c) The number of grid nodes required for the condition $p(N) = 1$ to hold for the function (4.1.8) at a fixed location of $x^* = 0.672$ and different values of peak widths δ.

Then, the value of integral will be estimated by applying statistical method (1.2.1) at each realisation of x_1. The integration error will be computed by using the relative error (3.1.8). By fixing the tolerance $\tau = 0.25$, the probability of achieving an accurate estimate at different values of grid node x_1 is calculated by equation (4.1.4), where in the probability equation $\hat{n_r}$ represents the number of realisations where the integration error is $E_{rel} \leq \tau$. The quadratic peak function (4.1.8) with $x^* = 0.672$ is depicted in Figure 4.12a. Figure 4.12b shows the curve $p(h)$ for the probability of obtaining an accurate answer at peak width $\delta = 0.06$. From this figure, it can be seen that at a very small number of grid nodes, the probability is $p(h) < 1$ and the integration error is a random variable with the high magnitude. The probability becomes equal to one at a critical number of grid nodes

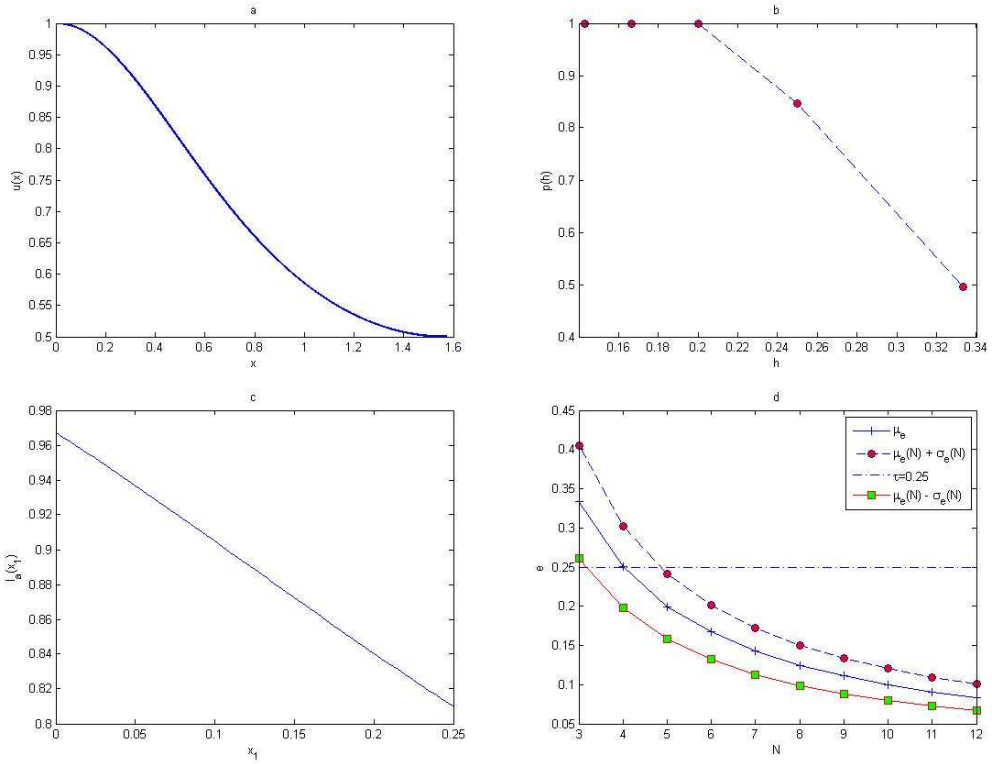

Figure 4.13: (a) The spatial population density distribution (4.3.1). (b) The probability $p(h)$ obtained by direct computation for the function (4.3.1). (c) The function $I_a(x_1)$ generated on a uniform grid of $N = 3$ nodes. (d) Analysis of the evaluation error for the spatial density distribution (4.3.1), where the mean of errors μ_e, $\mu_e(N) + \sigma_e(N)$ and $\mu_e(N) - \sigma_e(N)$ are computed on a sequence of regular grids.

$N^* = 22$. This required critical number of grid nodes N^* increases as the value of the peak width decreases (*i.e* a narrow peak) and it decreases as peak width increases.

Table 4.6 shows the number of grid nodes required to achieve an accurate estimate of the quadratic peak function at different values of peak widths δ. The numbers of grid nodes required for the condition $p(N) = 1$ to hold for the quadratic peak function (4.1.8) with peak widths $\delta = 0.03, 0.06, 0.12, 0.18, 0.24$, are displayed in Figure 4.12c. It can be deduced from Figure 4.12c that at different values of peak width, the required condition $p(N) = 1$ does not hold when the number of grid nodes N is very small, and the integration errors become random variables with high magnitudes, due to inadequate information about the integrand quadratic peak function. Furthermore, the relative error

when the number of grid nodes is very small tends to be quite large ($i.e\ E_{rel} \to 1$), so the accuracy in this case is probabilistic rather than deterministic, and achieving an acceptable level of accuracy becomes a matter of chance in this case. The findings obtained at regular computational grid nodes with random locations of the maximum peak x^* are almost identical to the results acquired at a fixed position of the maximum peak and a random location of grid node x_1, as shown in Figure 4.12b, and Figure 4.5b.

Let us now reproduce some of the results of the $1d$ test cases introduced in [134, 140] in order to compare them with the $2d$ test cases. we first consider the numerical test case that represents the $1d$ counterpart of the continuous front spatial density distribution [140] which is given by:

$$u(x) = \frac{1}{1 + sin^2 x},\ x \in [0, \pi/2]. \tag{4.3.1}$$

The function $u(x)$ shown in Figure 4.13a represents a spatiotemporal distribution spread throughout the domain from the left boundary of the region, where the population density is originally zero. The location of grid x_1 will vary with a fixed number of grid nodes N from $x_1 = 0$ to $x_1 = h$, where h is the grid step size. The exact value of the population size in the domain is given by $I = \sqrt{2}(\pi/4)$. Grid nodes must be generated within the region $[0, \pi/2]$ therefore, the last grid node x_N will be excluded. On a grid of $N = 3$ points, when the value of x_1 increases from 0 to h, the approximated total population size I_a is given by a continuous monotone function, as shown in Figure 4.13c. The range of variable $I_a(x_1)$ for $x_1 \in [0, h]$ occurs within the interval $[I_{min}, I_{max}]$, where $I_{min} = I_a(x_1 = h)$ and $I_{max} = I_a(x_1 = 0)$. Therefore, any random location of x_1 will generate a randomly approximated I_a taken from $[I_{min}, I_{max}]$ [140]. It can be seen from Figure 4.13b that at the fixed tolerance $\tau = 0.25$ and a random location of x_1, the probability $p(N) = 1$ of obtaining an accurate evaluation holds for a small number of grid nodes $N \geq N^* = 5$. Increasing or decreasing the value of tolerance τ will decrease or increase this threshold number of grid nodes N^*, as demonstrated in [140].

In the sequence of regular number N grid nodes with $n_r = 10^4$ random locations of x_1, the mean of variable error $\mu_e(N)$ and the standard deviation of variable error $\sigma_e(N)$ will be computed from equations (4.2.3), (4.2.4). The upper $\mu_e(N) + \sigma_e(N)$ and lower $\mu_e(N) - \sigma_e(N)$ bounds of the error will be computed in order to compare them with the tolerance τ. In order to approach the required level of accuracy, the following condition must hold:

$$\mu_e(N) + \sigma_e(N) \leq \tau. \tag{4.3.2}$$

For function (4.3.1), the condition (4.3.2) first holds at the number of grid nodes $N = 5$, as shown in Figure 4.13d. The results obtained from Figure 4.13d prove that a good accuracy can be acquired with a coarse grid of $N = 5$, where the upper bound of error falls below the tolerance $\tau = 0.25$. Therefore, for the distribution given by (4.3.1), at any grid node $N \geq 5$, the population density is considered to be a uniform distribution and the random variable e at any realisation will be within a desirable level of accuracy. Furthermore, I_a provides a good evaluation of the total population density for any realisation on a grid of $N \geq 5$ points.

Consider now a function with a more sophisticated pattern that represents the $1d$ counterpart to the ecological phenomenon of highly aggregated spatial distribution [78, 104]. Let us consider a single peak function introduced in the [140] which is given by :

$$u(x) = Aexp(-(x - x^*)^2/2\delta^2), \ x \in [0, L], \tag{4.3.3}$$

where A, δ, L and x^* are selected parameters. Figure 4.14a depicts the shape of function (4.3.3) where $A = 1000$, $L = 300$, and $\delta = 0.06$.By increasing the value of x_1 continuously from 0 to h on a grid of $N = 3$ points, the approximated total population size $I_a(x_1)$ is not a continuous monotone function of x_1, as shown in Figure 4.14c, and so differs from the monotone function 4.13c for function (4.3.1). It can be seen from Figure 4.14c that the

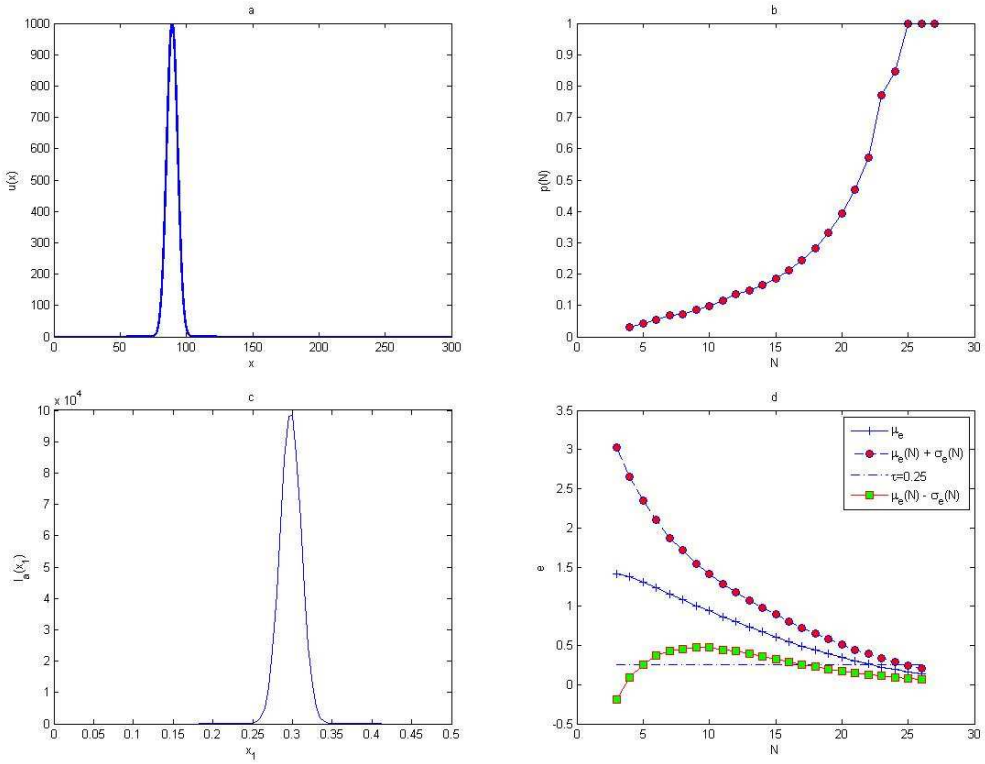

Figure 4.14: (a) The spatial population density distribution (4.3.3) at $x^* = 89.3$. (b) The probability $p(h)$ obtained by direct computation for the function (4.3.3). (c) The function $I_a(x_1)$ generated on a uniform grid of $N = 3$ nodes. (d) Analysis of the evaluation error for the spatial density distribution (4.3.3), where the mean of errors μ_e, $\mu_e(N) + \sigma_e(N)$ and $\mu_e(N) - \sigma_e(N)$ are computed on a sequence of regular grids.

range of random variable $I_a(x_1)$ is very large, and only for some realisations approximated values can provide a good approximation to the exact value of the population density $I = 1$. The accuracy can be improved by increasing the number of grid nodes to, for instance $N = 33$, to reduce the range of the random variable $I_a(x_1)$, as shown in work [140].

It can be seen from Figure 4.14b that at a fixed tolerance $\tau = 0.25$ and a random location of x_1, the probability $p(N) = 1$ of obtaining an accurate evaluation holds for a number of grid nodes $N \geq N^* = 24$, which is much larger than the number required for $p(N) = 1$ to hold for function (4.3.1). Increasing or decreasing the value of tolerance τ will decrease or increase the probability of achieving an accurate evaluation for a fixed number of nodes

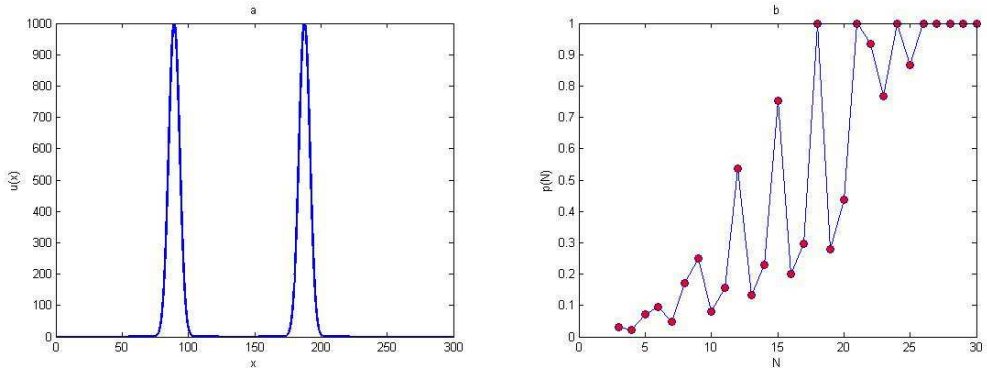

Figure 4.15: (a) The spatial population density distribution (4.3.4) at $P = 2$, where $A = 1000, \quad \delta = 0.06$. (b) The probability $p(h)$ obtained by direct computation for the function (4.3.4) at $P = 2$.

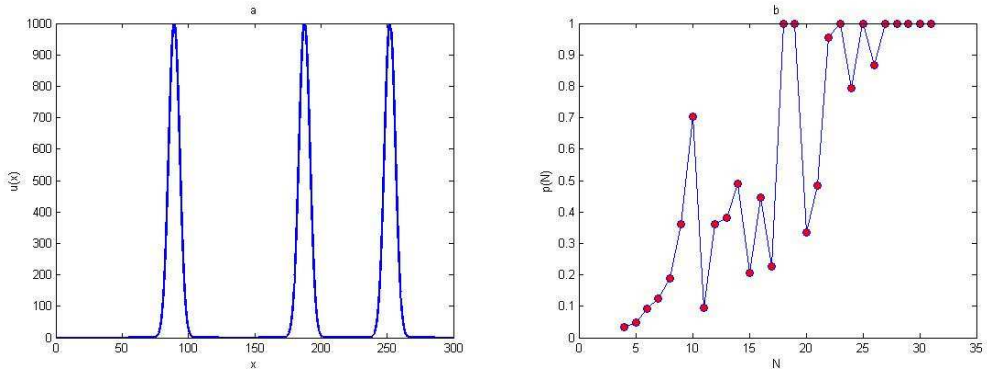

Figure 4.16: (a) The spatial population density distribution (4.3.4) at $P = 3$, where $A = 1000, \quad \delta = 0.06$. (b) The probability $p(h)$ obtained by direct computation for the function (4.3.4) at $P = 3$.

N, respectively, as demonstrated in [140]. It has been illustrated in [140] that the required threshold number of grid nodes N^* at $\tau = 0.4$ is $N^* = 22$, and at $\tau = 0.1$ is $N^* = 30$, which does not provide significant differences between them. The probability $p(N)$ of the condition $e \leq \tau$ holding depends on the shape of the population density distribution $u(x)$, as proven in [134].

The upper $\mu_e(N) + \sigma_e(N)$ and lower $\mu_e(N) - \sigma_e(N)$ bounds of the error are computed in order to compare them with the tolerance τ. In order to approach the required level of accuracy, condition (4.3.2) must hold. For the function (4.3.3), condition (4.3.2) first

holds at the number of grid nodes $N = 24$, as shown in Figure 4.14d. The results obtained from Figure 4.14d prove that we require a high number of grid nodes to provide an acceptable level of accuracy when the upper bound of the error falls below the tolerance τ. Therefore, for the distribution given by (4.3.3), the probability density is considered as a nonuniform distribution of the random variable e on coarse grid nodes. On a grid of N nodes, a significant number of realisations of the random variable e may satisfy the condition $e(N) \geq \tau$, even when the upper bound of the error is very close to the tolerance curve τ. Furthermore, significant values of random variable I_a will provide inaccurate approximations of the total population density on a grid of coarse points.

We want to extend the results in [140] and consider more computationally challenging examples. Let us now consider an exponent function with several peaks, by using the following equation:

$$u(x) = \sum_{i=1}^{P} Aexp(-(x - x_i^*)^2/2\delta^2), \ x \in [0, L], \tag{4.3.4}$$

where A, P, δ, and x^* are the chosen parameters. Let us fix the location of peaks x_i^* as $x_1^* = 89.3$, $x_2^* = 187.6$ and $x_3^* = 252.1$. At $P = 1$ in (4.3.4), the single peak function (4.3.1) will be obtained and two-peak and three-peak functions are generated by fixing $P = 2, 3$ respectively in (4.3.4). We now repeat the computations undertaken for (4.3.1) on the density distributions (4.3.4) of the two-peak and three-peak functions. The probabilities of obtaining an accurate estimation for in each case are shown in Figure 4.15b, and Figure 4.16b. The shape of the two-peak and three-peak functions are depicted in Figure 4.15a, and Figure 4.16a. It can be seen from Figure 4.15b, and Figure 4.16b that the probability curves $p(N)$ oscillate on a coarse grid nodes, and the condition $p(N) = 1$ to hold, a great deal of grid refinement is required. The probability of obtaining an accurate estimate of the total population size satisfies the condition $p(N) = 1$ at a significant number of grid

nodes, $N \geq N^* = 27$, in both cases, as shown in 4.15b, 4.16b. The results obtained from (4.3.4) support the belief that the probability p depends on the shape of the population density distribution, as demonstrated by [134], and the probability function $p(N)$ while dealing with exponent heterogeneous density distributions with more than one peak shows more sophisticated behaviour, as shown in Figure 4.15b, and Figure 4.16b.

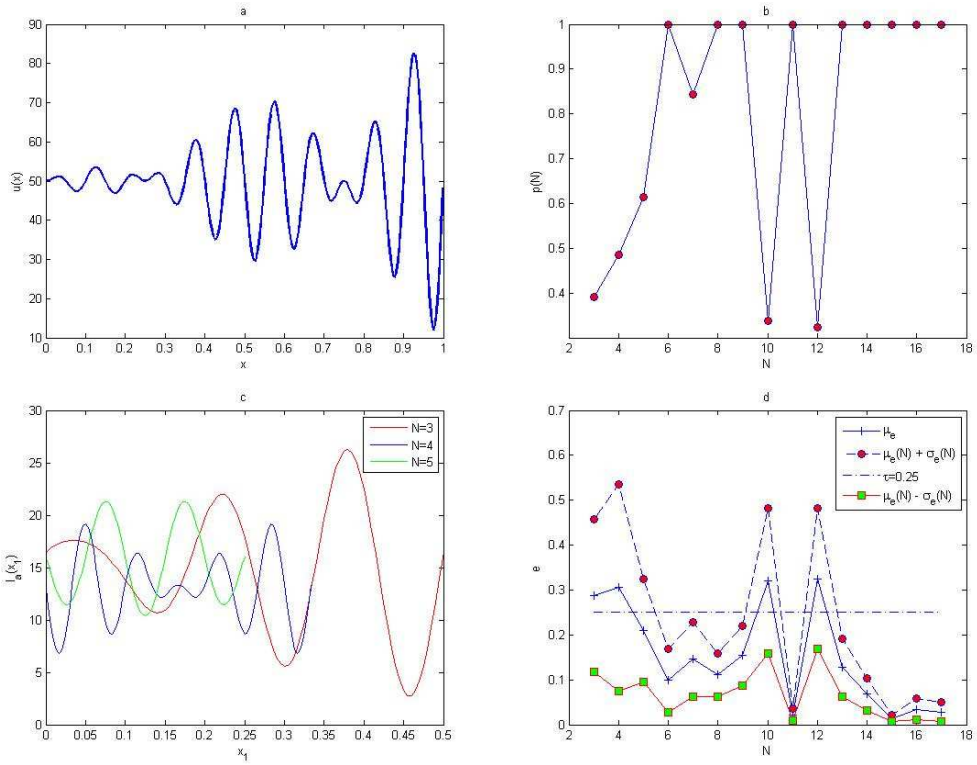

Figure 4.17: (a) The spatial population density distribution (4.3.5). (b) The probability $p(h)$ obtained by direct computation for the function (4.3.5). (c) The function $I_a(x_1)$ generated on a uniform grid of $N = 3, 4, 5$ nodes. (d) Analysis of the evaluation error for the spatial density distribution (4.3.5) where the mean of errors μ_e, $\mu_e(N) + \sigma_e(N)$ and $\mu_e(N) - \sigma_e(N)$ are computed on a sequence of regular grids.

Let us consider heterogeneous population density with multiple peaks to investigate its convergence behaviour. The $1d$ function that represents a pattern of patchy density distribution is considered as follows [140]:

$$u(x) = 4\pi^2 x sin(A\pi x) cos(B\pi x) + C, \ x \in [0, 1], \tag{4.3.5}$$

where parameters in (4.3.5) are taken as $A = 20.0,\ B = 2.0,$ and $C = 50.0$. The population density distribution given by function (4.3.5) is shown in Figure 4.17a. Changing the value of x_1 continuously from $x_1 = 0$ to $x_1 = h$ provides the range of variable $I_a(x_1)$, as shown in Figure 4.17c. It can be seen from Figure 4.17c that for an odd number of grid nodes $N = 3$, 5 the shapes of variable $I_a(x_1)$ are roughly similar which is different from the shape at even number of grid nodes $N = 4$. Therefore, we expect that the probability of obtaining an accurate estimate of population density for random variable I_a on coarse grids will show different behaviour for odd or even numbers of points. The shape of function $p(N)$ is depicted in Figure 4.17b. From Figure 4.17b, it can be noticed that $p(N)$ does not provide a monotone function, as presented in some previous test cases. Instead, it has oscillatory behaviour before hold the condition $p(N) = 1$ at critical number of grid nodes $N \geq N^* = 13$ then monotone behaviour will be acquired.

Now let us see how close the probability density distribution $p(N)$ is to the evaluation error e, by computing the mean of error μ_e and the standard deviation σ_e of random variable e on a sequence of regular grids. Figure 4.17d shows a comparison between μ_e and upper and lower bounds $\mu_e \pm \sigma_e$ for the tolerance $\tau = 0.25$. For the function with patchiness behaviour, the graph of mean error and bounds of error exhibit oscillating behaviour on coarse grids. Furthermore, as a result of increasing the number of grid nodes, the mean of error μ_e does not always guarantee an approximation close to the true population size, and the approximation of σ_e does not necessary to decrease. However, an interesting observation is that the graph of $p(N)$ shown in Figure 4.17b presents a strong form of synchronisation with the graphs of the mean of error μ_e and upper bound of error, as can be seen from Figure 4.17c, where the peaks of the probability $p(N)$ obviously correspond to troughs of the upper bound $\mu_e + \sigma_e$.

The probability $p(N)$ of obtaining an accurate evaluation will depend on the level of accuracy required, which is given by tolerance τ. It has been demonstrated in [134] that

choosing different values of tolerance τ will present different behaviours of the probability function $p(N)$ when a heterogeneous population density distribution (4.3.5) is considered. It has been shown in [134] that when the level of accuracy is low (*i.e* for a large value of tolerance $\tau = 0.4$) the critical number of grid nodes required for $p(N) = 1$ to hold is $N \geq N^* = 2$. The critical number N^* will increase as a result of increasing the level of accuracy to $\tau = 0.2$, where in this case $N^* = 12$. The probability $p(N)$ does not present a monotone function, but has oscillatory behaviour on a grid of coarse points, before the condition $p(N) = 1$ holds at $N^* = 12$. This oscillating behavior of $p(N)$ becomes more obvious when the value of required tolerance decreases to $\tau = 0.05$, where the probability $p(N)$ exhibits strongly oscillating behaviour on a coarse grid $N < N^* = 14$. Therefore, it has been proven in [134] that at $\tau = 0.4$, the evaluation procedure becomes deterministic and the required accuracy of over 40% is obtained even on a grid of coarse points. Meanwhile, this evaluation procedure is probabilistic rather than deterministic on coarse grid nodes at $\tau = 0.2, 0.05$ and a fine grid has to be used to fulfill the required level of accuracy. In the next section, we expand the results obtained in [134, 140] by considering $2d$ density distribution for the purpose of comparison, as mentioned previously.

4.4 $2d$ Probability Analysis

After analysing the probability of obtaining an accurate estimation for the $1d$ problems, our analysis is expanded to involve a more realistic $2d$ problem. Let us consider a peak function given by spatial distribution for the $2d$ counterpart of the normal distribution (4.1.2), where the function $f(x, y)$ is given by:

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-x_1^*)^2 + (y-y_1^*)^2}{2\sigma^2}),\ (x,y) \in D_u. \tag{4.4.1}$$

The domain of interest is represented by the unit square $D = [0,1] \times [0,1]$. In the equation (4.4.1), (x^*, y^*) provides information about the peak maximum and is chosen randomly. The sub-domain D_u of the peak represents a circular disc of radius R, which is centered at (x^*, y^*). The peak width is given by $\delta = 6\sigma = 2R$, which in the $1d$ distribution is fixed at $\delta = 0.06$. The $2d$ normal distribution given by(4.4.1) is depicted as a single peak in Figure 4.18.

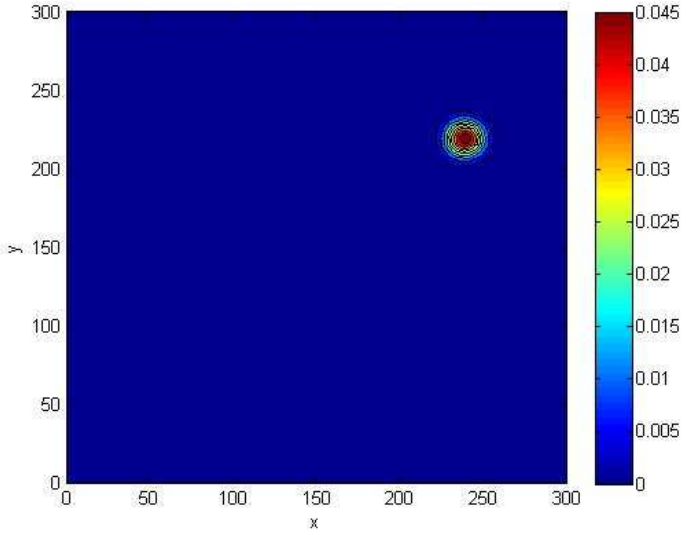

Figure 4.18: The spatial population distribution given by the $2d$ normal distribution (4.4.1) at fixed $\delta = 0.06$.

We have determined the critical number of grid nodes required to satisfy $p(N) = 1$ at $N^* = 13$ for a $1d$ normal distribution (4.1.2) with $\delta = 0.06$. The computation will now be extended to obtain N^* for a $2d$ normal distribution given by (4.4.1) with $\delta = 0.06$. Although we expect the probability graphs of $p(N)$ to be the same for both $1d$ and $2d$ problems, the results obtained for the $2d$ normal distribution are shifted from the probability graph $p(N)$ for the $1d$ normal distribution (4.1.2), as depicted in Figure 4.19a, where the critical number needed to gain $p(N) = 1$ is $N^*_{2d} = 24$, as can also be seen from Figure 4.19a. Table 4.7 shows how the critical number N^* depends on the values of the

chosen tolerances τ and σ in both $1d$ and $2d$ normal distributions. Furthermore, Table 4.7 supports the previous observation that the $1d$ probability graph for normal distribution shifts in the case of a $2d$ problem.

σ	$\tau = 0.25$	$\tau = 0.4$
4	N^*_{1d}= 25, N^*_{2d}= 47	N^*_{1d}= 22, N^*_{2d}= 41
6	N^*_{1d}= 17, N^*_{2d}= 31	N^*_{1d}= 15, N^*_{2d}= 28
8	N^*_{1d}= 13, N^*_{2d}= 24	N^*_{1d}= 11, N^*_{2d}= 21
10	N^*_{1d}= 11, N^*_{2d}= 20	N^*_{1d}= 9, N^*_{2d}= 18
12	N^*_{1d}= 9, N^*_{2d}= 16	N^*_{1d}= 8, N^*_{2d}= 15
15	N^*_{1d}= 8, N^*_{2d}= 13	N^*_{1d}= 6, N^*_{2d}= 12

Table 4.7: The number of grid nodes required to achieve an accurate estimate of the normal distribution functions (4.1.2) and (4.4.1) at several values of $\tau = 0.25, 0.4$ and several values of σ, alongside the number of grid nodes required for the condition $p(N) = 1$ to hold for both $1d$ and $2d$ problems.

From Figure 4.19, it can readily be noticed that the width of the peak and the value of

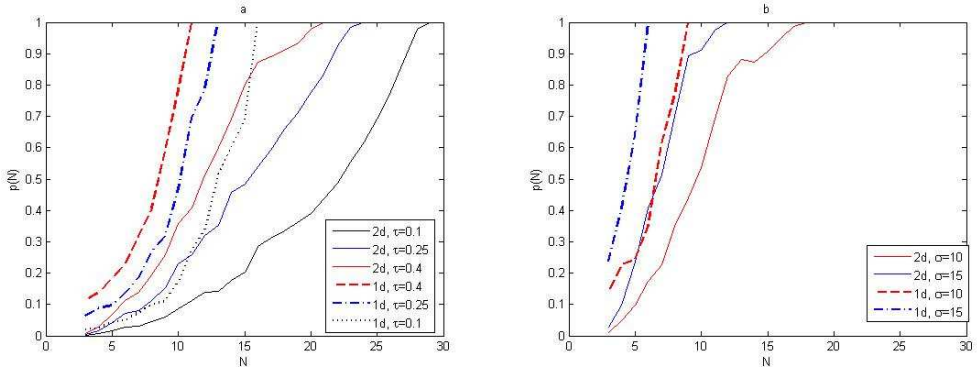

Figure 4.19: (a) The probability $p(N)$ obtained by direct computation for functions (4.1.2) and (4.4.1) at fixed $\delta = 0.06$ and different values of tolerance τ= 0.1, 0.25, 0.4. (b)The probability $p(N)$ obtained by direct computation for functions (4.1.2) and (4.4.1) at fixed $\tau = 0.25$ and different values of σ = 10, 15.

tolerance τ play important roles in deciding the critical number of grid nodes N^* required to satisfy $p(N) = 1$ in both dimensions for normal distributions. However, this critical number is higher in the case of $2d$ normal distributions, and it increases as the values of tolerances τ and σ decrease, as shown in Figure 4.19 and Table 4.7.

We now consider the superposition of $2d$ normal distributions to be given by:

$$u(x,y) = \frac{A}{4\pi\sigma^2}\sum_{j=1}^{P} exp(-\frac{(x-x_j^*)^2+(y-y_j^*)^2}{4\sigma^2}),\ (x,y)\in D_u, \tag{4.4.2}$$

where (x^*, y^*) denotes the random location of peaks, and P is the number of peaks. Let us first consider $P = 2$, then the spatial population distribution density appears as two peaks, as shown in Figure 4.20a.

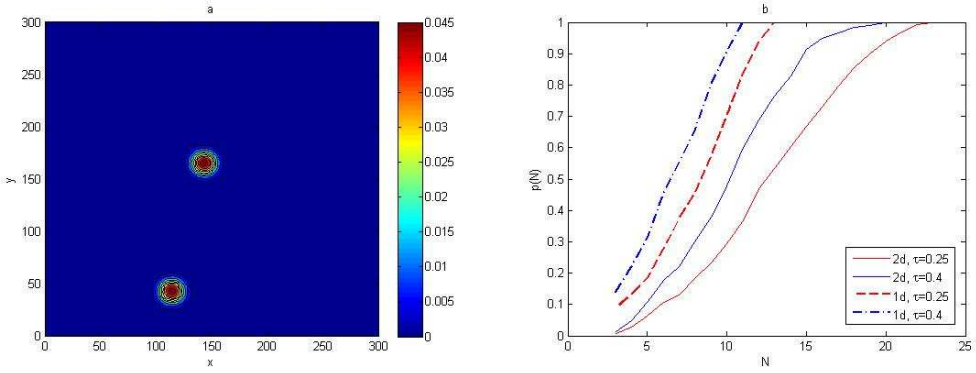

Figure 4.20: (a)The spatial population distribution given by a $2d$ normal distribution (4.4.2) at fixed $\delta = 0.06$ and $P = 2$. (b)The probability $p(N)$ obtained by direct computation for functions (4.4.2) and (4.1.5) at fixed $\delta = 0.06$ and different values of τ= 0.25, 0.4.

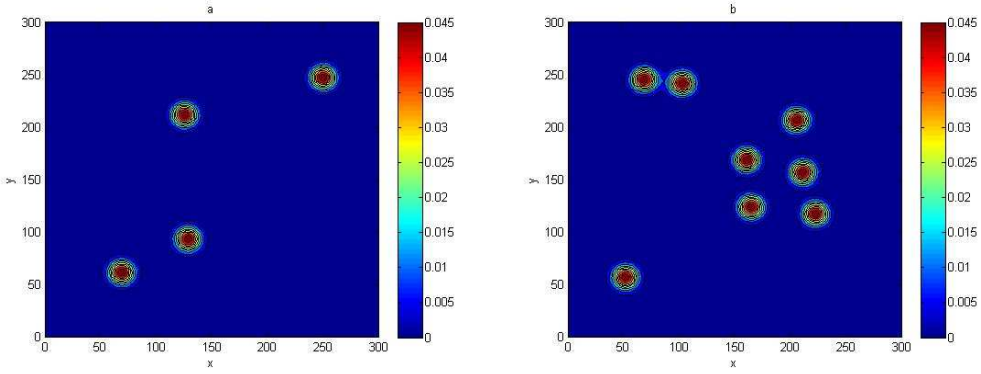

Figure 4.21: (a)The spatial population distribution given by a $2d$ normal distribution (4.4.2) at fixed $\delta = 0.06$ and $P = 4$. (b)The spatial population distribution given by a $2d$ normal distribution (4.4.2) at fixed $\delta = 0.06$ and $P = 8$.

It can be seen in Figure 4.20b that the value of critical number N^* for the condition

$p(N) = 1$ to hold for a $2d$ normal distribution of two patches $P = 2$ is similar to the value for one patch $P = 1$, where it is equal to N^*=24, for τ=0.25, and $N^* = 21$ for $\tau = 0.4$. Furthermore, the $1d$ normal distribution required fewer grid nodes to reach $p(N) = 1$ than in the $2d$ case, as depicted in 4.20b. The investigation was then expanded to involve $2d$ normal distributions with four and eight peaks $P = 4$, and $P = 8$, as shown in Figure 4.21. It can be seen from Figure 4.22a that the probability graph $p(N)$ of $1d$

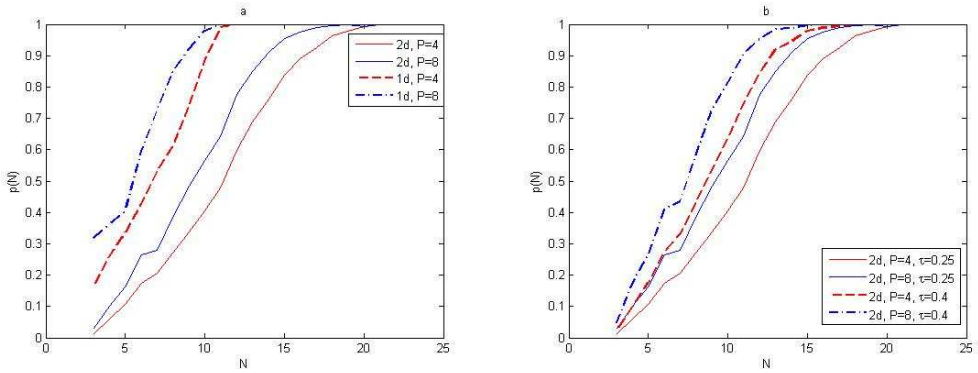

Figure 4.22: (a)The probability $p(N)$ obtained by direct computation for functions (4.4.2) and (4.1.5) at fixed $\delta = 0.06$ and fixed tolerance $\tau = 0.25$, with different numbers of patches P= 4, 8. (b) The probability $p(N)$ obtained by direct computation for function (4.4.2) at fixed $\delta = 0.06$, with different values of τ= 0.25, 0.4 and different numbers of patches $P = 4$, 8.

and $2d$ normal distributions at fixed peak width δ and fixed tolerance $\tau = 0.25$ has been computed for different numbers of peaks. By inspecting Figure 4.22a, it can be readily seen that the critical number N^* required to have $p(N)$ depends clearly on the number of peaks, where this number at $P = 4$ is bigger than N^* at $P = 8$. Moreover, the critical number needed for a $1d$ superposition of normal distribution is obviously lower than for the $2d$ problem. By varying the value of tolerance τ and number of peaks P for a $2d$ normal distribution with fixed $\delta = 0.06$, it can be seen from Figure 4.22b that a lower value of $\tau = 0.25$ with $P = 4$, requires more grid nodes $N^* = 23$ to achieve $p(N) = 1$, and this number decreases at $P = 8$, to be $N^* = 21$. When the degree of accuracy required decreases at $\tau = 0.4$, we expect that the critical number N^* will decrease accordingly, as

shown in Figure 4.22b, where it is equal to N^*=20, for $P = 4$ and $N^* = 17$ for $P = 8$.

4.5 $2d$ Probability Analysis at Random Locations of First Grid Node (x_1, y_1): Ecological Example

Let us now investigate the probability of obtaining an accurate evaluation of $2d$ ecological data generated from a system of diffusion-reaction equations given by (2.2.10), (2.2.11). Consider three spatial population density distributions in a squire domain $[0, 1] \times [0, 1]$, shown in Figures 2.7a, Figure 2.7c and Figure 4.23, where the spatial distributions shown in Figure 2.7a and Figure 2.7c represent the continuous front and patchy invasion density distributions generated in Chapter 2 from the system of equations (2.2.10), (2.2.11). The spatial distribution depicted in Figure 4.23 introduces a highly aggregated pattern of spatial distribution, which was obtained from equations (2.2.10), (2.2.11) for $m = 0.414$ and $t = 450$, where the initial conditions and other parameters were the same as in Figure 2.7c. The exact value of the population density is obtained from (1.2.1) at the finest

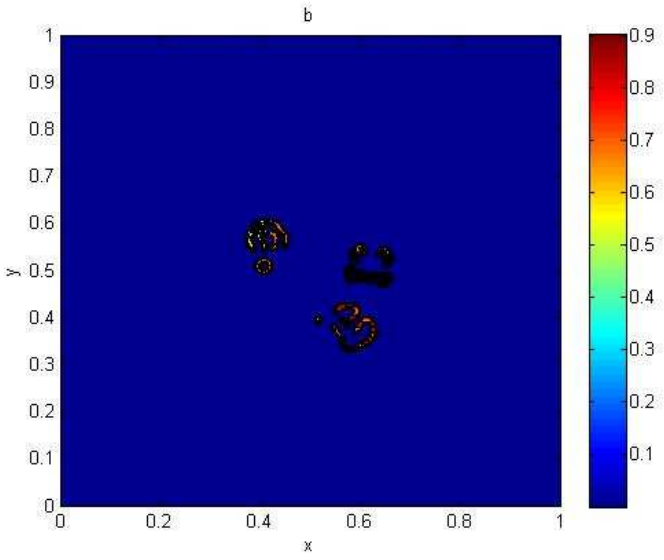

Figure 4.23: A highly aggregated spatial density distribution.

grid nodes $N = 1025 \times 1025$. Let us now consider a simulation procedure for sampling

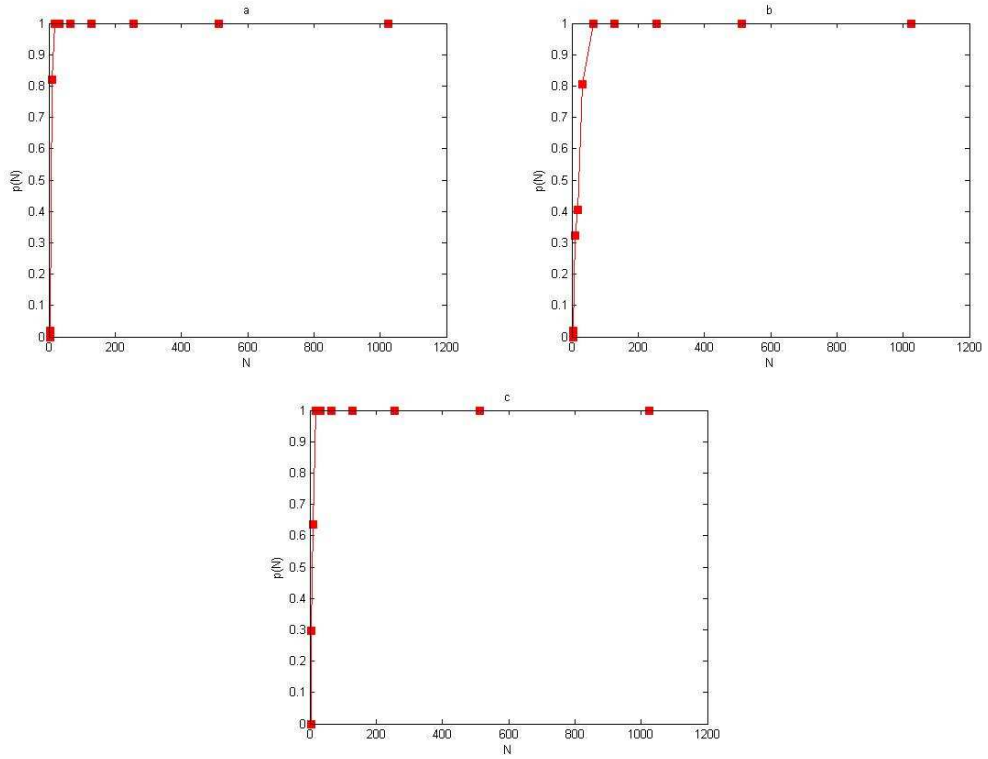

Figure 4.24: (a)The probability $p(N)$ obtained by direct computation for the continuous front spatial density distribution, with random locations of (x_1, y_1) at fixed tolerance $\tau = 0.25$. (b) The probability $p(N)$ obtained by direct computation for the highly aggregated spatial density distribution, with random locations of (x_1, y_1) at fixed tolerance $\tau = 0.25$. (c) The probability $p(N)$ obtained by direct computation for the patchy invasion spatial density distribution, with random locations of (x_1, y_1) at fixed tolerance $\tau = 0.25$.

in domain $D = [0, 1] \times [0, 1]$. The values of the population density are taken from the nodes of the regular grid sampling procedure, as taking sampling at the nodes of a regular grid is a common situation in ecological applications [50, 78]. The computational grid nodes in the x-direction, $x_i,\ i = 1, ..., N_1$ over interval $[0, L]$ are generated as follows: $x_{i+1} = x_i + h_1,\ i = 1, ..., N_1 - 1$, where we require that $x_1 = a > 0,\ x_{N_1} = b < L$ where L is the linear size of the domain and the grid step size $h_1 = (b - a)/N_1$. Let us now consider a set of points in the y-direction, $y_j,\ j = 1, ..., N_2$ over interval $[0, L]$, generated as $y_{j+1} = y_j + h_2,\ j = 1, ..., N_2 - 1$, where the grid step size $h_2 = (d - c)/N_2$ for some $0 < c < d < L$. We require the location of the new grid node x_1 to be suffi-

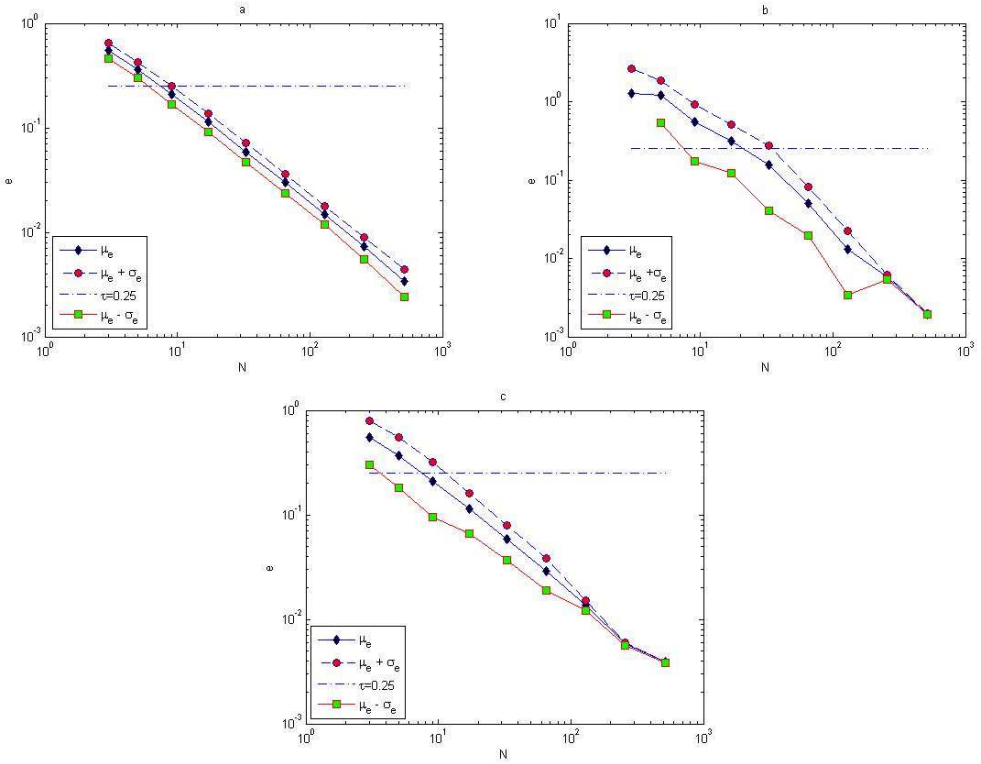

Figure 4.25: Analysis of the evaluation error for $2d$ spatial density distributions where the mean of errors μ_e, $\mu_e(N)+\sigma_e(N)$ and $\mu_e(N)-\sigma_e(N)$ are computed on a sequence of regular grids.(a) For a continuous front spatial density distribution. (b) For a highly aggregated spatial density distribution. (c) For a patchy invasion spatial density distribution.

ciently close to the original grid node $x_1 = 0$ of the domain D therefore, the conditions $0 < x_1 < h_1,\ 0 < y_1 < h_2$ must hold to make sure that the simulated sampling corner grid is located close to the corner in the domain D.

In this chapter, it has been shown that the probability of obtaining an accurate evaluation can be improved by choosing a random location of grid x_1 in $1d$ cases. Now, we will extend this investigation to involve the $2d$ ecological test cases introduced previously. The location of grid node (x_1, y_1) is randomly moved over the domain $[0, h_1] \times [0, h_2]$, where $h_1 = h_2 = \frac{1}{N-1}$, and $N = N_1 = N_2$ are the grid step size and the number of grid nodes in each direction respectively. The grid nodes must be generated within the unit domain $[0, 1] \times [0, 1]$; therefore, the last grid node (x_N, y_N) will be excluded. We provide $n_r = 10^4$

realisations of the random variable (x_1, y_1). As we are dealing with simulated ecological data, and the density is only known at grid points, we therefore need to find the closest node to the new grid node (x_1, y_1) from the original computational grid nodes in the domain $D = [0, 1] \times [0, 1]$. Once the data for the population density is extracted at the new first grid node (x_1, y_1), the value of the integral is estimated by applying the statistical method (1.2.1) at each realisation of (x_1, y_1). The integration error is computed by using the relative error (3.1.8). By fixing the tolerance $\tau = 0.25$, the probability of achieving an accurate estimate with $n_r = 10^4$ random locations of (x_1, y_1) will be calculated from equation (4.1.4). The computation starts with a coarse number of grid nodes $N = 3$. Then the number of grid nodes is doubled N= 3, 5, 9, 17, 33, 65, 129, 257, 513; and the above computations is repeated.

It has been shown in work [140] that, considering the bottom-left corner point of the sampling grid depicted in Figure 4.23 to be $A = (a, c)$, a very accurate evaluation of the integration error $e \approx 10^{-2}$ is provided for the continuous front population density distribution. However, an inaccurate estimation of the integration error for the same sampling grid is provided for the patchy invasion. The integration error becomes much greater for a highly aggregated distribution. The degree of accuracy decreases dramatically from $e \approx 1$ to $e \approx 10^{-3}$ for the patchy invasion and the highly aggregated distribution respectively, when the location of grid A is moved slightly within the limited domain given above. However, the evaluation of the continuous front distribution does not change with this new sampling grid, as it deals with a smooth function. Remarkably, choosing a random grid location (x_1, y_1) improved the degree of accuracy when dealing with all distributions of zero and non-zero patches. This procedure enables ecologists to increase the probability of catching high density value to gather these values with zero values, to eventually give us an accurate estimate of the population density.

The number of points required to achieve the condition $p(N) = 1$ depends on the level

of complexity within the ecological data. For the simplest pattern of ecological data that shown in Figure 2.7a, the number of grid nodes required is $N^* = 9$ in each direction to hold the probability condition, and this number is slightly increased to become $N^* = 17$ in each direction for ecological distributions 4.23 and 2.7c respectively, as depicted in Figure 4.24. A grid with a number of nodes below the value $N < N^*$ represent an ultra-coarse grid, where the evaluation of the integration error is probabilistic and is a matter of chance. Once the required number of grid nodes N^* satisfies the probability condition, the integration error becomes deterministic and the asymptotic error (3.1.10) holds for a fine number of grid nodes.

In the sequence of regular number N of grid nodes with $n_r = 10^4$ random locations of (x_1, y_1), the mean of variable error $\mu_e(N)$ and the standard deviation of variable error $\sigma_e(N)$ can be computed from equations (4.2.3), (4.2.4). The upper $\mu_e(N) + \sigma_e(N)$ and lower $\mu_e(N) - \sigma_e(N)$ bounds of error are computed to be compared with the tolerance τ. In order to achieve the required level of accuracy, the condition (4.3.2) must hold. It can be seen from Figure 4.25a, and Figure 4.25c that the condition (4.3.2) holds at $N = 9 \times 9$ and $N = 17 \times 17$ for the continuous front and the patchy invasion distributions, which are in fact the same number of grid nodes required to hold the probability condition $p(N) = 1$. For the highly aggregated distribution, the number of grid nodes required for the condition (4.3.2) to hold is larger than that required to achieve the probability condition to be $N = 33 \times 33$, due to the complicated pattern of distribution. It can be concluded from this that, for ultra-coarse grid nodes, it is not possible to provide the information required to provide a reliable degree of accuracy, therefore, the integration error will be probabilistic rather than deterministic. Hence, the probabilistic approach is used to estimate the size of population density for each ecological distribution.

4.6 Chapter 4 Conclusion

A simple evaluation of the population density, provides a deterministic level of accuracy when information about population density is sufficient. When dealing with a grid with a very small number of points, the information required to evaluate the population density is insufficient, which means that it is impossible to guarantee the required accuracy. Therefore, the probabilistic approach should be applied to evaluate population abundance. The probability p of achieving a sufficiently small error is used to assess accuracy, rather than considering the error itself; therefore, achieving the desired accuracy will be a matter of chance, rather than being deterministic, by considering the integration error as a random variable. The probabilistic approach requires series of statistical computations e.g. the mean and the standard deviation, that in turn require repetitive and resource-heavy data collection techniques, which render the sampling procedure impossible.

For homogenous density distributions, or density distributions slowly changing in space, any single realisation of a random variable obtained from the sampling procedure may possibly provide an accurate estimate of the true population density. In this case, any individual realisation can be considered to be a desired approximation of the population density, and the mean value will be close to the true population size. Moreover, at any realisation, the standard deviation value will tend to be small. However, when dealing with heterogeneous population density, the above observations are incorrect. It has been shown in this chapter that there is a significant difference between a single realisation and the true population size in heterogeneous distributions. The probability of producing an accurate evaluation of population density on ultra-coarse grids is low and is a matter of chance.

On ultra-coarse grids, the conventional approach of convergence analysis do not work, as it does not distinguish between the performance of different numerical integration methods.

Therefore, the probability approach has been considered as a measurement of accuracy instead of conventional convergence analysis. A desirable level of accuracy is achieved if the condition $p(N) = 1$ holds, where $p(N)$ is the probability of obtaining $e \leq \tau$. In order to achieve the above condition, the threshold number N^* of grid nodes is required, where $p(N^*) = 1$. Determining the threshold number of grid nodes N^* is an important process in order to utilise the probabilistic approach which depends on computations of random variables with the deterministic approach, which is based on the results obtained from the sampling procedure.

As a result of the above conclusions, the classification of computational grid nodes can be redefined according to the threshold number of grid point N^*. For a spatial pattern of population density distributions, a grid is considered ultra-coarse if the number of grid nodes $N < N^*$, the transmission from ultra-coarse to coarse grids will be at $N = N^*$ and the error will be deterministic. Finally, a fine grid is defined as a grid in which $N > N^*$ and the asymptotic error (3.1.10) always holds. Thus, reliable approaches to evaluating the threshold number N^* must be determined, especially as we are dealing with unknown features of spatial density distribution. In the numerical experiments, it has been proven that the estimation of the threshold number N^* is a reliable means of determining the minimum number of grid nodes required to achieve sufficient accuracy.

On coarse grids, the locations of essential features of the spatial density distribution are not known in advance, which is an intrinsic property of the problem. Therefore, this chapter has investigated the probability on a sample of coarse grid where a random location was provided for the first grid node, which reflects the lack of information about density patterns with several humps, patches or nonzero population densities. To deal with this uncertainty, the threshold number N^* for reliable spatial distribution data, available in real-world settings, must be determined.

The analysis of the probabilistic approach was expanded to involve $2d$ problems. The

$2d$ form of the normal distribution, with a varying number of patches and several values for the chosen tolerance, are considered with the purpose of comparing to the $1d$ problem. The critical number N^* required to achieve an accurate estimate in $1d$ cases was considerably lower than the number required in $2d$ cases for the condition $p(N) = 1$ to hold. Furthermore, it has been demonstrated that the value of the critical number N^* depends on the number of patches, where for a small number of patches, N^* increases, but it decreases gradually when the number of patches P increases from $P = 1$ to $P = 8$. The width of each patch and the values of the chosen tolerance τ obviously affect the required value of the critical number N^*. It has been explained in this chapter that with a wide patch, N^* becomes smaller than N^* at narrow patch. The required number N^* relies on the value of the chosen tolerance τ, where at a low level of accuracy ($i.e$ at a large value of τ), N^* is quite small; this number increases when greater precision is required. The process of slightly altering the position of the bottom left corner grid node (x_1, y_1) within a limited domain $[0, h_1] \times [0, h_2]$ in the application of $2d$ ecological distributions is beneficial and highly recommended for complicated distribution pattern. In a $2d$ ecological distribution, we often do not have any prior information about the properties of the pattern and the location of the domain of non-zero density. The optimal positions of the grid nodes are difficult to discern, hence the choice of point A will be random. In other words, at each random location of grid A, the evaluation of the total population size is considered as a single realisation of the random variable I_a, and different position of A will result in different values of I_a. Let us emphasise that, the location of A is not important on a sampling of very fine grid nodes, because uncertainty about the optimal location of grid nodes no longer applies.

Chapter 5

Investigating Spatial Synchronisation in the Generation of Ecological Data

When information about spatial distribution is only available at nodes of a discrete mesh, and the number of mesh nodes required to resolve the details of the heterogeneous distribution of the population density is not large enough, a spatial distribution system that uses sparse data should be considered. In several applications in science and engineering, the problem of evaluating the properties of a spatial distribution system from coarse grid nodes is known as an inherent problem [13, 23, 163]. The properties of the spatial distribution system are usually evaluated by integrating the available sparse spatial data. The evaluation of an integral property from sparse data is common in different fields: in ecology, to evaluate the total population size of dangerous pest insects [141, 163, 176]; in geology, the total mass of valuable minerals is evaluated [23, 197]; in medicine, it is used to evaluate the total mass of tumors [48]. However, due to dealing with sparse data, the above evaluations may appear within the range of unacceptable accuracy. Low accuracy intuitively leads to wrong decisions being made about control methods and treatments.

In ecology, information about the abundance of population is used for many purposes. One of these purposes is to address synchronisation between spatial patterns [9, 69, 132, 195]. Data is routinely collected from offload samples, which are taken from across the whole domain. The accuracy of any evaluation of population density in will, in turn, depend on the nature of the samples, which itself depends on the properties of the species being evaluated [18, 136, 137, 177]. Different factors affect the precision of the local population density evaluation and the number of samples taken. This chapter will investigate how the size of samples impact on the dynamics of population density evaluation (in particular synchronisation). Hence, we consider the available local data to be precise, and the error in evaluation is caused by considering coarse grid of samples.

The location and number of samples are both considered carefully in any census or population survey. The question that needs to be answered is this: when is reliable information about the population density distributions obtained from collected data? It has been demonstrated in previous studies that obtaining reliable information depends on the nature of spatial pattern of the population density [134, 136, 137, 141]. The main objective of this chapter is to illustrate how the quality of sampling data will affect the presence or absence of synchronisation between population density in different habitats. It is essential to fully understand that population densities change over time, either upward, downward or a combination of both. Spatial synchronisation is an ecological phenomenon that appears as a result of synchronisation between changes in ecological populations at different locations. There are several approaches that can be used to measure the degree of synchronisation among population densities. Therefore, in this chapter the phenomenon of spatial synchronisation will be investigated.

5.1 The Concept of Spatial Synchronisation

It has been noticed in different locations that the population densities frequently tend to fluctuate synchronously. This phenomenon is known as spatial synchronisation or population synchronisation. In large geographical areas, the behavioural pattern of spatial synchronisation between the population densities of many species has been examined, for instance, in a study in [148, 176, 178] on butterflies, forest defoliators [106, 120, 195] and acorn crops [35].

The degree of spatial synchronisation in the dynamics of population density has a high level of importance, as it enables biologists to answer questions regarding population and conservation biology and metapopulation dynamics [53, 54, 66, 191]. According to [67, 70], in a metapopulation that represents a group of population densities for the same species spread spatially with some level of interaction, increasing the degree of spatial synchronisation may lead to decreasing the life expectancy of the metapopulation. In other words, the expected lifespan of the metapopulation declines if the degree of spatial synchronisation is extensive. In conservation biology, calculating the ideal distance between nature reserves relies on recognising how strong spatial synchronisation decays significantly with increasing distance between conspecific populations. Although the degree of spatial synchronisation for some taxa has been determined, for instance selected butterflies in the United Kingdom [147, 148, 186], information regarding the degree of spatial synchrony in the dynamics of most species populations is in general scarce.

Two factors are most likely to generate spatial synchrony: migration or dispersal among population densities, and the impact of exogenous factors or regional stochasticity [68, 107]. An illustrative example of exogenous factors is weather conditions, which frequently influence population abundance. Where spatial synchrony exists, the most difficult challenge is not to decide whether this phenomenon occurs due to dispersal or regional stochas-

ticity, but to determine which one makes a difference to the population dynamics. It was demonstrated in [81] that dispersal may lead to spatial synchronicity in population dynamics, because increasing the population density at one location will produce new emigrants that move to adjacent locations. Exogenous factors, such as changes in weather conditions over a large geographical area, may affect various population densities in the same way [71]. It was shown in [118] that populations that have the same exogenous factors produce models with the same degree of linear autoregressive spatial synchronisation.

The widespread existence of spatial synchrony due to weather conditions is predictable, because to some degree most populations are influenced by variations in weather conditions. Other exogenous factors which may generate spatial synchronisation are natural enemies, that spread between locations of prey densities. It was suggested in [121] that between various locations of population densities, virus infections may occur and serve to introduce patterns of spatial synchrony in forest tent caterpillar densities.

On the one hand, according to [68], a high rate of reoccupation facilitates metapopulation persistence in spatial synchrony occurring through significant levels of migration, which is more than pursuit to compensate the negative impacts of synchronisation in the associated local extinctions. On the other hand, metapopulation persistence in spatial synchronisation due to regional stochasticity must decline in the same way as decreasing the life expectancy of a local population, as a result of increasing environmental stochasticity [62, 92, 183]. Then, it is easy to understand how spatial synchrony is generated in population dynamics by regional stochasticity. However, under the same conditions, the generation of the same results for spatial synchronisation due to migration is less clear.

5.2 Measures of Spatial Synchronisation

The phenomenon of spatial synchrony has been assessed by applying different measurement methods to two series of population densities. It has been proposed in [1, 57, 148]

that the most common and effective ways to measure synchrony between two series of populations are the Pearson and Spearman correlations. In order to measure a linear relationship between two series of populations, the Pearson correlation is applied. However, if the relationship tends to change with nonlinear transformations of the data, the Spearman correlation is fitted.

Another way to quantify synchrony among populations of different species, or between populations with no coincident peaks (disjunctive populations), is to use cross-correlograms. In some ecological applications, there is a lag in synchrony between population densities, which occur due to either different geographical responses to exogenous factors, or as a result of dispersal influences. Cross-correlograms depend on graphs of lag correlations between series of populations plotted against lag intervals (that represent intervals which vary from negative to positive values).

The next stage is calculating correlations between residuals. The most common strategy to compute correlations between residuals is to use suitable autoregressive model either AR(1) or AR(2). Autoregressive models represent a process in which some current values in linear population patterns depend on specific previous values in this population series. In terms of log (density), some studies used AR(1) model, such as studies by [69, 178], whereas others, such as [195] relied on the AR(2) approach. The aim of applying this approach is to measure synchrony in the noise that occurs after computing for local dynamics as a result of responding to some exogenous factors, for instance weather conditions. It has demonstrated in [88, 153] that to remove linear tendency from population series, the spatial synchrony among residuals will be calculated after detrending the series. In population densities that have periodic behaviour as a function of time, with a trigonometric function, spatial synchronisation will be computed after eliminating periodicity from the series [155]. When all values go up or down together, detrending is beneficial as it increases the probability of presenting synchrony between series. In con-

trast, removing periodicity from series may negatively affect the appearance of synchrony Another method of measuring spatial synchrony depends on changes in population dynamics. Spatial synchrony occurs in large areas that experience fluctuations, due to changes in population series of population in the same direction. Some researchers [181] prefer to measure synchronisation between series of populations by computing correlation coefficients according to changes in population densities. The simple equation to illustrate how the two series are changing together is A_{ij} = (number of series time (i), and (j) move in the same way)/$(T-1)$, where A represents the same value for either density or log (density) T is the time points [17]. The values of correlations A_{ij} range between 0 and 1, which can be transformed to measure correlations from (-1) to (1) by using the following transformation: $\tau_{ij} = 2A_{ij} - 1$.

Measuring peak coincidence is another approach to measuring synchrony. It has been used in [35] peak coincidence to measure correlations in the production of acorns in England. It produced in [120] another example by comparing several species of forest Lepidoptera, at different locations, during outbreak years of species against nonoutbreak years. In order to calculate synchrony by using peak coincidence, let us consider variable N, which represents the number of times series i, and j, where both series have peaks. However, this strategy failed to compute the regularity among the series of peaks. In order to overcome these obstacles, the term N will be replaced by the term $C = \frac{N}{M}$, where M_i and M_j are the number of peaks in series i and j respectively and $M = max[M_i, M_j]$. If the series of data increase and decrease together, then $C = 1$. The previous procedure can be applied to troughs instead of peaks, and it can be also applied to both peaks and troughs at the same time. The last approach is measuring spatial synchrony by using the coefficient of variation in population densities. It has calculated in [83] the coefficient of variation over each time, and over the years the average of these coefficients was computed to introduce values of spatial synchronisation. This procedure is a poor strategy, because

the possibility of dealing with different levels of population density is not considered. It is easy to construct different sets of series where for each collection of series, there is a specific pattern of synchrony: they either increase or decrease together. The coefficient of variation can be changed; by manipulating the overall means of several series, the average coefficient of variation may be made larger or smaller arbitrarily.

5.3 Synchronisation in a Mathematical Framework

In order to reveal synchronisation between series of population densities, let us consider the general case of a certain population in domain A, which is prescribed by the density $u(x,t)$. Our domain represents the species habitat, and for convenience invertebrate species will be considered to include insects, worms and slugs. Traps are usually used as sampling process when dealing with invertebrate species, and we assume that traps counts are available at certain times t [141, 151]. In mathematical term, available information about population density is given as a set of values, $u(x_1,t),\ u(x_2,t),\ ....,\ u(x_N,t)$ where x_i represents the position of installed traps, t is the time duration in which the census was taken, and N is the number of installed traps. As illustrated previously, all this information is used to evaluate the population density in the given area.

There are several numerical integration methods that can be used to estimate the population density in a given agricultural field with different degrees of accuracy. Let us consider the statistical approach given in (1.2.1) to calculate the average density by computing the arithmetic mean $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u(x,t)$ of the sampled values [163, 177]. We will have series of time $t_1, t_2, ..t_M$, if the census is done regularly, which we can use to obtain the average of population densities at series of time $S = \bar{u}^1,\ \bar{u}^2, ...,\ \bar{u}^M$. In order to investigate the synchronisation, the sampling procedure must be applied to a number of habitats, for instance habitats A and B, to produce sampling with two time series $S_A = \bar{u}_A^1,\ \bar{u}_A^2, ...,\ \bar{u}_A^M$ and $S_B = \bar{u}_B^1,\ \bar{u}_B^2, ...,\ \bar{u}_B^M$ respectively. The correlation coefficient will be used to reveal

the synchronisation [98, 151, 153, 154, 155, 176]:

$$\rho_{AB}(M,N) = \frac{\sum_{i=1}^{M}(\bar{u}_A^i - \mu_A).(\bar{u}_B^i - \mu_B)}{\sqrt{(\sum_{i=1}^{M}(\bar{u}_A^i - \mu_A)^2)(\sum_{i=1}^{M}(\bar{u}_B^i - \mu_B)^2)}}, \tag{5.3.1}$$

where μ_A and μ_B are the sample means:

$$\mu_A = \frac{1}{M}\sum_{i=1}^{M}\bar{u}_A^i \; ; \; \mu_B = \frac{1}{M}\sum_{i=1}^{M}\bar{u}_B^i.$$

The correlation coefficient is widely used to measure the degree of linear dependance between two series of variables. The concept of the correlation coefficient was introduced in [56], and has been developed in [129]. According to statistics, the Pearson product-moment correlation coefficient is a measure of the linear correlation between two variables $\bar{u}_A$ and $\bar{u}_B$, as presented in (5.3.1). In general, the value of correlation coefficient is between $-1 < \rho < 1$, where $0 < \rho \leq 1$ indicates to correlation and $-1 \leq \rho < 0$ indicates anti-correlation. According to [52, 55], if the value of correlation coefficient equals zero there is no linear relationship between variables. When the value of one variable increases and the value of the other variable increases as well, a perfect positive linear relationship will be acquired at $\rho = +1$, whereas from the exact linear rule, a perfect negative linear relationship is achieved as a result of one variable increasing, and the other variable decreasing. According to [98, 151, 153, 154] values for the correlation coefficient of between zero and ± 0.3 refer to a weak positive or (negative) linear relationship, and values for the correlation coefficient of between (± 0.3) and (± 0.7) indicate a moderate positive or (negative) linear relationship, whereas a strong positive or (negative) linear relationship is shown by values for the correlation coefficient of between (± 0.7) and (± 1).

Depending on the value of ρ, a decision regarding the existence or absence of synchronisation can be made.

The number of samples in any sampling process plays an essential role in the degree of accuracy of the evaluation of the population density. It can be seen from equation (5.3.1)

that there is no explicit dependence on the number of samples N; however, the value of the correlation coefficient strongly depends on the number N. The correlation coefficient requires the same number of samples as are used to calculate the average of population densities $\bar{u}_A^i$ and $\bar{u}_B^i$. The methods used to calculate the average $\bar{u}$ are required to determine the number of samples N, which is an inherent problem. If the value of N is quite small, a very low degree of accuracy is expected. The properties of the population distribution are largely responsible for determining the number of samples N required to provide an accurate evaluation of population abundance [136, 137, 141]. A uniform distributed population density does not require a large number of samples N to provide an acceptable level of accuracy, (*i.e* a very good level of accuracy is acquired with few samples; in some cases, with an exactly uniform population, one sample is enough to provide a good accuracy). Meanwhile, for heterogeneous or patchy population densities, a large number of samples is required to effectively resolve details of regarding fluctuations over area, and to achieve a reasonable accuracy [136, 139]. When the number of samples taken is quite small, in other words on coarse grid nodes, the accuracy of evaluation is poor, due to insufficient information about patch distribution. As a result, a large value is expected for the numerical integration error, due big a difference between the exact population density value and evaluated value . The values of the correlation coefficient will consequently be affected, and the conclusion regarding the existence or absence of synchronisation will also be affected.

5.4 Synchronisation Analysis: Ecological Example

Population densities changes over time, as mentioned previously. In order to investigate these changes, let us consider a population density distribution generated by a numerical solution to the Rosenzweig-MacArthur system (2.2.10), (2.2.11). The aim of this analysis is to see how the strength of synchronisation changes when increasing the distance

between areas for which we have ecological data. In this section, we will show how the population properties at a fixed point x^* may become uncorrelated by uniformly moving to points $x^* + \Delta x$ and computing their correlation coefficient. Let us first numerically generate ecological data by solving the Rosenzweig-MacArthur model for specific initial and boundary conditions. The equations (2.2.10), (2.2.11) are effectively used in order to generate different sets of $1d$ ecological data. Let us consider a computational grid of $N = 8193$ points over the interval $[0, L]$, where $L = 300$ is the right endpoint. The parameters in equations (2.2.10), (2.2.11) are: $m = 0.7$, $k = 2$, $H = 0.4$, $t = 5000$, and the diffusion coefficient $d = 1$. The solution for this system is obtained over a time series t, which starts at the initial time $t = 0$ and then increases as a time series according to the following form: $t_{i+1} = t_i + dt$, where $dt = 25$ and $i = 0, ..., k$, $k = 199$. We will record 200 solutions to the system (2.2.10), (2.2.11) for time series $S = (t_1, ...t_{200})$, where each solution represents the population densities at 8193 uniform grid points. The solution values at $x = 150$, $u_i(x = 150)$ where $i = 1, ..., 200$, will be taken. The solutions at moving points from the basic fixed point $x^* = 150$ will be extracted as follows:

$$x_{\tilde{i}} = x^* + \Delta x \text{ where } \tilde{i} = i + \Delta_i \text{ , } i = 4097, ..., 8192, \text{ and } \Delta_i = 1.$$

This will be uniformly increased each time. The extracted elements will be combined in one file that represents the solutions to the system (2.2.10), (2.2.11) at point $x^* + \Delta x$. Then, form (5.3.1) is applied to compute the correlation coefficient between the basic array at $u(x^*)$ and the array at $u(x^* + \Delta x)$. By increasing the value of Δ_x, we will have a newly extracted array of solutions at a new point $u(x^* + 2\Delta x)$. The correlation coefficient will be computed now between the basic array of solutions at $u(x^*)$ and the newly extracted array of solutions at the new point $u(x^* + 2\Delta x)$. This procedure will be repeated many times, until we have $u(x^* + 4096\Delta x)$. Then, all values of the correlation coefficient will be brought together in one file to draw a graph of the correlation coefficient and investigate the synchrony in this pattern of population densities. A graph of the correlation coefficient

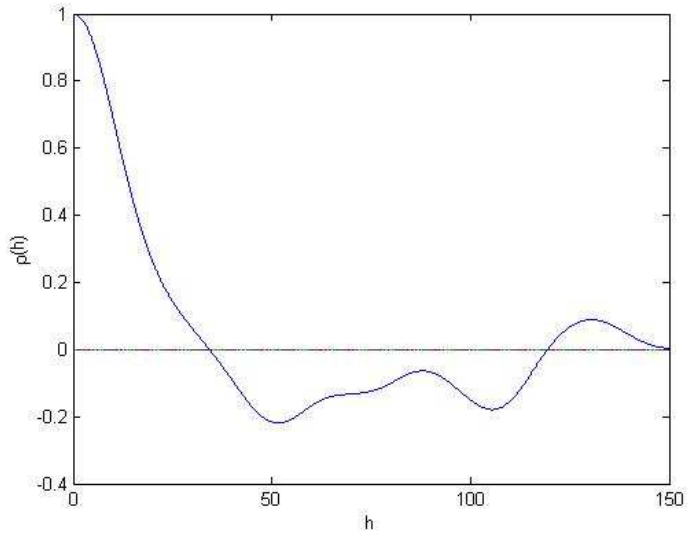

Figure 5.1: The correlation coefficient of the population density between fixed point x^* and moving points $x^* + i\Delta x$, where $i = 1, ..., 4096$.

is plotted against distance h to show how correlation coefficient values decrease with increasing distance between points, as explained above. It can be seen from Figure 5.1 that there is a strong linear relationship between solutions at the original location x^* and solutions at moving locations $x^* + \Delta x$ from the original fixed location, where the distance between them is very small and approximately equal $h = 0.0366$. Gradually, a weak linear relationship will be achieved when the distance between the original point x^* and moving points increases until this distance becomes $h \approx 32.50$; at this stage, the curve of the correlation coefficient crosses the x-axis i.e $\rho = 0$. According to Figure 5.1, the distribution of prey population density has changed gradually over time, which negatively affects the values of the correlation coefficient. By comparing values of the correlation coefficient, it can be noticed that ρ changed gradually from $\rho \approx 1$ to $\rho \approx 0$ when the distance between points becomes $h \approx 32.50$ and $h \approx 120$. Between these distances, values of the correlation coefficient become negative; from $h = 120$ to $h = 150$, values of ρ become positive again. As a result, the degree of synchronisation between the fixed location and other locations

in the series of population densities decreases gradually, as long as the distance increases. It can be concluded from this example that the properties of population densities are gradually changing over the domain, as can be seen from the values of the correlation coefficients in Figure 5.1. The strength of synchronisation between series of population densities declines as a result of the increasing distance between traps, which reflects the importance of installing traps within the whole domain in order to prevent, or at least reduce, any loss of important information about the distribution of population densities.

5.5 How Absence of Synchronisation Can Be Treated

The objective in this section is to investigate which numerical integration method is more correlated than others in a series of coarse ecological data. From the previously generated ecological data for 200 time series, various coarse grids will be extracted. The process of extracting a coarse number of grid nodes from a fine number of grid nodes has been explained in Chapter 3.

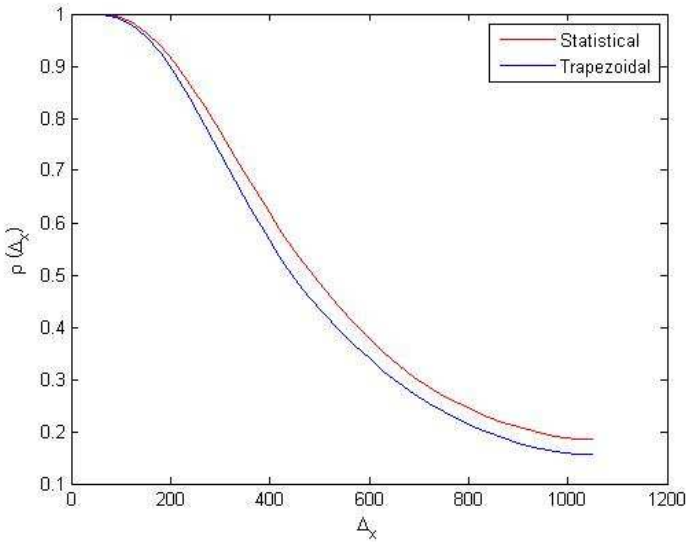

Figure 5.2: Synchronisation curves of the numerical integration methods between basic coarse grids and extracted midpoints at $N_c + 1 = 5$.

From these coarse grids, each midpoint between them will now be extracted. To

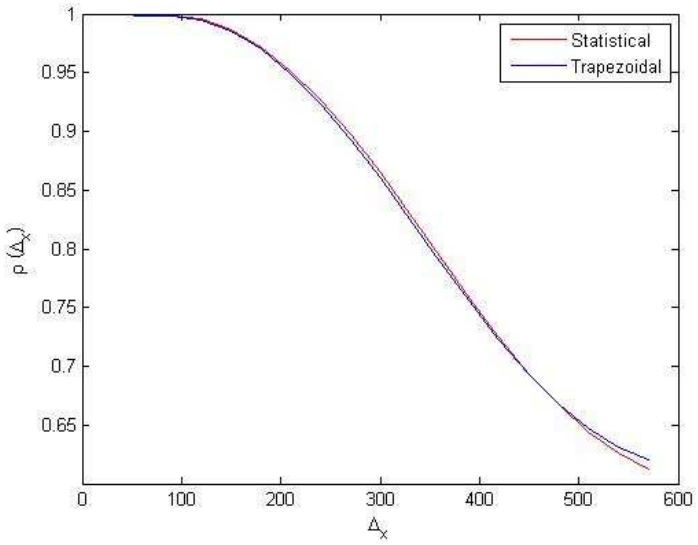

Figure 5.3: Synchronisation curves of the numerical integration methods between basic coarse grids and extracted midpoints at $N_c + 1 = 9$.

demonstrate this procedure, let us consider x_j^c denotes extracted coarse grids where $j = 1, 2, ..., N_c$, then each midpoint between them will be extracted as follows: $x_j^m = x_{j+l}^c$, where x_j^m represents midpoints, and $l = (N_f/N_c)/2$ is the formula to extract midpoints between coarse grids, where the distance between each original coarse grid equals N_f/N_c, which will be the same distance between each midpoint. For ecological data generated in the previous section for the time series $t = 1, 2, ...200$, the trapezoidal (3.1.13), and statistical (1.2.1) rules will be applied to evaluate the population density size for extracted coarse grids u_j^c. All integral values will be collected in one file for the trapezoidal rule and another file for the statistical rule. The same procedure will be applied to the extracted midpoints u_j^m, and the values of integrals will be recorded as well. Once the values of integrals have been recorded, the correlation coefficient (5.3.1) between them will be computed. Now the distance between midpoints and original extracted coarse grids will be reduced gradually as follows: $x_j^m = x_{\Delta_x}^c$, where $\Delta_x = (j + l) - h$ is the index of midpoint positions and $h = 30 * r$ denotes a reduced distance between midpoints and original ex-

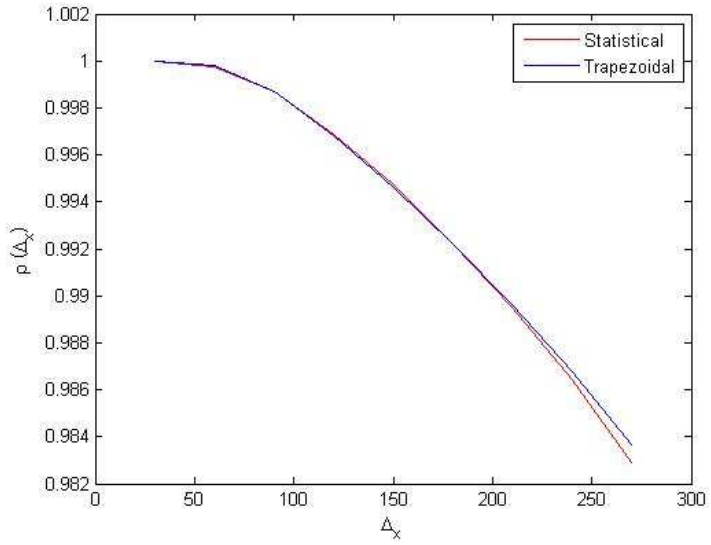

Figure 5.4: Synchronisation curves of the numerical integration methods between basic coarse grids and extracted midpoints at $N_c + 1 = 17$.

tracted coarse grids, and $r = 1, 2, ..,$ is the repetition required to become points identical to each other. This technique will be continued until these points become identical to each other at the largest value of h. The correlation coefficient of the numerical integration methods will be calculated at each new value of distance h. In order to compare trapezoidal and statistical rules, in terms of which method is more correlated, their correlation coefficient curves is drawn in one graph against distance Δ_x per coarse grid nodes.

After explaining the technique of extracting midpoints from coarse grids and computing the degree of synchrony for the numerical integration methods applied, different test cases will be introduced. The first test case represents the synchrony curves of numerical integration methods at $N_c + 1 = 5$. By investigating Figure 5.2, it can be noted that the values of the correlation coefficient between the basic extracted coarse grid nodes and the values of the integrals at the extracted midpoints depends on the value of parameter h. The strongest linear relationship has been achieved at $h = 1050$, and gradually this relationship becomes less than $\rho \approx 1$ when decreasing the value of h. However, the val-

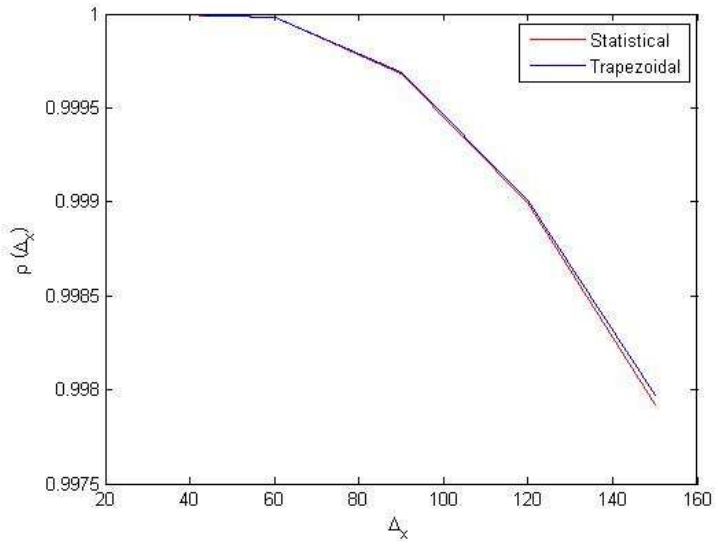

Figure 5.5: Synchronisation curves of the numerical integration methods between basic coarse grids and extracted midpoints at $N_c + 1 = 33$.

ues of the correlation coefficient reflect the existence of strong synchronisation between coarse grids and moving midpoints, where the lowest value of the correlation coefficient is achieved at the original position of the midpoints, which is approximately equal to $\rho \approx 0.1853$ for the statistical rule and $\rho \approx 0.1557$ for the trapezoidal rule. It can be noted from Figure 5.2 that the statistical rule has provided more correlated synchrony values than the trapezoidal rule for coarser grid nodes. As demonstrated previously, the degree of accuracy of numerical integration methods can be increased by increasing the number of grid nodes. Then, the degree of correlation between values of integrals may become more correlated as a result of increasing the number of extracted coarse grids. The number of extracted coarse grids will be increased as follows: $N_c + 1 = 9, 17, 33$. At each number of coarse grid, a graph of the correlation coefficient against distance, Δ_x will draw.

It can be seen from Figure 5.3 that at a number of coarse grids equal to $N_c+1 = 9$, the degree of synchronisation becomes greater as a result of increasing the number of grid nodes,

where the lowest value of the correlation coefficient at the minimum value of distance h is $\rho \approx 0.6204$ for the trapezoidal rule, which is, in fact, slightly more correlated than $\rho \approx 0.6127$ for the statistical rule. In addition, the amount of repetition r reduces as a result of dealing with fine grid nodes. It has been proven that the trapezoidal rule provides slightly more accurate estimations than the statistical rule for coarse grids; Therefore, values for the trapezoidal rule may be more correlated than for the statistical method. Increasing the number of extracted coarse grids leads to a reduction in the distance between the synchrony curves for the trapezoidal and statistical rules, as can be seen from Figure 5.4, and Figure 5.5, where the number of extracted coarse grids are $N_c+1 = 17, 33$ respectively. Figure 5.5 shows that the correlation coefficients curves of the trapezoidal and statistical rules become identical to each other at a number of coarse grids equal to $N_c + 1 = 33$. In this case, the lowest value of the correlation coefficient is $\rho \approx 0.9980$, which indicates a very strong synchrony between values of integrals. These observations reflect the previous result, which is the importance of dealing with fine number of grid nodes to obtain a very accurate and correlated estimation of population density. The degree of synchronicity can be improved by dealing with fine number of grid nodes.

5.6 Effect of Grid Coarseness on Synchronisation

Previous sections investigated synchronisation between ecological data generated by using a spatially explicit predator-prey system from the Rosenzweig-MacArthur model (2.2.10), (2.2.11), which was controlled by species diffusion. For different ecological test cases, it has also been demonstrated that the number of grid nodes affects the degree of synchrony. Perfect synchronisation was achieved at very fine number of grid nodes however, this deteriorates when dealing with coarse number of grid nodes [136]. In this section, synchrony with another classification of ecological data will be discussed. In some ecological applications, the population density is spread over a very small area of the agricultural

field. This distribution phenomenon is referred to as the highly aggregated density. A biological invasion is an example of a highly aggregated density distribution in the initial spread, the pest population density can be located in a very small area, but over time the pest species spreads throughout the entire field [133]. As a result of the spread of pest population densities over time, a strong, heterogeneous, patchy spatial distribution will be generated [168]. This pattern of population density distribution is known as a patchy invasion [168]. In order to prevent, or at least limit, damage to crops as a result of the spread of patchy spatial density, it would be beneficial to estimate the pest population before they spread throughout the field. However, this is a complex procedure, because the location of the patch is not known, and the traps are placed throughout the entire field by using regular computational grid nodes [135, 137]. One consequence of this method of sampling is that if the grid is very coarse, information about population density abundance can be missed entirely.

Now let us generate a peak function by using formula (4.1.2) over interval $[0, L]$, where $L = 300$, $\sigma = 8$, and the location of the peak will be altered randomly within $[a_I, b_I] = [40, 260]$, where $\epsilon = 40$. Different numerical integration methods, statistical rule (1.2.1), composite trapezoidal rule (3.1.13) and composite Simpsons rule (3.1.14) are applied to several series with an increasing number of grid nodes $N = 3, 5, 9, 17, 33, 65, 129$, in order to investigate synchronisation between peak functions. For each number of computational grid nodes, two different arrays of generated population densities for different habitats A and B are numerically integrated at times $t = t_1, t_2, ..., t_k$ to generate two time series:

$$S_A = (I_A^1, I_A^2, ..., I_A^k), \quad S_B = (I_B^1, I_B^2, ..., I_B^k).$$

Let us consider the peak function (4.1.2) for a fine number of grid nodes, where $\delta = 6\sigma$, $\sigma = 8$, as shown in Figure 5.6a. After generating peak functions over series of refined uniform grids $x_{i+1} = x_i + h_i, \quad i = 1, 2, ...N$, where $h_>0$ is the grid step size different numerical integration methods are employed over increasingly refined regular number of

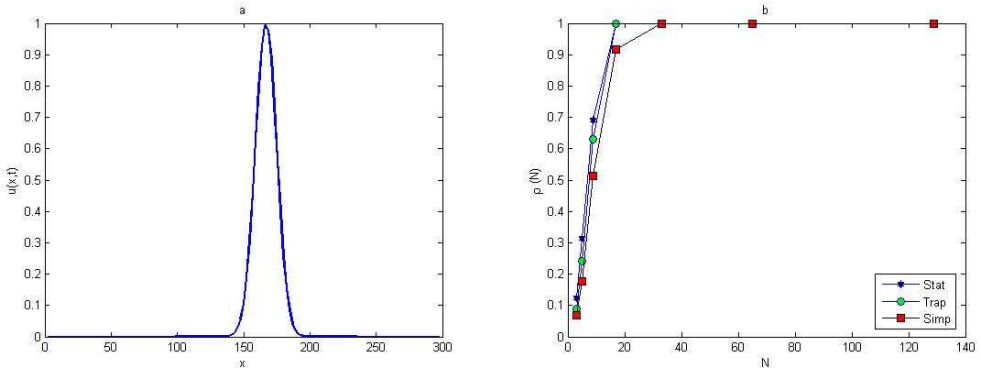

Figure 5.6: Numerical integration of a peak function. (a) The peak function (4.1.2), where the location of the peak maximum is $x^* = 166.88$ and the peak width is $\delta = 48$. (b) The correlation coefficient curves of the numerical integration methods between series of peak functions with increasingly refined regular grids.

grid nodes $N = 3, 5, 9, 17, 33, 65, 129$. Once the population densities have been obtained by different numerical integration approaches, the degree of synchrony will be computed form (5.3.1) to see how the degree of synchronisation changes with the numerical methods used. It can be readily seen from Figure 5.6b that the statistical rule has achieved slightly more correlated data, especially at coarse number of grid nodes, than the trapezoidal and Simpson's rules. Moreover, the composite trapezoidal rule is more synchronised than the composite Simpson's rule. Figure 5.6b shows that the degree of synchronisation between series of data can be improved by increasing the number of regular grid nodes. In addition, it can be noticed that methods which have low convergence rates may provide a better degree of synchronisation than methods with higher convergence. Therefore, it can be said that the correlation coefficient relies heavily on the number of available samples. In addition, in order to provide synchronised values so that $\rho = 1$, a sufficient number of grid nodes must be generated.

By computing the correlation coefficients between integrated peak functions over interval $[0, 300]$, and by varying the value of σ, let us now consider peak function (4.1.2) and the value of σ will be alerted so that $\sigma = 4, 6, 8, 10, 12, 15$. We numerically integrate the

peak function by applying a statistical approach to a series of increasingly refined grid nodes $x = 3, 4, 5, 6, ..., N$, then the correlation coefficients between values of integrals are computed. At each peak width, the critical number of grid nodes required to achieve the value of $\rho = 1$ is determined. It can be seen from Figure 5.7 that the threshold number of grid nodes required to fulfil the condition $\rho = 1$ becomes larger than that number of grid nodes obtained from the probabilistic approach used in Chapter 4 at different values of tolerance τ. It can be seen from Figure 5.7 that at a value of standard deviation $\sigma = 4$,

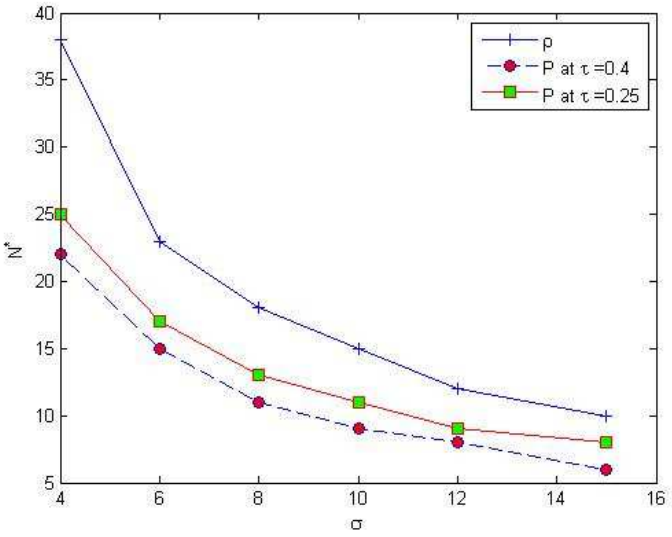

Figure 5.7: Comparison between critical number of grid nodes required to hold the conditions $\rho = 1$ and $p = 1$ for the probabilistic approach, at different values of σ and different values of τ.

which is the narrowest peak that has been chosen in our corresponding domain $[0, 300]$, the condition $\rho = 1$ holds for quite a large number of grid nodes $N^* = 38$, whereas this number becomes smaller, $N^* = 25$, when computing the probability of achieving an accurate answer at $\tau = 0.25$, and the smallest critical number, $N^* = 22$, is obtained at $\tau = 0.4$. The gap between these values becomes smaller when increasing the value of σ as $\sigma = 15$, which represents the widest peaks available. At $\sigma = 15$, the condition $\rho = 1$ holds at $N^* = 10$, and the probability $p = 1$ is obtained at $N^* = 8$ and $N^* = 6$ for $\tau = 0.25$, and

$\tau = 0.4$ (*i.e* the critical number of grid nodes becomes close to each other when dealing with a wider peak). The gap between threshold numbers of grid nodes for correlation coefficients and the probabilistic approach significantly relies on the tolerance τ, where at small $\tau = 0.25$, the gap is smaller than the value at $\tau = 0.4$, as shown in Figure 5.7. The main aim is to determine which tolerance τ we have to hold the conditions $\rho = 1$, and $p = 1$ at the same critical number of grid nodes. It can be deduced from above discussion that in order to achieve this target, the value of tolerance must decrease.

The next task is now to investigate whether increasing the number of peaks will affect the degree of synchronicity. In order to find this out, the previous form of standard normal distribution (4.1.1), will be generalised. The population density now is considered as a superposition of normal distributions given by equations (4.1.5), (4.1.6) and $w(t) = Ct$, where C is const. Equation (4.1.5) will be used to generate functions of four and eight

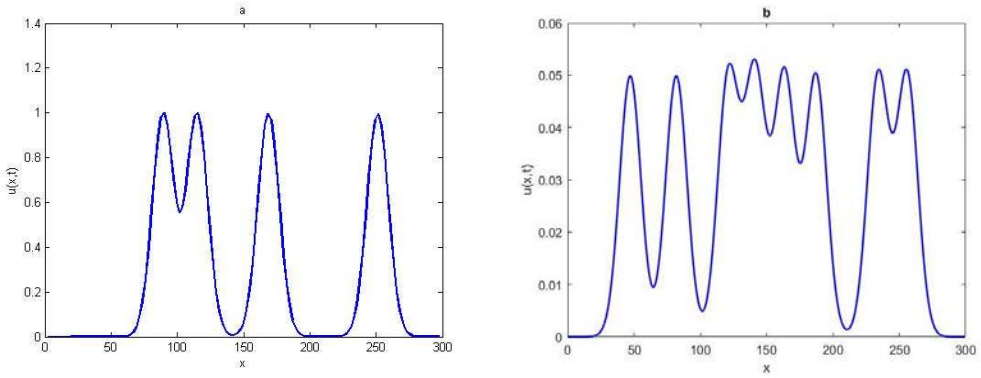

Figure 5.8: (a) The superposition of normal distribution (4.1.5), where the number of peaks is $P = 4$. (b) The superposition of normal distribution (4.1.5), where the number of peaks is $P = 8$.

peaks by substituting value of P with $P = 4$ and $P = 8$, as shown in Figure 5.8a, and Figure 5.8b. Then, the same procedures that were used with a single peak function will apply. It should be noted here that equation (4.1.5) can be used to generate a standard normal distribution function (peak function) by replacing the value of P with $P = 1$. Then, for each numerical integration method we will have three synchronisation curves,

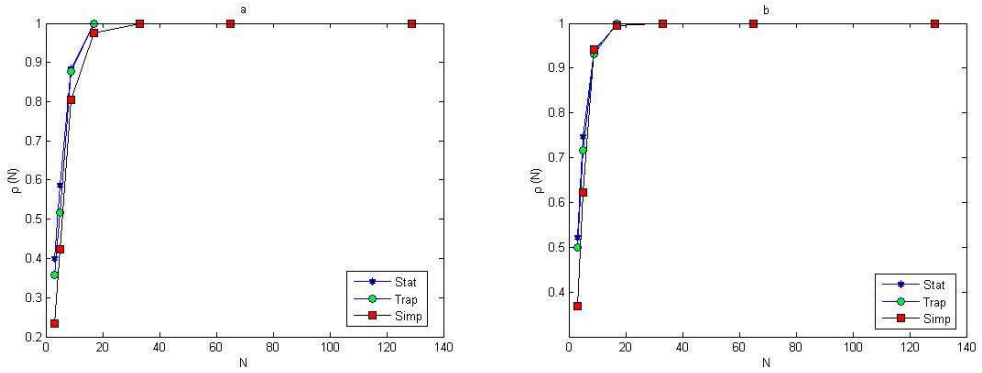

Figure 5.9: Numerical integration of peak functions. (a) The correlation synchronicity curves of the numerical integration methods between series of peak functions for increasingly refined regular meshes at $P = 4$. (b) The correlation coefficient curves of the numerical integration methods between series of peak functions for increasingly refined regular meshes at $P = 8$.

depending on the number of peaks.

By inspecting Figure 5.9a, and Figure 5.9b, it can be noted that for the statistical rule on coarse grids, more correlated estimation values have been achieved at a number of peaks equal to $P = 8$, and this degree of correlation decays gradually at $P = 4$, and $P = 1$. Furthermore, with coarse computational meshes the statistical rule is more correlated than the trapezoidal or Simpson's rules at $P = 1$, as shown previously, and this is true for $P = 4$ and $P = 8$, as shown in Figure 5.9a, and Figure 5.9b. From the application of the superposition of normal distributions, the degree of synchronisation depends on the number of peaks and is more correlated for numerical integration methods that have the lowest convergence rate on coarse computational grids. The degree of synchronisation is improved by increasing the number of grid nodes.

5.7 Sampling Coarse Grids and Ghost Synchronisation

It has been discussed above that in ecological studies information about the population density is usually obtained from data collected by sampling. Sampling over a coarse grid may lead to wrong conclusions about population dynamics. When dealing with population densities that have obvious fluctuations over areas, with sparse available data (*i.e* the data are collected from sampling over coarse grid nodes), the synchronisation is likely to be lost and the value of the correlation coefficient is always much smaller than the actual value. In this section, we are going to illustrate that sampling on a coarse grid may provide a value for the correlation coefficient close to the true value, whereas in reality the dynamics of population density is anti-correlated; this phenomenon is known as ghost synchronisation [142].

In order to investigate this phenomenon, let us consider a system that consists of two different one-dimensional habitats, A and B, where the population has a unimodal distribution, and species abundance is concentrated at the boundary of the domain as follows:

$$u_j(x,t) = \frac{w_j(t)}{\sigma_j\sqrt{2\pi}} exp(-\frac{x^2}{2\sigma_j t^2}), \quad 0 \leq x \leq L_j, \ j = A, B. \tag{5.7.1}$$

The population distribution given in (5.7.1) is specific. The population density decays monotonously with the distance from the domain border. This is common in ecosystems and agroecosystems. For instance, it may correspond to the invasion of pest insects to a farm field from neighbouring uncultivated areas [10]. The population size in each domain is affected by different factors; therefore, it evolves differently with time t, where the shape of spatial distribution is preserved. In order to illustrate ghost synchronisation in detail,

let us consider the following hypothetical situation:

$$w_A(t) = w_0(t+1)^a, \quad \sigma_A(t) = \sigma_0, \quad w_B(t) = w_0 - \gamma t, \quad \sigma_B(t) = \sigma_0(t+1)^b, \tag{5.7.2}$$

where $t \geq 0$ and a, b, γ, w_0 and σ_0 are the chosen parameters. At $t = 0$, the initial

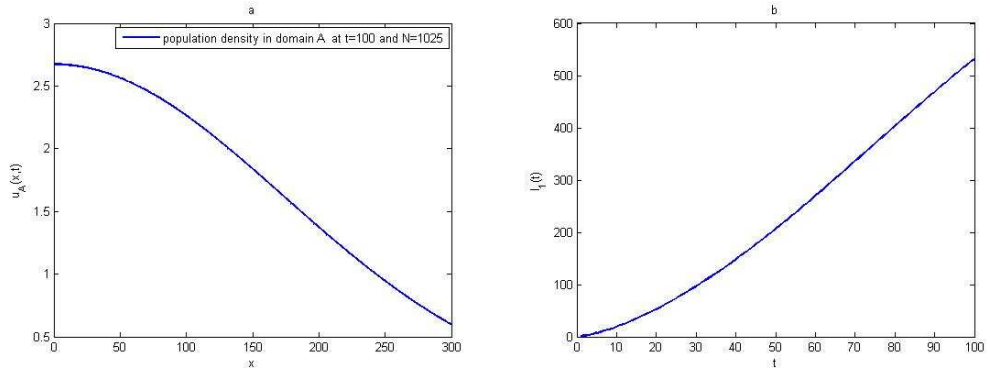

Figure 5.10: (a) Spatial population distribution in domain A obtained from equations (5.7.1), (5.7.2), at number of grid nodes $N = 1025$, with parameters $a = 0.5$, $b = 0.25$, $\gamma = 0.0001$, $w_0 = 2$ and $\sigma_0 = 3$. (b) The population density of domain A, computed by the trapezoidal rule (3.1.13) at time series $t = 0, 1, ..., 100$.

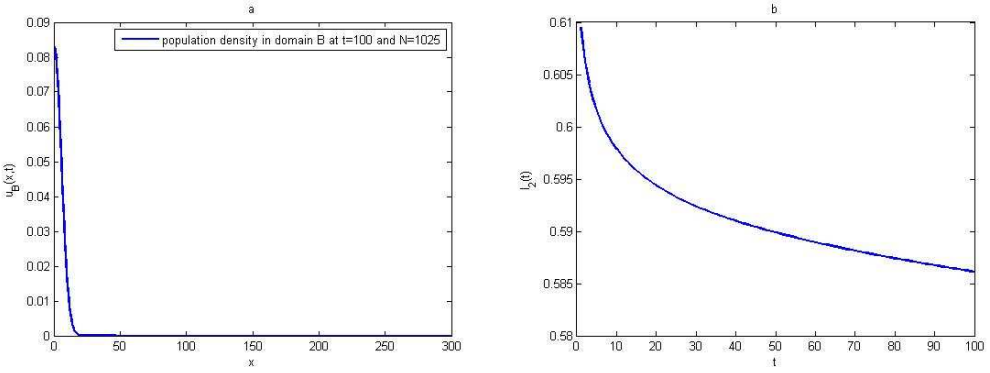

Figure 5.11: (a) Spatial population distribution in domain B obtained from equations (5.7.1), (5.7.2), at number of grid nodes $N = 1025$, with parameters $a = 0.5$, $b = 0.25$, $\gamma = 0.0001$, $w_0 = 2$ and $\sigma_0 = 3$. (b) The population density of domain B, computed by the trapezoidal rule (3.1.13) at time series $t = 0, 1, ..., 100$.

population distributions in both domains are the same, where $w_A(t) = w_B(t) = w_0$ and $\sigma_A(0) = \sigma_B(0) = \sigma_0$. While the initial population distributions in both domains are the

same, the dynamics, however are different: in region A, the size of the maximum population density increases over time, whereas it decays in domain B. Furthermore, in domain B, the width of the patch increases with time, but it does not change in domain A.

To generate the previous population distribution in domains A and B, let us assume that in the vicinity of the region boundary, the initial population distribution is aggregated, ($i.e\ \sigma_0 << [L_A, L_B]$), where the tail of the population distribution on the right-hand side of the areas is thin, due to following restrictions: $\sigma_A(t) << L_A$ and $\sigma_B(t) << L_B$. We therefore observe that the population size in domain A grows, but it decreases in domain B, as depicted in Figure 5.10b , and Figure 5.11b. The shapes of population densities in domains A and B at the fixed time $t = 100$ with a fine number of grid nodes $N = 1025$ are shown in Figure 5.10a , and Figure 5.11a. As a result of this change in the size of population densities in both domains, the dynamics will be anti-correlated and the value of the corresponding correlation coefficient tends to be negative and close to $\rho = -1$.

From equation (5.7.1), it can be seen from its properties that the population densities in both domains tend to increase simultaneously in a sub-domain, as shown in Figure 5.10b , and Figure 5.11b. If the samples of population densities are only taken from this sub-domain, the value of the corresponding correlation coefficient tends to be positive and is likely to be close to one. The dependence of the correlation coefficient on the number of samples, and also on their locations, is still unclear and requires further investigations.

Now let us consider the system of equations (5.7.1), (5.7.2) in order to address its properties in a more quantitative way. The value of the correlation coefficient will be computed over a sequence of generating grid nodes N to obtain $\rho(N)$ accordingly. In this simulation we use the following parameters $a = 0.5, \quad b = 0.25, \quad \gamma = 0.0001, \quad \sigma_0 = 3, \quad w_0 = 2$, and $L_A = L_B = 300$. The location of first grid nodes will be fixed as $x_1 = 10$, then the grid will be refined in two different ways. The first approach is that extra nodes will be added to x_1 towards the tail of population density distribution according to the following form

$x_i = x_1 + (i-1)\Delta_x$ where $i = 2, .., N$, and $\Delta_x = 10$.

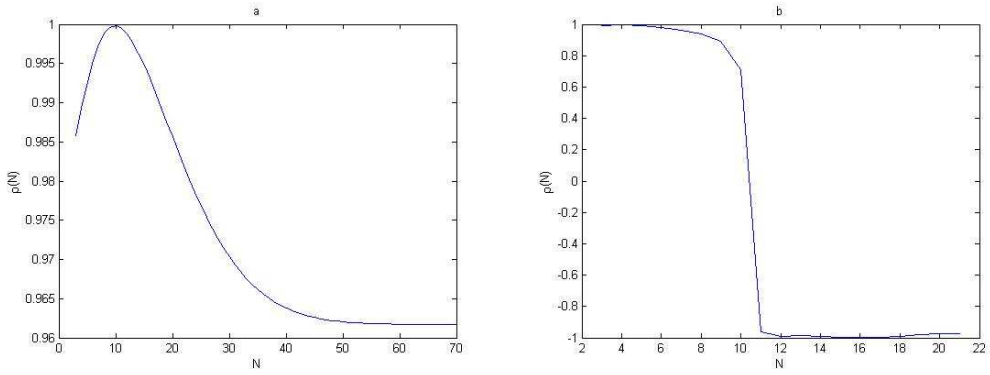

Figure 5.12: (a) Dependence of the correlation coefficient on the number of nodes in the sampling grid where the grid is refined by adding extra nodes to the tail of population density distribution where $x_1 = 10, \quad \Delta_x = 10$. (b) Dependence of the correlation coefficient on the number of nodes in the sampling grid where the grid is refined by adding extra nodes to the centre of population density distribution where $x_1 = 10, \quad \Delta_x = 1$.

Therefore, it can be seen from Figure 5.12a that, phenomenon of ghost synchronisation is acquired. The values of correlation coefficients change slightly from $\rho(N = 3) \approx 0.9857$ to $\rho(N = 10) \approx 1$, and then decays to $\rho(N = 59) \approx 0.9617$ where the value of ρ does not change later even at very fine number of grid nodes. Hence, the value of ρ in this situation does not depend on the samples size N ($i.e$ it is not sensitive to the number of grid nodes N). Furthermore, the strong correlated values obtained here it is in reality false estimation, as the population dynamics prescribed by equations (5.7.1), (5.7.2) is anti-correlation where $\rho \approx -1$.

In fact, the location of grid node x_1 strongly affects the value of ρ, where value of correlation coefficient ρ is very sensitive to position of x_1. As x_1 becomes bigger as probability of obtaining ghost synchronisation increasing. In order to prove this statement, let us consider the previous case of refining the grid nodes, where the values of parameters are the same just $x_1 = 5$ instead of $x_1 = 10$. The range of correlation coefficient tends to move towards actual value $\rho = -1$ when decreasing the value of x_1 as shown in Figure 5.13a, the values of correlation coefficients in this case provide more realistic results than

$x_1 = 10$, however x_1 needs to be minimised to reach $\rho \approx -1$. For the same value of $x_1 = 5$ but the grid step size Δ_x is chosen to be $\Delta_x = 1$, the values of $\rho(N)$ completely change to provide values close to actual value $\rho \approx -1$, where an accurate estimation of the population density size is obtained as a result of sufficient resolving to spatial details of the domain as shown in Figure 5.13b. Whereas, these vital details are missed to end up with ghost synchronisation with large value of $x_1 = 10$ and large value of $\Delta_x = 10$ as is shown in 5.12a.

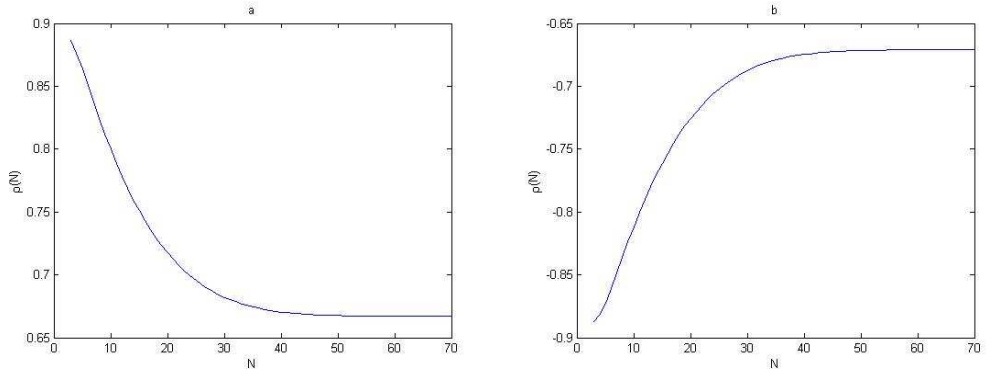

Figure 5.13: (a) Dependence of the correlation coefficient on the number of nodes in the sampling grid where the grid is refined by adding extra nodes to the tail of population density distribution where $x_1 = 5,\ \Delta_x = 10$. (b) Dependence of the correlation coefficient on the number of nodes in the sampling grid where the grid is refined by adding extra nodes to the tail of population density distribution where $x_1 = 5,\ \Delta_x = 1$.

The second approach of refining the grid nodes depends on adding extra nodes to the maximum of the spatial population distribution given by (5.7.1) (*i.e* towards the centre of distribution) according to the following form $x_i = x_1 - (i-1)\Delta_x\ ,\ \ i = 2, ..., N$, where $x_1 = 10,\ \Delta_x = 1$. In this case, the value of correlation coefficient shows strong dependence on the samples size N by presenting monotonous decays from false value $\rho(N = 3) \approx 1$ to the actual value $\rho(N = 12) \approx -1$ as depicted in Figure 5.12b. To end up, it is essential to mention here that, not just number of samples affects the degree of synchronisation weather negatively or positively, the locations of samples have the same effect unless more effect on existence of synchronisation [142]. It has been illustrated here and in [142] that

sampling on a coarse grid may emerge a ghost synchronisation. In the ghost synchronisation the value of correlation coefficient calculated from using coarse grid nodes has a value sufficient close to the true value $\rho \approx 1$ while the dynamics of population density is anti-correlated. The work [142] emphasis that, in order to obtain the actual value of the correlation coefficient the number and the location of samples are very important. In addition, it has been demonstrated that the information collected from the area close to the maximum of the distribution is apparently more important than that information obtained from area close to the tail of its distribution [142].

5.8 Factors That May Increase or Decrease Synchronisation for Aggregated Population Densities

Different factors may affect the degree of synchrony whether negatively or positively. The main factor, which plays a vital role in enhancing or decreasing synchrony between series of superposition normal distributions, is the width of the peak. As can be seen from previous Figure 5.6a, Figure 5.9a, and Figure 5.9b, different superposition normal distributions have been generated with different numbers of peaks, depending on the value of variable P in equation (4.1.5). Does the width of these peaks affect the degree of synchrony? The spatial population density given by (4.1.6) at $P = 1$ has been examined for different numerical integration methods, with peak width equal to $\delta = 48$ where $\sigma = 8$. Now, the width of peak will be reduced in accordance with equation (4.1.3), to become $\delta = 18$, where in this case $\sigma = 3$ instead of $\sigma = 8$. The statistical approach (1.2.1) will be applied to approximate the values of integrals for (4.1.6) over series of increasingly uniform grid nodes at $\delta = 3, 8$, and the correlation coefficients will be computed.

As shown in Figure 5.8, the synchronisation at a larger value of σ is better than a peak of smaller width on a coarse grid nodes. It can be readily seen from Figure 5.8 that the true value $\rho = 1$ is not obtained until $N = 65$ or larger in the case of $\sigma = 3$, and $N = 17$

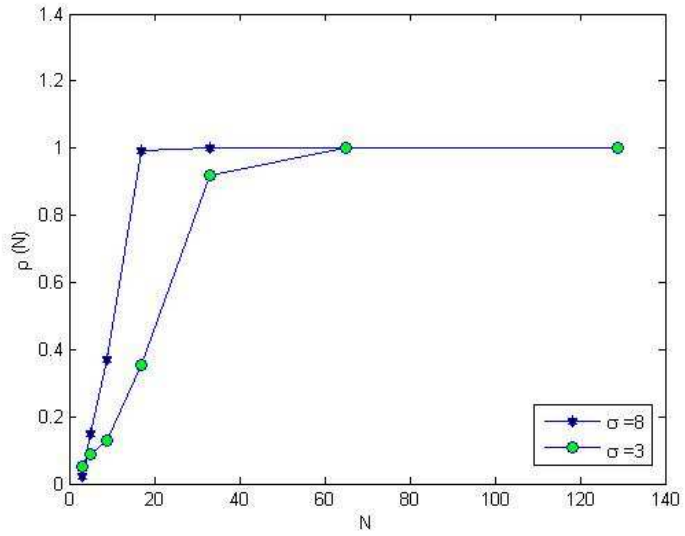

Figure 5.14: The correlation coefficient for various values of σ for the function of a single peak $P = 1$.

for $\sigma = 8$, which proves the above statement. Furthermore, before we achieve the critical number of grid nodes required to achieve true value $\rho = 1$, synchronisation appears to be lost, the domain shows a weak correlation or at very coarse grid nodes it is very weakly correlated. This observation reflects the importance of dealing with fine number of grid nodes in order to provide an accurate estimation of population density, which in turn provides more correlated results. When dealing with coarse number of grid nodes, the percentage of important information about population densities missed is high, because the location of the peak was unknown previously, as illustrated before, and the number of installed traps, with large distances between them, is not enough to cover the total area of the agricultural field in order to produce an accurate evaluation of population abundance. Therefore, the probability of providing a false synchronisation on sampling of coarse number of grid nodes becomes greater than for fine number of grid nodes.

Another factor that may affect the degree of synchronisation between series of superposition normal distributions is the location of peaks, which relies on the truncated sub-

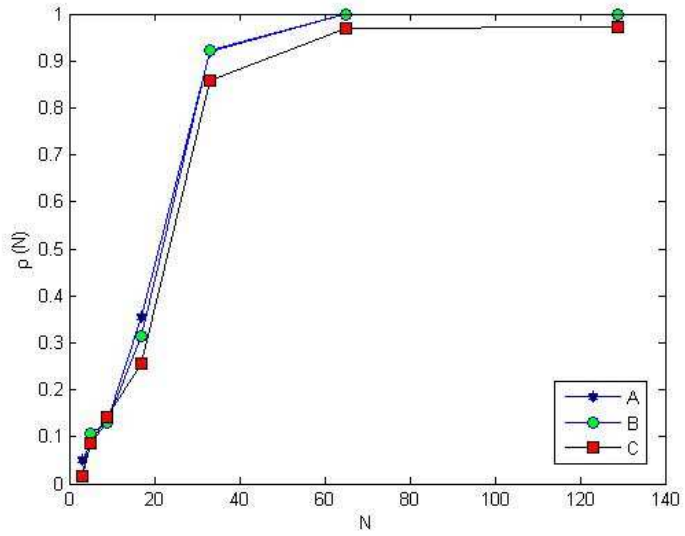

Figure 5.15: The correlation coefficient for various values of truncated sub-intervals $A = [40, 260]$, $B = [15, 285]$, $C = [0, 300]$ for the function of a single peak $P = 1$.

interval. The length of the truncated sub-interval has a very important impact on the degree of synchrony. To demonstrate the effect of truncated sub-interval, let us consider three different truncated sub-intervals, as follows: $A = [40, 260]$, $B = [15, 285]$ and finally $C = [0, 300]$, which represents the original domain, where $\epsilon = 0, 15, 40$ respectively. The correlation coefficient curves for each value of the sub-interval are combined in one figure for the purpose of comparison. As can be seen from Figure 5.8, at truncated sub-interval A, the estimation of population density is slightly more correlated than the evaluation at truncated domain B for coarse number of grid nodes, whereas the degree of synchronisation becomes worse when the peak could be located anywhere within the original domain C. The peak here can be placed identical at the endpoints of original interval $[a, b]$ or very close to it, which leads to missing data and a false evaluation of the synchronisation between the sampling data on a coarse number of grid points. Furthermore, the true value $\rho = 1$ is never approached over domain C, even for a large number of grid nodes, as mentioned previously.

5.9 Chapter 5 Conclusion

The population density of a given species needs to be estimated based on discrete data collected from the grid nodes of a certain domain [13, 23, 163, 177]. It is common in ecology to encounter heterogeneous patterns of spatial population density distributions, where the amount of data collected may not always be enough to resolve all the details of the given spatial configuration [37, 94]. The exact position of peaks or patches is usually not known in advance, which makes the problem more sophisticated; it must be determined from an analysis of the sampling data [78]. Therefore, the number of samples used to evaluate the total size of the population density and its location will be a matter of guesswork; moreover, some external factors may negatively affect the procedure of sampling, such as a limited budget. Consequently, a coarse grid points (traps) will be installed, which provides insufficient information to resolve the details of the spatial population distribution. An evaluation of the total size of the spatial population density on a coarse grid normally provides an estimation with a low level of accuracy [21, 22, 136, 137, 141, 163]. The evaluation of the population size is used as input information to calculate the correlation coefficient between two habitats or areas, where if this input information is inaccurate, the level of synchronisation will be affected as a result.

The phenomenon of synchronisation and its properties have been discussed in this chapter, through some instructive examples. It has been proven that the degree of synchronisation depends strongly on the distance between traps over the series of population densities; and the degree of synchronisation decays gradually as a result of increasing distance between them as depicted in Figure 5.1. Furthermore, traps must be distributed over the whole domain in order to prevent any loss of important information, especially when coping with a heterogeneous spatial pattern of population density distributions, where at each grid node the estimated value for the population density differs from another node.

The strength of synchronisation relies on the numerical integration methods used to evaluate the population abundance. It has been demonstrated that the composite trapezium rule achieved more correlated estimation values than the statistical rule on a coarse grid; meanwhile, on the fine grid nodes, both methods converge to the true value at the same time, as depicted in Figures 5.2, 5.3, 5.4, and 5.5. Hence, the degree of synchronisation among series of ecological data can be improved by refining the coarse grid to become sufficiently fine.

For the heterogeneous population density given by a function of superposition of the normal distribution, the value of correlation coefficient ρ, which is computed over coarse sampling data, often does not reach the true value $\rho = 1$ until the grid has been refined sufficiently. It is readily seen from Figure 5.6b that the correlation coefficient ρ for a highly aggregated single peak distribution depends strongly on the number of samples N. In order to approach the true value $\rho = 1$, the number of grid nodes must be large enough to resolve all the details of the spatial distribution, which can be seen from Figure 5.6b to be $\rho = 1$ if $N \geq 33$. Before this critical number of grid nodes, the values of correlation ρ varied between moderate and weak synchronisation. Furthermore, for $N < 6$, the synchronisation is lost. Therefore, for the aggregated population density distribution, a sufficient number of grid nodes is required to see a strong synchronisation pattern.

The same results are acquired for distributions that are not highly aggregated, which consist of several patches or a multi-peak distribution. It can be seen from Figure 5.9a, and Figure 5.9b that the decay in the calculated value of ρ is observed for a small value of N, and it still has significant difference from the true value, but it is less dramatic than a distribution of a highly aggregated population. As the number of peaks P affects the strength of synchronisation, using different values of standard deviation σ influences the degree of synchronisation too. For a single patch distribution, it is readily seen that at $\sigma = 8$ is the critical number $N = 17$ to approach the true value $\rho = 1$ and $N = 65$ for

$\sigma = 3$, which reflects again the importance of providing a large number of samples, as shown in Figure 5.8. Therefore, according to the above results, it is important to understand that information concerning synchronisation is often lost when dealing with coarse sampling data. The ideal way to prevent, or at least reduce, the risk of losing important information is to refine the grid of nodes adequately.

Chapter 6

Concluding Remarks

This study has considered the evaluation of population abundance for an ecological monitoring problem set in an agricultural field. The decision to implement a control action, in order to prevent the spread of pests, depends on such an evaluation. On one hand, the most common means of pest control is pesticides, which obviously have many side effects both at the economic and environmental levels. On the other hand, the damage caused to crops due to pest attacks is significant and needs to be controlled immediately. In many ecological applications, obtaining knowledge about the population size is essential and has to be used to its fullest extent. Therefore, an accurate pest population density estimate is highly required to enable ecologists and farmers to make correct decisions on implementing pest management actions. Furthermore, a correct conclusion about the ecological problem obviously relies on an accurate evaluation of the total population size. It is known that a variety of environmental communities exist in any ecosystem, which exhibit different patterns of interaction between species. We have shown that a variety of ecological population samples can be generated from any population community. In Chapter 2, we considered the spatially explicit predator-prey system given by the Rosenzweig-MacArthur model. The main properties of the system's dynamics were investigated and a numerical solution of the system was provided. We focused on ecological

data generated by propagating population fronts in the Chapter 2. The front's wake is essentially attributed to the excitation of chaotic oscillations.

The total population abundance is typically estimated using a standard statistical approach (1.2.1). However, this does not always provide a sufficient degree of accuracy. Therefore, we used numerical integration methods instead of the statistical approach, in order to yield a more accurate evaluation of the problem. Numerical integration methods depend on the chosen weight coefficients to exceed the convergence rate of the statistical approach.

Ecological problems differ greatly from standard numerical integration problems, due to the complexity and constraints imposed by nature on the ecological problem. The conventional approach to increasing the level of accuracy relies on the availability of large quantities of data, which is not the case with ecological data, and in pest monitoring the number of samples N is particularly limited. Therefore, from a very limited quantity of data, an accurate approximation must be gained.

It was shown in [136, 137, 138] and confirmed in this thesis that, for the limited available data, the asymptotic error does not hold, on coarse grids. The definition of coarse grids depends on the nature of the integrand function. For one integrand function, the grid may be considered to be coarse, whereas the same grid may be considered to be a fine grid for another function, depending on the spatial heterogeneity of the function. On the one hand, for a homogenous population density spread over the entire domain, the asymptotic error often holds for a small number of grid nodes. On the other hand, a high number of samples is required for the asymptotic error to hold when a more heterogeneous distribution is considered. It proven that numerical integration methods with higher convergence rates obviously provide a more accurate evaluation of the population density. In reality, the process required to handle a large number of nodes in order to obtain the asymptotic error may become an obstacle; this is how the coarse grid problem

appears. Consequently, the conventional method, which depends on the application of a numerical integration approach with a higher convergence rate, does not work on coarse grids. It has been shown in this research that the Richardson extrapolation did not provide a satisfactory degree of accuracy when applied to the ecological problem, especially in coarse grids.

There are some other approaches that can be applied to improve the accuracy of standard numerical integration problems, but unfortunately these approaches are not available. Adapting the sampling grids is one approach that can be used to acquire more accurate results; it is applicable if we have prior knowledge of the pattern of spatial density distribution, so that, for the area that requires a high level of resolution, more data can be collected to fulfil the accuracy requirement. However, in our circumstances information about the pattern of population density is not available and an accurate evaluation of the population density is only considered on fixed grid nodes over a set time duration. As can be seen, the processes for improving accuracy in either grid refinement or variety of sampling plan can not be applied, and in pest monitoring all of these processes require the repetition of the sampling procedure. Pests are living creatures, so over time the population density function changes, which makes the repetition of initial conditions impossible. For a fixed and very limited amount of data, accuracy depends on the nature of the population density distribution (*i.e* the integrand function). When the population density is propagated over the entire domain, a more accurate evaluation will be achieved, even with a heterogeneous function. However, if the pattern of population density is severely aggregated (*i.e* we have a patchy distribution where we face areas of zero and non-zero abundance), we predict more inaccurate estimates. Furthermore, as a sampling procedure is applied to collect the data, the non-zero patch may be missed completely. Therefore, our aim is to find an appropriate approach to assess and improve the accuracy of evaluation on a coarse grids, where the asymptotic error estimates do not hold.

It has been demonstrated in [134, 140] and proven in this thesis that for highly aggregated population density distributions, not only estimates of the asymptotic error do not hold, but a desirable accuracy of integration error becomes a matter of chance. In other words, the integration error will behave as a random variable. The probabilistic approach is used to estimate the evaluation accuracy for the population density when the integration error is a random variable. The probability of achieving an accurate estimate of the integration error depends on determining the required number of grid nodes for the probability condition $p(N = N^*) = 1$ to hold within a certain tolerance τ. As a result, a new classification of grid nodes has been introduced when the number of grid nodes is less than N^*, which is defined as ultra-coarse grid points where $p(N) < 1$. By modelling a single peak population as a quadratic function, and applying the probabilistic approach when the integration error is a random variable, the number of grid nodes N^* needed to reach the condition $p(N^*) = 1$ at random locations of the maximum peak is roughly the same at random locations of the first grid point x_1 with a fixed position of the maximum peak. In our work, a $2d$ problem has been investigated in detail to conclude about the efficiency of probabilistic approach previously developed in [134, 140] for a $1d$ case. For the $2d$ ecological distributions, a random location of the corner left point (x_1, y_1) provided more accurate estimates of the integration error for very complicated population density distributions ($i.e$ the highly aggregated distributions), and affects accuracy of integration when a continuous front function is considered, due to slight changes in the population density abundance values. In $2d$ applications, the number of grid nodes required to achieve $p(N^*) = 1$ is larger than the corresponding number in $1d$ cases. Furthermore, N^* is affected by different factors, such as how many patches or peaks there are, the values of certain tolerances and the width of the peak. Another accuracy criterion is available to find the number of nodes N^* required to guarantee a sufficient level of accuracy in terms of the condition of the upper error bound $\mu_e + \sigma_e \leq \tau$. Here, the condition of the

upper error bound in some sophisticated functions on coarse grid nodes is misleading ($i.e$ it gives values of the boundary $\mu_e + \sigma_e$ sufficiently close to τ, even though, the values of the integration error are still larger than τ). Consequently, we recommended the use of another criterion for accurate evaluation in this case.

As demonstrated previously, in reality a sampling grid is often coarse and does not resolve all the required details of the spatial population density. As a result, differences in the estimates of the integration error can be significant. The values of correlation coefficients make an impact on the estimation of the population abundance synchronisation between two habitats. In this research, it has been demonstrated that incorrect estimate can provide irrelevant values for the correlation coefficient on coarse grid nodes. Several population models were considered to demonstrate that on coarse sampling grids, the correlation coefficient often does not exceed $\rho = 0.5$, even when the true value is extremely close to one. Different factors impact the abundance estimate: spacing between traps, number of population density patches, width of patch and number of available samples. It has been shown in [142] and proven in this work that the phenomenon of ghost synchronisation can be observed on coarse grid nodes where the value of $\rho \approx 1$, which reflects a very high value for the correlation coefficient; however, in reality the population dynamics are not correlated. Further research is required to investigate whether the phenomenon of ghost synchronisation happens only on coarse grids or occurs more frequently. The applications must be extended to involve $2d$ ecological data in future work. Another important factor that should be examined carefully is the effect of environmental heterogeneity on the degree of synchronisation.

It has been shown in this work that any prior information regarding spatial population density must be used to its fullest extent. Higher-order numerical integration methods are recommended when the amount of available data is sufficient to resolve all details concerning the spatial heterogeneity of the population density distribution. When dealing

with population density on coarse sampling grids, for instance a patchy invasion spread over the entire domain, more research is needed before a recommendation can be made as to which numerical integration methods should be applied. For a more sophisticated density pattern known as aggregated distribution, more investigation is required to determine the best approaches to evaluating the total population size. The probabilistic approach is suggested as a method to use to assess the performance of statistical methods on ultra-coarse grids. The probabilistic approach must be extended to involve other numerical integration methods of a high order. Our examinations of pest population density focused on the evaluation of accuracy at a fixed point in time; further studies must extend this to cover the problem of accuracy when time is a dependent variable.

It is important to emphasise again that different numerical integration methods are used to assess the accuracy of population density; however, we cannot yet state which method is the best and is ready for use in pest monitoring. Much future research is needed before this recommendation can be made.

appendix A
Appendix

The application of the numerical integration methods to ecological data from a sequence of irregular grid nodes has been discussed in Chapter 3. In this section, this application will be extended to involve some of the twenty-four mathematical test cases shown in Table 3.1. The convergence curve behaviour for different numerical integration methods on regular computational grid nodes may differ from the convergence curve on irregular computational grid nodes for the same methods. Therefore, the performance of each numerical integration method on irregular grid nodes will be investigated separately for some test cases from the twenty-four mathematical test cases in Table 3.1.
For the trapezoidal rule, the convergence curves of almost all of the mathematical test cases that have simple behaviour on irregular grid nodes are very close to, and parallel, the behaviour obtained on a regular computational grid. Furthermore, the differences in convergence rates are very small, as can be seen from Figures A.1, A.2, A.3, A.4, A.5a and A.6b. Although in some test cases the differences become prominent for very coarse grids, after increasing the number of grid nodes, the convergence curves become parallel to each other, as shown in Figure A.6a. When dealing with more sophisticated test cases, the differences in the degree of accuracy are evident, as can be seen in Figures A.7, A.8, A.9, A.10 and A.11. However, when applying the statistical rule, the difference in the

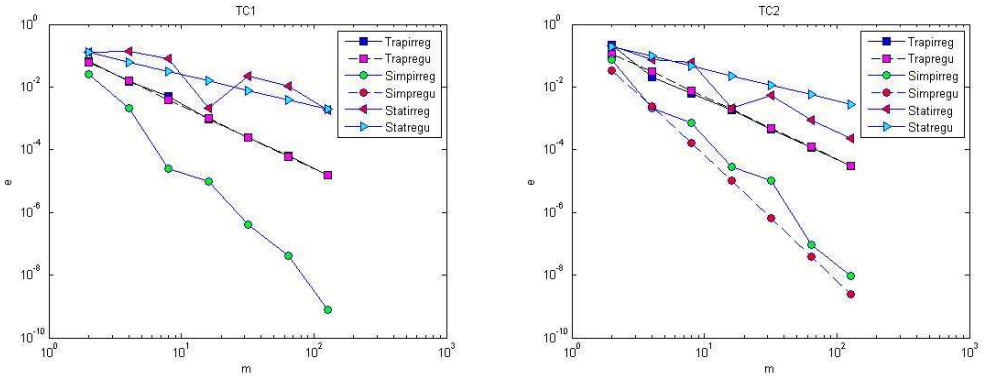

Figure A.1: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC1. (b) TC2.

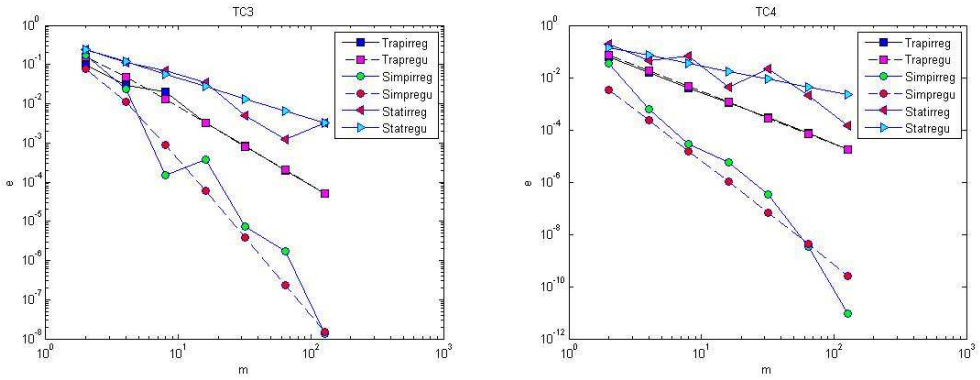

Figure A.2: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC3. (b) TC4.

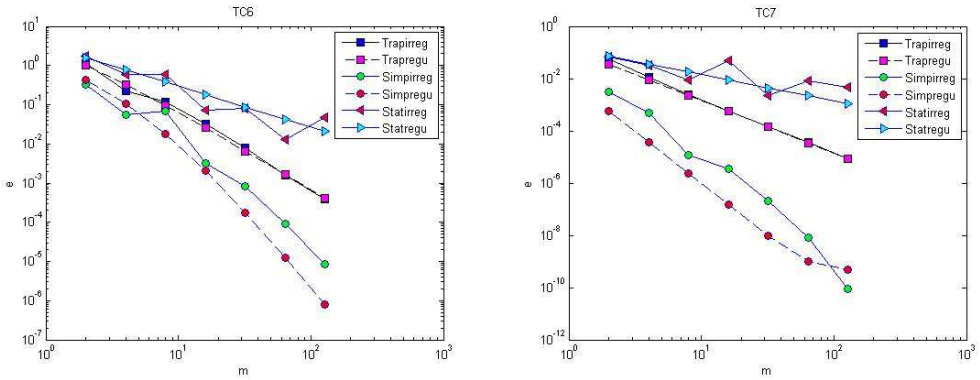

Figure A.3: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC6. (b) TC7.

convergence curves on regular and irregular grid nodes is prominent for some test cases: TC1, TC2, TC3, TC4, TC6, TC7, TC8, TC11 and TC12. Furthermore, the convergence rate for irregular grid nodes is poorer than that on regular grid nodes, as shown in Figures A.1, A.2, A.3, A.4a and A.5b. For functions that may have more complicated behaviour patterns, such as presented in the mathematical test cases: TC9, TC10, TC13, TC15, TC16, TC18, and TC19, the integration error on a sequence of regular grid nodes and irregular grid nodes is quite large, and the convergence curves are parallel to each other. However, in other mathematical test cases it can be noted that the convergence curves of the statistical method were oscillating from the beginning whether on regular grid nodes or irregular grid node as shown in Figures A.8a, A.9b, A.10, A.11. Therefore, in order to achieve a satisfactory degree of accuracy, in this case the grid has to be refined sufficiently.

The convergence curves of Simpson's rule on a sequence of irregular grid nodes are different from those on regular grid nodes for all twenty-four test cases. This difference becomes prominent when dealing with more sophisticated test cases, as shown in Figures

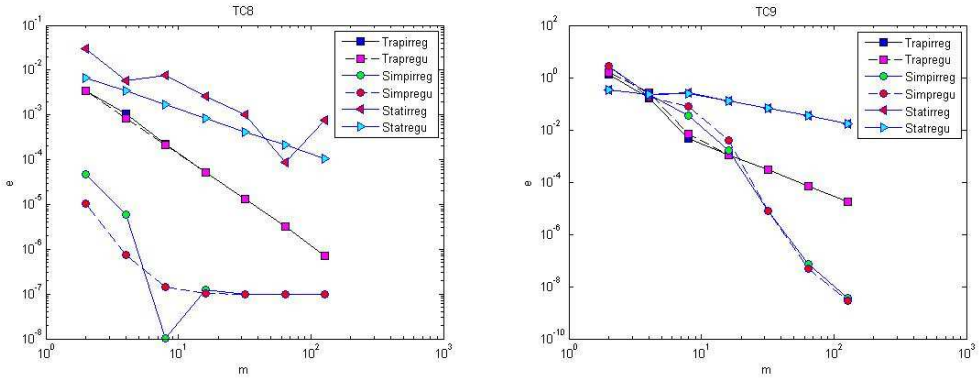

Figure A.4: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC8. (b) TC9.

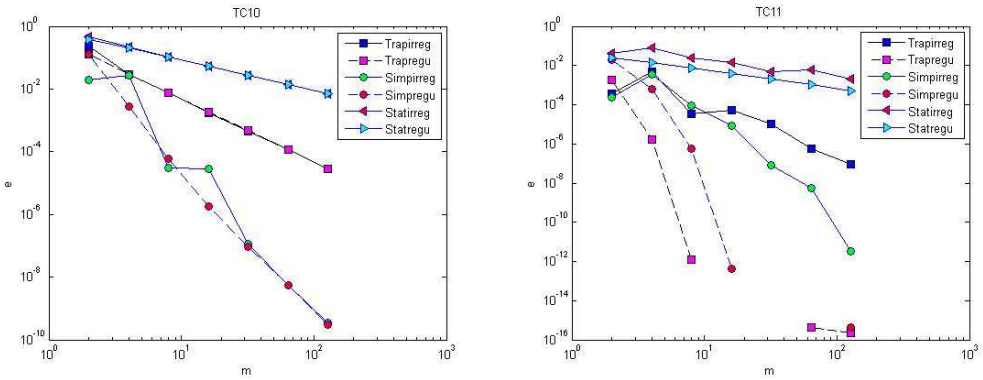

Figure A.5: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC10. (b) TC11.

A.7, A.8, A.9, A.10 and A.11.

As a result of the previous discussion, it can be concluded that randomisation affects the degree of accuracy of the different numerical integration methods. The results obtained from applying the trapezoidal rule on irregular grid nodes are very close to those obtained on regular grid nodes. With regards to the statistical rule, when the integration error is quite large on regular grid nodes, the convergence curves on irregular grid nodes are parallel to those on regular grid nodes. However, a prominent difference has been noted of the degree of accuracy in the convergence rate when applying the statistical rule to sequences of irregular and regular grid nodes. The application of Simpson's rule to sequences of irregular and regular grid nodes has shown that the most accurate results are obtained on regular computational grid nodes, rather than irregular grid nodes. Furthermore, the convergence curves on regular grid nodes always lie below those for irregular grid nodes. This shows that numerical integration methods used on regular grid nodes achieve much accurate results than when used on irregular grid nodes.

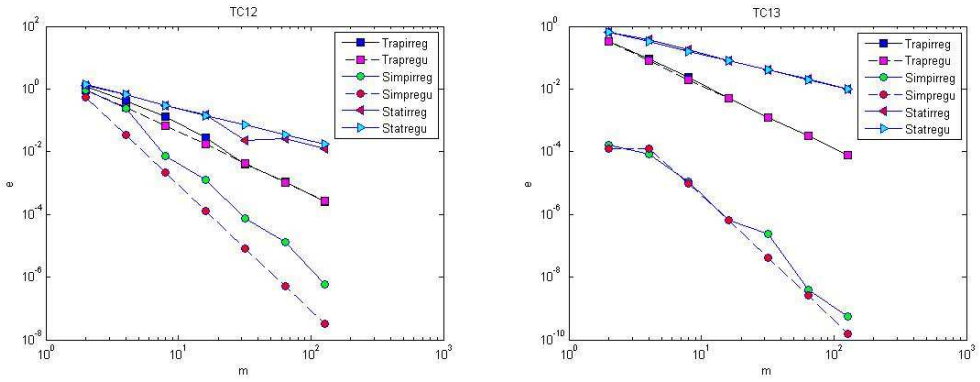

Figure A.6: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC12. (b) TC13.

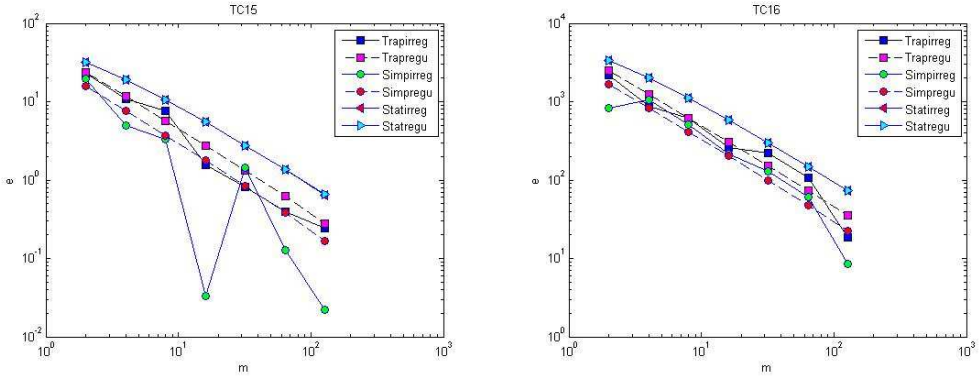

Figure A.7: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC15. (b) TC16.

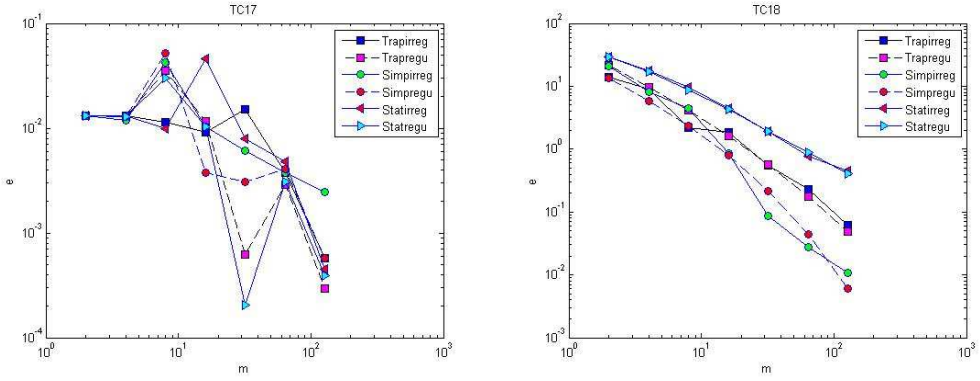

Figure A.8: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC17. (b) TC18.

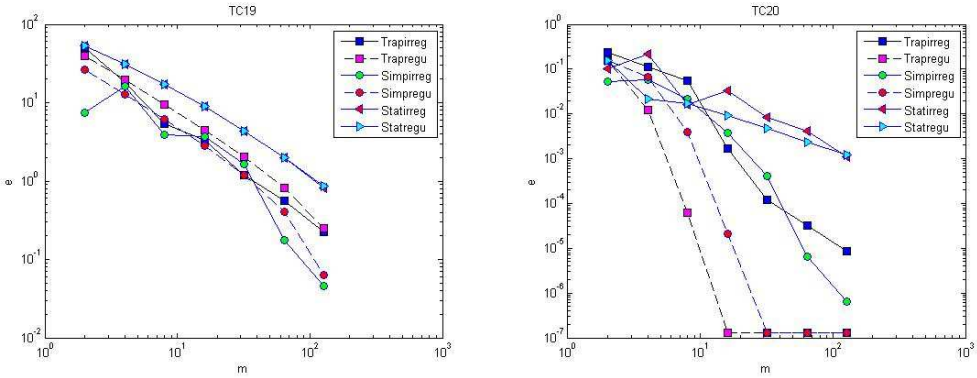

Figure A.9: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC19. (b) TC20.

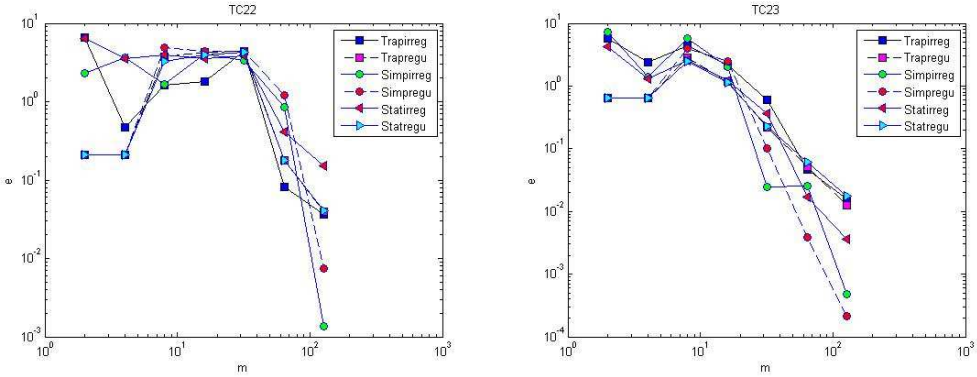

Figure A.10: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes. (a) TC22. (b) TC23.

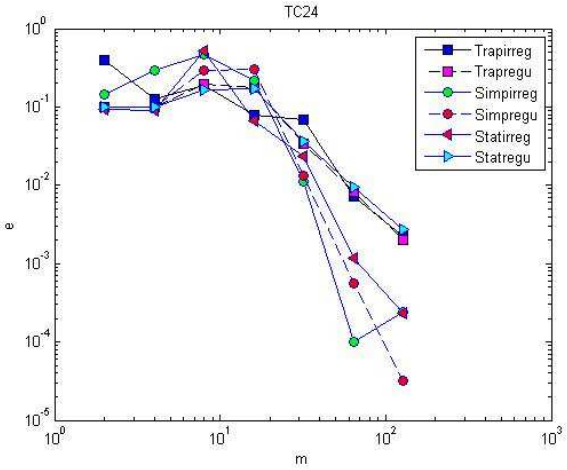

Figure A.11: Comparison of the convergence curves of the numerical integration methods on a sequence of irregular and regular grid nodes for TC24.

List of References

[1] G. Adler. Tropical forest fragmentation and isolation promote a synchrony among populations of a frugivorous rodent. *Journal of Animal Ecology,* 63: 903-911, 1994.

[2] C. Alexander, J. Holland, L. Winder, C. Woolley, and J. 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] V.I. Arnold, Ordinary Differential Equations. (translated by RA Silverman). *MIT Press, Cambridge, Massachusetts,* 2, 196-213, 1973.

[5] P. Amarasekare. Interactions between local dynamics and dispersal: Insights from single species models. *Theoretical Population Biology,* 53(1): 44-59, 1998.

[6] M. Ausden. Invertebrates. In: *Ecological census techniques: a handbook,* (ed. W. Sutherland), pp. 139-177, Cambridge University Press, 1996.

[7] S. Bates, J. Zhao, R. Roush, and A. Shelton. Insect resistance management in gm crops: past, present and future. *Nature Biotechnology,* 23(1): 57-62, 2005.

[8] G. Batzli, H. Jung, and G. Guntenspergen. Nutritional ecology of microtine rodents: Linear foraging-rate curves for brown lemmings. *Oikos,* 37(1): 112-116, 1981.

[9] D. Bearup, S. Petrovskii, R. Blackshow, and A. Hastings. The impacts of terrain and weather conditions on the metapopulation of Tipula Paludosa in South-Western Scotland: linking pattern to process. *American Naturalist*, 182: 393-409, 2013.

[10] D. Bearup, N. Petrovskaya, and S. Petrovskii. Some analytical and numerical approaches to understanding trap counts resulting from pest insect immigration. *Mathematical Biosciences*, 263: 143-160, 2015.

[11] J. Beddington, and C. Free. Age structure effects in predator-prey interactions. *Theoretical Population Biology,* 9(1): 15-24, 1976.

[12] M. Begon, L. Harper, and C. Townsend. *Ecology: Individuals, Population and Communities*. Blackwell Scientific publications, 1986.

[13] M. Benjamin, and D. Lawler. *Water quality engineering: Physical/chemical treatment processes*. John Wiley & Sons, 2013.

[14] A. Berryman, and P. Kindlmann. *Population systems: a general introduction.* Springer Science & Business Media, 2008.

[15] M. Binns, J. Nyrop, W. Van Der Werf, and W. Werf. *Sampling and Monitoring in Crop Protection: the theoretical basis for designing practical decision guides.* CABI, 2000.

[16] A. Birmingham, E. Kovacs, J. Lafontaine, et al. A new trap and lure for drosophila melanogaster (Diptera: Drosophilidae). *Journal of Economic Entomology,* 104(3): 1018-1023, 2011.

[17] O. Bjornstad, R. Ims, and X. Lambin. Spatial population dynamics: Analyzing patterns and processes of population synchrony. *Trends in Ecology and Evolution* 14(11): 427-432, 1999.

[18] R. Blackshaw. The annual leatherjacket survey in Northern Ireland, 196582, and some factors affecting populations. *Plant Pathology,* 32(3): 345-349, 1983.

[19] A. Blair, and S. Zahm. Cancer among farmers. *Occup. Med.,* 6(3): 335-354, 1991.

[20] C. Bliss. Statistical problems in estimating populations of Japanese beetle larvae. *Journal of Economic Entomology,* 34(2): 221-232, 1941.

[21] B. Boag, L. Deeks, A. Orr, and R. Neilson. A spatio-temporal analysis of a New Zealand flatworm (Arthurdendyus triangulatus) population in western Scotland. *Annals Applied Biology,* 147(1): 81-88, 2005.

[22] B. Boag, K. Mackenzie, J. McNicol, and R. Neilson. Sampling for the New Zealand flatworm. *Proceeding Crop Protection in Northern Britain,* 45-50, 2010.

[23] J. Boissonnat, and S. Nullans. Reconstruction of geological structures from heterogeneous and sparse data. *In Proceedings of the 4th ACM international workshop on Advances in geographic information system*, ACM, 172-179, 1996.

[24] C. Brezinski, and M. Zaglia. *Extrapolation methods. theory and practice*. Elsevier, 2013.

[25] C. Budd, W. Huang, and R. Russell. Adaptivity with moving grids. *Acta Numerica,* 18(111-241): 79-83, 2009.

[26] A. Burn, T. Coaker, and P. Jepson. *Integrated pest management.* Academic Press, New York, 1987.

[27] R. Burden, and J. Faires. *Numerical Analysis.* 6th ed. Brooks/Cole Publishing Company, Pacific Grove, USA, 1997.

[28] R. Butt. *Introduction to Numerical Analysis Using MATLAB.* Jones & Bartlett Learning, 2009.

[29] J. Byers, O. Anderbrant, and J. Lqvist. Effective attraction radius: a method for comparing species attraction and determining densities of flying insects. *Journal of Chemical Ecology,* 15(2): 749-765, 1989.

[30] J. Byers. Simulation and equation models of insect population control by pheromone-baited traps. *Journal of Chemical Ecology,* 19(9): 1939-1956, 1993.

[31] G. Caughley. Eruption of ungulate populations, with emphasis on Himalayan thar in New Zealand. *Ecology* 51(1): 53-72, 1970.

[32] P. Christou, T. Capell, A. Kohli, et al. Recent development and future prospects in insect pest control in transgenic crops. *Trends in Plant Science*, 11(6): 302-308, 2006.

[33] E. Coddington, & N. Levinson. *Theory of Ordinary Differential Equations*, Tata McGraw-Hill Education, New York, 1955.

[34] J. Crank. *The Mathematics of Diffusion*, volume 1. Clarendon Press Oxford, 1979.

[35] M. Crawley, and C. Long. Alternate bearing predator satiation and seedling recruitment in Quercus robur L. *Journal of Ecology,* 83: 683-696, 1995.

[36] G. Dahlquist, and A. Bjrck. Numerical Methods in Scientific Computing, volume 1. *Society for Industrial and Applied Mathematics*, 2008.

[37] P. Davis. Statistics for describing populations. In:*Handbook of sampling methods for Arthropods in Agriculture,* (eds. L. Pedigo, and G. Buntin), CRC Press, London, pp. 33-54, 1994.

[38] P. Davis, and P. Rabinowitz. *Methods of numerical integration.* Courier Corporation, 2007.

[39] P. Davis. Interpolation and Approximation. *Blaisdell Publishing Company, New York*, 1963.

[40] P. Davis, and P. Rabinowitz. *Methods of Numerical Integration.* Dover Publications INC, New York, 1975.

[41] P. Davis. Error of numerical approximation for analytic functions.*Journal of Rational Mechanics and Analysis,* volume 2, pp. 303-313, 1953.

[42] P. Davis. Error of numerical approximation for analytic functions. In *Survey of Numerical Analysis,* New York: McGraw-Hill, pp. 468-484, 1962.

[43] D. Dent. *Insect pest management.* CABI Publishing, Wallingford, 2000.

[44] S. Dunn, A. Constantinides, and P. Moghe. *Numerical Methods in Biomedical Engineering.* Academic Press, 2005.

[45] S. Dunbar. Travelling wave solutions of diffusive Lotka - Volterra equations. *Journal of Mathematical Biology,* 17(1): 11-32, 1983.

[46] N. Embleton, and N. Petrovskaya. On numerical uncertainty in evaluation of pest population size. *Ecological Complexity,* 14: 117-131, 2013.

[47] N. Embleton. *Handling sparse spatial data in ecological applications.* PhD Thesis, University of Birmingham, 2015.

[48] M. Eppstein, D. Hawrysz, A. Godavarty, and E. Sevick-Muraca. Three-dimensional Bayesian image reconstruction from sparse and noisy data sets: Near-infrared fluorescence tomography. *Proceedings of the National Academy of Sciences*, 99(15): 9619-9624, 2002.

[49] A. Ester, and K. van Rozen. Monitoring and control of Agriotes lineatus and A. obscurus in arable crops in the Netherlands. *IOBC Bulletin Insect Pathogens and Insect Parasitic Nematodes: Melolontha,* 28: 81-86, 2005.

[50] A. Ferguson, Z. Klukowski, B. Walczak, et al. 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 Tichomalus perfectus. *Entomologia Experimentalis et Applicata,* 95(2): 161-171, 2000.

[51] A. Ferguson, Z. Klukowski, B. Walczak, et al. Spatial distribution of pest insects in oilseed rape: implications for integrated pest management. *Agriculture, ecosystems & environment,* 95(2-3): 509-521, 2003.

[52] R. Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4): 507-521, 1915.

[53] S. Fllner, and A. Shmida. Why are adaptations for long-range seed dispersal rare in desert plants? *Oecologia* 51(1): 133-144, 1981.

[54] M. Gadgil. Dispersal: Population consequences and evolution. *Ecology,* 52(2): 253-261, 1971.

[55] A. Gayen. The frequency distribution of the product-moment correlation coefficient in random samples of any size draw from non-normal universes. *Biometrika,* 38(1/2): 219-247, 1951.

[56] F. Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15: 246-263, 1886.

[57] S. Garber, and J. Burger.A 20-Yr Study Documenting the Relationship Between Turtle Decline and Human Recreation. *Ecology Applications,* 5(4): 1151-1162, 1995.

[58] A. Gatehouse, N. Ferry, M. Edwards, and H. 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.

[59] M. Gilpin, and M. Rosenzweig. Enriched predator-prey systems: Theoretical stability. *Science* 177(4052): 902-904, 1972.

[60] L. Ginzburg, and P. Inchausti. Asymmetry of population cycles: Abundance-growth representation of hidden causes in ecological dynamics. Oikos, 80: 435-447, 1997.

[61] P. Glendinning. *Stability, instability and chaos: an introduction to the theory of nonlinear differential equations.* Cambridge university press, 1994.

[62] D. Goodman. The demography of chance extinction. *Viable Population for Conservation* , pp. 11-34, 1987.

[63] G. Grimmett, and D. Stirzaker. *Probability and random process.* 3rd ed. Oxford University Press, Oxford, 2001.

[64] J. Gwinner, R. Harnisch, and O. Mck. *Manual on the prevention of post-harvest grain losses.* GTZ, Eschborn, 1996.

[65] J. Hagler and C. Jackson. Methods for marking insects: Current techniques and future prospects. *Annual Review of Entomology,* 46(1): 511-543, 2001.

[66] I. Hanski. Four kinds of extra-long diapause in insects: A review of theory and observations. *In Annales Zoologici Fennici.* Finnish Academy of Sciences, Societas Scientiarum Fennica, Societas pro Fauna et Flora Fennica and Societas Biologica Fennica Vanamo, 25: 37-53, 1988.

[67] I. Hanski. Metapopulation dynamics: does it help to have more of the same? *Trends in Ecology & Evolution,* 4(4): 113-114, 1989.

[68] I. Hanski, and I. Woiwod. Delayed density dependence. *Nature*, 350, 28, 1991.

[69] I. Hanski, and I. Woiwod. Spatial synchrony in the dynamics of moth and aphid populations. *Journal of Animal Ecology,* 62: 656-668, 1993.

[70] S. Harrison, and J. Quinn. Correlated environments and the persistence of metapopulations. *Oikos,* 56: 293-298, 1989.

[71] D. Haydon, and H. Steen. The effects of large- and small-scale random events on the synchrony of metapopulation dynamics: a theoretical analysis. *Proceedings of the Royal Society of London B, Biological Sciences,* 264(1386): 1375-1381, 1997.

[72] P. Hennig, M. Osborne, & M. Girolami. Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. A,* 471(2179), 20150142, 2015.

[73] P. Hennig, & C. Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research,* 13(Jun), 1809-1837, 2012.

[74] F. Hildebrand. Finite-difference equation and simulations. *Prentice-Hall, Englewood Cliffs, New Jersey,* 1968.

[75] M. Hirsch, & S. Smale. Differential equations, *Dynamical systems and linear algebra, Academic Press, New York*, 1974.

[76] J. Hoffman, and S. Frankel. *Numerical Methods for Engineering and Scientists.* CRC press, 2001.

[77] H. Hokkanen. Trap cropping in pest management. *Annual Review of Entomology,* 36(1): 119-138, 1991.

[78] J. Holland, J. 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.

[79] C. Holling. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *The Canadian Entomologist,* 91(5): 293-320, 1959a.

[80] C., Holling. Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist,* 91(7): 385 398, 1959.

[81] M. Holyoak, and S. Lawler. The role of dispersal in predator-prey metapopulation dynamics. *Journal of Animal Ecology*, 65: 640-652, 1996.

[82] S. Hutchins. Techniques for sampling arthropods in integrated pest management. In *Handbook of sampling methods for arthropods in agriculture, CRC Press, London,* pp. 74-96, 1994.

[83] R. Ims, and H. Steen. Geographical synchrony in microtine population cycles: A theoretical evaluation of the role of nomadic avian predators. *Oikos,* 57: 381-387, 1990.

[84] V. Jansen. Regulation of predator-prey systems through spatial interactions: A possible solution to the paradox of enrichment. *Oikos,* 74: 384-390, 1995.

[85] V. Jansen. The dynamics of two diffusively coupled predator-prey populations. *Theoretical Population Biology,* 59(2): 119-131, 2001.

[86] F. Jeltsch, C. Wissel, S. Eber, and R. Brandl. Oscillating dispersal patterns of tephritid y populations. *Ecological Modelling,* 60(1): 63-75, 1992.

[87] P.Jepson, and J. Thacker. Analysis of the spatial component of pesticide side-effects on non-target invertebrate populations and its relevance to hazard analysis. *Functional Ecology,* 4: 349-355, 1990.

[88] W. Koenig. Spatial autocorrelation in California land birds. *Conservation Biology,* 12(3): 612-620, 1998.

[89] M. Kogan. Integrated pest management: Historical perspective and contemporary developments. *Annual Review of Entomology,* 43(1): 243-270, 1998.

[90] I. Korostil, G. Peters, J. Cornebise, & D. Regan. Adaptive Markov chain Monte Carlo forward projection for statistical analysis in epidemic modelling of human papillomavirus. *Statistics in medicine,* 32(11), 1917-1953, 2013.

[91] D. Legg, and R. Moon. Bias and variability in statistical estimates. In: *Handbook of Sampling Methods for Arthropods in Agriculture, CRC Press, Boca Raton, FL,* 55-69, 1994.

[92] E. Leigh Jr. The average lifetime of a population in a varying environment. *Journal of Theoretical Biology,* 90(2): 213-239, 1981.

[93] H. Levy, and F. Lessman. *Finite difference equations.* Dover Publications, Inc., New York, 1961.

[94] S. Levin. Physical and biological scales and the modeling of predator- prey interactions in large marine ecosystems. *Large marine ecosystems: patterns, processes and yields,* Washington: American Association for the Advancement of Science, pp. 179-187, 1990.

[95] M. Lewis, and P. Kareiva. Allee dynamics and the spread of invading organisms. *Theoretical Population Biology* 43(2), 141-158, 1993.

[96] Z. Li. High convergence rates of digital image transformation by numerical integration using spline functions. *Computers & Mathematics with Applications,* 41(1): 229-255, 2001.

[97] S. Liao, and M. Pawlak. On image analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18(3): 254-266, 1996.

[98] A. Liebhold, W. Koenig, and O. Bjrnstad. Spatial synchrony in population dynamics. *Annu. Rev. Ecol. Evol. Syst*, 35: 467-490, 2004.

[99] M. Liebman, and E. Dyck. Crop rotation and intercropping strategies for weed management. *Ecological Applications* 3(1): 92-122, 1993.

[100] A. Lotka. Elements of physical biology. Williams and Wilkins, Baltimore. D. Ludwig, D. Jones, and C. Holling (1978). Qualitative analysis of insect outbreak systems: The spruce budworm and forest. *Journal of Animal Ecology* 47: 315-332, 1925.

[101] F. Louws, C. Rivard, and C. Kubota. Grafting fruiting vegetables to manage soilborne pathogens, foliar pathogens, arthropods and weeds. *Scientia Horticulturae,* 127(2): 127-146, 2010.

[102] L. Markus, & H. Yamabe. Global stability criteria for differential systems. *Osaka Mathematical Journal*, 12(2), 305-317, 1960.

[103] H. Malchow, S. Petrovskii, and E. Venturino. *Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation. Mathematical and Computational Biology Series.* Boca Raton: Chapman & Hall/CRC, 2008.

[104] H. Malchow, and L. Schimansky-Geier. *Noise and diffusion in bistable nonequilibrium systems*, Volume 5 of Teubner-Texte zur Physik. Leipzig: Teubner-Verlag, 1985.

[105] H. Malchow, and S. V. Petrovskii. Dynamical stabilization of an unstable equilibrium in chemical and biological systems. *Mathematical and Computer Modelling*, 36(3): 307-319, 2002.

[106] R. Mason. Synchronous patterns in an outbreak of the Douglas-fir tussock moth. *Environmental Entomology* 7(5): 672-675, 1978.

[107] J. Maynard-Smith. *Models in Ecology.* Cambridge University Press, Cambridge, 1978.

[108] J. Mayor, and M. Davies. A survey of leatherjacket populations in south-west England, 1963-1974. *Plant Pathology,* 25(3): 121-128, 1976.

[109] R. May. Limit cycles in predator-prey communities. *Science*, 177(4052): 900-902, 1972.

[110] R. May. *Stability and complexity in model ecosystems.* Volume 6, Princeton: Princeton University Press, 2001.

[111] M. Mahsereci, & P. Hennig. Probabilistic line searches for stochastic optimization. *In Advances in Neural Information Processing Systems*, 181-189, 2015.

[112] R. Metcalf, and W. Luckmann (eds). *Introduction to insect pest management.* John Wiley & Sons, 1994.

[113] A. Milne. The centric systematic area-sample treated as a random sample. *Biometrics*, 15(2): 270-297, 1959.

[114] J. Milnor. Dynamics in one complex variable. second edition, 2000.

[115] A. Morozov, S. Petrovskii, and B. L. Li. Bifurcations and chaos in a predator-prey system with the Allee effect. *Proceedings of Royal Society of London B: Biological Sciences,* 271, 1407-1414, 2004.

[116] A. Morozov, S. Petrovskii, and B. L. Li. Spatiotemporal complexity of patchy invasion in a predator-prey system with the Allee effect. *Journal of theoretical Biology,* 238(1): 18-35, 2006.

[117] K. Morton, and R. Richtmyer. Difference methods for initial value problems. 2nd Edition, *Wiley-Interscience, New York,* 1967.

[118] P. Moran. The statistical analysis of the Canadian Lynx cycle. *Australian Journal of Zoology* 1(3): 291-298, 1953.

[119] J. Murray. *Mathematical Biology: An Introduction,* volume 1. Springer-Verlag Berlin,1989.

[120] J. Myers. Synchrony in outbreaks of forest Lepidoptera: A possible example of the Moran effect. *Ecology,* 79(3): 1111-1117, 1998.

[121] J. Myers. Can a general hypothesis explain population cycles of forest Lepidoptera? *Advances in Ecological Research, Academic Press,* 18: 179-242, 1988.

[122] I. Newton. *Mathematical Differentials.* (ed. W. Jones) London, 1711.

[123] 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.

[124] E. Oerke. Crop losses to pests. *The Journal of Agricultural Science* 144(1): 31-43, 2006.

[125] A. Okubo. Diffusion and ecological problems: Mathematical models. *Volume 10 of Biomathematics Texts. Berlin: Springer*, 1980.

[126] L. Ott, and M. Longnecker. *An Introduction to Statistical Methods and Data Analysis.*Nelson Education, Duxbury Press, 2008.

[127] M. Owen, and M. Lewis. How predation can slow, stop or reverse a prey invasion. *Bulletin of mathematical biology,* 63(4), 655-684, 2001.

[128] M. Pascual, and P. Kareiva. Predicting the outcome of competition using experimental data: Maximum likelihood and Bayesian approaches. *Ecology,* 77(2): 337-349, 1996.

[129] K. Pearson. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58: 240-242. 1895.

[130] S. Pearce, and M. Zalucki. Do predators aggregate in response to pest density in agroecosystems? Assessing within-field spatial patterns. *Journal of Applied Ecology,* 43(1): 128-140, 2006.

[131] L. Pedigo, and m. Rice. *Entomology and pest management.* Pearson Prentice Hall, New Jersey, 2009.

[132] M. Peltonen, A. Liebhold, O. Bjrnstad, and D. Williams. Variation in spatial synchrony among forest insect species: Roles of regional stochasticity and dispersal. *Ecology,* 83: 3120-3129, 2002.

[133] S. Petrovskii, A. Morozov, and E. Venturino. Allee effect makes possible patchy invasion in a Predator-Prey system. *Ecology Letters,* 5(3): 345-352, 2002.

[134] N. Petrovskaya, and N. Embleton. Evaluation of peak functions on ultra-coarse grids. In*Proc. R. Soc. A.*Vol. 469. No. 2153. The Royal Society, 2013.

[135] N. Petrovskaya, N. Embleton, and S. Petrovskii. Numerical study of pest population size at various diffusion rates. In: *Dispersal, Individual Movement and Spatial Ecology: A Mathematical Perspective,* Springer Berlin Heidelberg, pp. 355-386, 2013.

[136] N. Petrovskaya, and S. Petrovskii. The coarse-grid problem in ecological monitoring. In:*Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* The Royal Society, 466: 2933-2953, 2010.

[137] N. Petrovskaya, S. Petrovskii, and A. Murchie. Challenges of ecological monitoring: estimating population abundance from sparse trap counts. *Journal of The Royal Society Interface* 9(68): 420-435, 2012.

[138] N. Petrovskaya, and E. Venturino. Numerical integration of sparsely sampled data. *Simulation Modelling Practice and Theory,* 19(9): 1860-1872, 2011.

[139] N. Petrovskaya, and N. Embleton. Computational methods for accurate evaluation of pest insect population size. *Ecological Modelling Applied to Entomology*, Springer International Publishing, pp. 171-218, 2014.

[140] N. Petrovskaya. Catch me if you can: Evaluating the population size in the presence of a spatial pattern. *Ecological Complexity* 2017.

[141] S. Petrovskii, N. Petrovskaya, and D. Bearup. Multiscale approach to pest insect monitoring : Random walks , pattern formation, synchronization, and networks. *Physics of life reviews,* 11(3): 467-525, 2014.

[142] N. Petrovskaya, and S. Petrovskii. Catching ghosts with a coarse net: Use and abuse of spatial sampling data in detecting synchronization. *Journal of The Royal Society Interface,* 14(127), 2017.

[143] D. Pimentel. *Techniques for reducing pesticide use: Economic and environmental benefits.* John Wiley & Sons, 1997.

[144] D. Pimentel. Pesticides and pest control. In:*Integrated pest management: innovation-development process,* Springer Netherlands, pp. 83-87, 2009.

[145] D. Pimentel. Amounts of pesticides reaching target pests: Environmental impacts and ethics. *Journal of Agricultural and Environmental Ethics,* 8(1): 17-29, 1995.

[146] D. Pimentel, and M. H. Pimentel, eds. *Food, Energy, and Society.* CRC press, 2007.

[147] E. Pollard, M. Hall, and T. Bibby. *Monitoring the abundance of butterflies 1976-1985.* Research and Survey in Nature Conservation No. 2, 1986.

[148] E. Pollard. Synchrony of population fluctuations: The dominant influence of widespread factors on local butterfly populations. *Oikos*, 60: 7-10, 1991.

[149] T. Powell. Physical and biological scales of variability in lakes, estuaries and the coastal ocean. *Ecological Time Series. Chapman & Hall, New York,* pp. 119-139, 1995.

[150] A. Quarteroni, R. Sacco and F. Saleri. *Numerical Mathematics.* Volume 37. Springer Science & Business Media, 2010.

[151] S. Raimondo, A. Liebhold, J. Strazanac, and L. Butler. Population synchrony within and among Lepidoptera species in relation to weather, phylogeny, and larval phenology. *Ecological Entomology,* 29(1): 96-105, 2004.

[152] S. Ralston, and H. Williams. Numerical integration of daily growth increments: An efficient means of ageing tropical fishes for stock assessment. *Fishery Bulletin,* 87(1): 1-16, 1989.

[153] E. Ranta, J. Lindstrom, and H. Linden. Synchrony in tetraonid population dynamics. *Journal of Animal Ecology,* 64: 767-776, 1995.

[154] E. Ranta., V. Kaitala, K. Lindstrom, and E. Helle. The Moran effect and synchrony in population dynamics. *Oikos*, 78: 136-142, 1997.

[155] E. Ranta, J. Lindstrom, V. Kaitala, and H. Linden. Synchrony in population dynamics. *Proceedings of the Royal Society of London*, 262(1364): 113-118, 1995.

[156] D. Raworth, and M. Choi. Determining numbers of active carabid beetles per unit area from pitfall-trap data. *Entomologia Experimentalis et Applicata,* 98(1): 95-108, 2001.

[157] J. Readshaw. The numerical response of predators to prey density. In: Hughes,Ed., Quantitative Evaluation of Natural Enemy Effectiveness. *Journal Applied Biology,* 10: 342-351, 1973.

[158] L. Real. The kinetics of functional response. *The American Naturalist,* 111(978): 289-300, 1977.

[159] W. Reisen, and H. Lothrop. Effects of sampling design on the estimation of adult mosquito abundance. *Journal of the American Mosquito Control Association,* 15(2): 105-114, 1999.

[160] M. Rosenzweig. Paradox of enrichment: Destabilization of exploitation ecosystems in ecological time. *Science,* 171(3969): 385-387, 1971.

[161] M. Rosenzweig, and R. MacArthur. Graphical representation and stability conditions of predator-prey interactions. *The American Naturalist,* 97(895): 209-223, 1963.

[162] J. Ruberson. *Handbook of Pest Management.* CRC Press, 1999.

[163] G. Seber. The estimation of animal abundance and related parameters. London: Charles Griffin, 1982.

[164] J. Sherratt, and M. Smith. Periodic traveling waves in cyclic populations: Field studies and reaction- diffusion models. *Journal of The Royal Society Interface,* 5(22): 483-505, 2008.

[165] J. Sherratt, M. Lewis, and A. Fowler. Ecological chaos in the wake of invasion. *Proceedings of the National Academy of Sciences,* 92(7), 2524-2528, 1995.

[166] J. Sherratt. Irregular wakes in reaction-diffusion waves. *Physica D: Nonlinear Phenomena,* 70(4): 370-382, 1994.

[167] J. Sherratt. On the evolution of periodic plane waves in reaction-diffusion equations of lambda-omega type. *SIAM Journal on Applied Mathematics,* 54(5): 1374-1385, 1994.

[168] N. Shigesada, and K. Kawasaki. *Biological invasions: Theory and practice.* Oxford, UK: Oxford University Press. 1997.

[169] A. Shoffner, and J. 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.

[170] J. Silver. *Mosquito Ecology: Field Sampling Methods*. Springer Science and Business Media, 2007.

[171] A. Smigocki, S. Ivic-Haymes, H. Li, and J. Savi. Pest protection conferred by a beta vulgaris serine proteinase inhibitor gene. *PLoS one*, 8(2), e57303, 2013.

[172] G. Snedecor, and W. Cochran. *Statistical methods.* The Iowa State University Press, Ames, 1980.

[173] M. Solomon. The natural control of animal populations. *The Journal of Animal Ecology*, 18: 1-35, 1949.

[174] T. Southwood, and P. Henderson. *Ecological methods*. Third edition, Blackwell Science Ltd., 2000.

[175] S. Sudman, M. Sirken, and C. Cowan. Sampling rare and elusive populations. *Science* 240: 991-996, 1988.

[176] O. Sutcliffe, C. Thomas, T. Yates, and J. Greatorex-Davis. Correlated extinctions, colonizations and population fluctuations in a highly connected ringlet butterfly metapopulation. *Oecologia,* 109(2): 235-241, 1997.

[177] W. Sutherland. *Ecological census techniques: A handbook.* Cambridge University Press, 2006.

[178] O. Sutcliffe, C. Thomas, and D. Moss. Spatial synchrony and a synchrony in butterfly population dynamics. *Journal of Animal Ecology* 65: 85-95, 1996.

[179] V. Stern. Economic thresholds. *Annual review of entomology,* 18(1): 259-280, 1973.

[180] V. Stern,R. Smith, R. Bosch, and K. Hagen. The integration of chemical and biological control of the spotted alfalfa aphid: the integrated control concept. *Hilgardia,* 29(2): 81-101, 1959.

[181] H. Steen, R. Ims, and G. Sonerud. Spatial and temporal patterns of small-rodent population dynamics at a regional scale. *Ecology,* 77(8): 2365-2372, 1996.

[182] J. Stoer, and R. Bulirsch. *Introduction to Numerical Analysis.* Springer Science & Business Media, 1980.

[183] D. Strebel. Environmental fluctuations and extinction-single species. *Theoretical Population Biology,* 27(1): 1-26, 1985.

[184] A. Taboada, C. Prez-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.

[185] L. Taylor, I. Woiwod, and J. Perry. The density-dependence of spatial behaviour and the rarity of randomness. *The Journal of Animal Ecology,* 47: 383-406, 1978.

[186] C. Thomas. Spatial and temporal variability in a butterfly population. *Oecologia,* 87(4): 577-580, 1991.

[187] L. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review,* 50(1): 67-87, 2008.

[188] P. Turchin. *Complex Population Dynamics: A Theoretical/Empirical Synthesis*, Volume 35. Princeton University Press, 2003.

[189] A. Turing. On the chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 237(641): 37-72, 1952.

[190] H. Van Der Werf. Assessing the impact of pesticides on the environment. *Agricultural, Ecosystem & Environment,* 60(2-3): 81-96, 1996.

[191] D. Venable, and L. Lawlor. Delayed germination and dispersal in desert annuals: escape in space and time. *Oecologia,* 46(2):272-282, 1980.

[192] F. Verhulst. *Nonlinear differential equations and dynamical systems.* Springer Science & Business Media, 2006.

[193] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature,* 118(2972): 558-560, 1926.

[194] M. Wang, and M. Kot. Speeds of invasion in a model with strong or weak Allee effects.*Mathematical biosciences,* 171(1), 83-97, 2001.

[195] D. Williams, and A. Liebhold. Influence of weather on the synchrony of gypsy moth (Lepidoptera: Lymantriidae) outbreaks in New England. *Environmental Entomology* 24(5): 987-995, 1995.

[196] B. Woodcock. Pitfall trapping in ecological studies. In: *Insect Sampling in Forest Ecosystem,* (Ed. S. Leather), *Wiley Online Library*, pp. 37-57, 2005.

[197] Q. Wu, H. Xu, X. Zou. An effective methods for 3D geological modeling with multi-source data integration. *Computers & Geosciences*, 31(1): 35-43, 2005.

[198] S. Yachi. Spatial patterns of propagating waves of fox rabies. *Forma,* 4: 3-12, 1989.

[199] L. Yaroslavsky, A. Moreno, and J. Campos. Frequency responses and resolving power of numerical integration of sampled data. *Optics Express,* 13(8): 2892-2905, 2005.