

Modelling and Classifying Stochastically Episodic Events

by
Colin Bellinger

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfilment of
the requirements for the degree of
Master of Computer Science

Carleton University
Ottawa, Ontario

© 2010, Colin Bellinger

Abstract

Pattern Recognition (PR) in its standard form, which involves training and testing classifiers on a representative set of data drawn from a domain of interest, has been applied to automate an immense number of classification tasks. More recently, a challenging set of so-called “one-class” classification problems have been identified and explored. In this thesis, we introduce a further challenging class of PR problems, involving the recognition of Stochastically Episodic (SE) events, and present a first attempt at classifying them within their characteristic fields of background noise. More specifically, this class of problems is characterized by the presence of an overwhelming number of background measurements, which are acquired in the form of a time-series. The time-series is, however, interwoven with a minute number of random (in time, space and magnitude) SE events, which are deemed to be of considerable interest and require classification. The rarity and random nature of the SE events, along with their presence within a time-series of noise-like measurements, renders the learning of their corresponding distribution extremely difficult, if not entirely impossible. By extension, the classification of the SE events is an extremely interesting and ambitious undertaking.

Since the acquisition of a sufficient number of SE events is, by definition, unachievable, we propose a flexible framework for the modelling and simulation of such events, as they propagate through a field of background noise. In practice, the initiator of the SE event may take many forms, such as an earthquake, tsunami, erroneous release of pollution into the environment, etc..

The thesis, thereafter, considers the PR of these events from two perspectives; namely those of binary and one-class classification. The thesis contains an empirical demonstration of all these concepts, based on the exemplary scenario that is suggested by the verification of the Comprehensive Test-Ban-Treaty (CTBT).

Acknowledgements

This profoundly rewarding work would not have been possible without the encouragement, patience, advice, and support of a great number of friends and colleagues. Thus, I cannot conceive of a more appropriate way to open this thesis than with a series of heartfelt acknowledgements.

First and foremost, I must thank my loving parents, grandparents, and brother as they started me on this journey, and continue to be there for me, taking whatever role is required of them.

I am also indebted, and forever grateful to Dr. Oommen for the countless hours he has dedicated to advising me, including time during his well-earned vacation. In addition, I would like to thank him for being a positive and truly inspirational role model over the past two years.

I am, of course, extremely grateful to Dr. Japkowicz for her time and insight into numerous issues concerning machine learning. I am additionally thankful to Dr. Japkowicz and Dr. Ungar for their financial support, and the introduction to the intriguing demonstration domain that so frequently appears in the thesis.

I record my gratitude to my Thesis Committee, as they have taken time out of their undoubtedly taxing schedules to provide me with their insightful comments.

Where would I, and indeed the School of Computer Science as a whole, be without the dedication of the support staff. I would like to thank them for everything they have done during my time as a student at Carleton University. I would also like to thank the many wonderful people I have met along the way, include Cesar, Amir, Raghad, and most recently JC, to name but a few.

On a more personal basis, I would like to thank all of my wonderful friends for their support over the last two years. Most importantly, I would like to thank Gillian, as she has been a committed partner throughout the long days and late nights, and has patiently read through more work on pattern recognition and simulation than she could possibly have imagined at the outset of this journey. She has supported me emotionally, and without her expertise my writing could not have evolved to its current state.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Goal	3
1.3	Problem Statement	3
1.4	Contribution	4
1.5	Scope	5
1.6	Organization	5
2	Background	7
2.1	Introduction	7
2.2	Atmospheric Processes	8
2.3	Energy Sources	10
2.4	Air Stability	10
2.5	Implications for Pollutants	12
2.6	Local Wind Structure	13
2.7	Frames of Reference	15
2.8	Statistical Descriptions of Advection and Diffusion	16
3	Literature Review	19
3.1	Data Generation	20
3.1.1	Radioxenon Emitters	20
3.1.2	Estimating the State of the Troposphere	23
3.1.3	Modelling Atmospheric Transport	29

3.2	Classification	46
3.2.1	Density Estimation	51
3.2.2	One-Class Extensions to Binary Classifiers	53
3.2.3	Nearest Neighbour	56
3.2.4	Combined Density and Class Probability Estimation	57
4	Simulating Stochastically Episodic Events	58
4.1	Introduction	58
4.2	Motivation	58
4.3	Modelling System	60
4.3.1	Modelling the Background	62
4.3.2	Derivation of Fundamental Equation of K-Theory	63
4.4	Demonstration of the Dispersion Models	68
4.5	Experimental Setup	76
4.5.1	Motivation	76
4.5.2	Modelled System	78
4.6	Results	81
4.6.1	Background	81
4.6.2	SE Events	82
4.7	Discussion	86
4.8	Conclusion	88
5	Classifying Stochastically Episodic Events	90
5.1	Introduction	90
5.2	Characteristics of the Domain of Problems	92
5.3	Pattern Recognition Algorithms	94
5.3.1	Binary Classification	94
5.3.2	One-Class Classification	109
5.4	Experimental Setup	116
5.4.1	Application Domain	116
5.4.2	Simulation Aspects: Procuring Data Collection	117
5.4.3	Source Position Recalculation	118

5.4.4	Simulation Setup	120
5.4.5	Classification Scenarios	121
5.4.6	Classifier Assessment Criteria	123
5.5	Results: Scenario 1	123
5.5.1	General Performance	124
5.5.2	Performance on Short- and Long-Range Detonations	126
5.5.3	Performance as a Function of Distance	128
5.5.4	Expanded Feature-Space	132
5.6	Results: Scenario 2	137
5.6.1	General Performance	137
5.6.2	Performance on Short- and Long-Range Detonations	139
5.6.3	Performance as a Function of Distance	140
5.6.4	Expanded Feature-Space	142
5.7	Discussion	142
5.7.1	Results: S1	142
5.7.2	Results: S2	145
5.8	Conclusion	147
6	Conclusion and Future Research	148
6.1	Contribution	148
6.1.1	Modelling and Simulation	149
6.1.2	Pattern Recognition	149
6.2	Future work	150
6.2.1	Modelling and Simulation	150
6.2.2	Pattern Recognition	151
	Bibliography	153

List of Tables

2.1	Atmospheric stability related to lapse rates	12
3.1	Background cumulative ^{133}Xe concentrations	23
5.1	One-class classification domain comparison	93
5.2	Overview of S1 classification results	126
5.3	Overview of S2 classification results	139

List of Figures

2.1	Vertical temperature profile	9
2.2	The internal structure of the troposphere.	9
2.3	Plume shape	14
3.1	Demonstration of overfitting	48
3.2	Specialized model	49
3.3	Optimal model	49
3.4	Feedforward artificial neural networks	53
4.1	Gaussian plume demonstration	71
4.2	Effect of wind speed on background noise	73
4.3	Effect of plume height on background noise	74
4.4	Effect of seasonal variations background noise	75
4.5	Effect of wind speed on SE events	77
4.6	Background distribution	83
4.7	Background time-series scatter plot	84
4.8	SE event time-series scatter plot	85
4.9	SE event propagation	87
5.1	Feedforward artificial neural networks	103
5.2	Maximum Margin Hyperplane for the SVM	107
5.3	Demonstration of the one-class Nearest Neighbour	112
5.4	Neural structure of the autoassociator classifier	114
5.5	Plume Trajectory Map	119

5.6	Iterative simulation scenario	121
5.7	Box plot of general classifier performance	124
5.8	Box plot of classifier performance over a long-range	128
5.9	Plot of classifier performance (AUC) as a function of Distance	130
5.10	Plot of classifier performance (FPR) as a function of Distance	131
5.11	Expanded feature-space (binary learners)	134
5.12	Expanded feature-space (one-class learners)	135
5.13	Expanded feature-space (boxplot comparison)	136
5.14	Box plot of general classifier performance	138
5.15	Box plot of classifier performance over a long-range	140
5.16	Plot of classifier performance (AUC) as a function of Distance	141
5.17	Plot of classifier performance (FPR) as a function of Distance	141
5.18	Expanded feature-space (one-class learners)	143
5.19	Expanded feature-space (boxplot comparison)	144

Chapter 1

Introduction

1.1 Introduction

Pattern Recognition (PR) systems, from simple instance-based learners [85] to the state-of-the-art in support vector machines and artificial neural networks [49, 97], have been widely applied to automate processes that previously required the active involvement of human agents. This acceptance of PR systems within both the public and private sectors stems from a long standing recognition of the fact that computer automation can:

- Free humans from tedious tasks.
- Lower costs through a reduced reliance on domain experts.
- Amongst a host of other benefits, significantly increase accuracy and efficiency.

In general, the objective of PR, as a field of study within the broader domain of machine learning, is to “learn” to distinguish between, what in the literature are commonly referred to as classes, states-of-nature, and concepts [28, 68, 126].

In their standard formulations, PR systems assume the existence of two classes of data, cancerous tissue and non-cancerous tissue, for example, which must be discriminated between via the so-called learned discriminate function. The specifics of the

physical discriminate function are ultimately dependent upon the particular classification problem to be solved, and are formulated during the “learning” (or “training”) process. This process is largely unique to the individual PR algorithms; however, the essential commonality that exists in all standard PR systems is their reliance on a labelled set of training instances, which are assumed to be drawn independently and identically from both states-of-nature. It is, indeed, such training sets that facilitate the derivation of the discriminant function.

Recently, a new, and more challenging domain of PR problems have been identified. This new class of problems has previously been denoted as outlier detection, novelty detection and one-class classification [34, 42, 111]. The key feature within this class of problems is that a strong representative set of training instances can only be acquired from a single class within the binary problem, thus, inhibiting the training of a standard PR system.

The recognition of this problem has motivated significant research into the so-called domain of one-class classification, with considerable success in spite of the challenges introduced by the absence of a second class for use during training.

More in-depth discussion of PR, both binary and one-class PR, is conducted in Chapter 3.

In the most extreme case, in particular, where not only is the data insufficient to train standard binary classifiers, but where the supply of data drawn from the second class is insufficient to test and validate one-class classifiers, alternative solutions for model selection must be explored.

Indeed, this is the case in the newly defined sub-category of one-class learning problems that we present and explore in this thesis. We observe that the characteristics of this sub-category are such that the vast majority of instances extracted from the application domain form a time-series of so-called background noise. However, the background noise is, under extremely rare and unpredictable circumstances, interrupted by a minute number of random (in time, space and magnitude) Stochastically Episodic (SE) events, which are of considerable interest, such as earthquakes, tsunamis and erroneous releases of pollution into the environment.

It is, indeed, a first attempt at the classification of such events, which we have submitted for publication in [7], and that composes the body of this thesis. In addition, due to the rarity of these events, we explore and develop a strategy to model and simulate SE events as they propagate through a field of background noise. Most recently, a portion of this research on modelling and simulation was published in [8]. The framework is subsequently applied to generate artificial datasets suitable for the exploration of PR within this new sub-category of one-class learning.

1.2 Goal

In this thesis, we aim to present a first attempt at classifying rare SE events, which, as we previously noted, are interwoven in a field of background noise.

In doing so, we recognize that the limited presence of the SE events creates a natural draw towards the use of artificial data. Thus, in the absence of an existing framework, we propose a complementary objective. In particular, the derivation of a modelling and simulation framework that simultaneously facilitates the scenario-specific exploration of the SE events as they propagate through the system of background noise, and the generation of labelled PR datasets¹.

1.3 Problem Statement

In this thesis, two very distinct, and yet complementary problems are solved. In particular, we demonstrate how, through a framework that utilizes a divide-and-conquer strategy, this relatively unexplored domain, which is characterized by a time-series of well-defined background noise that is interwoven with a minute number of unpredictable, random SE events, can be modelled and simulated.

The problem that is inherent in the modelling and simulation of such systems is that the vast majority of measurements compose the so-called background noise.

¹While Health Canada had kindly released data for the exploration of our demonstration domain [106], our investigation showed that this data is neither appropriate for the exploration of the problem in terms of SE events, nor does it, even remotely, reflect the *a priori* class probabilities of the problem.

Thus, when combined with the random and unpredictable nature of the SE events, few assumptions can accurately be made about the distribution of the class of such events. This information is typically essential when modelling and simulating any system.

In terms of PR, we have identified two scenarios in this particularly challenging sub-category of problems. In the first scenario, we assume that although the SE events are rare and difficult to identify, it is possible through clustering/PR means to separate the SE events in the acquired time-series from the background data, and to assign the appropriate labels to the instances of the training set. Thus, in this first scenario, both binary and one-class classification have been explored.

More appropriately, given the reality of many SE events, the second scenario recognizes the extreme challenges that are inherent in manually identifying the presence of a SE event in the acquired time-series. Thus, we propose a one-class PR task, which insists that the PR system be trained only using noisy data. More specifically, the one-class classifiers must learn from training data, which, while largely composed of background instances, is distorted by a minute number of *unidentifiable* instances of the SE event class.

1.4 Contribution

In summary, this thesis contains two substantial contributions; one related to the field of modelling and simulation, and the other to PR.

In particular, we propose a novel framework that facilitates scenario-specific modelling and simulation of SE events, both in terms of their unpredictable occurrences within the system, and their subsequent propagation through the system. This framework is applicable to the generation of labelled time-series datasets on which the PR of SE events can be explored. In addition, it facilitates the exploratory modelling and simulation of SE events, such as earthquakes, tsunamis and erroneous releases of pollution into the environment, for preparedness and policy-making purposes.

Finally, we demonstrate how under the appropriate conditions, binary learners can be applied to classify SE events. Moreover, we show how in SE event circumstances,

such as in the verification of the Comprehensive Test-Ban-Treaty (CTBT), some standard one-class classifiers can be trained on mislabelled data drawn directly from the application domain to identify SE events. Furthermore, we modify the one-class nearest neighbour algorithm [23] to significantly reduce its susceptibility to noise, which considerably improves its overall applicability.

1.5 Scope

This thesis encompasses a broad swath of topics. As a result, we have limited our exploration and assessments of some of the areas.

In particular, while we emphasize the temporal aspect of the SE event domain to be of considerable importance from a PR perspective, we do not explore it in this work. Instead, we opt to set it aside for future consideration, where it can be approached with considerable care by a possible time-series analysis.

In addition, due to the constrained time requirements, we do not aim to optimize the performance of the various classifiers studied, as this would require the rigorous fine-tuning of numerous parameters. Rather, we aim to produce a set of baseline scores by employing the “accepted” parameters of the software suite.

With respect to modelling and simulation, we submit that validation is an important and time consuming, process. Fortunately, the atmospheric transport models to be applied in this work have seen considerable application, and have been widely validated. Thus, their strengths and weaknesses are well understood in general terms. While our application is, in many ways, a novel one, for purpose of data generation, and the exploration of SE event recognition, we consider the existing studies to have sufficiently validated the models. However, specific validation for the SE event aspects of this problem domain will be of interest in the future.

1.6 Organization

The remainder of this thesis is structured as follows. Chapter 2 reviews important background material used in this thesis; in particular, that regarding meteorology.

We discuss the state-of-the-art in related work in Chapter 3, highlighting research areas in atmospheric transport modelling (ATM) as it applies to the modelling and simulation framework, and in PR. Chapter 4 describes and demonstrates the proposed modelling and simulation framework. In Chapter 5, we demonstrate two strategies for classifying SE events. Finally, we conclude the thesis with a summary of the key contributions made, and discuss potential future work in Chapter 6.

Chapter 2

Background

2.1 Introduction

This chapter provides a brief review of the primary atmospheric forces that affect the dispersion of pollutants in the lower atmosphere. It is our objective that this material provides the foundation for a better understanding of the remainder of this thesis. The explanations provided are not, however, intended to form a comprehensive description of the system. Thus, we encourage the reader to refer to the many excellent texts on the subject (for example, [21, 48, 64]) in addition to the references provided below.

The remainder of this chapter is organized as follows. Section 2.2, discusses the conceptual layering of the atmosphere and the characteristics that affect dispersion in the lower regions. The role of energy in the transport and diffusion of pollutants, along with the sources of that energy, are considered in Section 2.3. The concept of air stability, and in particular, the question of how it can be defined and also leads to an estimation of dispersion, is highlighted in Section 2.4. Section 2.5 illustrates the effects of air stability and inversion layers on airborne pollutants. Section 2.6 provides a brief review of the local wind structure. The final section, Section 2.7, illustrates both the Lagrangian and Eulerian frames of reference.

2.2 Atmospheric Processes

Once emitted, pollutants enter the atmosphere and become subjected to the various forces present at the time. These forces, which are the results of energy exchanges in the atmosphere, along with the atmosphere's interaction with the earth's surface, ultimately determine the fate of pollutants emitted into the lower portion of the atmosphere.

Generally speaking, pollution is emitted into the lowest layer of the atmosphere. Together, these layers, which have been defined on the basis of the prevailing temperature gradient, are displayed in Figure 2.1. The lowest layer of the atmosphere is referred to as the troposphere and is highly complex due to the fact that its lower boundary is at the surface of the earth. The energy exchange that occurs at this margin and propagates outward is largely responsible for the heat and moisture profile present throughout [64]. These profiles subsequently effect the transport of the pollutant. In addition, convective eddies, which are a key factor in the diffusion of pollutants, are also produced as a result of the interplay between the lower atmosphere and the earth's surface.

For the above reasons, the troposphere is a highly complex slice of the atmosphere. Fortunately, the balance struck within it suggests some further subdivision, which proves informative in the development of an understanding of atmospheric dispersion. The subdivided structure of the troposphere is presented in Figure 2.2. At the bottom of this is the atmospheric boundary layer. Above this lies the convection boundary layer, which is the region that contains the convective eddies. The lowest five to ten percent of the convective boundary layer is most effected by the topography of the the earth. Within this range, strong lapse rates occur along with significant wind shear [100]. The convection boundary layer is bounded on top by an inversion layer. This upper bound serves to limit the degree to which vertical mixing can take place. As a result, when conditions in the atmosphere are such that the inversion layer is relatively high, there is a greater degree of diffusion, resulting in lower pollution levels on the ground. Conversely, when the inversion layer is relatively low, observers at the surface of the earth will experience higher levels of pollution.

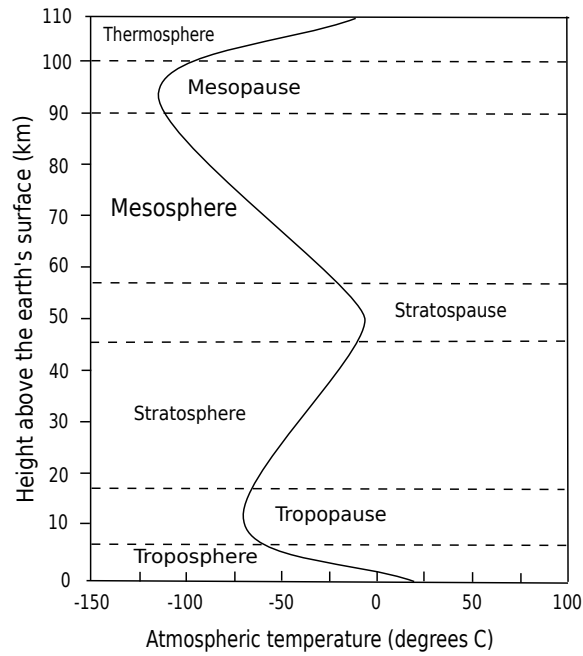


Figure 2.1: Vertical temperature profile of the atmosphere (recreated from [21].)

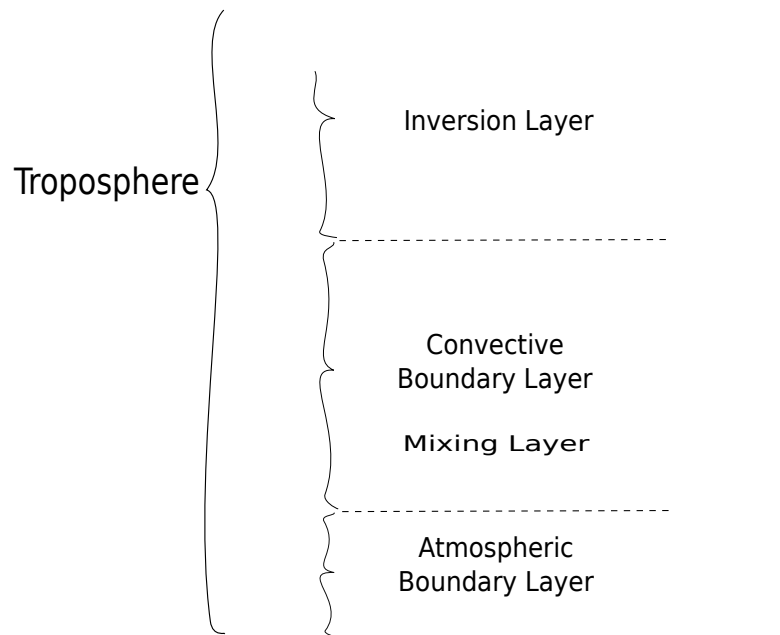


Figure 2.2: The internal structure of the troposphere.

2.3 Energy Sources

The primary source of energy for the earth's atmosphere is solar radiation, transmitted directly from the sun. This energy facilitates the transport and diffusion of pollutants within the atmosphere. Thus, it is informative to discuss the various forms that this energy takes and how it is propagated through the system before moving on to consider the forces that explicitly transport and diffuse pollutants within the earth's atmosphere.

In addition to transferring energy directly to the atmosphere, the sun also produces an influx of energy to the surface of the earth. Energy received at the earth's surface is subsequently transmitted to the atmosphere in the form of sensible heat, latent heat and long-wave radiation. These energy exchanges are the secondary forces behind both the local and global weather systems that transport and diffuse airborne pollutants. The sea breeze is an example of a local phenomenon that is produced as a result of the preferred heating of the earth's surface in coastal areas. The warm surface transfers heat to the parcels of air directly above the ground, causing them to warm and rise. As they rise, they allow the cooler denser air, which resides over the sea, to advect inward. As the cool parcels shift in over the warm surface, they are heated and the process repeats while the ground remains warm. It is this cycle of cool dense parcels of air replacing warmer ones that constitutes the sea breeze. On a global scale, the cumulonimbus clouds of the inter-tropical zone, which are the driving forces behind global circulation, acquire their energy from tropical oceans, which, in turn, emit latent heat from their warm surfaces.

2.4 Air Stability

Within the atmosphere, stability refers to the degree to which individual parcels of air shift vertically and horizontally, in addition to the transfer of individual particles between their encompassing parcels. This movement can be classified as either advection, situations where the fluid flows as a whole, or convection, where the fluid

doubles back on itself to form eddies. The latter is important as it enhances the mixing of pollutants between adjoining parcels of air, producing a shift down the gradient scale, while the former transports the pollutant in the direction of the mean wind. In theory, the smaller, highly random eddies, are separated out of the larger class of convective flows and referred to as turbulence. However, in practice, this is difficult to do, as larger cells have a tendency to break down into smaller ones.

Convection may be induced by a number of sources. For example, it may occur as a result of the flow over obstacles or rough surfaces. Alternatively, it may be a product of instability resulting from density and buoyancy differences in the fluid, or as a result of a heated surface. In practice, the disposition of these forces is to work in concert to produce convective flows.

It is informative to consider the stability of the atmosphere in terms of adiabatic processes, which provide insight into the amount of heat emitted from a parcel as it rises and expands. In particular, a parcel is said to have undergone dry adiabatic conditions, denoted by Γ_d , in the case where the temperature decreases at a constant rate of $9.8Kkm^{-1}$. Under more realistic conditions, the parcel will contain some moisture, resulting in an alternate rate of change in temperature.

The lapse rate, Γ_e , which is the measured or inferred rate of temperature change in the environment, is defined by the following equation:

$$\Gamma_e = -\frac{dT}{dz}, \quad (2.1)$$

where Γ_e is the lapse rate, T is the temperature, z is the altitude and the negative sign indicates a decrease in temperature with height. By comparing the lapse rate to the rate of change for dry adiabatic conditions, $9.8Kkm^{-1}$, we can estimate the state of stability in the lower atmosphere [64]. The conclusions that may be drawn from this comparison are displayed in Table 2.1

Stability in the lower atmosphere has a strong effect on the vertical temperature profile. During unstable conditions, characterized by large convective flows, mixing destroys the pre-existing vertical temperature gradient. As stability in the lower atmosphere increases, a well-defined vertical temperature gradient becomes more apparent.

Lapse Rate	State of Stability	Remarks
$\Gamma_e = \Gamma_d$	Neutral or adiabatic	Common conditions in the UK, light winds and cloudy skies.
$\Gamma_e > \Gamma_d$	Strong lapse rate or super-adiabatic	Unstable conditions, good vertical mixing of pollution.
$0 < \Gamma_e < \Gamma_d$	Weak lapse rate or sub-adiabatic	Stable conditions, vertical mixing is limited.
$\Gamma_e = 0$	Isothermal	Stable conditions, vertical mixing is very limited.
$\Gamma_e < 0$	Thermal inversion	Very stable conditions, generally there is no vertical mixing.

Table 2.1: Relationship between the lapse rate and the stability of the atmosphere (recreated from [21]).

Generally speaking, a cyclical trend can be seen in the daily temperature profile. Overnight, facilitated by the absence of solar energy, radiation inversion occurs. As the hour of 10.00 approaches, and the sun's energy is increasingly felt on the surface, the inversion begins to break up. At approximately 16.00 hours a super adiabatic lapse rate occurs resulting in a well-mixed lower atmosphere. Stability is re-established as the sun recedes and the surface cools once again.

In addition to solar energy, humidity, cloud formation and wind are all factors in the establishment of a strong vertical temperature gradient.

2.5 Implications for Pollutants

The stability of the atmosphere, and hence the prevailing meteorological conditions affect the fate of pollutants emitted into the atmosphere. In Figure 2.3, the influence that atmospheric stability has over a plume emitted from a stack is displayed. Each sub-figure relates a particular lapse rate of interest (Γ_e), indicated by a solid line, to the expected plume shape for that lapse rate. Figure 2.3(a), provides an example of *looping*. This occurs during super-adiabatic (unstable) conditions where the presence of convective flows stretch the plume both upwards and back down upon

itself. Figure 2.3(b), displays a *coning plume*, which is characterized by neutral conditions where gentle winds allow the plume to spread in all directions. The *fanning plume* of Figure 2.3(c), occurs when the plume is emitted into a zone bounded to the top and bottom by inversion layers. The plume is, thus, restricted in its vertical expansion but free to spread across the horizon. Finally, Figures 2.3(d) and (e), display the results of a plume being emitted below and above an inversion layer. Clearly, the former will produce significantly higher levels of pollution at the surface.

In the case of a puff, which is a short term release of pollution from industry and/or an explosion, the inversion layers act much the same forming a vertical barrier. However, the distinction between mean and turbulent motion is rather delicate and entirely dependent on the size of the system being investigated. That is to say, for fluctuations in wind larger than the puff itself, we tend to see them as contributing to mean motion. Alternatively, smaller fluctuations tend to contribute to dispersion.

2.6 Local Wind Structure

The wind is a primary force in determining the fate of airborne pollution, as well as being influential in the construction of heat and moisture profiles in the lower atmosphere and the production of mechanical turbulence.

When fluids, such as air, flow across a surface, their momentum is impeded by viscous forces. Thus, parcels of air coming in direct contact with a surface, such as the earth's surface, are obstructed and their velocity approaches zero. Shearing forces then cause the parcels directly above these to be slowed, which in-turn affects a decrease in the velocity of the next layer of parcels. As the elevation is increased, the effect of the surface is reduced, and eventually the shearing forces completely dissipate, leaving the horizontal pressure gradient and Coriolis forces to strike an equilibrium. Through these forces, the wind speed becomes a function of elevation. Generally speaking, the wind speed increases considerably in the first 10 meters, after which the rate of change slows. Beyond 1 kilometre, shearing, or drag, due to the surface is effectively zero. In practice, large-scale pressure distributions and weather systems add a great deal of complexity to the system by altering both the speed of

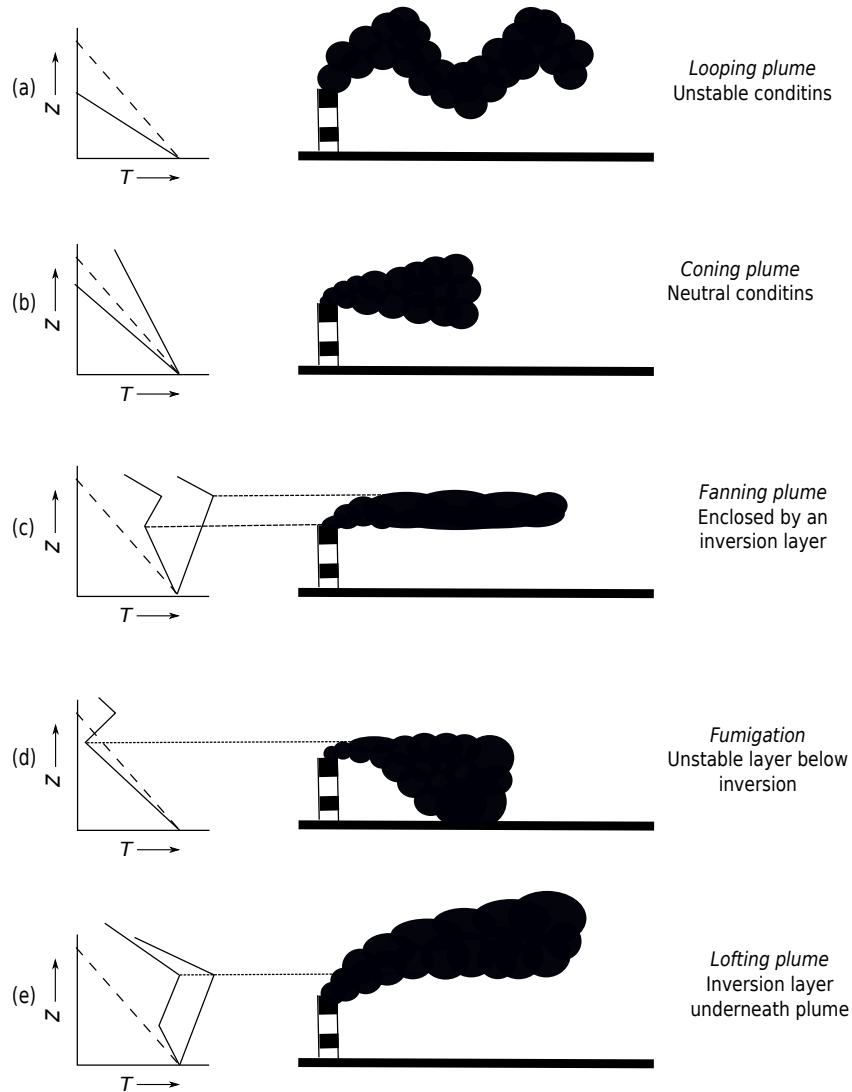


Figure 2.3: The relationship between the stability in the lower atmosphere and the ultimate shape of plumes emitted from elevated stacks [10]. In the inset z vs T diagram, the dashed line represents Γ_d and the solid line represents Γ_e .

the wind and its direction with height. As a result, it can be difficult to build a completely accurate model of the local wind structure.

2.7 Frames of Reference

There are two conceptual frames of reference inherent in the collection of data from a fluid matter. When measurements are taken from a fixed position, the reference frame is considered Eulerian. In regards to the study of atmospheric flows and airborne pollutants, this approach implies successive and sequential observations of the meteorological conditions and pollutant volumes, taken with respect to individual parcels of air, as they are advected past the observation point. Data measurements collected within the Eulerian frame of reference is the simpler of the two techniques, however, it is limited in its descriptiveness.

A particularly ubiquitous example of a meteorological instrument that operates in the Eulerian frame of reference is the weather vane, which is commonly seen perched on barns in the countryside, but also upon high-tech weather observatories spread around the globe, feeding data to centralized computer models.

Alternatively, measurements may be extracted relative to a particular parcel of air. This technique requires that the observer “ride” with the advecting parcel of air. Thus, measurements of this form illustrate the evolution of an individual parcel over time. The most notable meteorological instrument operating in a Lagrangian reference frame is the weather balloon, which is a free-floating neutrally buoyant device equipped with meteorological sensors.

Lagrangian measurements are particularly attractive, as they simplify the modeling of turbulent diffusion by capturing the three-dimensional path of air parcels. However, the realization that air parcels will inevitably traverse a largely unique and convoluted route as they are acted upon by eddies in the turbulent flow, suggests the difficulty of acquiring Lagrangian observations. As a result, a great deal of capital has been invested in an effort to develop theoretical and empirical techniques to translate the more accessible Eulerian statistics into Lagrangian statistics. While the theoretical means exist for doing so, their application in real flows is a significant

challenge.

2.8 Statistical Descriptions of Advection and Diffusion

There are three relevant approaches to averaging when concerned with the modelling of advection and diffusion in the atmosphere. Each of these is relevant to a particular set of scenarios and modelling objectives. In particular, the averaging techniques are temporal, spacial and ensemble averaging. Both temporal and ensemble averaging are of significant interest in this thesis. As such, they are discussed in greater detail below. While spacial averaging can be considered under much the same guise as temporal averaging, interested readers should solicit further detail on spacial averaging, in addition to temporal and ensemble averaging, from Arya, [5].

Temporal, or time, averaging is commonly applied during the analysis of data collected from a fixed point in space, and gathered over a time period, T . Data acquired in this way is of the Eulerian form, however, data acquired of a Lagrangian process may also be analyzed in this manner. In particular, both the temporal and spacial averages are the product of the specified sampling time, T , and possibly a smaller averaging or interval time, t . The former refers to a period of continuous measurement, which is averaged to produce a single observation point, while the latter is indicative of a delay between the instantaneous discrete measurements taken over T . In both the discrete and continuous circumstances, the result is a set of observations that are combined to form the temporal average.

Time and space averaging are somewhat subjective, as they are invariably dependent upon the selection of the sampling time, in the case of temporal averaging, and length, where space is concerned. While the ideal value of these variables is not clear, the effect may be pronounced. Indeed, the choice of sampling time may have great implications on the result of the modelling phase. If, for example, the sampling period is too long, the standard deviation will increase, and important trends, such as the diurnal variations or urban/rural differences, inherent in the evolution of the

component, may be masked. Conversely, it is necessary that the specified period be large enough to both ensure a stable mean and to incorporate the effect of the largest turbulent and diffusive features of interest to the modeller.

The degree to which diffusion occurs within the atmosphere is related to the standard deviation of the above measurements. Moreover, it is also related to the frequencies or the range of frequencies in the spectrum that compose the greatest contributions to the total standard deviation [64]. This intuitively results from the fact that distinct scenarios may produce the same or similar standard deviations. To illustrate this, consider two environments, one of which is subjected to a few large oscillations, while the second witnesses numerous small oscillations. While their resulting standard deviations may be similar, their effects on an airborne puff are quite distinct. Indeed, the former scenario will tend to transport the pollutant, while the latter breaks it apart. This effect must be carefully considered when selecting the appropriate averaging time.

Arya [5] describes the typical averaging time, T , for the ABL as ranging from 10^3 seconds to 10^4 seconds, and as being dependent upon the observation height, ABL depth and the stability of the prevailing atmosphere during the period under consideration. Furthermore, studies report that where air pollution and dispersion are of concern, the sampling length applied to the derivation of the wind speed, wind direction and the statistics of turbulence, must be consistent with those of the concentration and dispersion parameters.

The ensemble average, which is sometimes referred to as a probability average, is used almost exclusively as a theoretical tool. It is defined by a large number, one which is approaching infinity, of realizations of a variable achieved through independent, identical experiments. By definition, the derivation of an ensemble mean for components of the earth's atmosphere, is essentially impossible due to the high degree of variability inherent in the atmosphere. However, this approach to averaging is often favoured in theoretical work, as it is not dependent on a subjectively chosen sampling period. A particularly desirable feature of the ensemble average, is that it conforms to Reynold's averaging rule, where other approaches may not.

Due to the divergence in application of the ensemble mean and the time and space

averages, it is important to recognize the statistical relationship between them. For the case of a stationary random variables, where stationarity indicates independence from all mean variables and the turbulence statistics on time, the time average converges to the ensemble average, as the number of observations goes to infinity [70]. Similarly for space, homogeneous field variables tend towards the ensemble mean as the number of observations goes to infinity, where, homogeneity, in this case, refers to independence from mean flows and the turbulence statistics on distance. The natural variability of the atmosphere renders the necessary conditions of stationarity and homogeneity infeasible in the strictest of senses. At best, quasi-stationary and homogeneous conditions can be expected during short time periods in the overnight hours and during the day, in situations where an individual large weather system dominates the area. While these conditions remain, typically over a course of one to three hours, approximate equivalence between the temporal and spacial averages, acquired through experimentation or observation, and the ensemble averages can be assumed.

Chapter 3

Literature Review

Researchers in machine learning have long relied on artificial data sets when considering the performance and stability of new and existing algorithms. A particularly motivating factor behind the use of artificially generated data, is the flexibility which it provides the researcher, and the accompanying power to acutely understand the behaviour of the algorithm in question, along with the ability to rigorously test theoretical hypotheses. Alternatively, some research efforts are impelled to use artificial data due to the limited availability of “real” data for their particular domain of interest. Examples of the former justification can be viewed in [4, 56, 129], and of the latter in [26].

The remainder of this chapter is bisected by the two fundamental topics of this thesis. The processes that were informative to our formalization of a modelling and simulation framework are described in the first half, specifically Section 3.1, and is followed by a discussion on the state-of-the-art in PR, in Section 3.2. The section on data generation contains a number of sub-sections, which are intended to inform the reader on the choices made and the techniques used in simulating the dispersion of radioxenon in the lower atmosphere. A short description of the individual objectives of these sub-sections is provided below.

In subsection 3.1.1, we consider the important question of radioxenon emitters. The purpose of this section is to provide the reader with an understanding of the various sources of radioxenon, which include industrial emitters and the detonation

of nuclear weapons. While this body of science is not fully developed, the discourse includes a best estimate of emission rates from the primary sources. With respect to subterranean nuclear detonations, the overlying material has an effect on time lapsed before venting, the rate of venting and the percentage of radioxenon vented. These points are considered, as they are crucial in developing our understanding of how much and when pollution will be witnessed downwind. In addition, slow venting rates and/or long lapses prior to venting may produce higher levels through a process commonly referred to as chaining, or cumulative yields.

An integral part of this study involves estimating the state of the troposphere, thus, approaches for doing so are presented in Subsection 3.1.2. While Chapter 2 described the atmospheric processes involved in dispersing airborne pollutants, dispersion models rely on a variety of parameters that describe the state of the atmosphere in order to predict how a pollutant will behave. In this chapter we will dedicate a sub-section to highlight the practices used to determine the value of parameters such as atmospheric stability, the effective stack height, the wind speed profile, the mixing height and the variances for the Gaussian diffusion models.

In Subsection 3.1.3, we discuss relevant approaches to modelling atmospheric transport. In particular, we articulate three approaches to modelling atmospheric dispersion, the first of which is the modelling technique implemented in this thesis. Although they are not necessarily the most accurate approaches, collectively they sufficiently demonstrate the considerable range of techniques appropriate for the modelling and simulation aspects of this thesis.

3.1 Data Generation

3.1.1 Radioxenon Emitters

In conducting this research on the PR of, what we have identified as SE events (a phenomenon that is described in greater detail in subsequent chapters), we recognized that the CTBT verification problem presents itself as an exemplary demonstration domain. In its essence, the CTBT bans all nuclear explosions in the environment,

and its verification is based on the detection, identification and localization of possible detonations [94]. More specifically, verification has been proposed based on the quantification of four xenon radioisotopes in the mass range of 131 to 135. Thus, consideration of both the source and the nature of radioxenon is warranted.

Noble gases are, in general, produced in large quantities from nuclear fission and are inherently difficult to contain [16]. This is particularly true for xenon. Thus, if radioxenon were to be measured, it could indicate an earlier detonation of a nuclear weapon. In that regard, Bowyer *et al.*, in [16], specified the independent yields that would result from a nuclear blast equivalent to one kiloton of TNT as follows:

$$\begin{aligned} {}^{131m}\text{Xe} & - 5 \times 10^9 \text{ Bq/kt}, \\ {}^{13m}\text{Xe} & - 2 \times 10^{13} \text{ Bq/kt}, \\ {}^{133m}\text{Xe} & - 2 \times 10^{14} \text{ Bq/kt}, \\ {}^{135}\text{Xe} & - 2 \times 10^{11} \text{ Bq/kt}. \end{aligned}$$

The cumulative yields, which result as a combination of the initial production and chaining (specifically, decay from parent isotopes), are stated to be several orders of magnitude higher than the independent yields alone. The cumulative yields are considered, for example, when a detonation takes place in a containment cavity, from which the release is delayed for a considerable amount of time. During the initial period of containment, specifically while parent isotopes exist, the concentrations of the four radioxenon isotopes will rise. Given that since 1963 most detonations have occurred underground in order to limit the release of radioactive fission products into the atmosphere [74], this accumulation could have a significant influence over the total release. However, only a portion of the total production is expected to escape. This is dependent upon a series of events subsequent to the test. Figuring prominently in the degree to which accidental venting occurs, are blast induced fractures in the containment facility and the overlying surface material. However, the inert property of radioxenon dictates that a portion is likely to be vented from even the soundest of containment facilities [15]. Indeed, as seen in the past, the containment of the pollutants produced during subterranean nuclear detonation is not a straightforward task [92].

The four radioxenon isotopes that have been identified as forming a suggestive fingerprint have radioactive half-lives of 11.9 days, 5.2 days, 2.2 days and 9.1 hours, respectively. These durations have been identified as favourable because they are long enough that when released into the atmosphere they can be dispersed and measured at considerable distances. In addition, they are short enough that they do not linger in the atmosphere, and hence, they have a relatively low ambient concentration.

Saey, in [94], noted however, that Medical Isotope Production Facilities (MIPF) are the most prominent emitters of radioxenon into the earth's atmosphere, with other nuclear industries, such as power generation, taking ancillary roles in the overall background levels.

There is little information known about the explicit rates of emission from stacks associated with MIPFs. However, Saey reported an eight month study of ^{133}Xe stack emissions from a facility that utilized neutron irradiated highly enriched uranium (this is reported to be the most common practice in MIPFs), and five years of emissions data from a facility that utilized low enriched uranium, in [95]. The former site released between 10^{12} and 10^{13}Bq/day , with a mean of $4.6 \times 10^{12}\text{Bq/day}$. Alternatively, the longer study indicated that the latter site released between 10^{11} and 10^{12}Bq/day , with a mean of $7.5 \times 10^{11}\text{Bq/day}$.

More commonly, the literature reports radioxenon concentrations measured at specific receptor sites. However, due to the complexity of the "airshed", which is responsible for the ultimate path traversed by an emitted plume, and, by extension, the pollutant concentration at the measurement site, it is very difficult to associate these levels with their sources. However, such studies do provide interesting insight into the background (emission from "peace" nuclear industries) concentration distribution for various locations across the globe. In Table 3.1, we have listed the mean 12 hour cumulative results, which were acquired from continuous sampling between the years of 2003 and 2008, at four cities in Europe and four cities in North America. These results were originally reported in [95].

	Mean (mBq/m^3)	Outliers (mBq/m^3)	NPP	MIPF
Freiburg, Germany	6.0	up to a few 100	30	+1
Dubna, Russia	3.0	up to a few 100	30	+1
Stockholm, Sweden	0.4	up to a few 100	6	0
Spitsbergen, Norway	0.2	up to a few 100	1 [†]	0
Ottawa, Canada	300.0	frequent spikes between 1-10	13	1
St. John's, Canada	3.0	-	1 [†]	1
Charlottesville, USA	4.0	-	24	24
Yellowknife, Canada	0.2	-	-	1 [‡]

Table 3.1: This table reports the mean ^{133}Xe concentrations, which were measured cumulatively over consecutive 12 hour intervals between the years of 2003 and 2008. In the two instances where the number of Nuclear Power Plants (NPP) are accented by the [†] symbol, the closest plants are at a distance of 1,300 km. Alternatively, the [‡] symbol indicates that the nearest MIPF is 2,000 km away.

3.1.2 Estimating the State of the Troposphere

The challenges inherent in accurately describing the behaviour of pollutants in an environment as complex as the earth's lower atmosphere are grand, indeed. For this reason, despite considerable effort, no single physical model is capable of completely articulating the significant aspects of the problem [64]. Alternatively, many workers have concentrated on a particular aspect, or process, within the broader atmospheric system in order to model the subsystem, which can ultimately be included as components in the overall atmospheric transport models.

Of primary importance to the degradation, or diffusion, of a pollutant body within the earth's atmosphere is the mixing height, which caps vertical dispersion, stability, the wind speed and the elevation from which the pollutant is emitted. These four essential topics are discussed in the following subsections.

Mixing Height

The mixing height is an extremely important parameter, as it places a ceiling on vertical diffusion. Physically, the top of the mixing layer is restricted by an overlying stable body of air, which impedes vertical motion. It has been demonstrated that atmospheric transport models are extremely sensitive to specified mixing heights [24, 25]. More specifically, an erroneous estimation of the mixing height can result in a significant under or overestimation of the downwind concentrations. Therefore, modellers attempting to recreate specific scenarios require extremely accurate estimations of the mixing height. One particularly favourable approach, is to rely on physical measurements of the mixing height [3, 19]. In the absence of physical measurements, the bulk Richardson number can be utilized to estimate the mixing height under specific conditions [57, 105, 109]. Alternatively, empirical observations of a mixing heights, under various atmospheric conditions, can be compiled and utilized as a simple approximation. Clark, in [20], defines a table that relates the mixing height to atmospheric stability and wind speed, while Slade, in [103], and Hanna *et al.* in [39], articulate that mixing heights can range from 1 to 2 kilometres above the surface.

Stability Classification

Estimations of the prevailing stability and/or turbulence within the atmosphere are generally required by dispersion models in order to simulate an appropriate degree of diffusion in the vertical and cross-wind directions. The most appropriate parametrization of overall stability in the ABL is the ratio,

$$\frac{h}{L} = -\frac{hk}{u_*^3} \frac{g}{T_o} \frac{H_o}{\rho c_p} = \frac{hk}{u_*^2} \frac{g\theta_*}{T_o}, \quad (3.1)$$

which is based on similarity theory in the ABL, and is recommended by most boundary layer and air pollution meteorologists [5]. The individual components of the ratio are described as follows,

- L : Obukhov (buoyancy) length,
- h : boundary layer depth,

- k : von Karman constant,
- u_* : frictional velocity,
- g : acceleration due to gravity,
- T_o : air temperature at the reference height,
- H_o : sensible heat flux at the surface,
- ρ : mass density of the air, and
- c_p : specific heat at constant pressure.

In scenarios where the required measurement of the surface fluxes of momentum and heat are unavailable, Arya, in [5], suggests the application of the bulk Richardson number for the ABL,

$$Ri_h = \frac{g}{T_o} \frac{\Delta\bar{\theta}h}{\bar{V}_h^2}, \quad (3.2)$$

in which $\Delta\bar{\theta}$ is the potential temperature difference across the ABL depth and \bar{V}_h is the mean wind speed at the top of the ABL.

In the absence of on-site turbulence measurements, the standard deviations of the wind velocity fluctuations can be utilized. For the ABL as a whole, the following empirical similarity relations,

$$\frac{\sigma_u}{u_*} = 2.5 \left(1 - \frac{z}{h}\right)^a \quad (3.3)$$

$$\frac{\sigma_v}{u_*} = 1.9 \left(1 - \frac{z}{h}\right)^a$$

$$\frac{\sigma_w}{u_*} = 1.3 \left(1 - \frac{z}{h}\right)^a, \quad (3.4)$$

where $\sigma_{u,v,w}$ are the standard deviations of wind fluctuations in the three component directions, and a ranges from 0.5 to 1 under neutral and stable conditions, according ABL data [61, 62, 72, 73, 80], is recommended. Alternatively, for unstable and

convective scenarios, the following form is suggested,

$$\frac{\sigma_u}{w_*} = \frac{\sigma_v}{w_*} = \frac{\sigma_w}{w_*} \simeq 0.6. \quad (3.5)$$

The most prevalent characterization of the stability and turbulence for dispersion model parametrization within the regulatory realm, is the stability categorization proposed by Pasquill, in [77], and subsequently modified by Turner, in [121]. Pasquill's classifications, commonly referred to as Pasquill stability classes, are favoured for their simplicity, and in particular, their applicability in models with analytical solutions. Unlike the continuous function defined by $\frac{h}{L}$ and Ri_h , Pasquill defines six discrete categories, each of which result as a function of routinely available meteorological data, including wind speed, insolation and solar radiation. The major weakness inherent in the stability classes is the vast range of physical conditions within each stability class.

Diffusion as a Function of Stability

When incorporating horizontal diffusion in a dispersion model, Ayra recommends a formulation based on the statistical theory relation as a best estimate,

$$\sigma_y = \sigma_v t f_y \left(\frac{t}{T_L} \right), \quad (3.6)$$

or its equivalent form with the Taylor transformation, $t = x/\bar{u}$,

$$\sigma_y = \sigma_\theta x f_y \left(\frac{x}{\bar{u} T_L} \right), \quad (3.7)$$

where T_L is the Lagrangian time scale, which may be specified as a function of stability.

Due to the uncertainty that exists around quantification of T_L , Pasquill, in [79], suggests an abbreviation of σ_y , and specifies the function, $f(x)$, in tabular form, which simply requires the measurement of σ_v or σ_θ at the release site.

The best practices, when estimating vertical dispersion are described as those considering the ratio of the ABL height, h , to the Obukhov length, L , or the bulk Richardson number, based on the height of the ABL, and the potential temperature difference across the ABL.

In scenarios characterized by diffusion from an elevated source, where σ_z is less than the source height, the standard deviation should be estimated from the statistical theory relationship,

$$\sigma_z = \sigma_w t f_z \left(\frac{t}{T_L} \right) \quad (3.8)$$

or

$$\sigma_z = \sigma_\theta t f_z \left(\frac{x}{\bar{u} T_L} \right), \quad (3.9)$$

where the function, f_z , is presented by Draxler, in [27], over a large range of x for elevated sources and unstable stratification.

The Pasquill-Gifford (P-G) scheme is widely used in the estimation of dispersion within Gaussian models. The technique, originally proposed by the British Meteorological Office, and published by Pasquill, in [77], defines lateral plume spread, θ_p , and vertical plume spread, z_p , as functions of the downwind distance and the prevailing atmospheric stability, which is categorized according to the Pasquill stability classes defined above. Gifford, in [36], converted the original parameters to a representation of the standard deviation of the concentration in the horizontal and vertical directions as follows,

$$\begin{aligned} \sigma_y &\cong \frac{x \tan(\theta_p/2)}{2.15} \\ \sigma_z &\cong \frac{z_p}{2.15}. \end{aligned} \quad (3.10)$$

In [78] and [35], Pasquill and Gifford, respectively, review the major limitations of the P-G approach to the parametrization of diffusion. The primary difficulty with P-G estimation of diffusion is that the functions are derived from data gathered during a small-scale field experiment, and are thus not widely applicable. In addition, the utilization of discrete stability classes induces abrupt shifts in diffusivity as the atmosphere moves between stability classes.

Because the P-G dispersion scheme is an empirical result stemming from a study of dispersion from a continuous point source, there is no theoretical justification for its application in instantaneous point source models. Indeed, few field experiments have been conducted to study the dispersion of pollutant clouds emitted from instantaneous point sources. However, Slade, in [103], provides a table which summarizes horizontal

and vertical diffusion as a function of downwind distance. In addition, a power law is suggested based on three discrete stability classes.

Wind Speed Profile

Within the lower levels of the ABL, a wind speed profile arises, which is dominated by the frictional drag inflicted on the overriding atmosphere by the earth's surface. This effect is often characterized by a mathematical power law, in pollutant dispersion models [5, 120], which takes the following form,

$$\frac{\bar{V}}{\bar{V}_r} = \left(\frac{z}{z_r} \right)^m, \quad (3.11)$$

where \bar{V}_r represents the wind speed at the reference height of z_r , and m is an exponent, which takes a value between zero and one. Empirical studies suggest that the value of m is dependent upon surface roughness and stability, with the value increasing during stable conditions and over rough surfaces. However, the power law is most applicable in the lower regions of the ABL. At higher elevations the rate may be linear or even constant, and moreover, the physical direction may change.

Izumi and Caughey, in [47], presented a graphical comparison of observed wind speed profiles and their corresponding m values at the various sites. For neutral stability conditions over smooth water, snow or ice, the m is assumed to range from 0.15 to 0.4 over well-developed urban areas. Over moderately rough areas, the value ranges from 0.4 in unstable or convective conditions and approaches 1 as stability increases. While the power law, in general, is not based upon a theoretical foundation, it has been known to produce satisfactory results.

Plume Rise

Plume rise is particularly important due to its effect on the location of the maximum downwind concentration. More specifically, the maximum concentration is generally expected to occur at a distance inversely proportional to the square of the plume centre-line. In general, plume rise can be expected to have a large elevating effect on

the plume's centre-line, increasing it by as much as ten times the stack height when winds are light [5].

Plume rise calculations in Gaussian models are generally conducted separately, and subsequently added to the source height. Briggs, in [17], derived a set of plume rise equations for a variety of effluents in various atmospheric conditions, each of which relies on a different empirical constant. Briggs' 1968 formalization has proven quite consistent in comparison with plume rise data, and has been utilized in a number of Gaussian models, including [6]. In addition, his proposed plume rise model was applied by the USEPA, in [123].

3.1.3 Modelling Atmospheric Transport

By definition, a mathematical model is an abstraction. As such, some components, or features, considered to be of less importance to the problem under examination are excluded from the modelled system. Mathematical models have long been used in a variety of fields, both to help humans understand complex processes and to predict future behaviour of elaborate systems. Their application spans from computer networks to global climate change and well beyond. Indeed, mathematical models have been integral in our ability to predict the atmospheric transport and diffusion of dangerous and undesirable airborne pollutants. Within the domain of environmental assessment, models are deployed to aid in the understanding of the fall-out from, and resulting health effects of, industrial accidents [25, 46, 104]. Furthermore, governments regularly apply models in order to estimate the effects of proposed emitters as a part of the regulatory process [71, 99].

Atmospheric transport models take on a variety of forms and entail a wide range of complexities. The most general of these apply the Gaussian assumption to the dispersion of airborne pollutants. While models based on the Gaussian principle are not as accurate as other, typically, more complex models, the relative ease of implementation and the satisfactory results, which have been observed to produce an acceptable upper bound, make Gaussian models a popular choice among researchers and government regulators alike [64]. Indeed, Batchelor surmised that the Gaussian

function may provide a general description of the average plume diffusion because of the essential random nature of the phenomenon, by analogy with the central limit theorem of statistics [103]. For these reasons, the Gaussian assumption has been applied to the study of pollutants emitted from a range of scenarios including point, area and line sources. Emissions from these sources may be considered either instantaneous or continuous (i.e., short term or long-term releases of pollutants from a source of emissions). Examples of these are on the one hand, the massive cloud resulting from the detonation of a nuclear weapon, and on the other, the relatively continuous plume emitted from nuclear power plants along with other industries.

Gaussian Models

The Gaussian model, and its derivatives, have been widely applied in order to develop an understanding of the transport and diffusion of radionuclides, in particular, and pollution in general, within the lower atmosphere (see [14, 9, 122]). Awasthi *et al.* [6], developed a model based on the standard advection-diffusion equation, which was described in [51]. The resulting Gaussian Plume model was assessed by comparing the predicted SO_2 levels at four receptors surrounding the Dadri thermal power plant in Uttar Pradesh, India, with the physical measurements being recorded by ground-level receptors at the sites. Furthermore, the USEPA's Gaussian Plume-based, namely the Industrial Source Complex Long Term (ISCLT) model, has been widely applied in the field. For example, a commercial variant of the model was utilized to predict the concentration of radionuclides emitted from a nuclear fuel fabrication plant in northwest England [3].

Despite its widespread acceptance and its considerable application, it is broadly recognized within the community that the Gaussian model suffers from some significant limitations. Specifically restricting, is the requirement of homogeneous, stationary conditions within the modelled scenario. This limitation enables the diffusion problem to be stated as a straightforward Fickian differential equation [64, 69]. Implicit in this requirement are the following assumptions:

A1: Continuous emission from the source, or emission times equal to or greater than

travel times to the downwind position of interest. This assumption is required in order to facilitate the disregard of diffusion in the direction of mean transport, which is a requirement in the derivation of the Gaussian Plume diffusion equations. The Gaussian Plume model is not applicable in the case of short release or travel times, and instead, the puff model must be applied.

- A2: The material released is a stable gas or aerosol.
- A3: The equation of continuity holds. Specifically, this assumes that none of the emitted pollutants are removed from the plume and that there is complete reflection at the earth's surface. Reflection can also be added at the inversion layer.
- A4: Unless otherwise indicated, the pollutant body is assumed to follow the Gaussian form in the vertical and horizontal directions.
- A5: The Gaussian model assumes steady-state conditions during its application. Typically, this suggests steady-state conditions over the period of one hour.
- A6: Constant wind speeds and direction with height is assumed. This assumption, however, is never realized in the lower atmosphere.
- A7: The wind-shear effect on the horizontal diffusion of the pollutants is not considered. This is a good approximation over short distances, but it may introduce significant error at distances greater than 10 Km [130].
- A8: The dispersion parameters σ_y and σ_z are assumed to be independent of the elevation above the earth's surface and, instead, assumed to be functions of mean wind speed alone, which is not reflective of actual atmospheric flows.
- A9: The averaging time of all the pertinent quantities, and in particular, of the wind speed, the diffusion parameters and the predicted concentrations, are the same.

Bearing the above concerns in mind, an approach that treats the plume emitted from a continuous source as a series of puffs has been developed. This formulation allows the wind speed and its direction to be heterogeneous within the horizontal space

of interest. In practise, this amounts to dividing the plume into a series of puffs, which are emitted according to a predefined, generally small, time interval. Once emitted into the lower atmosphere, the transportation and diffusion of that particular puff is able to reflect the current atmospheric conditions. In this manner, models based on a series of puffs, are able to simulate complex wind and diffusion fields. For example, the authors of [69] applied the “series of puffs” approach in order simulate the dispersion of non-reactive pollutants from a combination of line sources and a single area source under non-homogeneous wind conditions in the topographically complex city of Lisbon, Portugal. Hargreaves and Baker applied this modelling approach to the dispersion of pollution emitted from motor vehicles in an urban street canyon [41]. In this approach, vehicles proceed in time steps based on the respective speeds. At the beginning of each step the vehicle emits a single puff, which is subsequently dispersed based on the complex wind and turbulence fields present within the street canyon at the time.

The momentum of air parcels travelling at the interface between the atmosphere and the earth’s surface is significantly impeded by frictional forces inherent in the relationship, although, these forces decrease with altitude. Thus, air parcels at greater heights generally have considerably more forward momentum. This creates a wind speed profile which runs counter to assumption A6 of the Gaussian model, which requires a constant wind speed with height. The logical result of such a wind speed profile on a body of advecting pollution in the lower atmosphere, is that particles presently at the top of the body will have a greater rate of advection than those nearest to the surface. As a result, the cross-section of a cloud of particulate matter should appear skewed, to some degree, in the direction of the mean wind [124]. Models that simulate this process are commonly referred to as Skewed Puff models. An application of the skewed puff model was verified against both the Prairie Grass data set and the Copenhagen data set [119]. In addition, a skewed puff models was applied, by Oliveira, to investigate the dispersion of radionuclides from a nuclear facility in Brazil[75].

Recently, Ainslie and Jackson [2] developed a method to identify the average influence of regional waste wood, “slash”, burning in a $40km^2$ area centred at the city

of Prince George, British Columbia, Canada, for a range of meteorological conditions, burn locations and dispersion characteristics.

The Prince George region is both complex in terms of geography and meteorology, as it is situated in central British Columbia at the confluence of the Fraser and Nechako rivers valleys. These characteristics, combined with considerable emissions from local industry, affect levels of $PM_{2.5}$ that often exceed the national 24 hour standard.

The dispersion model, CALPUFF, in conjunction with the meteorological model, CALMET, which was supplemented by three years of weather data, gathered at six ground-level sensors and one upper air sensor, was utilized to calculate the regional effect of burning slash piles in the regions outlying Prince George.

CALPUFF is described in [99], as a multi-layer, multi-species non-steady-state puff dispersion model. Much like the approach discussed above, CALPUFF divides the pollutant stream emitted from a continuous source into a series of puffs, each of which contains the total mass of the pollutant emitted over the short time period that composed the puff. In this way, the model is able to simulate the effect of time- and space-independent meteorological conditions on dispersing pollutants. More specifically, the evolution of each puff takes place in a stepwise fashion. For the duration of each time step, the simulated puffs evolve in a manner that is particular to the conditions in the current time- and space-interval.

DERMA, the Danish multi-puff, long-range atmospheric transport model, was examined by Sorensen in [105]. The DERMA model linearly interpolates meteorological data generated by numerical weather prediction models as a means of determining the paths of a series of puffs released from a point source. Upon release, the puffs are assumed to be well-mixed below the ABL in the vertical direction. Horizontally, the puffs diffuse according to the Gaussian equation illustrated in [128]. During the dispersion simulation, pollutant concentrations are calculated at individual locations based on the combined effect of the modelled puffs.

In experimentation with the ETEX dataset, DERMA was found to be quite capable of predicting the arrival time of tracer clouds, along with the duration of the

tracer cloud at individual locations. In addition, when high resolution data was utilized, the model was able to capture the effect of a mesoscale anti-cyclonic eddy on the tracer gas.

Gradient-Transfer Models

Pollution models developed on the basis of K-theory, adhere to the hypothesis of gradient-transfer. This hypothesis assumes a net shift of material down the gradient scale, at a rate proportional to the magnitude of the gradient itself. The proportionality factor is analogous to the coefficients of viscosity or conductivity in the laws of molecular transfer of momentum or heat in laminar flows. A particular strength of the K-theory approach is that it provides the modeller with the ability to produce a comprehensive description of the movement of pollutants through the atmosphere. By design, it can account for variations in the wind and diffusivity fields in the three component directions. Put differently, K-theory offers an anisotropic description of diffusion in the lower atmosphere. In addition, this approach allows the modeller to account for spacial variations in the diffusivity. However, Lyons and Scott [64] note, that its inability to account for an increase in K_z over time as the plume expands, is a significant limitation of the model. The implication of this restriction is that in the absence of stable conditions, the simulated dispersion near the source may be overestimated.

Some further objections are raised by Kamst and Lyons [50] and Corrsin [22]. Corrsin observes that the length- and time-scales of the transporting action should be sufficiently uniform and small compared with the length- and time-scales of the variation in the mean field gradients of the property undergoing transport. Furthermore, he contends that such a scenario is highly unlikely in the real atmosphere. Kamst and Lyons noted that a plume cross-section affected by eddies larger than itself cannot be described by the gradient-transfer relation. However, during stable conditions, characterized by small eddies, gradient-transfer is justifiably applicable. Despite this, Pasquill [78] notes that the approach which is known to give useful results for momentum transfer in the atmospheric boundary layer.

Alternatively, a notable strength of the K-model is its ability to describe diffusion,

in the three coordinate directions, $K_{x,y,z}$, not only as a function of stability, but of the surface roughness, space and time as well. As a result, researchers have taken considerable interest in defining the diffusivity parameters for both particular and general atmospheric conditions.

Shir and Shieh [102], for example, noted that horizontal diffusion appeared to have no significant effect on model prediction. In addition, they highlight the fact that the process of horizontal dispersion is not well understood, particularly during stable conditions. As a result, their solution to the advection-diffusion equation assumed horizontal diffusion, K_y , to be equal to the diffusion in the direction of the mean wind, specifically, $K_x = 500 \text{ m}^2 \text{ s}^{-1}$. Furthermore, $K_{x,y}$ were presumed to be independent of space and time. Diffusion in the vertical direction was calculated based on the turbulent transport model described in [101]. This model was assigned to predict SO_2 levels in the metropolitan area of St. Louis, Missouri, USA, over a twenty-five day period. The results were subsequently compared with measurements from the same version and found to be an acceptable approximation. The strongest predictions occurred during scenarios characterized by either strong or light winds. In general, the authors found the model to be an improvement on the Gaussian approaches for which the experiment was originally established.

The air quality model developed by MacCracken, [65] was designed to simulate numerical dispersion in the highly complex topography and meteorology of the San Francisco Bay Area. Due to the limited understanding of the vertical structure present during inversion conditions, they apply a version of the advection-diffusion equation that has been integrated vertically through the well-mixed region. In particular, the area between the surface and the inversion base was assumed to be well-mixed in their simulation of horizontal dispersion. By assuming the region below the inversion base to be well-mixed, the authors incorporate a simplification which particularizes the model to the most interesting atmospheric condition. Specifically, this involves inversion conditions which trap pollution and are thus characterized by a significant degradation of air quality. Subsequent testing of the model, demonstrated promising agreement with measurements taken in the San Francisco Bay Area [29].

A slightly less restrictive solution to the advection-diffusion equations is the probabilistic assessment model proposed by Lauritez and Mikkelsen [60], in which they aimed to provide a first estimation of European-scale risk. The derived model was applied to estimate the fallout of radionuclides from the Chernobyl accident, and although not as accurate as implementations that incorporate complex real-time models, and in particular, those that utilize numerical weather prediction schemes, it does provide an “ensemble mean” without the considerable overhead inherent in the more complex approaches.

Langner in [57], describes an operational emergency response model, termed MATCH (Multiscale Atmospheric Transport and Chemistry model), used to predict the impact of industrial accidents. MATCH is fundamentally a K-model that has been extended to utilize a Lagrangian particle approach in the initial ten hours of a simulated dispersion process. Subsequent to the preliminary dispersion process, the particles are introduced into a Eulerian grid as instantaneous point sources. The system requires that atmospheric weather data be provided by an external source, typically, a numeric weather prediction model. MATCH was evaluated on the European scale against both the ETEX dataset and Chernobyl data. The latter included concentration and deposition measurements taken after the accident, while the ETEX data is the result of a controlled experiment which involved the release, and the succeeding measurement at a variety of sites across Europe, of the released inert, non-depositing tracer. The standard version of the model was found to overestimate the ETEX data, however, when compared with the previous attempts to model deposition and the air concentration of radionuclides released from the Chernobyl accidents, the authors determined that their system produced sharper and more realistic gradients.

The CANadian Emergency Response Model (CANERM), which is similar in design to the above model, was found to produce an analogous agreement with the ETEX data [30]. In particular, CANERM demonstrated the ability to recreate the major features of the dispersing cloud, although, also remnant of the MATCH, it had a tendency to overestimate the concentration of pollutants on the ground.

Probability Risk Assessment (PRA), is an area within pollutant transport modelling that aims to address the impact of long- and short-term releases into the environment for regulatory and defensive purposes. A quantitative estimate of risk for a particular scenario can be extracted from a generated Probability Density function (PDF). Consider, for example, an accidental release of radionuclides from a point source. In this case, the PDF would produce an estimate of risk by indicating the total concentration or deposition over an ensemble of weather conditions.

Lauritzen *et al.* in [59], introduce a simplified approach for the estimation of PDFs for long-range dispersion scenarios, which is less demanding on computer resources and provides more generic results than numerical approaches. The technique utilizes a two-dimensional solution to the advection-diffusion equation in order to produce ensemble-average (climatological mean) concentration fields. In comparison with DERMA [60, 58], the Danish numerical dispersion model, the model of Lauritzen *et al.* is found to be flexible in terms of parametrization, and to also provide sufficient approximations of the mean results.

Subterranean Venting

The hybrid Eulerian-Lagrangian, Atmospheric Diffusion Particle-In-Cell (ADPIC) model [55] was utilized by Rodriguez and Peterson, in [92], to simulate the dispersion of the radionuclide, ^{131}Xe , vented from a subterranean nuclear explosion. More specifically, the authors aimed to unambiguously determine the source strength of the underground detonation, engineered by the former Soviet Union at Semipalatinsk in February of 1987, through the comparison of modelled and measured levels of radioxenon isotopes in Freiburg, West Germany.

As a means of assessing the source strength, the simulation was repeated numerous times, commencing at 0500 on the day of the detonation. Each iteration of the simulation utilized the archived weather data for the period and varied the rate of release at intervals and vertical distributions of the vented radioxenon. Each release in the simulation lasted for twelve hours, the rates of release (venting rates) were normalized to be uniform, the distribution of the radioxenon was assumed to be Gaussian with cut-offs at one standard deviation (± 500), and the cloud heights were

set to be either 500 metres or 1,500 metres. These variations produced six distinct simulation scenarios. Subsequent to the six dispersion experiments, the predicted concentrations were compared with the measured, twenty-four hour samples. An analysis of the results suggested the detonation of a 40 kiloton yield bomb. This result assumed the following: (a) that the pollutant was dynamically vented to an elevation of 1,000 metres over a period of 36 hours; (b) that there was no delay in the migration of the precursor; (c) that the in-growth by the decay of the ^{131}Xe precursor was total; and, (e) that no ^{131}Xe was incorporated into the melt.

Schwarz *et al.*, in [98], simulated the dispersion of CO_2 leaked from an underground reservoir, by applying a prevalent analytical solution to the standard advection-diffusion equations. The following encompasses but a few examples of this approach in previous literature, [5, 80, 110].

In a comparison with TOUGH2 [82, 83], an integral finite difference code, which simulates both the initial dispersion of CO_2 within the earth's surface, its venting and its subsequent dispersion throughout the atmosphere, the authors found the solution to the advection-diffusion equation to be simple and useful as a predictor of pollutant plumes for specific application domains. In particular, they noted that the approach applies to dispersion over a flat terrain where the wind and diffusivity profiles are known, and where the problem involves steady-state conditions.

Lagrangian Dispersion Models

As introduced in Chapter 2, the Lagrangian frame of reference represents an alternative to the fixed position Eulerian view, which, when considering the turbulent dispersion of pollutants, conceptually forms a more natural approach. Lagrangian stochastic models are particularly interesting techniques which can be applied to model dispersion with a Lagrangian frame of reference. In particular, models of this form utilize a statistical definition of the random velocity field in order to describe the path of individual particles in the modelled system [125]. In terms of atmospheric dispersion, each particle in the modelled system is analogous to an advecting and diffusing parcel of the pollutant in question.

In broad terms, Zannetti in [128] describes particle models as a set of algorithms

applied to the generation of realistic trajectories that simulate atmospheric flow. At each time step, the particles are moved at a pseudo-velocity, which accounts for transport, due to the mean fluid velocity; turbulent fluctuations within the wind components; and, molecular diffusion if it is considered a significant factor. An additional benefit of particle models is that each particle within the larger simulation may be tagged with domain specific information. When modelling dispersion from a continuous point source, the particles are analogous to puffs of the particular pollutant of interest. In this case, each particle is assigned an initial pollutant mass, which may be adjusted for chemical decay and other transformations over time. Thus, particle models enable the development of three-dimensional concentration fields. Furthermore, each particle can easily be attributed to a specific emitter within the modelled scenario, thus, facilitating superior analyses and understanding of the simulation results.

Lagrangian models have been applied to a wide variety of domains, indeed, those as minuscule as molecular dynamics and as astronomical in scale galaxy dynamics [43].

Taylor, in [113], initiated the application of Lagrangian frame methods to the description of pollutant dispersion within the turbulent atmosphere. However, this application was restricted to scenarios characterized by homogeneous turbulence. More recent developments in this field have facilitated considerable advances in the descriptive capabilities of Lagrangian stochastic models. In particular, Thomson, in [117], articulated the extremely helpful criteria for modelling neutral tracers; this effort resolved many of the outstanding difficulties with Lagrangian stochastic models.

Wilson and Sawford, in [125] describe the significant advantage inherent in the utilization of the Lagrangian frame. In particular, they postulated that the Lagrangian framework benefits from the fact that, unlike its Eulerian counterpart, the time derivative following particle motion implicitly includes the nonlinear advection term, without requiring an approximation. As a result, closure approximations for velocity do not require the nonlinear terms. In the case of the scalar concentration, and specifically where the Reynolds number is approaching infinity, molecular diffusion can be ignored. The resulting conservation equation for scalar concentration

simplifies to the trivial specification that the fluid particle retains its original concentration as it travels through the fluid; this is mathematically formalized as the following differential equation,

$$dc/dt = 0. \quad (3.12)$$

This result dictates that particles conserve their initial mass, and thus, changes in the mass distribution within the concentration field must result from the re-distribution of the particles within the fluid. Therefore, the concentration statistics that summarize the pollutant body in question are directly related to the displacement statistics that characterize the general path of the particles. The ensemble mean can thus be articulated as [115],

$$\langle c(\mathbf{x}, t) \rangle = \int_{-\infty}^t \int_{\mathbb{V}} P(\mathbf{x}, t; \mathbf{x}', t') S(\mathbf{x}', t') dt \mathbf{x}' dt', \quad (3.13)$$

where \mathbb{V} represents the volume of the entire fluid, $P(\mathbf{x}, t; \mathbf{x}', t')$ is the probability density function describing the particle transition from point \mathbf{x}' , at time t' to point \mathbf{x} and $S(\mathbf{x}', t')$ indicates the pollutant's distribution at the source.

Numerical approaches are often applied to generate the trajectories of marked particles in Lagrangian stochastic simulation models in order to estimate concentration statistics such as that of Equation (3.13). The overriding challenge is to generate Lagrangian particle statistics from the available Eulerian measurements that are both physically satisfying and self-consistent. With that in mind, individual particle trajectories are calculated by assuming that the particle velocity within the system can be simulated as a Markov process [37, 86, 116, 125], which implicitly evolves according to the Langevin equation described below.

The Langevin equation initially resulted from the study of Brownian motion, thus, it formalizes the motion of a particle subject to a retarding force and random acceleration [125]. In its general form, the Langevin equation is,

$$du/dt = -a_1 u + b\xi(t), \quad (3.14)$$

where u is the particle velocity, t represents time, and a_1 forms a damping coefficient associated with the viscous drag on the particle. The final term, $b\xi(t)$, in which b

is a coefficient and $\xi(t)$ is a random function, is a rapidly fluctuating acceleration component. An in-depth examination of the equation and its evolution is provided in [91].

Most relevant to a description of particle paths of this nature are the techniques founded on the idea of first-order Random Flights, or Markov-chains. As indicated earlier, these models assume that the particles within the system evolve as a Markov process. Moreover, the theory of first-order Random Flights postulates that the position vector, \mathbb{X} , and the velocity vector, \mathbb{U} , are jointly Markovian. Although, the hypothesis of a joint Markovian process for (\mathbb{X}, \mathbb{U}) cannot be rigorously justified, it is arguably a reasonable assumption for dispersion models characterizing a turbulent atmosphere with a high Reynolds number [125].

Based on this Markovian assumption, the most general equation applied to articulate particle velocity is the Stochastic Differential Equation (SDE):

$$d\mathbb{U}_i = a_i(\mathbf{x}, \mathbf{u}, t) + b_{ij}(\mathbf{x}, \mathbf{u}, t)d\omega_j, \quad (3.15)$$

where $d\omega_j$ is a component of Gaussian white noise, which is uncorrelated with the other components and also uncorrelated with time. The particle's position can subsequently be determined through the integration of the velocity equation. Equation (3.15) comprises the foundation of the Lagrangian particle models, which we will subsequently discuss. The functions a and b , however, remain an outstanding question, for which a variety of approaches have been applied. Notably, Thomson in [117] formulated the two functions through the implementation of two fundamental consistency conditions. Thomson's approach continues to be broadly applied and can be assumed throughout. Where counter examples exist, however, they will be discussed individually.

For a simplistic example of a Lagrangian particle model, consider a geographic region in which lies a single industrial emitter whose effect on regional air quality is to be determined. One particular solution to this modelling task would be to coordinate the geographic area by the superimposition of a three-dimensional concentration grid. Subsequently, particles are generated above the point source and through a Markov-chain or random walk process, and are then dispersed downwind. At any point

during the execution of the simulation, the pollutant concentration can be calculated by simply computing the summation of the masses of the individual particles in each cell. However, a more rigorous approach to this calculation, one involving weighted contributions from each particle based on its duration in the cell at each time step, is suggested by Lamb [54].

Gariazzo *et al.*, in [32], apply SPRAY, a three-dimensional Lagrangian particle dispersion model, to assess the impact of harbour, industrial and urban activities in the topographically complex region of Taranto, Italy. In the interest of capturing seasonal variations in the local meteorology, the experiment was conducted both in the winter and the summer.

The SPRAY dispersion model [33, 118], more specifically, is a three-dimensional solution to the Langevin stochastic equation with Gaussian random forcing [117]. This solution is applicable in complex scenarios involving non-homogeneous and non-stationary conditions. The time- and space-specific meteorological inputs, which are required by SPRAY, were calculated by the meteorological model, MINERVE [114], and SURFPRO, a turbulence pre-processor, in conjunction with data from a network of weather sensors.

The model was used to calculate the three-dimensional hourly concentration of NO_x , CO , SO_2 and primary PM_{10} within a $35\text{ km} \times 35\text{ km} \times 1.5\text{ km}$ cube superimposed on the greater Taranto region.

SPRAY was evaluated via comparison with hourly observations, the result of which indicates a tendency on the part of the model to underestimate pollutant concentrations. The notable exception was SO_2 , which was overestimated by the model. While the authors suggested some factors that were likely to have degraded the accuracy of the model, they indicate that the SPRAY model was successful in capturing general trends in the concentration distribution. For example, seasonal variations in exposure resulting from cyclical alterations in local meteorological conditions, were highlighted.

In addition, the model was applied to test the effects of individual sources and found that the harbour, whose regional effect had previously been unidentified, was not the main contributor of airborne pollutants in the city.

In the aftermath of the nuclear accident at the Chernobyl power plant in 1986, researchers and government regulators amassed meteorological data along with measurements of the airborne and ground-level concentrations of radionuclides from across Europe and the former USSR. This data was critical in the development of an expedient and realistic risk analysis as a mean of assessing the potential impact on the citizenry. Subsequent to the period of assessment and defensive reaction, this data has been utilized in the development and validation of new atmospheric dispersion models. In particular, researchers have employed the extensive meteorological data from the period to simulate the dispersion of ^{137}Cs and ^{131}I from the Chernobyl power plant across Europe. A number of exemplars are briefly discussed below.

Izraehl *et al.* in [46], for example, utilized a Lagrangian particle model to investigate the intermediate and regional scale fallout from the Chernobyl nuclear accident. In particular, the trajectory model accounted for the ^{137}Cs and ^{131}I emitted from a continuous point source which was dispersed in different directions by variable winds. Furthermore, the model simulated diffusion both vertically and horizontally.

A statistical comparison was undertaken in order to assess the predicative capability of the model in terms of cumulative fallout and maximum concentration. For the majority of receptor sites, the authors report quite acceptable agreement between the simulated and observed values.

Bonelli *et al.* in [13], applied STRALE, a Lagrangian puff trajectory model, to simulate the three-dimensional transport and diffusion of ^{137}Cs released during the accident, on the European scale. During simulation, the STRALE model creates a series of puffs above the source at each time step. Each puff is defined by a mass percentage, which is a portion of the total mass released over the experiment, and vertical depth over which the pollutant is uniformly distributed. Subsequently, the puffs are dispersed based on routinely available meteorological data. This data is utilized in the derivation of time- and space-independent, three-dimensional wind fields employed to update the positions of the puffs at each time step.

By varying the number of puffs emitted at each time step, together with the lower and upper bound of the emission layer, practitioners are able to recreate any emission profile as a function of the puff mass and the effective height reached above the source.

The concentration of ^{137}Cs is computed at each time step within each set of cubes superimposed on the domain of interest. In particular, the cubes are defined by regular latitude and longitude grids, and a height determined by the top of the mixing layer over a given time interval. Within a grid, the concentration is determined through the summation of the masses of the puffs within.

Through statistical comparisons, the authors cite a correlation coefficient of 0.63 between the predicted and measured results, which they suggest, implies good agreement for a model of this type.

APOLLO, a long-range Lagrangian particle model, was applied by Desiato in [25] to predict the ground and air concentrations of ^{137}Cs . The results from this validation exercise indicate similar success as the STRALE model. In particular, the APOLLO model simulates dispersion through the release of a large number of passive particles tracked by a three-dimensional wind field, on which turbulent motion is superimposed.

At each time step, the position of each particle is updated in order to capture dispersion under the current meteorological conditions. The single advection component is determined by the space and time interpolation of the three-dimensional wind field. Alternatively, diffusion is simulated both in the vertical and horizontal directions. Vertical diffusion is considered both above and below the mixing height, however, only particles below the mixing height contribute to the ground-level air concentration. Conversely, cross-wind diffusion is simulated by assigning each particle a Gaussian kernel density distribution σ , which is dependent on time.

The NAME model, for long-range transport and dispersion, developed by the U.K. Meteorological Office, was found to produce marginally satisfactory results by Maryon and Best in [67]. The model was applied to reconstruct the dispersion process as it occurred in the first 108 hours after the Chernobyl accident.

NAME is a Lagrangian model, in which plume spread is simulated by the release of a large number of passive puffs into the modelled atmosphere above the point source. After release, the puffs are advected by the mean winds extracted from the weather data generated by the Meteorological Office's operational NWP model. Diffusion results both from a "random walk" process, and the interaction between the local

wind shear and the degree to which vertical mixing exists. While the former perturbs the particle position at each time step, the latter affects a horizontal fanning of the plume. Vertically, it is assumed that the plume is well-mixed within the mixing layer.

More recently, Suh *et al.* in [109] performed sensitivity analysis on the Long-range Accident Dose Assessment System (LADAS), developed for the purpose of radiological emergency preparedness in Korea. LADAS is a three-dimensional Lagrangian particle model; it was designed to estimate air concentrations, wet deposition and dry deposition over a field of several thousand kilometres from a point source. The model was evaluated for a variety of different parameter inputs, based on a statistical comparison of the predicted dispersion of ^{137}Cs and measurements taken following the Chernobyl accident. In general, the results showed good agreement with levels witnessed around Europe in the days following the incident. In addition, the authors found the model to produce superior concentration distributions when the Richardson number was utilized in the calculation of the mixing height, rather than the simplified scenario where a constant mixing height over the computation domain was assumed.

Beyond nuclear disasters, long-range dispersion models have been applied to simulate the transport and diffusion of a variety of airborne pollutants. The FLEXPART package is one such example. The model, in general, was designed to simulate the transport and dispersion of non-reactive tracers over long distances. In particular, it is a Lagrangian particle model that treats advection and diffusion calculating the trajectory of a multitude of particles. Stochastic fluctuation of the three wind components, obtained by solving the Langevin equation [108], are superimposed on the grid-scale winds interpolated from ECMWF data to simulate transport by turbulence. Tracer concentrations on the three-dimensional grid are calculated by applying a kernel method, which is equivalent to, but more accurate than a simple summation of the masses of all the particles within the grid cell and the subsequent division by the total volume of the cell.

Previously Stohl, *et al.* in [107], validated FLEXPART with data resulting from three large-scale tracer experiments in North America and Europe to very satisfactory ends.

More recently, Forster *et al.* in [31], applied FLEXPART to recreate burning

scenarios in order to assess the breadth of CO emitted from forest fires in Canada. The results from their study unambiguously demonstrated that emissions from fires in the boreal forests of Canada can be dispersed across the Atlantic ocean, affecting a large-scale layer of haze over Europe. In addition, their simulations indicate that forest fire emissions from Canada are capable of influencing the concentrations of CO and other trace gases at the surface station in Mace Head, Ireland over a period of weeks.

3.2 Classification

This section serves to motivate PR as a particularly interesting discipline within computer science. In addition, it presents an overview of the field based, primarily, on the informative work of Duda, *et al.*, in [28]. In that regard, Duda *et al.* describes pattern recognition as follows:

“The act of taking in raw data and taking an action based on the ‘category’ of the pattern.”

It is, indeed, natural that we should desire to ‘teach’ machines to recognize sets of patterns that are easily recognizable to humans, such as handwritten characters, speech and faces, as computers present the possibility of increased efficiency and do not become tired of mundane tasks. Furthermore, the benefits of training machines to classify complex patterns, typically left to doctors and scientists with considerable specialization in the domain, are equally apparent. Thus, researchers have continued to push the state-of-the-art in Pattern Recognition (PR) systems since the advent of the modern computer.

Components of PR Systems

From the practitioner’s perspective, a PR system can be viewed as a simple black box, into which the subject of the categorization process is inserted, and from which a classification decision is produced.

Internally, PR systems can broadly be decomposed into four components:

- a) pre-processing component, which aims to reduce noise and to focus the input without losing important information;
- b) feature extraction component, which derives or calculates the values that compose the feature vector of the current input;
- c) classification component, which suggests the encompassing category of the input, based on a learned discriminating function; and,
- d) post-processor, which combines the classification decision suggested by the classifier with an assessment of the costs and/or risks associated with the particular decision, in the derivation of the final determination.

Supervised Learning

Prior to application, the PR system must be trained to discriminate between the objects of interest in its particular application domain. For multi-class problems, such as discrimination between handwritten characters, the PR system is said to learn a mapping that discriminates between the individual inputs by directing them to their corresponding categories. Alternatively, in the special scenario, which is of primary interest in this work, termed one-class learning, instances of a single target category are available for the training of the PR system. As a result, the system takes a recognition-based approach, and attempts to learn a function that maps novel instances of the target category to the target class, and all others to the outlier class.

Broadly speaking, standard PR systems for supervised learning are trained on datasets drawn from their prospective application domains, in which each feature vector has been accented with its corresponding class label. The objective of the training process is the derivation of a set of models that articulate the individual characteristics of the classes. Thus, while the performance on the training set is of little interest, rather, the focus shifts to the selection of a model that will perform well on novel instances in the future. The derivation of these models is algorithm-specific, however, there exists commonalities between all learners. Generally speaking, regardless of the learning strategy, the accuracy of the derived model on novel instances will

increase with the size of the training set. In addition, all learners strive to optimize the balance between specialization and generalization.

The decision boundary described by a highly specialized model becomes extremely complex as the classifier learns the minute details of the training set. This scenario is typically referred to as over-fitting. A highly specialized model of a two-dimensional feature space is depicted in Figure 3.1.

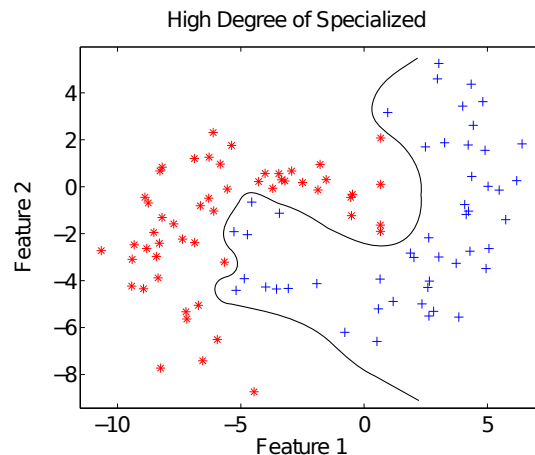


Figure 3.1: This figure demonstrates the results of a training process that focuses on reducing error on the training data.

Models of this nature are not expected to perform well during future application, because the minute details captured are unlikely to be characteristic of the category in general. At the other extreme, exists the fully generalized model seen in Figure 3.2, which, in this particular circumstance, is the result of a linear classifier. Over generalization, which is alternatively referred to as under-fitting, reduces the decision boundary to a straight line in a two-dimensional space or a hyperplane in higher dimensional spaces. Excessive generalization, like specialization, can also be expected to produce a faulty model, as it fails to depict the important characteristics of the data.

Superior classifiers are expected to achieve an optimal balance between specialization and generalization, i.e., to produce a model that ignores the irregularities of a particular training set, while remaining specific enough to distinguish the important characteristics of the data in general. One such potentially optimal decision boundary

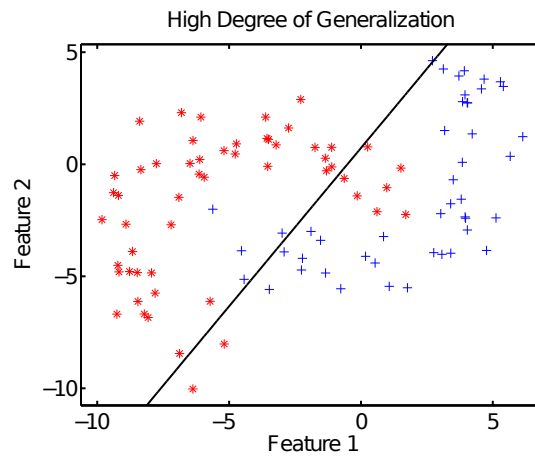


Figure 3.2: This figure demonstrates the result of a training process that focuses on generalization.

for the above demonstration scenario is displayed in Figure 3.3.

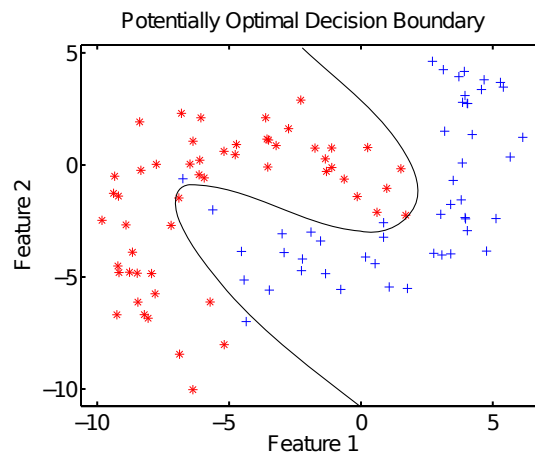


Figure 3.3: This figure demonstrates the result of a training process that attempts to find an optimal balance between specialization and generalization.

Under ideal circumstances, the training procedure for a binary learner is able to rely on an ample supply of data that has been uniformly drawn from both classes. As a result, increasingly accurate models of the classes in question can be constructed, and therefore, an effective classifier of novel instances is produced. In order to assess, or validate, the derived model's predictive capabilities, a typical procedure is to set a certain number of the training instances aside, or held-out, for a validation phase. The

standard hold-out approach was demonstrated by Lubinsky in [63], in an imbalanced scenario. The use of novel instances is essential to the validation phase in order to assess the performance in a manner that is independent of the training process. When the hold-out technique is applied to imbalanced classification problems, the results depend on extremely small training and/or validation sets, thus, limiting the confidence which can be placed on the future performance of the PR system. In general, but particularly where data is difficult to acquire, the more sophisticated technique of cross-validation [12] is preferred to the simple hold out method, for assessment during the validation phase. However, cross-validation alone, cannot account for the absence of a representative supply of training data.

As previously stated, there exists a particularly difficult family of classification tasks, namely one-class problems, in which an accurate description of one or more of the two classes is exceptionally difficult to acquire.

One-class learning problems characteristically involve scenarios in which the available class is easily acquired and exists in abundance, while the second class is exceptionally difficult to acquire, or naturally rare [112]. During extreme class imbalance, the majority class can be expected to compose as much as ninety-five percent of the data. In such scenarios, it is typical that the class we are most interested in identifying is the minority class, as is the case in automated mammogram scans and many other medical disciplines [111]. Japkowicz, in [49], and Kubat *et al.*, in [53], demonstrate scenarios in which acquiring instances is both difficult and expensive. In particular, the challenge of Kubat *et al.* requires the hand-labelling of satellite imagery, while the former involves fault detection in helicopter gearboxes, which are expensive to run. Moreover, the derivation of the outlier class would require the destruction of the gearboxes in an infinite number of ways. Alternatively, under certain conditions, the second class might be so large as to render the accumulation of a sufficient supply a seemingly insurmountable challenge. This scenario is well illustrated by the continuous typist recognition problem described by Hempstalk *et al.*, in [42]. The objective of the depicted classification challenge is to distinguish the sole legitimate terminal user from all other users. A proper training set, therefore, would be drawn uniformly from the set of all people, which is clearly infeasible.

A variety of approaches have been applied to one-class classification. The more traditional of these involve extensions to existing binary classifiers or density estimations. The density estimation approach fits a statistical distribution, such as Gaussian, to the target data, and classifies novel instances based on the learned probability of their occurrences. Such a technique has been applied by the authors of [11, 88, 111]. Techniques that extend existing classifiers typically modify the inner structure of the classifier to fit boundaries around the target class, and classify those novel instances falling outside the boundary as outliers, as is demonstrated by [49, 96]. These two approaches, in addition to some alternative approaches to one-class learning, such as [42], which is a combination of these two techniques, are discussed in the sections to follow.

3.2.1 Density Estimation

Density estimation is, perhaps, the most elementary of all approaches to one-class classification. The fundamental idea behind this one-class classification technique is the estimation of a Probability Density Function (PDF), $\hat{P}(\mathbf{x})$, based on a training set, $\mathbf{D}^n = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, drawn independently and identically from the underlying distribution, $P(\mathbf{x})$, of the target class. Subsequent to the estimation of the PDF, novel instances are classified according to a predefined target threshold or statistical tests.

Under ideal circumstances, and in particular, where sufficient training data is accompanied by a substantial understanding of the background distribution, or a flexible density estimation technique, density estimation-based classifiers are known to produce strong results [112]. However, a significant quantity of training data is required to overcome the curse of dimensionality, as is described by Duda *et al*, in [28].

Bishop, in [12], discusses three approaches to PDF estimation; the first of these techniques requires the modeller to provide an initial specification of the functional form of the underlying distribution, such as Gaussian or Poisson. An iterative process based on the predefined distribution, is applied to fit the density function to the

training data through the optimization of the corresponding PDF parameters. The application of the parametric method is significantly limited by the fact that in many cases the specified PDF may be incapable of describing the training data.

Non-parametric estimation techniques represent a more flexible approach, as they do not assume a particular functional form, and instead allow the training data to completely specify the PDF. As a result, the PDF is not limited to a small set of standard distributions, and does not have to be provided at initialization. However, the fact that the number of parameters to be optimized expands quickly as the dataset increases in size, can prove to be prohibitive.

Yeung and Chow, in [127], applied the non-parametric method for probability density estimation, introduced by Parzen, in [76], to the development of an intrusion detection system. More specifically, their approach utilized the Parzen-window estimation of $\hat{P}(\mathbf{x})$, with a Gaussian kernel, on a dataset composed of normal network activities. The generalized Parzen-window estimation of $\hat{P}(\mathbf{x})$, based on an n element dataset \mathbf{D} , takes the following form:

$$\hat{P}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^i \delta_n(\mathbf{x} - \mathbf{x}_i), \quad (3.16)$$

where $\delta_n(\cdot)$ is the kernel function (in this case, Gaussian in form), the exact form of which depends upon the number of instances in the training set. Subsequent to the training process, novel instances are classified based on their log-likelihood.

In addition, the Parzen-window approach was previously applied by Tarassenko *et al.*, in [111], to the classification of anomalous mammograms.

A final approach, sometimes referred to as semi-parametric estimation, attempts to strike a balance between the previous two methods. This approach enables a general class of functional forms, in which the number of adaptive parameters is increased systematically to build a progressively more flexible model.

The mixture of Gaussians approach is a particular category of semi-parametric estimation, which has received considerable application, as it is analytically attractive. This approach to semi-parametric estimation was applied in [89, 90], to the detection of novel instances in a series of medical datasets and as a procedure of noise removal

on an image processing task.

In its essence, the mixture of Gaussians method is composed of a linear combination of j Gaussian distributions, each of which is uniquely parametrized according to its particular mean, μ_j , and covariance, Σ_j , such that

$$\hat{p}_{MoG}(\mathbf{x}) = \frac{1}{N_{MoG}} \sum_j \alpha_j pN(\mathbf{x}; \mu_j, \Sigma_j), \quad (3.17)$$

where the α_j s are the mixing coefficients.

3.2.2 One-Class Extensions to Binary Classifiers

Autoassociator

An autoassociator is an example of a feedforward Artificial Neural Network (ANN). However, unlike its more prevalent binary counterpart, the Multi-Layer Perceptron (MLP), which aims to produce a classification decision at the output layer, the autoassociator is trained to *reconstruct* the input vector at the output layer [93]. The general architectures for both forms of ANNs are illustrated in Figure 5.1.

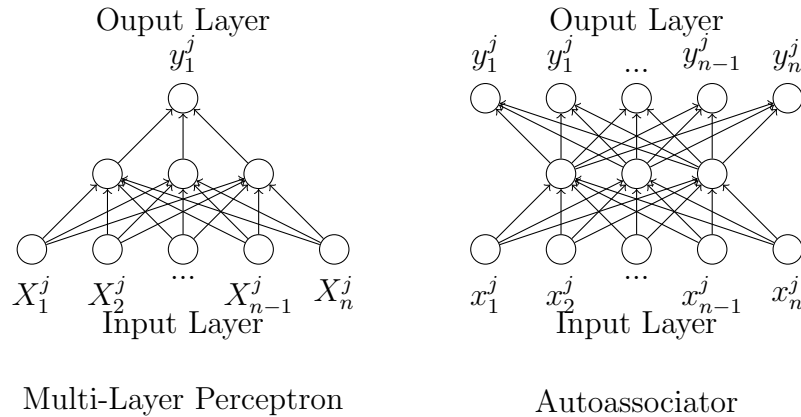


Figure 3.4: This figure demonstrates two possible feedforward artificial neural network architectures. Subfigure (a) illustrates the general form of the Multi-Layer Perceptron (MPL). In Subfigure (b), the essential structure of an autoassociator is displayed.

The theoretical basis for the autoassociator relies on the fact that it is trained

to compress and decompress instances of the concept class exclusively. Thus, during application, novel instances of the concept class should compress and decompress successfully. More specifically, the reconstruction error resulting from a novel member of the concept class, during application, is expected to be small. Alternatively, non-members of the target class are characterized by large reconstruction errors. Therefore, the classification procedure entails a comparison of the reconstruction error and a user-defined threshold. All instances reproduced with an error less than the threshold are considered to be members of the concept class, while the remainder are labelled as outliers, or non-members.

The one-class classifiers of the above form have been applied in a number of domains with considerable success. Hanson and Kegl, in [40], introduced an autoassociator system, namely PARSNIP, developed to reconstruct syntactically correct sentences using the backpropagation procedure described by Rumelhart *et al.*, in [93]. The PARSNIP system, trained on the Brown University Corpus of Present-Day American English, in which the words of each sentence are tagged with their active syntactic category, learned to accurately identify sentences that were syntactically correct and reject those that were incorrect.

Subsequently, Petsche *et al.*, in [81], developed a system similar to a fuel gauge based on the principle of the autoassociator. The system described in that work, learned to predict the impending failure of a motor. Its intended application domain is characterized by a high cost associated with failures, such as the fire pump on navy vessels.

More recently, Japkowicz, [49], examined the performance of the autoassociator in comparison with a variety of binary learners on three domains. In particular, the case studies utilized a CH46 Helicopter gearbox dataset, with the objective of predicting the failure of the gearbox based on vibration time signals, and the sonar and DNA promoter datasets from the U.C. Irving Repository of Machine Learning. The recognition task in the former was to distinguish mines from rocks in sonar data, while the objective in the DNA promoter dataset was to classify promoters in a DNA sequence. The autoassociator was found to be robust relative to the other classifiers in all three case studies, and more accurate on both the helicopter gearbox and DNA

promoter tasks.

One-Class Support Vector Machines

Schölkopf *et al.*, in [97], proposed a one-class extension to the existing support vector techniques, for the estimation of support in high-dimensional spaces. In general terms, their approach maps the training data into a dot product feature space, and inserts a hyperplane in a manner that separates the origin from the data with maximal margin.

In their work on one-class SVMs, Schölkopf *et al.* explored their implementation on both artificial and real-world data. For the latter category, the US Postal Service's handwritten digits dataset was utilized.

The handwritten digits dataset was converted to facilitate two distinct set of experiments. In the first experiment, tests that specified a random set of instances drawn from a single class to composed the target data, and left all instances from the remaining nine classes to form a large set of outliers, were conducted. On this experiment, the one-class SVM was found to correctly identify the target class 91% of the time, and had a false positive rate of 7%.

In the second experiment, ten binary features were added to the handwritten digits dataset; one new feature for each of the possible digits. These features were included to identify the class to the classifier during train, with the notion that the classifier would learn to recognize what each digit should look like. For this experiment, the one-class classifier was trained on instances drawn from each class, with the additional features. The authors found that the one-class SVM learned to accurately identify anomolous patterns, and erroneously labelled instances.

Similar implementations of the one-class SVM have subsequently been applied to a large number of problems. Manevitz and Yousef, in [66] for example, applied the one-class SVM to discover text documents of similar topics to those in the training set, and compared the results to a set of alternate one-class classifiers. They concluded that, while the one-class SVM is very sensitive to parametrization, with the right parameter set, it outperformed the other classifiers, with the exception of the one-class ANN consider in the study.

Some further examples of previous applications of one-class SVMs are to classify

yeast gene regulation predictors in [52], and for image retrieval in [18].

3.2.3 Nearest Neighbour

The standard Nearest Neighbour (NN) algorithm is a binary classifier that takes a non-parametric approach to PR. More specifically, in its simplest form, the training process involves “remembering” all of the training instances and their corresponding labels. During application, a novel instance \mathbf{x} is classified according to a majority vote rule, in which the k (k is an odd number specified by the user) NNs of \mathbf{x} , in the training set, are polled for their respective classes. The novel instance is subsequently assigned to the class that is occupied by the majority of its neighbours [28].

The NN algorithm has seen considerable application, Horton and Nakai, for example, compared NN to naïve bayes and decision trees, in [44], on the problem of predicting the cellular localization sites of proteins in yeast and ecoli. In their study, Horton and Nakai reported favourable results for the NN classifier.

More recently, modifications have been made to the NN classifier to facilitate one-class classification. Datta, in [23], adapted the standard NN algorithm to preform one-class classification through the utilization of a threshold learned during the training phases. More specifically, the algorithm searches the training set for the pair of NNs that are separated by the greatest distance, which is denoted τ . When classifying a novel instance, the distance between it and its NN is compared to the learned parameter. If the distance is less than or equal to τ , the novel instance is assigned to the positive class. Otherwise it is assigned to the negative class.

The author applied this implementation of the one-class NN algorithm to a number of UCI datasets and found it to be comparable with other one-classifiers. It was additionally found to be comparable with the binary C4.5 decision tree classifier on some classes of the Breast Cancer Wisconsin, Pima Indian Diabetes and Wine domains.

Tax, in [112], provided a comprehensive survey of the performance of one-class classifiers on a number of artificial domains, in which an alternate adaptation of the standard NN algorithm for one-class learning was included. Notably, the author

identified the one-class NN algorithm as a poor performer in a general analysis of robustness against outliers. This is, indeed, a problem that we identify and address in regard to Datta's one-class adaptation of the NN algorithm in Chapter 5.

3.2.4 Combined Density and Class Probability Estimation

Hempstalk *et al.*, in [42], introduced a technique for converting one-class classification problems into binary tasks, based on a two-fold strategy. The initial phase of the strategy involves an examination of the training data for the concept class in order to determine its distribution. This knowledge is subsequently utilized in the generation of a non-concept, or outlier, class. In the second phase, a standard binary classifier is trained based on the concept class and the generated class. Most standard classification techniques are applicable here. The single limiting factor in the selection of a binary classifier is the requirement that the classifier of choice can produce a class probability estimate at prediction time. Using Bayes' rule, the authors demonstrate how the class density function can be combined with the class probability estimate to yield a description of the concept class.

The performance of the combined density and class probability estimation technique was examined on a multitude of datasets, the bulk of which result from the U.C. Irving Repository of Machine Learning. In addition, the performance was gauged on the very interesting task of recognizing a "continuous typist". This latter application required the validation of individual computer terminal users based on their learned typing patterns. With considerations founded upon these experiments, the authors concluded that the combination of the density function with a classification model can produce an improvement in accuracy beyond that which resulted from the density function or the classification model alone.

Chapter 4

Simulating Stochastically Episodic Events

4.1 Introduction

Three categorical strategies for modelling the dispersion of pollutants in the earth's atmosphere were introduced in Section 3.1.3. In this chapter, we initially motivate the application of simulated data in the training, testing and validation of PR systems¹. In particular, we focus on an intriguing subset of classification challenges, which are oriented to the identification of events that are stochastically episodic in nature. The subsequent sections, Sections 4.3 through 4.5, which compose the bulk of the chapter, articulate both the applied modelling techniques and the simulation framework. Finally, in Section 4.6, the results the simulation process are demonstrated.

4.2 Motivation

Throughout history, modelling and simulation have been critical aids in human progress. Indeed, they have proven essential in our ability to understand, predict and often benefit from the behaviour of complex systems, natural or otherwise. Early models, derived to articulate the motion of celestial bodies, for example, enabled generations

¹Some of the work presented in this chapter has been published in [8].

of farmers to predict annual planting and harvesting cycles. In addition, such models were utilized to aid travellers with regional and global navigation. Over the past half century, advanced computer models have been developed to explore technological systems themselves, such as TCP and telephone networks, in addition to natural phenomena ranging from seismic waves to the dispersion of pollutants and climate change. Phenomena of the latter classes are the focus of this chapter. Indeed, as we shall argue, they possess certain unique characteristics, which make them hard to both simulate and study.

A typical focus of existing modelling experiments has been to simulate a particular hypothetical event, or to re-create actual scenarios as accurately as possible. Motivating the derivation of these models is often the requirement for (policy-making) regulatory regimes and safety guidelines, or to facilitate effective reaction to ongoing events, such as the failure of a node on a network, or more seriously, a nuclear accident. Alternatively, modelling and simulation have occasionally been utilized in order to generate classes of data for the training and validation of PR, and for example, Disease Contagion Prediction (DCP) systems, as was fundamental to the study of the SARS crisis.

While it is customary for PR and DCP systems to be trained, tested and validated on data sets drawn directly from the domain of interest, there are scenarios where simulated data is advantageous. Additionally, in many cases, where real-life data is unavailable, such generated data is actually mandatory. In the general case, arbitrary probability density functions can be applied to model a particular set of classes in a PR problem, and subsequently, utilized in the exploration of PR systems or to demonstrate their performance.

Alaíz-Rodríguez and Japkowicz, in [4], utilized an artificially generated medical domain in their study of the effects of changing environments on classifier performance. By generating artificial data, Rendell *et al.* in [87], avoided the uncontrolled scenarios presented by natural data sets, and described the speed and accuracy of the classifier as a precise function of a predefined set of data characteristics. Alternatively, Quilan, in [84], examined the effect of noise on the performance of the classifier by generating artificial noise in class membership and attribute instances.

Data generation is conceivably beneficial for the training and validation of classifiers in highly imbalanced classification tasks, such as the automated satellite detection of oil spills at sea [53] or the continuous typist recognition problem [42]. The concept of modelling artificial data as per specific domain characteristics was initially proposed by Aha in [1]. This was conducted for the purpose of generalizing from case-studies, and thereby producing a better understanding of overall performance of the classifier. Dietterich *et al.*, in [26], extended Aha’s generalizing strategy to derive an artificial dataset that was characteristically similar to data extracted from the target domain, as a means of overcoming the deficient supply of positive instances. These studies, however, relied on an underlying understanding of the class data being generated, which is typically elusive in domains composed of SE events.

The above-mentioned notion of data generation for imbalanced scenarios leads to a very interesting, and yet considerably less studied topic. Particularly pertinent, is the relationship between the measured characteristics (or features) of the SE events under examination, such as earthquakes or the massive short-term releases of pollutants into the environment, and the background levels of these measurable characteristics, which exist as noise, and are expelled from alternate sources. In terms of modelling and simulation, this can be conceptualized by the existence of two classes of data, namely the background data and the SE event data. The background class is considered to be relatively well understood, and in particular, strong estimates of its distribution are assumed to be known. Alternatively, the SE events, which are characteristically random and unpredictable in time, space and magnitude, rarely occur. Thus, the details of their distributions are extremely difficult – if not impossible – to estimate in general terms. Moreover, the relationship between the two phenomena, and the effect of one on the other is inherently difficult to determine.

4.3 Modelling System

As previously mentioned, this chapter considers a relatively new field, namely that of modelling SE events, which are characteristically random in space, time and magnitude. It also explores the relationship between these SE events and the well-defined

background data. The difficulty with such a modelling process is that most of the observations appear as noise. However, when the SE event does occur, its magnitude and features far overshadow the background, as one observes after a seismic event. In particular, we demonstrate how modelling and simulation can be applied to superimpose SE events on, and propagate their effects through, the background noise.

In doing this, we divide the modelling process into two phases, the effects of which are subsequently merged, in order to project the impact of the various processes on the receptor site. Indeed, the division into two distinct modelling scenarios is required as both the physical processes responsible for, and the duration of, such events, are inevitable novel occurrences, particularly in comparison to the generally consistent nature of the background noise. The details of the two phases, and in particular, the modelling of the background and the SE events, are discussed in detail in the following subsections. Before proceeding, however, it is important to note that while this particular simulation scenario is optimized for the airborne dispersion of pollutants, such as radionuclides emitted from the nuclear industry and the detonation of nuclear weapons, the theoretical concepts extend to any scenario characterized by SE interludes into a well-defined background distribution.

We can summarize our hypothesis by the following:

1. The simulation of background noise-like non-SE pollutants is best modelled by the Gaussian *plume* model;
2. The simulation of SE contaminants is best modelled by the Gaussian *puff* model.

These issues are clarified in the following sections.

Modelling SE Events

The SE events modelled in this simulation are highly discernible from the background noise effected by the series of industrial emitters. This is specifically a result of the divergent emission scenarios, and requires a distinct set of modelling principles.

In contrast to the generally constant and predictable nature of the industrial sources, the SE events in this simulation scenario, are representative of short-term

massive releases of pollutants into the environment from a random point in space and time. Following the release, however, the pollutant cloud is propagated through the atmosphere by the same forces effecting the pollutant plumes emitted by industry. Thus, at certain locations, the influence of the puff may eventually be observed at the receptor site as a deviation from the background distribution, which is physically realized as a sharp spike in the background concentrations.

The Gaussian puff technique is employed in the simulation of the dispersing pollutant cloud. While the puff model is fundamentally applicable to the dispersion of neutrally buoyant trace materials resulting from an instantaneous point source, its notoriety, and indeed, the vast majority of its application, has resulted from the simulation of dispersing pollutants emitted from continuous sources as a “series of puffs”, as explained in [2, 41, 69, 75]. The particular advantage of the puff model is that it frees the modeller from the steady-state requirement of the plume model, and allows the simulation to model the effects of time- and space- varying meteorological conditions. However, for the purpose of the present simulation, the Gaussian puff model is most desirable by virtue of its traditional function of modelling dispersion from a single instantaneous point source.

4.3.1 Modelling the Background

In this particular application of the theoretical model alluded to above, the background data is modelled as per pollutant observations made at a receptor site. More specifically, we assume the existence of one or more industrial emitters, which are commonly referred to as continuous point sources. This reflects the fact that their emissions are continuous, if not constant, over some time period, t . Each industrial emitter is positioned at a static location and characterized by an emission rate that is subjected to Gaussian fluctuations. The effect of the various sources on the receptor site is calculated based on the widely applied Gaussian plume model, which was initially introduced in Chapter 3.

In the subsequent sections, we follow the results of [21] in demonstrating the

derivation of both the Gaussian plume equation for continuous sources, and the Gaussian puff equations for instantaneous sources. We achieve this first by deducing the K-model for atmospheric dispersion from the main equation of advection-diffusion, and then, through a series of simplifying assumptions and an integration with respect to time, we settle on the two Gaussian equations.

4.3.2 Derivation of Fundamental Equation of K-Theory

The derivation of the fundamental equation of K-theory begins with what Slade [103] identifies as the starting point for most mathematical treatments of diffusion from a source. Specifically, this refers to a generalization of the classic equation for the transfer of heat in a solid, and is essentially a statement of the conservation of the suspended material. If χ represents the local concentration, quantified in terms of *mass per unit volume* of fluid, and assuming the fluid to be incompressible, then the transport of the conservative material can be described with the following Partial Differential Equation (PDE) [64]:

$$\frac{\partial \chi}{\partial t} = - \left(\frac{\partial(u\chi)}{\partial x} + \frac{\partial(v\chi)}{\partial y} + \frac{\partial(w\chi)}{\partial z} \right), \quad (4.1)$$

where the quantities u, v, w and χ are represented as the sum of a mean and eddy fluctuation. By expanding the terms and averaging the PDE, we get:

$$\frac{\partial \bar{\chi}}{\partial t} + \bar{u} \frac{\partial \bar{\chi}}{\partial x} + \bar{v} \frac{\partial \bar{\chi}}{\partial y} + \bar{w} \frac{\partial \bar{\chi}}{\partial z} = - \left(\frac{\partial \overline{u'\chi'}}{\partial x} + \frac{\partial \overline{v'\chi'}}{\partial y} + \frac{\partial \overline{w'\chi'}}{\partial z} \right). \quad (4.2)$$

Finally, replacing the eddy flux terms by the simple gradient-transfer forms and exchanging the left-hand side of the equation with the Lagrangian time derivative (cf., Equation 4.4), produces the general form of the K-theory model, commonly referred to as the K-model. The model is depicted in Equations (4.3), and 4.4) as:

$$\frac{D\bar{\chi}}{Dt} = \frac{\partial}{\partial x} \left(K_x \frac{\partial \bar{\chi}}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial \bar{\chi}}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial \bar{\chi}}{\partial z} \right) \quad (4.3)$$

$$\frac{D\bar{\chi}}{Dt} = \frac{\partial \bar{\chi}}{\partial t} + u \frac{\partial \bar{\chi}}{\partial x} + v \frac{\partial \bar{\chi}}{\partial y} + w \frac{\partial \bar{\chi}}{\partial z}, \quad (4.4)$$

where χ is the predicted mean concentration at point x, y, z ; t is time; and the $K_{x,y,z}$ terms represent the diffusion in each of the coordinate directions. Both analytical and numerical solutions to Equation (4.3) have been utilized in regulatory processes around the world [38].

In order to simplify this equation to its Gaussian form, it is typical to either assume that diffusivity is constant and independent of spacial direction (i.e. isotropic as opposed to anisotropic), or for further simplification, to apply the Fickian Diffusion assumption. For isotropic diffusion, Equation (4.3) is simplified to the following form:

$$\frac{D\bar{\chi}}{Dt} = K_x \frac{\partial^2 \chi}{\partial x^2} + K_y \frac{\partial^2 \chi}{\partial y^2} + K_z \frac{\partial^2 \chi}{\partial z^2}. \quad (4.5)$$

Alternatively, under the Fickian assumption, Equation (4.3) is further simplified to:

$$\frac{D\bar{\chi}}{Dt} = K \nabla^2 \chi, \quad (4.6)$$

with:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \quad (4.7)$$

and:

$$K = K_x = K_y = K_z.$$

Equation (4.6) can be applied to model atmospheric transport in one, two or three dimensions. For example, Equation (4.8) is the one-dimensional realization of the Fickian equations. In this equation, the first term represents the rate of change of the mean concentration, $\bar{\chi}$, with respect to time, and the second term corresponds to the advection of $\bar{\chi}$ at velocity \bar{u} .

$$\frac{\partial \bar{\chi}}{\partial t} + \bar{u} \frac{\partial \bar{\chi}}{\partial x} = K \frac{\partial^2 \bar{\chi}}{\partial x^2} \quad (4.8)$$

Gaussian Puff Model

The solution to the one-dimensional form of the Fickian equation, derived previously as Equation (4.8), for an instantaneous point source of emission strength Q is given by:

$$\bar{\chi}(x, t) = \frac{Q}{(4\pi Kt)^{\frac{1}{2}}} \exp\left[-\frac{(x - \bar{u}t)^2}{4Kt}\right]. \quad (4.9)$$

For a three-dimensional model, the solution to the Fickian equation is expanded to the following form:

$$\bar{\chi}(r, t) = \frac{Q}{(4\pi Kt)^{\frac{3}{2}}} \exp\left[-\frac{r^2}{4Kt}\right], \quad (4.10)$$

where:

$$r^2 = (x - \bar{u}t)^2 + y^2 + z^2, \quad (4.11)$$

and:

$$K = K_x + K_y + K_z. \quad (4.12)$$

Finally, for non-isotropic diffusion in the three coordinate directions, the Gaussian puff equation expands as follows:

$$\bar{\chi}(x, y, z, t) = \frac{Q}{(4\pi t)^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}} \exp\left[-\frac{(x - \bar{u}t)^2}{4K_x t} + \frac{y^2}{4K_y t} + \frac{z^2}{4K_z t}\right]. \quad (4.13)$$

This solution to Equation (4.3) is obtained according to the following boundary conditions:

- i The concentration at all points approaches zero as time goes to infinity, and thus,

$$\bar{\chi} \rightarrow 0 \text{ as } t \rightarrow \infty, \quad \forall x, y, z, \quad (-\infty < x, y, z < +\infty)$$

- ii The concentration goes to zero for all points except the source location, as time approaches zero, and so,

$$\bar{\chi} \rightarrow 0 \text{ as } t \rightarrow 0, \quad \forall x, y, z, \quad x, y, z \neq 0$$

- iii The total mass of the pollutant present at any particular time is equal to the total mass release, implying that,

$$\int_{-\infty}^{\infty} \bar{\chi} dx dy dz = Q$$

Gaussian Plume Model

While the Gaussian plume model has seen considerable application in the past, it is regularly noted that, in the strictest of terms, the Gaussian model is limited in its applicability, as it requires large diffusion times and homogeneous, stationary conditions. However, we cite Batchelor's supposition that the Gaussian function may provide a general description of the average plume diffusion because of the essential random nature of the phenomenon, by analogy with the central limit theorem of statistics [103]. Batchelor's assessment, in combination with the ease of implementation inherent in the Gaussian models, provide significant justification for the application of both Gaussian models in our modelling task. Moreover, as we aim to both simulate the variability and the mean tendencies at hypothetical receptor sites, within extremely complex, fabricated environments, and for utilization in PR experiments, the Gaussian models present themselves as well-suited frameworks.

The Gaussian plume dispersion equation has its foundation in the basic advection-diffusion equation, which through the series of assumptions previously described, can be solved analytically to produce the Gaussian puff equation, Equation (4.13). The Gaussian puff equation models the three-dimensional advection and diffusion of a neutrally buoyant cloud of tracer material in the atmosphere from the source to a receptor. By considering the continuous plume exiting from a source stack as an infinite number of Gaussian puffs, one arrives at the Gaussian plume model. Mathematically speaking, this implies integrating the Gaussian puff equations from $t = 0$ to $t = \infty$. After making a few simplifying assumptions, the Gaussian plume equations takes the following form, which was articulated by Lyons in [64] as:

$$\chi(x, y, z, t) = \frac{Q}{2\pi\sigma_y\sigma_z\bar{u}} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right) \right], \quad (4.14)$$

and describes the air pollutant concentration, χ , in *mass units* m^{-3} , at the receptor location, (x, y, z) , where the x -axis is assumed to be parallel to the mean direction of the wind. The parameter Q in the above equation represents the pollutant emission

rate from the source in *mass units* s^{-1} , the value \bar{u} takes the mean wind speed in $m s^{-1}$, and the values of sigma represent the crosswind and vertical dispersion as a function of the downwind distance, in meters. Finally, H takes the effective value of the pollutant plume's centre-line.

Turner, in [120], conveniently expressed the equation in terms of four distinct factors, which are combined to produce the final estimate. These factors represent the dependence upon emissions released from the source and the time-averaged atmospheric conditions. The emissions factor, Q , indicates that the concentration at the receptor site is directly proportional to the emissions. The downwind factor, $\frac{1}{\bar{u}}$, specifies that parallel to the x -axis, the concentrations are inversely proportional to the wind speed. Parallel to the y -axis, the crosswind factor,

$$\frac{1}{(2\pi)^{1/2}\sigma_y} \exp\left[-\frac{y^2}{2\sigma_y^2}\right],$$

indicates that the concentrations are inversely proportional to the crosswind spreading, σ_y , of the plume. The greater the downwind distance, the greater the horizontal spreading, implying a lower concentration. The exponential involving the ratio of y to σ_y provides a correction factor for the distance of the receptor from the center of the distribution – quantified in terms of the number of standard deviations. Finally, parallel to the z -axis, the vertical factor,

$$\frac{1}{(2\pi)^{1/2}\sigma_z} \left\{ \exp\left[-\frac{(z-H)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+H)^2}{2\sigma_z^2}\right] \right\},$$

specifies that the concentrations are inversely proportional to the vertical spreading, σ_z , of the plume. Once again, as the downwind distance increases, so does the vertical spreading, implying a lower concentration of the pollutant. The sum of the exponential terms in the vertical factor represents how far the receptor height, z , is from the plume's center-line, H , in the vertical direction. The first term represents the direct distance, $H - z$, of the receptor to the center-line. The second term represents the reflected distance, the distance from the plume's center-line to the ground and back up to the receptor. The last term accounts for the reflection of the spreading plume off the earth's surface.

Through iterative evaluations of Equation (4.14) over the experiment, with some location-specific fluctuations in the parameters over time, it becomes apparent that the background distribution at the receptor site is a function of the receptor's location relative to the industrial emitters, along with the mean tendencies of the individual industrial emitters and the overlying atmosphere.

4.4 Demonstration of the Dispersion Models

Before preceding to a demonstration of the simulation system in Section 4.5, we first examine the individual “plume” and “puff” components of the model, in order to understand the effect of the various atmospheric forces acting on them.

Regardless of the source type, plume or puff, the atmospheric forces play an integral role in determining their fate. Moreover, these forces often have similar effects on the various pollutant bodies. Thus, whenever possible, in the subsequent sections we explore the effect of these forces on a pollutant plume, and implicitly extrapolate the results to the puff scenario.

Atmospheric Stability

Chapter 2 first introduced the notion of atmospheric stability and its effect on pollutant concentrations. In this simulation, we characterize atmospheric stability based on the five Pasquill-Gifford (P-G) stability categories. The P-G class six signifies a stable atmosphere, which affects little dispersion, while category one signifies a highly unstable environment, and one that is very conducive to the dispersion of pollution.

In addition, to the influence of wind speed and plume height, Figures 4.1 through 4.3 illustrate the effect atmospheric stability on pollutant concentrations. In regard to the Gaussian plume model, there are two key diffusion parameters, the crosswind and vertical variances, which are determined as a function of the stability and the downwind distance. These parameters ultimately determine the degree to which the pollutant plume is diffused in the horizontal and vertical directions.

In Figure 4.1, the effect of stability on crosswind diffusion is most clearly demonstrated. Each plot in this figure has a constant, and equivalent, wind speed, receptor

height, and plume height. Thus, the only variable is the stability, measured in terms of the P-G categories. The emission rate Q , was not selected to emulate any particular nuclear point source, but for demonstration purposes. The first two plots in this figure, along with those in the two following figures, illustrate simulation results produced during P-G classes one and two. Alternatively, the third and fourth plots correspond to stability classes of five and six. More specifically, the results presented in the first two plots of Figures 4.1 through 4.3 were produced during stable atmospheric conditions, while the latter two are the results of unstable conditions.

Internally, the individual plots demonstrate the simulated concentration at six horizontal distances from the plume's centre-line, ranging from zero metres (solid red line) to 25 metres (dashed pink line). Intuitively, as the observations occur further from the plume centre-line, the overall concentration decreases. Once again, the short distances were selected for demonstration purposes. The general shape of the curves, however, extend to longer distances with large emission rates.

Generally speaking, two key results are depicted by these figures. The first of these is that as stability increases, so does the maximum detected concentration. This is demonstrated by the increased magnitudes realized in the successively more stable plots. The second result that is demonstrated is that the pollutant concentration remains at elevated levels for longer durations during stable atmospheric conditions. Once again, this is visualized in the successive plots. However, this notion is seen by the increasing width of the curves.

An additional result is realized through the comparison of the consecutive plots in this figure. The unstable environment present during the first simulation produced a quick dispersion of the pollutant concentration. Thus, there is only a minute delay in terms of downwind distance prior to each successive inward plot reaching its peak concentration. In contrast, the centre-line plot arrives at its maximum magnitude at approximately 150 metres downwind, while the most inward does not arrive at its maximum until after 250 metres.

Vertical dispersion is contrasted in Figures 4.1 and 4.3. In the former, the plume height is elevated 25 metres above the ground, while in the latter, it is elevated 50

metres. As previously demonstrated, during stable conditions, diffusion is significantly slower. Because there is a greater distance between the plume height and the receptor height in the latter figure, we expect that during similar conditions, the peak concentrations will occur at greater downwind distances. This is, indeed, confirmed by comparing the two figures.

Wind Speed

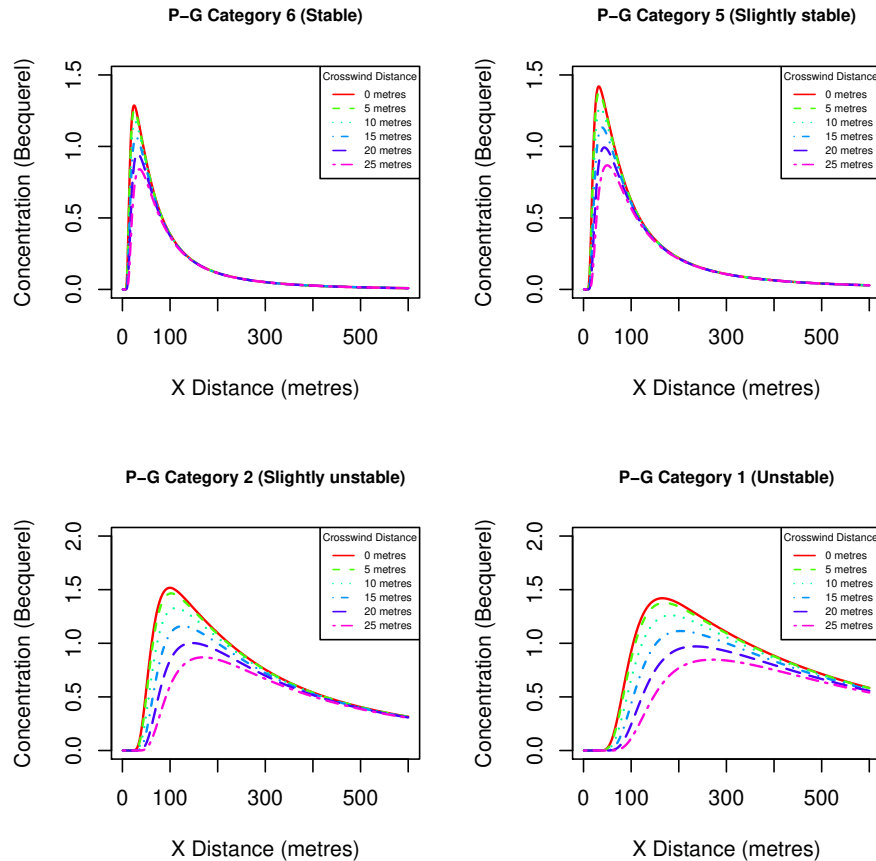
Wind speed, very clearly, has a profound effect on pollutant bodies in the earth's atmosphere. We introduced, for example, the concept of mechanical turbulence in Chapter 2. Mechanical turbulence has a mixing effect, and, in the atmosphere, is generated as the wind traverses rough surfaces. Thus, a decrease in concentration is observed with increased wind speeds.

The shift down the gradient scale, which is affected by an increase in the rate of advection on a stack emitted plume at various downwind distances, can be observed through careful examination of Figures 4.1 and 4.2. A considerable amount of information is provided in these figures. However, for the moment, consider only the solid red lines in the first plot of Figures 4.1 and 4.2. These plots illustrate the results of simulations that were identical in all manners, except for the rate of advection. As expected, the curves are very similar in shape, while the second figure has a lower magnitude, resulting from increased mixing due to the elevated wind speed.

Indeed, the primary effect of the rate of advection on a pollutant body is the speed at which the parcel is transported downwind. In terms of a pollutant plume, the effect can be explored under two scenarios, with and without chemical decay. If decay is a factor, an increase in the wind speed facilitates an increase in the area of influence of the emission source. In particular, each parcel is able to travel greater distances before chemical decay becomes a significant factor in removing the pollutant from the environment.

The previously explored figures provide examples of the latter scenario, in which the only change is a proportional decrease in the pollutant concentration due to mixing. This is an intuitive result, when one recalls the earlier analogy of a plume as an infinite number of puffs, or parcels. In a homogeneous environment, the wind

Plume Concentration with Distance



$$\mu_{ws} = 2 \text{ m/s, } z_r = 10 \text{ metres, } H_e = 25 \text{ metres}$$

Figure 4.1: The successive plots in this figure demonstrate the effect of an increasingly stable atmosphere on downwind pollutant concentrations. The individual lines plotted in each graph depicted the downwind concentrations at six specific crosswind distances. In each simulation, the receptor height was 10 metres above the ground, the plume centre-line was at 25 metres, and the wind was blowing at 2 metres per second.

effects each parcel of air identically. Thus, as advection increases, the analogous train of identical parcels, p_i , emitted from an industrial stack, increase in momentum. However, since each p_i has an identical concentration, for a homogeneous environment

the magnitude decreases. In our simulation, though, we did not experience the sharp rise, and subsequent fall, that is synonymous with an advecting pollutant cloud, as it passes over a receptor site.

As previously remarked, wind speed has a significantly more visible effect on a puff of pollution. This is illustrated in Figure 4.5. The bottom graph in this figure represents an increase in wind speed from 2 metres per second to 5 metres per second. The x -axis in this figure specifies the time lapsed since the release of the puff. The individual line plots articulate the change in the measured concentration at five distinct downwind receptor sites over time. The two effects of wind speed on a pollutant puff are soundly demonstrated in this figure. As the rate of advection increases, the pollutant reaches the receptor sites quicker and its presence is detectable for less time. This is visualized as both a shift to the left in the lower graph, and as a narrowing of the bell shape.

Plume Height

Some interesting results appear in Figure 4.3 as a result of increasing the plume height. The plume height in this figures is 50 metres, 25 metres above the previous plume heights. In comparison with Figure 4.1, it is noticeable that the concentration is consistently and considerably lower than seen in the previous figure. This results from the fact the the pollutant must diffuse a greater vertical distance to reach the receptor. An additional effect of the increased distance, is that the maximum magnitude is achieved at a greater distance. Finally, in terms of the crosswind concentrations, unlike the earlier plots, there is very little visible difference in the plotted crosswind concentration. This results from the fact that the vertical distance is greater than all of the horizontal distances plotted.

Seasonal Variations

To re-iterate our goal, we briefly explain the effect of seasonal variations in the atmospheric means, such as the wind speed and direction, on pollutant concentrations at the receptor site.

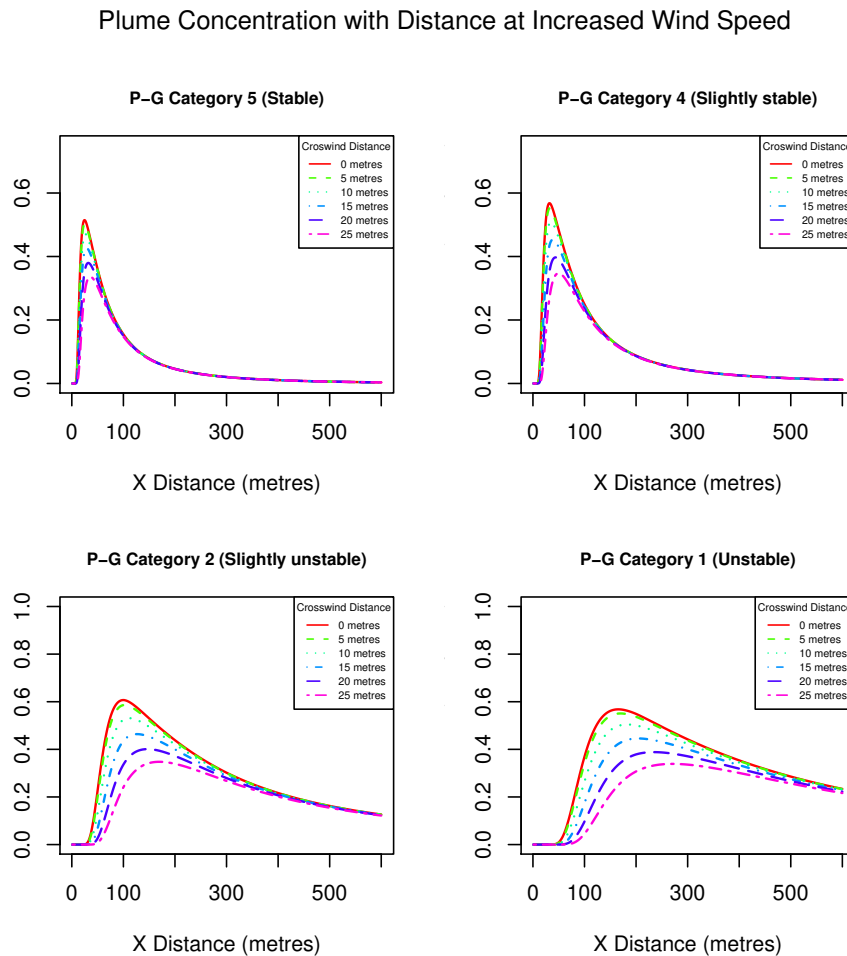
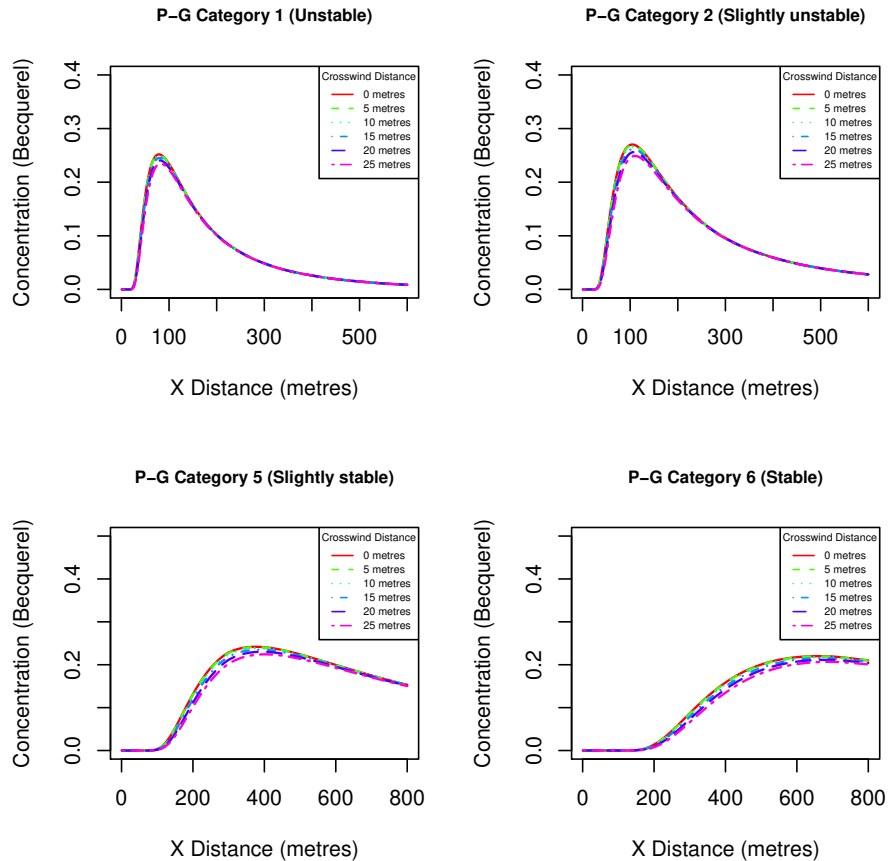


Figure 4.2: The successive plots in this figure demonstrate the effect of an increasingly stable atmosphere on downwind pollutant concentrations, at an elevated wind speed of 5 metres per second. The individual lines plotted in each graph depict the downwind concentrations at six specific crosswind distances. In each simulation, the receptor height was 10 metres above the ground, the plume centre-line was at 25 metres and the wind was blowing at 5 metres per second.

Unlike the three previous sections, which considered short-term meteorological averages, long-term season variability is only of interest when considering continuous source emissions. This is because the SE events, which we model as instantaneous

Plume Concentration with Distance at Increased Plume Height



$$\mu_{ws} = 5 \text{ m/s, } z_r = 10 \text{ metres, } H_e = 50 \text{ metres}$$

Figure 4.3: The successive plots in this figure demonstrate the effect of an increasingly stable atmosphere on downwind pollutant concentrations, when the plume centreline is elevated from 25 metre to 50 metres. The individual lines plotted in each graph depict the downwind concentrations at six specific crosswind distances. In each simulation, the receptor height was 10 metres above the ground, and the wind was blowing at 2 metres per second.

sources, are both rare and random (in time, space and magnitude) with relatively short lifespans. Thus, no seasonal trend, such as that depicted in Figure 4.4, will appear.

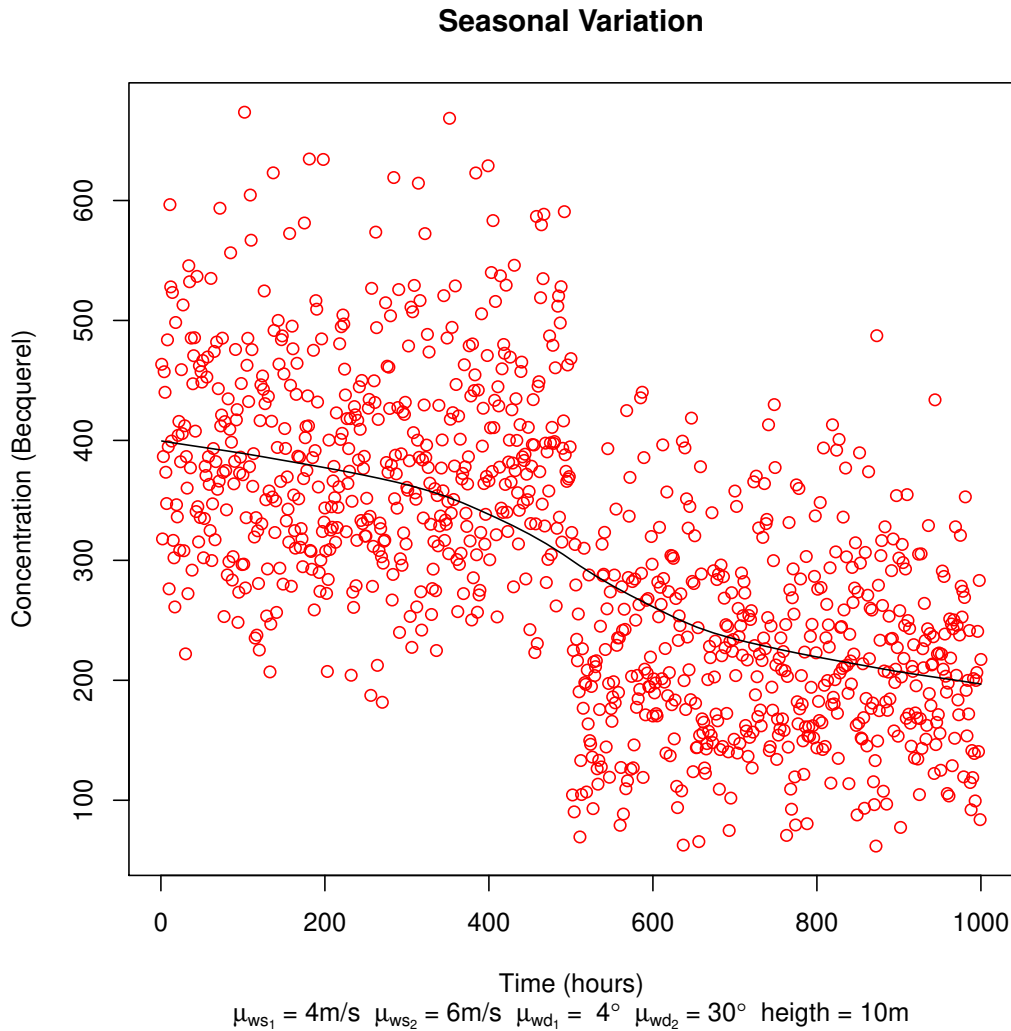


Figure 4.4: This figure illustrates the influence of seasonal meteorology on remote receptor sites. In particular, two seasons, each with distinct mean wind speeds and directions are simulated. The first five-hundred hours of the plot correspond to the first season, which more directly effect the receptor site. Thus, the first five-hundred hours display, on average, high pollutant concentrations.

Figure 4.4 demonstrates the simulated downwind pollutant levels in a hypothetical environment with two characteristic seasons. During the first season, the mean wind blows more directly at the receptor site and has a lower average velocity. Thus, in general, higher concentrations are observed during the first half of the plot, which

corresponds to the first season. We find it particularly interesting to simulate such scenarios, not only for its practicality, but also for the effect such variability may have on PR systems. It is notable that in this figure there exists a discrete shift from one season to the next. The current approach was utilized to emphasize a possible affect of changing seasons. In practice, however, when simulating seasonal changes, we utilize a process of randomization to emulate both gradual, and less predictable shifts.

4.5 Experimental Setup

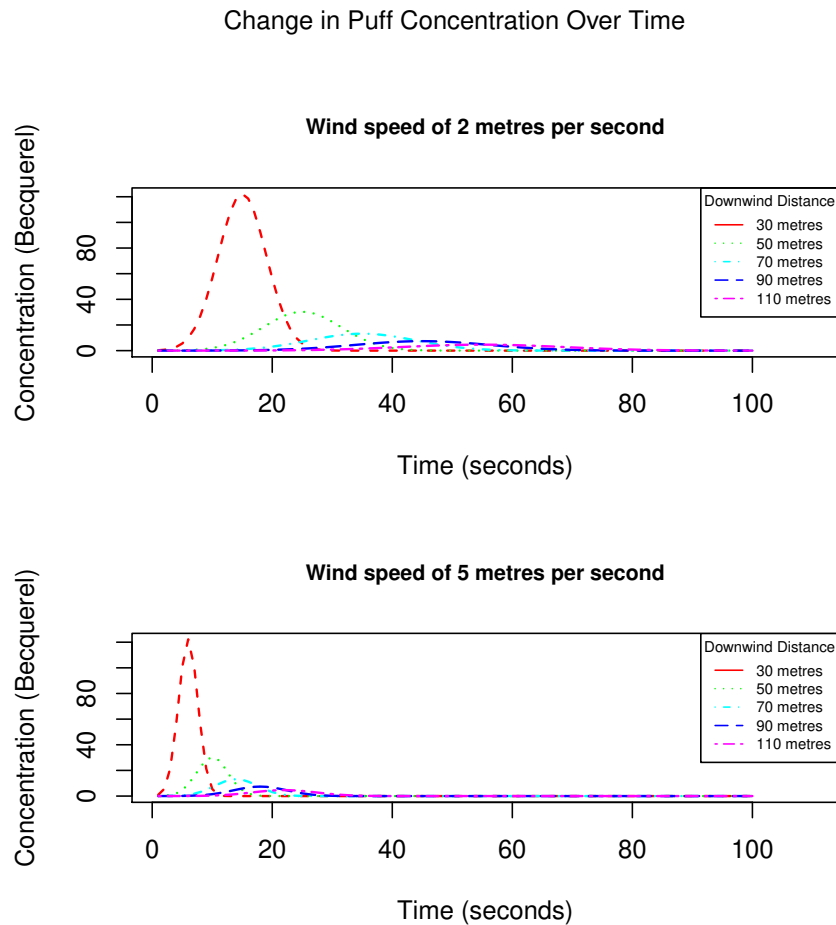
In this study, we propose a modelling technique designed to facilitate the simulation of SE events propagating through a modelled system. As a means of demonstrating this theory, we utilize the particularly interesting scenario suggested by the verification of the United Nations' CTBT. The remainder of this section provides essential details of the modelled system.

4.5.1 Motivation

The CTBT is a United Nations treaty, which, when it enters into force, will prohibit the detonation of nuclear weapons by member nations. As a result, a number of verification strategies are currently under study, aimed at ensuring the integrity of the treaty. The primary verification technique being explored relies on the quantity of radionuclides measured at sampling stations, otherwise referred to as “receptors”, distributed throughout the globe [106].

In general, it can be assumed that radionuclides are present within the atmosphere for one of two reasons, the primary being emissions from the nuclear industry. Alternatively, the detonation of nuclear weapons are known to release mass quantities of radionuclides into the atmosphere. This is particularly the case for surface and airborne detonations, but is also true, although to a lesser extent, for subterranean detonations.

Indeed, even in the most frightening of scenarios, it is expected that the testing



P-G Category = 4, zr = 10 metres, He = 10 metres

Figure 4.5: This figure illustrates the effect of wind speed on a pollutant puff. In both simulations, the P-G category was 4, the receptor height was 10 metres, and the puff elevation was centred at 10 metres. The wind speed in the second simulation was increased from 2 metres per second to 4 metres per second.

of nuclear weapons will be characteristic of an SE event. Alternatively, the operation schedules for individual nuclear industries are typically defined into the distant future, and thus, in the medium-term, it can be assumed that their emission rates are relatively consistent. We make this assumption with the above-defined modelling and simulation task in mind, and consequently do not aim to articulate fluctuations

that may result from cyclical production cycles within individual plants or resulting from unexpected shutdowns. Instead, our objective is to model the general effect of local industries on a particular receptor site, and to subsequently simulate the effect of SE explosions propagating through the system.

4.5.2 Modelled System

For the purpose of this demonstration, we assume a simplified environment. In particular, we apply a few simplifying assumptions to the process of atmospheric dispersion, industrial emissions and the SE emissions, in order to illustrate the general effects of such events at the receptor site. In addition, to maintain simplicity in this illustration, we utilized a small simulation domain with appropriately low emission rates. However, the results are extensible to larger domains. Indeed, a significantly larger domain is utilized in Chapter 5.

Atmosphere

Once emitted, an airborne cloud of pollutants becomes subjected to a complex array of interdependent forces, which stretch, pull and fold the pollutant body. Theorists and practitioners have attempted to classify the diffusive effect that results from these processes in general, and in particular, from the over-riding stability of the atmosphere, which we described in Chapter 2 as ranging from super-adiabatic to inversion, in a variety of ways. The Pasquill-Gifford stability classes (P-G classes) [35, 77] discretize atmospheric stability into classes ranging from A through D, where A indicates unstable conditions, G, strongly stable, and D, represents a neutral, or nearly adiabatic stability. The P-G classes were formulated based on an empirical study, and are presented in tabular form with the corresponding class specified as a function of wind speed, incoming solar radiation and cloud cover.

The Pasquill-Gifford dispersion parameters [99], which specify vertical and horizontal diffusion as a function of downwind distance and stability, have experienced

considerable favour within Gaussian dispersion models, and are applied for the determination of the K and σ terms in Equation (4.13) and Equation (4.14). More recently, new techniques have been proposed, which are subjected to fewer restrictions. However, for the purpose this study, the former strategy is applied, as it effectively captures the general nature of this study, and its can be efficiently implemented.

Diffusion aside, we choose to view both the atmosphere, and, indeed, the broader environment, from a purely statistical perspective. Thus, we assume that over the course of an experiment the wind speed and direction take independent Gaussian forms with user-defined means and variances. In addition, we facilitate the user definition of n seasons, each with a possibly distinct set of atmospheric tendencies. The potential for season variability is an intuitive notion when considering the earth's atmosphere. Seasonal fluctuations are, of course, a regular phenomenon in regions around the world. Such variations can have a significant effect on pollutant levels, as demonstrated by Gariazzo in [32], where season variations in meteorology are shown to transport, more or less, pollution from industrial sources depending on the season. These fluctuations compose possibly complicating factors when it comes to the detection of episodic events.

Background Emissions

Within this experiment, the background distribution is modelled as per the real-time measurements of radioxenon at a selected receptor site. Also, as previously indicated, in general, the background levels of radioxenon can be attributed to the nuclear industry, with the primary sources being the production of medical isotopes and the generation of nuclear power.

For the purpose of this experiment, we forgo the more realistic notion of cyclical emissions cycles, which may result from refuelling, planned maintenance, safety inspections, etc., as a means of concisely achieving our overall objective. Furthermore, the emissions rates and plume rise phenomena are assumed to take independent Gaussian forms, with user-defined means and variances. Moreover, subsequent to exiting the emissions stack, the plume is assumed to instantaneously reach its determined center-line, and to begin its dispersion in the downwind direction under steady-state

conditions over the course of a reasonable amount of time, say, one hour. Subsequent to the hour of dispersion, the model parameters are randomly recalculated within their respective Gaussian curves. In-line with the enabling assumption of homogeneous, stationary conditions for Gaussian dispersion models, the recalculated wind speed and direction are assumed to hold for the entire model over the duration of this hour.

SE Emissions

The SE events in this simulation take the form of clandestine nuclear explosions, the magnitudes of which are assumed to be more than an order higher than the background rate of emission. Thus, the detonation is somehow contained in an attempt to both conceal any visual evidence from satellites and other flyovers, which would induce suspicion, and to restrict the release of radionuclides. As seen in the past, however, the containment of the pollutants produced during a subterranean nuclear detonation is not a straightforward task [92]. Moreover, the inert property of radioxenon dictates that large quantities are likely to be vented from even the soundest of containment facilities, after a detonation [15].

Based on the evidence presented above, and consistent with [92], we assume that a random portion of the produced radioxenon is immediately vented and subsequently dispersed downwind. Furthermore, the total radioxenon produced by the detonation is assumed to be large relative to the background levels, with a substantial portion of it being vented into the lower atmosphere.

Being SE in nature, the occurrence of an explosion at any particular point in space and time, is at best estimated by a uniform random event. Thus, at any specific point in the simulation, it is unlikely that an explosion will occur, however, the remnant cloud of an early explosion is expected to be present within the modelled system, at trace levels, for some time. When a detonation occurs at a particular time, t , for example, it promptly becomes subject to the atmospheric condition present at that time, and continues to disperse until it exits the modelled domain or completely decays.

4.6 Results

This section details the results produced by the simulation processes and scenarios described above. We begin, in the subsection that immediately follows, by demonstrating how a single industrial emitter effects the radionon levels at a set of regional receptor sites. The accompanying subsection, Subsection 4.6.2, illustrates how the episodic events propagate through the otherwise consistent background distribution.

4.6.1 Background

In order to illustrate the propagation of episodic events through background noise, we simulate a simplified version of the CTBT scenario. In particular, the modelled environment is a five-hundred square meter site with a single industrial emitter and four downwind receptor sites.

The receptor sites are situated 100, 130, 160 and 190 meters downwind from the industrial emitter, at elevations of one meter. Upwind, the industrial pollutants are emitted from a stack elevated to twenty-five meters, and are expelled at a mean rate of 10,000 units per second and with a standard deviation of 10 units per second.

The overriding atmosphere is assumed to maintain a homogeneous, steady-state condition for a period of one hour, at which time the key atmospheric parameters are recalculated around their individually defined means. For the purpose of this experiment, the mean wind speed has been specified at a velocity of 7 meters per second, with a standard deviation of 5 meters per second. On average, the wind direction is assumed such that it transfers the pollutant plume emitted from the industrial source towards the receptor sites. Beyond this, it is subject to a standard deviation of 5 degrees. Within this simplified model of the atmosphere, diffusion is assumed to be approximately isotropic. Therefore, the $\sigma_{x,y,z}$ values in Equation (4.14) and Equation (4.13) assume a specified mean value of 15 with a standard deviation of 5. However, we use the term “approximately isotropic” because, each σ value oscillates independently about the mean, and therefore, they do not necessarily realize the same values at any particular point in the simulation.

Based on the above details, ten experiments were run, each over a period of 1,000

hours, which is approximately forty-one days. Unlike “real world” scenarios, which are subject to considerable uncertainty, and thus, require significantly longer experiments, the duration of this experiment is sufficient to ensure stable results. This is specifically a consequence of the fact that the means and higher order moments are known, and is validated based on the stability displayed in Figure 4.6. During each experiment, the mean hourly pollutant concentration was recorded. The resulting background probability distributions for the four receptor sites are displayed in histogram form in Figure 4.6.

By calculating the hour-on-hour mean over the ten iterations of the experiment, we derive an ensemble average for each hour in the forty-one day period. For another perspective, we also present the ensemble averages as illustrated in Figure 4.7. This figure gives a scatter plot with the successive hours plotted on the x -axis and the ensemble mean volume plotted on the y -axis.

4.6.2 SE Events

Unlike SE events, such as earthquakes and tsunamis, which radiate outward from the epicentre in all directions, the body of a pollutant cloud is advected in the direction of the mean wind. Moreover, the rate of advection often overshadows diffusion in each of the three coordinate directions. Therefore, only those receptors located in the general downwind direction can expect to be effected by a pollutant cloud. Thus, events that occur upwind are of primary interest in this experiment.

Being cognisant of the above fact, and for demonstration purposes, we model four upwind detonations at varying distances, thus, maximizing their effects on the receptor sites. In particular, detonations are simulated at distances of 200, 150, 100 and 50 meters upwind. Their individual effects are superimposed on the time-series scatter plot of the hourly ensemble means previously described in Figure 4.7. In this figure, in order to illustrate the shape of the peaks, the effect of radioactive decay was omitted from the simulation that produced these results. Alternatively, the effects of the decay are demonstrated in Figure 4.9. In this figure Figure 4.8, the solid blue circles indicate the background levels discussed earlier. The four peaks in

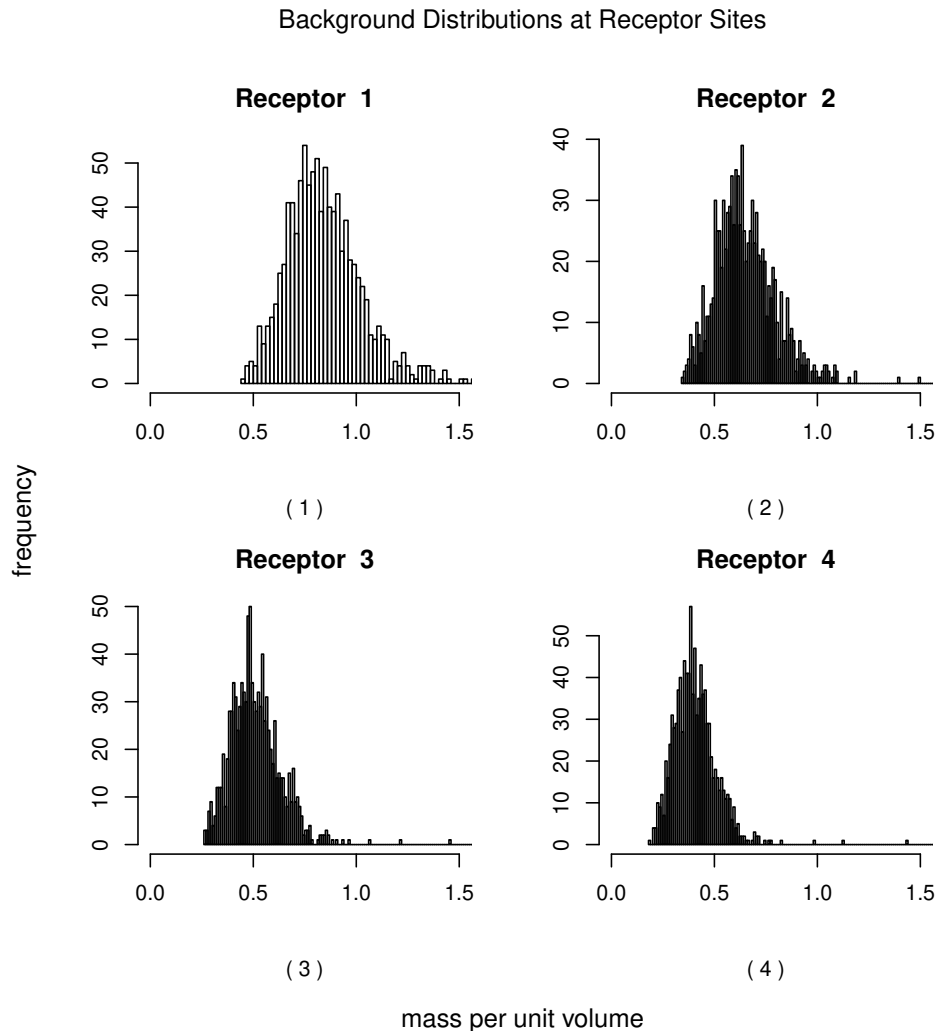


Figure 4.6: The calculated probability distribution for pollutant levels resulting from a single regional industrial emitter at each of the four receptor sites located 100, 130, 160 and 190 meters downwind. These results were calculated over a forty-one day period.

this plot result from four subterranean detonations, each approximately four orders of magnitude larger than the industrial emission rate. The four successively larger peaks, demonstrate the effect of moving the denotation incrementally closer to the receptor site. In particular, the taller, narrower peaks resulted from detonations that occurred closer to the receptor site, and thus, had less time to diffuse. As a result, the

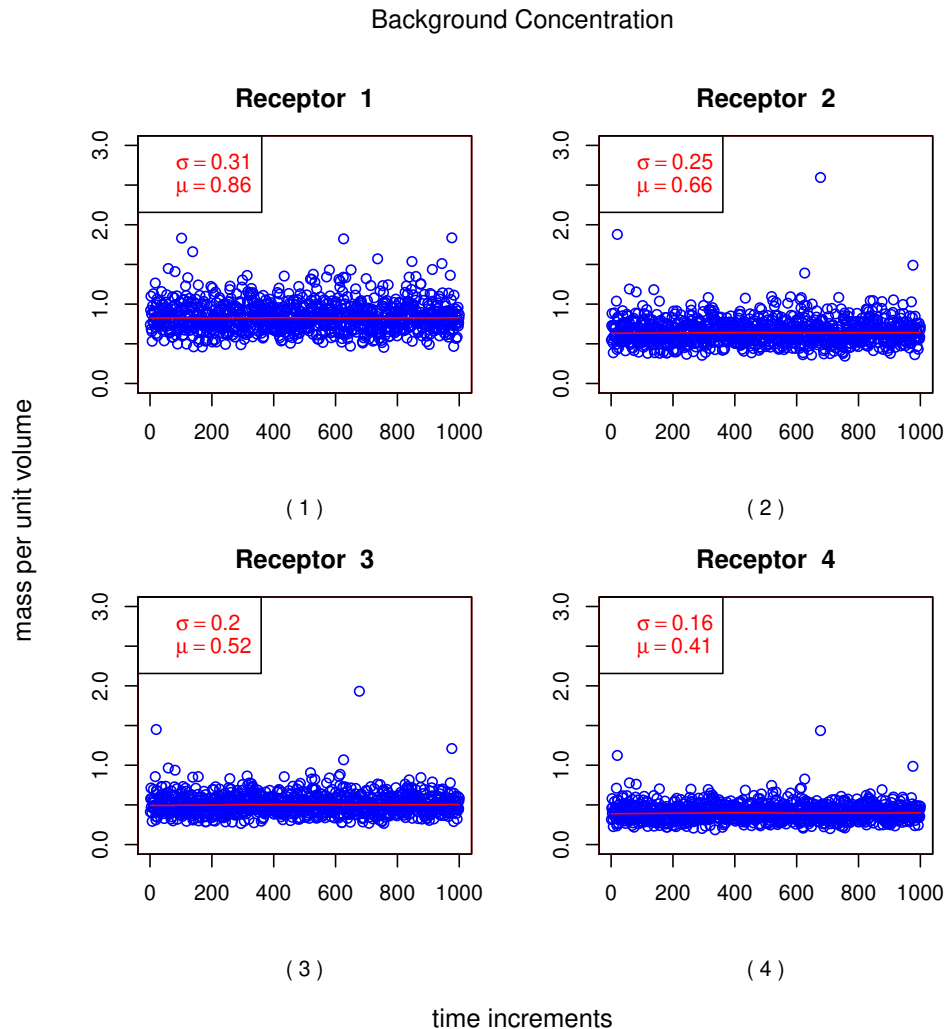


Figure 4.7: The time-series scatter plot illustrating the mean hourly background pollutant concentrations resulting from an upwind industrial emitter, for each of the four receptor sites located 100, 130, 160 and 190 meters downwind. These results were calculated over a forty-one day period.

majority of the pollutant concentration is witnessed over a shorter period of time and at higher levels. Alternatively, the pollutant clouds characterized by the lower, wider, peaks have travelled a longer distance, and thus, have had a greater opportunity to diffuse. Consequently, the receptor site depicts lower pollutant levels although their peaks are maintained for slightly longer durations. For detonations at greater

distance, this effect is further amplified.

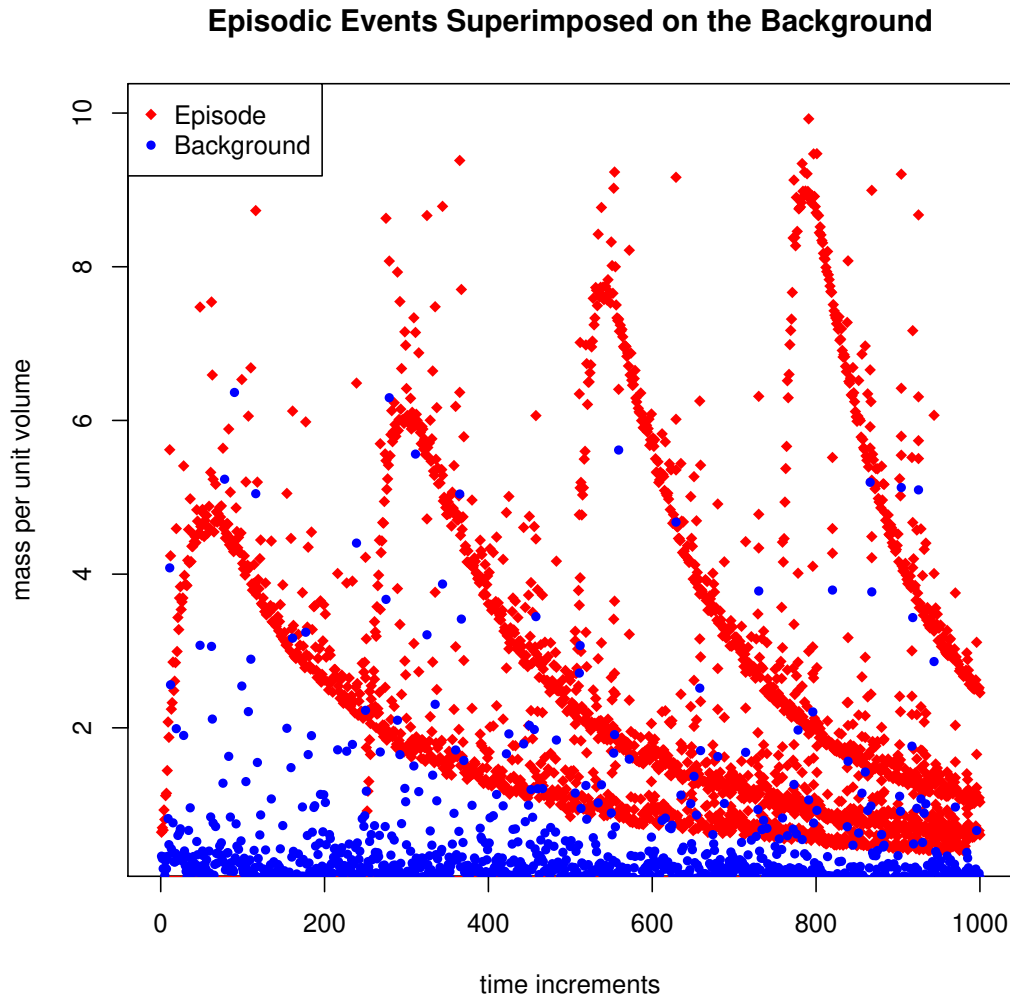


Figure 4.8: The time-series scatter plot illustrating the effect of four episodic events on pollutant levels at the receptor site nearest to the industrial emitter. The details about the figure and its legend are found in the body of the paper.

The downwind propagation of an instantaneous vented cloud of radioxenon, resulting from a subterranean detonation is depicted in Figure 4.9. The vented cloud of radioxenon was four orders of magnitude larger than that resulting from the industrial emitter. In particular, this figure demonstrates the effects of both dispersion and radioactive decay. By comparing the SE event's spike in Figure 4.9 (Sub-figure

1) with the third episodic spike in Figure 4.8 (which corresponds to a similar detonation), it becomes apparent that the radioactive decay plays a fundamental role in the removal of radionuclides from the atmosphere. In particular, the SE event peaks slightly above a concentration of 2 in Sub-figure 1 of Figure 4.9, although it reaches elevations greater than 8 in the previous figure, where the decay was not accounted for. Dispersion and decay are also visible in the fading of the spikes at the successive receptors depicted in the four sub-figures.

4.7 Discussion

An analysis of the accuracy of the Gaussian dispersion models presented above is beyond the scope of this report. Indeed, as earlier indicated, these models, and their variants, have received considerable application and analysis in the past. For in-depth analyzes of these techniques, interested readers are directed to the earlier citations.

Of primary interest in this work, is the relationship between the background data and the SE events, and in particular, how this relationship can be modelled. We theorize that over the long-term, in many instances, the background sources of features deemed to be of particular interest, such as air polluting industries or regular p- and s- wave generating small-scale tectonic movements, can be modelled in a probabilistic manner based on a knowledge of their emission rates and the medium in which the propagation takes place. Radionuclide emissions from the nuclear industry are used to demonstrate this theory, and the results displayed in Figure 4.6 indicate that our objective of demonstrating how the general characteristics of the background data can be modelled has been realized. As expected, the model captures the key feature system under study. In particular, the simulation demonstrates the successive movement of the distribution to the left of the histogram at greater distances from the source. Furthermore, we witness a narrowing of the shape of the histogram at a greater distance, which is indicative of the pollutant body becoming increasingly well-mixed over an expanding area. Both of these results are confirmed in Figure 4.7 through the sequentially decreasing mean values and standard deviations. Larger-scale and longer experiments, which include seasonal variations in the atmosphere,

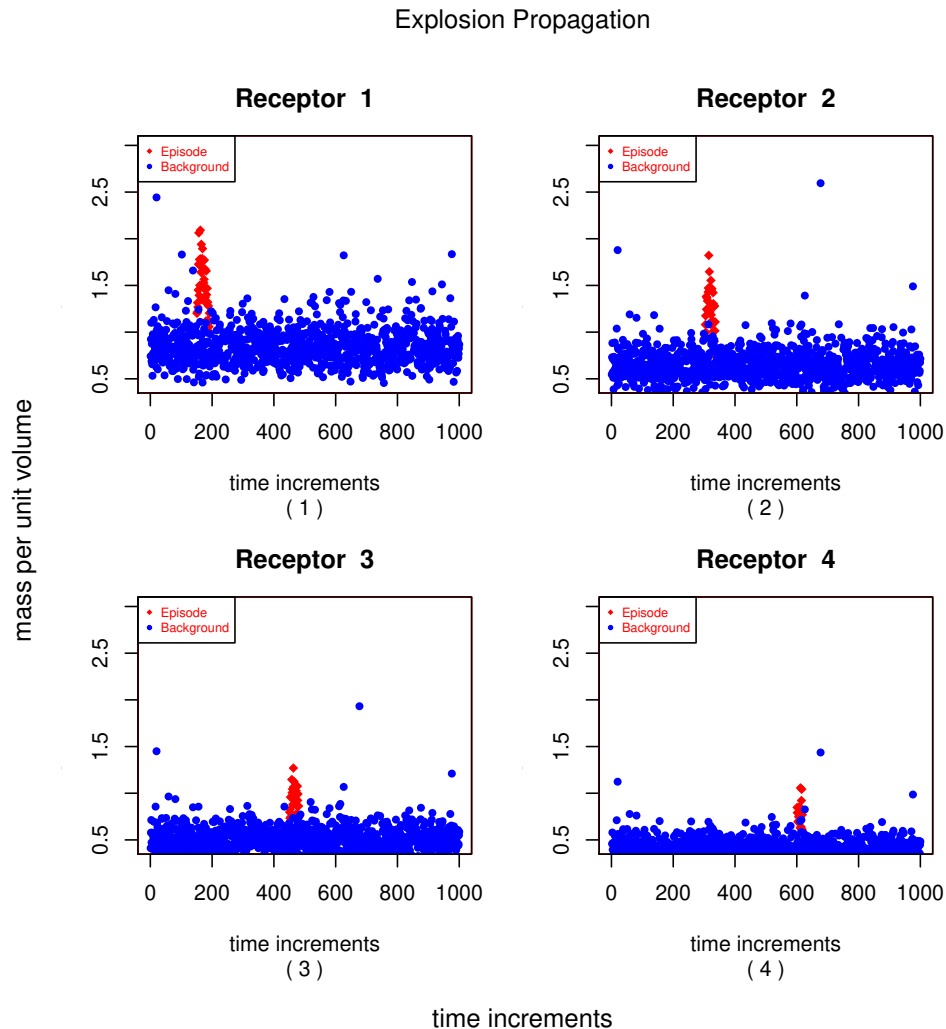


Figure 4.9: This figure illustrates the downwind propagation of a cloud of vented radioxenon. In particular, the effects of diffusion and radioactive decay are demonstrated through the incrementally decreasing concentrations of radioxenon (modelled as smaller red spikes) measured at each successive receptor site.

and for more emitters at greater distances, will help to further validate this in the future. However, we believe that the current model presents a solid foundation for any future study.

SE events, which are characteristically random in a multitude of ways, are inherently difficult to simulate. Moreover, the successful recreation of one event within a

modelled domain does not imply a generally applicable model. A great deal of effort must be applied to the proper tuning of most models in order for them to recreate events with minimal error. However, often, as is the case for the generation of data for PR systems, our desire is to produce a large number of plausible scenarios that capture the general relationship between the background and SE events. Moreover, we are interested in how the rare SE events, which possess largely unknown distribution in time, space and magnitude, can be simulated, and their effects subsequently propagated through the model. Figure 4.8 and Figure 4.9, provided excellent depictions of how a series of probabilistic choices can be applied to generate SE events in space, time and magnitude, in addition to propagating the resulting phenomena through the transmission medium. Indeed, the demonstrated model captures the subsequent rise and fall that occurs in the feature space of an individual downwind receptor affected by the episode. In particular, the slumping spikes in Figure 4.8 illustrates the relationship between the dispersion and distance travelled, as does Figure 4.9, which additionally demonstrates the modelling of radioactive decay.

Our proposed modelling framework, thus, provides an effective means by which a researcher can execute the initial exploratory phase analysis within in a large number of domains and scenarios, in addition to facilitating a domain-specific data generation scheme for the training, testing, validating and debugging of PR systems.

4.8 Conclusion

In this chapter, we have considered a relatively new field, namely that of modelling *SE events* such as earthquakes, nuclear explosions etc. The difficulty with such a modelling process is that most of the observations appear as noise. However, when the SE event does occur, its magnitude and features far overshadow the background, as one observes after a seismic event.

In particular, we present a straightforward theory, which states that in the long-term, measurable features produced by characteristically noisy background sources, take a relatively consistent and recognizable form. Moreover, by using the knowledge of the propagation medium and a general description of the background sources, the

major features can be modelled sufficiently for the purpose of exploring the effects of SE events. Given the largely random and sporadic nature of these SE events, we argue that for exploratory purposes, they proceed through a series of probabilistic decisions.

In the spirit of the radionuclide monitoring challenge suggested by the Comprehensive Nuclear Test-Ban-Treaty, we demonstrate how the nuclear industry can be assumed to take the role of the background source, thus, affecting relatively consistent levels of radioxenon at a set of receptor sites. Subsequently, we have demonstrated how the consequence of detonations of nuclear weapons can be generated and propagated through the modelled system.

Chapter 5

Classifying Stochastically Episodic Events

5.1 Introduction

A common assumption within supervised learning is that the distributions of the target classes can be learned, either parametrically or non-parametrically. Moreover, it is assumed that a representative set of data from these classes is available for the training of supervised learning algorithms; indeed, the latter implies the former¹.

Beyond this commonly reported method of classification, there exists a special form of PR. This “exceptional” category of binary classification is noteworthy in lieu of the significant challenge that it presents. Escalating the difficulty, is the fact that drawing a representative set of data to compose the second class, which is fundamental to the derivation of a binary discriminant function, is abnormally arduous, if not altogether impossible. The difficulty of acquiring a sufficiently symbolic set may arise because of (a) the natural *imbalance* in the classification task, (b) the difficulty of acquiring samples of the minority class, and (c) the task of obtaining representative samples of the minority class is overwhelming.

PR tasks of this nature have previously been constituted as involving outlier (or

¹Some of the research that is presented in this chapter has been submitted for publication in [7].

novelty) detection due to the fact that the vast majority of the data takes a well-defined form that can be learned, and that samples from the ω_2 class will appear anomalously – outside the learned distribution. Examples of such problems have traditionally motivated research into the so-called area of “one-class” classification [34, 42, 49, 53, 111]. Although these problems are significantly more difficult than those that involve well-defined binary problems with ample data, the results in these papers report state-of-the-art PR solutions.

To expand the horizon of the field, we observe that there exists a further, and yet more challenging subset of the one-class classification domain of problems, which remains unexplored. Typically, the accessible class, and in particular, the data on which the one-class classifier is trained, is considered to be well-defined. Thus, it is presumed that this data will enable the classifier to generalize an adequate function to discriminate between the two conceptual classes. This, for example, was demonstrated in [111], where the training set consisted exclusively of images of non-cancerous tissue. Similarly, in [42], a representative set of the target computer user’s typing patterns, which are both easily accessible and verifiable, were utilized in the training processes.

The classification task becomes considerably more difficult when deriving a strong estimate of the target class’s distribution is unfeasible, and when the validity of instances drawn from the target class are suspect, or if the occurrences of the minority class are temporally (i.e., with respect to the time-axis) interwoven with the data from the majority class. More specifically, the derivation of a suitable classifier is further obstructed when the target class presents itself in the form of a time sequence that is interwoven with a *minute* number of SE events, which must be distinguished.

Under these circumstances, we envision two possible techniques for discriminating between the target class and the SE events of interest. If the incoming training data contains a sufficient quantity of accurately identifiable, SE events, a standard clustering/PR algorithm could be applied to label both the classes appropriately. Subsequent to the labelling procedure, a standard binary classifier could be trained and utilized to achieve the classification of novel instances. In this body of work, we refer to this scenario as S1, and the subsequent scenario as S2.

Alternatively, and more applicable in scenarios in which the SE events are extremely rare, all of the training data can be assigned to the target class, and a one-class classifier can be applied. The details of, and justification for, this approach are described in the subsequent sections. Our primary objective in this research is to illustrate how standard supervised learning algorithms can be applied to discriminate rare SE events, which apart being unanticipated, are random in magnitude and position within the sequence of background data.

The subsequent section, Section 5.2, proceeds to describe the target domain in general terms, and to draw a conceptual distinction between it and those to which one-class classifiers have traditionally been applied. The state-of-the-art in PR is described in some detail in Section 5.3. In Section 5.4, we describe an experiment based on the exemplary task of verifying the Comprehensive Nuclear Test-Ban-Treaty (CTBT). The results of the two classification experiments are contained in Sections 5.5 and 5.6. Finally, a discussion of the results is undertaken in Section 5.7, and Section 5.8, consists of our concluding remarks.

5.2 Characteristics of the Domain of Problems

In the introductory section, we presented an interesting sub-category of novelty detection. We contend that this particular class of problems can be viewed in a manner that separates it from the larger set of one-class classification tasks. To reiterate this, we have identified a new sub-category of PR problems. This sub-category has a set of characteristics that collectively distinguish it from its more general counterparts. The characteristics of this category can be best summarized as follows:

- The data presents itself as a time sequence;
- The state-of-nature is dominated by a single class;
- The minority class is challenging to identify, even for domain experts, and occurs both rarely and randomly within the data sequence.

To accentuate the difference between the problems that have been studied and the type of problems investigated in this research, we refer the reader to Table 5.1.

Dataset	Temporal	ID Challenge	Imbalance <i>Type I</i>	Imbalance <i>Type II</i>	Interwoven
Mammogram	No	Low	Yes	Medium	No
Continuous typist recognition	No	Low	Yes	Medium	No
Password hardening	No	Low	Yes	Medium	No
Mechanical fault detection	No*	Low	Yes	Medium	No
Intrusion detection	No*	High	Yes	High	No
Oil spill	No*	High	Yes	Medium	No*
CTBT verification	Yes	High	Yes	High	Yes

Table 5.1: A comparison of well-known one-class classification problems. The explanation about the entries is found in the text.

This table displays an assessment of six classification problems that have previously appeared in the literature on one-class classification. In addition, we include the CTBT verification problem that characterizes the problems analyzed in this research. The first column indicates whether the problem has traditionally been viewed as possessing an important *temporal* aspect. The three entries with an asterisk require special consideration. In particular, we note that while traditionally these domains have not been studied with a temporal orientation, they do indeed contain a temporal aspect. The subsequent column signals whether the manual labelling of data drawn from the application domain is a significant challenge. This is, for example, considered to be a very difficult task within the field of computer intrusion detection, where attacks are well disguised in order to subvert the system.

The following two columns quantify the presence of class imbalance. In the first of these, we apply a standard assessment of class imbalance, one which relies on the determination of the *a priori* class probabilities. Our subsequent judgement departs slightly from the standard view, and considers class imbalance that arises from the difficulty of acquiring measurements (due to cost, privacy, *etc.*). The final column specifies if the minority class occurs rarely, and randomly (in time and magnitude), and if it occurs within a time sequence dominated by the majority class.

To summarize, in this section we have both demonstrated the novelty of this newly introduced sub-category of PR problems, and positioned the CTBT verification task within it. We additionally note that the fault detection, intrusion detection, and

oil spill problems could be reformulated to meet the requirements of our proposed category. This, indeed, suggests a new angle from which these problems can be approached.

5.3 Pattern Recognition Algorithms

In this section we present, in some detail, the algorithmic composition of five binary classifiers and four one-class classifiers, which were respectively applied to the classification scenarios previously described as S1 and S2.

5.3.1 Binary Classification

Standard PR problems typically assume the existence of data that was drawn independently and identically from the application domain, and that the data can be divided upon class lines into the (in our setting, two) representative sets. The availability of such data facilitates the training of binary classifiers, which have been demonstrated to be proficient at learning class distributions, and ultimately labelling novel instances.

Bayesian Classification

Bayesian decision theory is the fundamental statistical approach to PR [28]. It assumes both that the problem is stated in probabilistic terms, and that the relevant parameters are known. More specifically, it is assumed that the *a priori* class probabilities, $P(\omega_i)$, for the i states-of-nature are known, and these are in practice, estimated. Likewise, the *class-conditional* probability densities, $p(\mathbf{x}|\omega_i)$, are also assumed to be known, in practice, they are derived from the training data.

With this essential information, Bayesian classifiers are able to assign novel instances to the most probable state-of-nature [68]. In particular, a novel instance, \mathbf{x} , where \mathbf{x} is a d -dimensional attribute vector, $\{x_1, x_2, \dots, x_d\}$, is assigned to the class with the *Maximum A Posteriori* (MAP) probability, formally denoted Ω_{MAP} .

The Ω_{MAP} decision is arrived at via Bayes' rule, and is formulated as follows:

$$\begin{aligned}\Omega_{MAP} &= \arg \max_{\omega_j \in V} P(\omega_j | \mathbf{x}) \\ &= \arg \max_{\omega_j \in V} \frac{p(\mathbf{x} | \omega_j) P(\omega_j)}{p(\mathbf{x})}.\end{aligned}\tag{5.1}$$

Since the $p(\mathbf{x})$ term is assumed for all the classes, it can subsequently be omitted, to yield:

$$\Omega_{MAP} = \arg \max_{\omega_j \in V} p(\mathbf{x} | \omega_j) P(\omega_j),\tag{5.2}$$

While the required *a priori* probability can be easily estimated from a sufficient training set, procuring the class-conditional probabilities turns out to be an obstacle to the application of the Bayesian classifier. The challenge results from the large number of training instances required in order to derive a strong estimate of the entire density for the class-conditional distributions. To demonstrate the magnitude of this challenge, consider a binary classification problem in a d -dimensional feature space, where each of the d attributes can take one of k nominal values. In this scenario, the practitioner is faced with the problem of estimating $\Theta(k^d)$ probabilities, each requiring a considerable number of instances if one aims to obtain a strong estimate.

Naïve Bayes

The Naïve Bayes (NB) classifier is founded on Bayes' decision theory; however, it simplifies the decision making process by involving an assumption that the attributes are conditionally independent. Therefore, given class ω_j , the probability of observing the feature vector $\{x_1, x_2, \dots, x_d\}$, is the product of the probability of witnessing the individual attributes, calculated as:

$$p(\mathbf{x} | \omega_j) = \prod_i p(x_i | \omega_j).\tag{5.3}$$

Generally, the NB classifier takes the following form:

$$\Omega_{NB} = \arg \max_{\omega_j \in V} P(\omega_j) \prod_i p(x_i | \omega_j).\tag{5.4}$$

As a result of the simplifying assumption of conditional independence, the NB classifier, in accordance with our previous example, only requires the estimation of $\Theta(dk)$ class-conditional probability densities.

Estimating Class-Conditional Probabilities

Generally, one of two approaches are utilized in the estimation of the class-conditional densities during the training of the NB classifier. In the case of numeric attributes, a statistical distribution, such as the Gaussian, is assumed to characterize the instances of the attribute under consideration [126]. Thus, learning/training involves estimating the required parameters for the underlying distribution, which must be achieved using the set of instances.

If a Gaussian distribution is assumed (similarly for other distributions), the mean, μ , and the standard deviation, σ , are estimated from the training instances. Subsequently, the probability density of the variables x_i take the form

$$p_{\mathcal{N}}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} - e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}, \quad (5.5)$$

where x_i is the attribute value of interest.

When nominal attributes are to be considered, the class-conditional distributions, are calculated by observing the occurrence frequencies, within the class, as a proportion over the total number of training instances. Although this approach generally produces strong estimates, a good estimate is unlikely if the nominal value appears only rarely in the training data. In particular, such scenarios produce biased underestimates, and the classifier becomes denominated by the infrequent nominal value, due to the product found in Equation (5.4).

Mitchell, in [68], describes an approach to avoid this issue based on the so-called *m-estimate*:

$$\frac{n_c + mp}{n + m}, \quad (5.6)$$

where n_c is the number of times the value occurs in the training set, and n is the number of training instances. The quantity p denotes the estimated *a priori* probability of the nominal value, and finally, the quantity m is the equivalent sample size, which dictates how heavily p is weighted relative to the observed data.

Instance-Based Learning

Instance-based (IB) learning is, perhaps, the simplest form of machine learning. In its generic form, all training instances are stored, or memorized, for use during testing, and when the classifier is deployed. More specifically, when a trained IB classifier is presented with a novel instance, the instance is labelled according to the class of the most “similar” instance(s) in memory. As a result, it can be said that the IB learners use the training data to represent knowledge rather than inferring a decision boundary or a set of classification rules. Thus, the outstanding question is that of deciding how one should approach the question of similarity.

The k -Nearest Neighbour (k NN) approach to IB learning assesses similarity according to the distance between instances in their d -dimensional feature space. Thus, for the base-case, where $k = 1$, a novel instance is assigned to the class of its nearest neighbour in the training set. More generally, k NN classification takes the form of a majority vote rule, where the novel instance is assigned to the class occupied by the majority of its k -nearest neighbours.

Distance Measures

Within the feature space, distance is typically assessed as a Euclidean measurement in k NN implementations [68]. The Euclidean distance between an arbitrary training instance, \mathbf{x} , and a novel instance, \mathbf{q} , is measured as:

$$d(\mathbf{x}, \mathbf{q}) = \sqrt{\sum_{i=1}^d (x_i - q_i)^2}. \quad (5.7)$$

Since we are only interested in performing comparisons, within the k NN framework, the square root can be omitted.

Alternative approaches to assessing distance include increasing the power factor in the p -norm, where the exponent is replaced by a constant ‘ p ’. Such norms have the effect of elevating the influence of larger spreads within the feature space on the final result. Alternatively, the Manhattan distance has also been applied to measure distance in k -NN implementations.

Normalization

Prior to quantifying the measured and recorded distance, the attributes generally require normalization, as they typically exist on different scales. Without normalization, attributes whose instances span larger ranges, acquire a disproportionate influence over the resulting distance measure. The standard normalization procedure, as described in [126], for an instance v of the i^{th} attribute is:

$$a_i = \frac{v_i - \min(v_i)}{\max(v_i) - \min(v_i)}, \quad (5.8)$$

where the minimum and maximum are taken over the set of training instances for the attribute in question.

Decision Trees

Decision trees are one of the most widely-researched and applied forms of machine learning utilized in the field of data mining [126]. The iterative process of deriving a decision tree from a training set, \mathbf{D} , can be concisely described as follows:

1. Select an attribute, A_i , to form the root;
 - (a) Make one branch for each nominal value, or numeric range of the attribute;
2. Repeat the process for each child node, N_i , while the instances of N_i are not drawn from a single class.

The choice of which attribute to split on at each iteration of the tree's construction is fundamental to the ultimate form of the decision tree. Thus, the goal when selecting A_i , is to find the attribute that will most effectively split the training instances along class lines, thereby producing the most compact tree. Unfortunately, the determination of the smallest possible tree for a given training set is an NP-hard problem [45]. Therefore, one seeks a method that can be applied so as to estimate the best attribute upon which the tree can be split. This process is described in next section.

A more accurate description of the stopping criteria for the iterative process that was described above, contains two possibilities. The process either halts when each

leaf is pure, or when splitting offers no improvements. Due to the inherent challenge of selecting the “best” A_i to split on, the iterative process often results in a complex tree that “over-fits the data” by inferring more structure than is warranted [85]. As a means of reducing the effect of over-fitting, and in particular, to limit the degradation of performance during application, the tree is pruned to a more general form in the final stage. The process of pruning a learned tree is described in a subsequent section.

Calculating Information Gain

As a means of assessing the benefit of splitting on a particular attribute, the authors of [126] describe a process that calculates the *Information* content of the attribute, which measures the number of bits required to specify the class of a novel instance. In particular, the information (or more formally the *entropy*) of each branch resulting from a split on the attribute under consideration is calculated as follows:

$$Entropy(p_1, p_2, \dots, p_n) = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_n \log p_n, \quad (5.9)$$

where, in a binary problem, the p_i terms specify the ratio of instances from each class at a particular branch. Therefore, if five instances (2 from class ω_1 , and 3 from class ω_2 , for example) fall under the branch being considered, then the information for that branch is estimated as:

$$\begin{aligned} Information([2, 3]) &= -\frac{2}{5} \log \frac{2}{5} - \frac{3}{5} \log \frac{3}{5} \\ &= 0.971. \end{aligned}$$

In continuing with the example presented in [126], we assume three daughter nodes are produced by branching on the attribute A_i , each with the following information:

$$Information([2, 3]) = 0.971,$$

$$Information([4, 0]) = 0.0,$$

$$Information([3, 2]) = 0.971.$$

In order to derive the final assessment of information gain produced by A_i , the three

values are first combined to form the *average information*:

$$\begin{aligned} \text{information}([2, 3], [4, 0], [3, 2]) &= \frac{5}{14} \times 0.971 + \frac{4}{14} \times 0.0 + \frac{5}{14} \times 0.971 & (5.10) \\ &= 0.693, \end{aligned}$$

where the previously calculated information scores are scaled according to the ratio of instances in the branch over the total number of instances in the daughter nodes. This value indicates the number of bits required to classify a novel instance.

Finally, to estimate the amount of information gained by splitting on A_i , the average information calculated in Equation (5.10) is subtracted from the information of the parent node. Thus, if we assume that the parent node contains nine instances of class ω_1 and five instances of class ω_2 , the *information Gain* (IG) takes the following form:

$$\begin{aligned} IG(A_i) &= \text{Information}([9, 5]) - \text{Information}([2, 3], [4, 0], [3, 2]) \\ &= 0.940 - 0.693 = 0.247. \end{aligned}$$

The IG quantity, thus, provides the necessary tools to estimate the best attribute on which the tree can be split, based on the information available at the node under consideration. In particular, the best attribute to split on is found by choosing the attribute A_i that produces the largest IG.

This calculation, however, is biased towards attributes with a large number of possible values, or ranges. In the worst case, for example, each instance has a unique value, and therefore, the resulting average information equals zero. Thus, the IG is equal to the information of the parent node, which is, by definition, the highest possible information gain for the node. The structure that would result from this choice, however, would not be informative.

This bias is neutralized by taking into consideration the number and size of daughter nodes produced by the split. More specifically, the information gain is divided by the information for each branch, disregarding all class information, and is referred to as the *IG Ratio*. This penalizes attributes that produce a large number of splits and/or few instances per split.

Finally, to prevent an overcompensation, the attribute identified by the IG Ratio is selected if, and only if, its IG is at least as large as the average IG as calculated over the set of alternate attributes.

Pruning

The pruning of the decision tree can either be commenced during the building of the tree or after the complete tree has been learned from the training data. These two approaches to generalizing decision trees are referred to as *pre-pruning* and *post-pruning*, respectively. The widely applied C4.5 decision tree routine [85] utilizes a specific form of pruning, referred to as *subtree raising*. The post-pruning process begins with the leaves, and traverses up the tree, at each node considering whether or not a pruning phase should be involved. This paradigm aims to find nodes, and their subtrees, that can be replaced by their descendent nodes and their subtrees. While this approach is implemented in C4.5, Witten and Frank note, in [126], that the benefits of this complex replacement procedure are not obvious.

Alternatively, post-pruning can take the form of *subtree replacement*. This is a considerably more routine procedure. The objective of this post-pruning procedure is to iteratively progress from each of the leaves to the root, finding ascendants that can be replaced by leaves.

Regardless of the approach, the pruning decision is made according to an estimate of the error at the node. The standard verification technique, which holds back some of the training data for use as an independent verification set, is referred to as *reduced-error* pruning. Error estimation is subsequently performed according to standard statistics, and is based on the error incurred during validation. The inherent disadvantage of this technique is that the complete set of training data is not utilized in the learning process.

The alternative approach, which is utilized in the C4.5 algorithm, performs error estimation based on the training data itself. While this approach is based on weak statistical reasoning, it generally works well [126]. In order to estimate the error at a node without having to set aside a portion of the training data, the class of each instance at the node is considered, and the node is assigned to the majority class.

This assignment produces an error, E , for that node as a rate out of the total number of instances, N . In estimating the error rate, we assume q is the true probability of the error, and that the N instances are generated as Bernoulli random events, with the parameter q , of which E are errors.

Finally, given a particular confidence limit, c ($c = 0.25$ in C4.5), the upper confidence limit for z is found in the standard statistical way:

$$Pr \left[\frac{f - q}{\sqrt{q(1 - q)/N}} > z \right] = c,$$

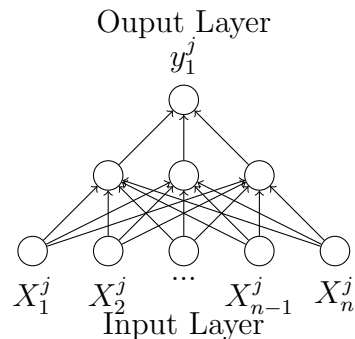
where f is the observed error rate. The upper confidence limit is then applied as a pessimistic estimate of the error rate, e , for the node under consideration:

$$e = \frac{f + \frac{z^2}{2N} + z \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}.$$

Multilayer Perceptron

The feed-forward Multilayer Perceptron (MLP) is a nonlinear classification strategy, which is composed of perceptron-like models arranged into a hierarchical structure with directed edges. A generic form of the MLP classifier with one hidden layer is depicted in Figure 5.1.

Individually, perceptrons, which for the building blocks of MLP classifiers, are capable of learning a linear discriminant function (linear discriminant functions are commonly referred to as hyperplanes) to bisect the feature space, and separate the class ω_1 from class ω_2 . In terms of propositional logic, simple perceptron classifiers can represent the AND, OR, and NOT connections. The XOR connection, however, is not linearly separable, and thus, cannot be represented by a perceptron. This limitation is overcome by the MLP classifier through the introduction of at least one hidden layer composed of an arbitrary number of hidden perceptron units. An example with a single hidden layer and three hidden units is illustrated in Figure 5.1. The first and last layers of the figure are standard to all MLP classifiers. In particular, the first layer is referred to as the input layer, it has one unit per component of the input vector. The last layer contains a single output unit, which specifies the predicted class.



Multi-Layer Perceptron

Figure 5.1: This figure demonstrates the general form of the Multilayer Perceptron (MPL).

Backpropagation

MLP classifiers learn through a process formally known as backpropagation. In essence, this process enables the determination of the appropriate connection weights for each edge in the network. The learning process deviates from that utilized by the simple perceptron because the appropriate output for the inner units is unknown, and thus, cannot be applied to verify the output and initiate an adjustment of the connection weight when necessary. As a result, the necessity of weight adjustments must be inferred from the success or failure of the output units prediction.

Subsequent to a miss-classification during training, the recalculation of the connection weights is treated as a problem of mathematical optimization, known as *gradient descent*. This process, however, utilizes differential calculus to find the most direct path (i.e., steepest slope) to the minimum error. Thus, the error function must be differentiable. The square error loss function of the network's output is commonly applied to quantify error in the network. For a single instance, the loss function takes the following form:

$$E = \frac{1}{2} \left(y - f(x) \right)^2, \quad (5.11)$$

where $f(x)$ is the predicted class value, and y is the class label (0 or 1). The direction that most significantly reduces the error on the training set can be found by

differentiating Equation (5.11).

In addition to the error function, the prediction function, $f(x)$, must be differentiable. The standard step function:

$$f(x) = \begin{cases} 1 & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

which predicts class ω_1 if the output is greater than zero, and class ω_2 otherwise, is utilized by the perceptron classifier. However, since this function is not differentiable, one resorts to using an alternate function. The sigmoid function:

$$f(x) = \frac{1}{1 + e^{-y}}, \quad (5.12)$$

is differentiable and behaves similar to the step function, and thus, it is a good replacement.

To reiterate, Witten and Frank describes the gradient descent process, in [126], as one that iteratively adjusts the functions parameters in an organized manner, thereby producing successive movements towards the optimal values. More specifically, the result of the derivative is multiplied by the *learning rate*, which is a small constant value (e.g., 0.1 [68]), and finally, the combined result is subtracted from the current weight.

The purpose of the learning rate is to specify the size of the step taken in the direction of the steepest descent at each successive iteration. Therefore, assigning the learning rate to be a small value induces a slow progression towards convergence. Alternatively, larger learning rates can reduce the time taken to converge to the minimum error. However, with larger steps there is a risk of over-stepping the minimum error, and possibly oscillating about it.

The gradient descent strategy for minimizing error can be shown to converge to the local minimum. When a series of local minima exist, it is often necessary to pass through one or more hills in the error curve in order to reach the global minimum. In such scenarios an additional *momentum* parameter is required to propel the gradient descent process through the valleys formed by the local minima to settle at the global minimum [68].

Gradient Descent

As previously described, the gradient descent algorithm iteratively differentiates Equation (5.11) in order to update each weighted edge, w_i , in the network. Thus, the derivative acquires the following form:

$$\frac{dE}{dw_{ij}} = (y - f(x))f'(x)f(x_i), \quad (5.13)$$

where $f(x_i)$ is the output of the i^{th} unit, x is the weighted sum of the input, and w_{ij} is the weight of the edge from the input unit j to the i^{th} hidden unit. The w_i term denotes the weight of the i^{th} hidden unit to the output layer, and finally, $f(x)$ and $f'(x)$, represent the classification decision, determined as a result of the sigmoid function, and the derivative to the sigmoid function.

Updating the weights between the inner units involves a slightly more complex derivative:

$$\frac{dE}{dw_{ij}} = (y - f(x))f'(x)\frac{dx}{dw_{ij}}, \quad (5.14)$$

where, since

$$x = \sum_i w_i f(x_i)$$

$$\frac{dx}{dw_{ij}} = w_i \frac{df(x_i)}{dw_{ij}}.$$

Furthermore:

$$\frac{df(x_i)}{dw_{ij}} = f'(x_i)\frac{dx_i}{dw_{ij}} = f'(x_i)a_i. \quad (5.15)$$

Thus, the equation for the derivative of the error function with respect to the w_{ij} weights is:

$$\frac{dE}{dw_{ij}} = (y - f(x))f'(x)w_i f'(x_i)a_i. \quad (5.16)$$

Equation (5.16) is evaluated for every training instance, and the modifications associated with each w_{ij} are combined, multiplied by the learning rate, and subtracted from the current value at the end of each iterative cycle.

For networks with multiple hidden layers, the process is easily extended. In the case of two hidden layers, for example, the previous strategy is applied a second

time. The weights of the second hidden layer are adjusted in the manner previously described, and those of the first hidden layer are updated by propagating the error from the output layer through the second hidden layer to the first hidden layer.

Support Vector Machines

Support Vector Machines (SVM) utilize linear models in the derivation of nonlinear classifiers. This is achieved by mapping the input onto a high-dimensional space via a nonlinear mapping. Through the mapping, a function that is linear in the original feature space become nonlinear. Conversely, a decision boundary, which is linear in the high-dimensional space, appears nonlinear in the original feature space.

A variety of possible mappings between lower and higher dimensional spaces exist, one approach is to expand the feature space by replacing the original set of features with all the products of n factors that can be constructed:

$$x = w_1 a_1^n + w_2 a_1^{n-1} a_2^1 + \dots + w_n a_1^1 a_2^{n-1} + w_{n+1} a_2^n, \quad (5.17)$$

where x is the outcome, \mathbf{a}_1 and \mathbf{a}_2 are the original features in a 2-dimensional space, and the w_i terms are the weights that must be learned by, for example, a multi-response linear regression classifier.

Polynomial mappings that are of this sort, and are of sufficiently high order can approximate arbitrary decision boundaries of any required accuracy [126].

As it stands, two significant challenges effect the application of a classifier of this form:

- a) Computational Complexity: Mapping the feature space onto a higher dimension causes the number of weight coefficients that must be learned to rapidly expand. In a 10-dimensional feature space, for example, if the original feature set is mapped according to all of the products with 5 factors, one must learn more than 2,000 coefficients.
- b) Over-fitting: In any scenario where the number of features is large relative to the number of training instances, the derivation of a discriminant function that overfits the training data is a risk.

The fundamental strength of the SVM is that it reduces the computational complexity and minimizes the risk of over-fitting through the application of a special form of linear model, namely, the Maximum Margin Hyperplane (MMH). The MMH is a linear model that separates the two classes of data in the training set, coming no closer to either class than is absolutely necessary. More specifically, of the set of separating hyperplanes that are appropriate for the training data, the MMH is the hyperplane possessing the greatest distance from the convex hulls that encompass each class. Recall that the two classes are linearly separable as a result of a mapping to a high-dimensional space, thus, the convex hulls are linearly separable by the MMH. The MMH is further distinguishable by the fact that it is perpendicular to the shortest path between the two convex hulls. Subsequent to the determination of the

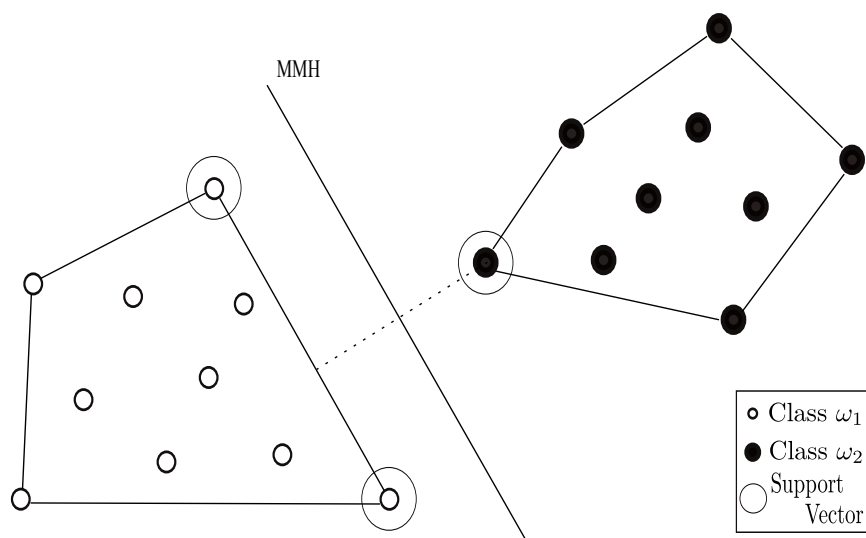


Figure 5.2: This figure demonstrates the calculation of the Maximum Margin Hyperplane (MMH) for the SVM.

MMH, the support vectors are recognized as those vectors that reside closest to the MMH. By definition, each class has at least one support vector, and often more than one.

In a two dimensional space, the separating hyperplane can be defined as:

$$x = w_0 + w_1 \mathbf{a}_1 + w_2 \mathbf{a}_2. \quad (5.18)$$

More generally, the MMH can be written in terms of its support vectors as follows:

$$x = b + \sum_{i \in SV} \alpha_i y_i \mathbf{a}(i) \cdot \mathbf{a}, \quad (5.19)$$

where SV is the set of support vectors, y_i is the class value, and $\mathbf{a}(i)$ is the i^{th} support vector. The b and α_i are numeric terms that must be learned, \mathbf{a} denotes a test instance, and $\mathbf{a}(i) \cdot \mathbf{a}$ signifies the dot product of the two terms.

The determination of the support vectors and the corresponding b and α_i values can be viewed as an optimization problem, and in particular, can be solved via *constraint quadratic optimization*. However, the complexity of the problem can be reduced with SVM-specific training procedures [96].

Overfitting

In general, overfitting is associated with variability in the finally learned decision boundary. In particular, over-fitting, as we previously discussed in Chapter 3, occurs when the discriminant function incorporates the intricacies of the training data. Thus, with high probability, when the training data changes, so does the discriminate function. This is largely avoided, however, in SVM classifiers, because the support vectors are inherently stable across training sets. More specifically, the support vectors are unlikely to change, and as a result, the MMH is unlikely to change.

Computational Complexity

The computation complexity that results from the application of Equation (5.19) in high-dimensional spaces is significant due to the number of components in vectors of the dot product, $\mathbf{a}(i) \cdot \mathbf{a}$, which is computed both during training and testing. However, much of this complexity can sufficiently be avoided by computing the dot product prior to the nonlinear mapping. In doing so, Equation (5.19) is modified to:

$$x = b + \sum_{i \in SV} \alpha_i y_i (\mathbf{a}(i) \cdot \mathbf{a})^n, \quad (5.20)$$

where the terms in the dot product have not been mapped to the high-dimensional space, and the n term denotes the number of factors in the transformation.

When expanded, $(\mathbf{a}(i) \cdot \mathbf{a})^n$ contains all of the terms that were present in the nonlinear transformation described by Equation (5.17). In addition, the expansion contains some binomial coefficients, however, these are inconsequential to the final result. Therefore, the constraint quadratic optimization problem can be evaluated for $(\mathbf{a}(i) \cdot \mathbf{a})^n$, where n is initialized to the value of 1, and is incremented until the estimated error stabilizes.

This approach utilizes what is commonly referred to as the *polynomial kernel*, $(\mathbf{a}(i) \cdot \mathbf{a})^n$. Alternate techniques apply the *radial basis function*, or the *sigmoid function* as the kernel [96].

5.3.2 One-Class Classification

As previously indicated, so-called one-class classifiers rely on instances drawn independently from a single class in the derivation of a discriminant function. A broad set of one-class classifiers exists in the literature, each of which applies a slightly different strategy for the construction of a binary classifier using the data from a single class. However, in simple terms, the process can be articulated as one in which the selected classifier learns to recognize, in some general terms, novel instances that are similar to those viewed during the training process. Should a novel instance be recognized, it is subsequently labelled as a member of the positive class. In particular, this refers to the class on which the one-class classifier was trained. Alternatively, if the novel instance does not appear to fit into the learned distribution, it is identified as a member of the outlier class.

As previously indicated, the standard motivation behind the use of a one-class classifier is the presence of an underlying difficulty that hinders the acquisition of a representative set of instances of the second class. Thus, the primary class, often referred to as the target class [42], or the positive class [23], which is importantly considered to be well-defined, is utilized alone in the derivation of the binary classifier. This is, indeed, often the case in the first scenario (namely S1) that we described in Section 5.2. However, the challenge inherent in the second scenario (S2) dictates that the primary class will not be well-defined, as it is likely to contain erroneously labelled

instances of the outlier class. This is a direct result of the impracticality of manually identifying and labelling them.

In such a scenario, we propose the application of one-class classifiers as unsupervised classifiers, that are trained on datasets in which the vast majority of instances have correctly been extracted from the background class. However, the impracticality of identifying outliers (rare SE events that are interwoven in the time-series of background data), implies the probable presence of some naïvely labelled training instance.

We propose that by utilizing estimates of the state-of-nature acquired, possibly, from domain experts, the noisy training set can be overcome by setting the error-rate (or rejection rate) parameter, which exists in many one-class classifiers, to account for the *a priori* probability of the outlier class.

In the subsections that follow, we describe, in reasonable detail, the various reported approaches for achieving one-class classification.

Nearest Neighbour

Datta, in [23], introduced a modified version of the standard nearest neighbour classifier, which was described in Section 5.3.1, for conducting one-class classification.

The proposed one-class Nearest Neighbour (ocNN) algorithm deviates from its binary counterpart by searching the training set, \mathbf{X} , for the *target rejection rate* threshold, τ where:

$$\tau = \arg \max_{\mathbf{x}_i \in X} \text{distance} \left(\mathbf{x}_i, NN(\mathbf{x}_i) \right), \quad (5.21)$$

where \mathbf{x}_i is a training instance, and $NN(\mathbf{x}_i)$ returns the nearest neighbour of the instance \mathbf{x}_i . Thus, τ is the maximum distance between any two nearest neighbours in the set of nearest neighbours constructed from the training data.

Given a novel instance, \mathbf{x} , and the learned parameter, τ , \mathbf{x} is classified by first finding its nearest neighbour in the training set, and then comparing the distance that separates them to the learned target rejection rate threshold.

Formally, the ocNN classification decision is defined as:

$$f(\mathbf{x}) = \begin{cases} 1 & \text{distance}(\mathbf{x}, NN(\mathbf{x})) \geq \tau \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if \mathbf{x} deviates too much from the training data, it is assigned to the outlier class (ω_2).

Scaled Nearest Neighbour

In recognition of the proposed sub-category of the so-called domain of “one-class” classification that we have introduced in this chapter, a corresponding modification has been made to the ocNN classifier.

The modification was motivated by the second classification scenario, namely S2. In particular, this scenario is characterized by a series of rare SE events where:

- The data exists as a time-series.
- The state-of-nature is dominated by a single class (the ω_1 class composes more than, for example 90% of the instances).
- The minority class is nearly impossible to manually identify. Thus, it naïvely takes the ω_1 label in the training set.

In this scenario, we have stated that due to the rarity of the outlier class, and the extreme challenge of manually labelling those instances in the training set, it can be naïvely issued to a one-class classifier. Moreover, this can be done with considerable confidence, provided that an estimate of the *a priori* probability of the outlier class can be acquired. This hypothesis relies on the availability of a so-called rejection rate, which ensures a portion of the training set will be misclassified after the derivation of the discriminant function.

However, the standard ocNN algorithm is intuitively unable to learn a threshold capable of discriminating most of the erroneously labelled outliers, and is inherently ineffective in the presence of noise. The problem, which is embedded in the ocNN algorithm, is depicted in Figure 5.3. By definition, the naïvely labelled instances of

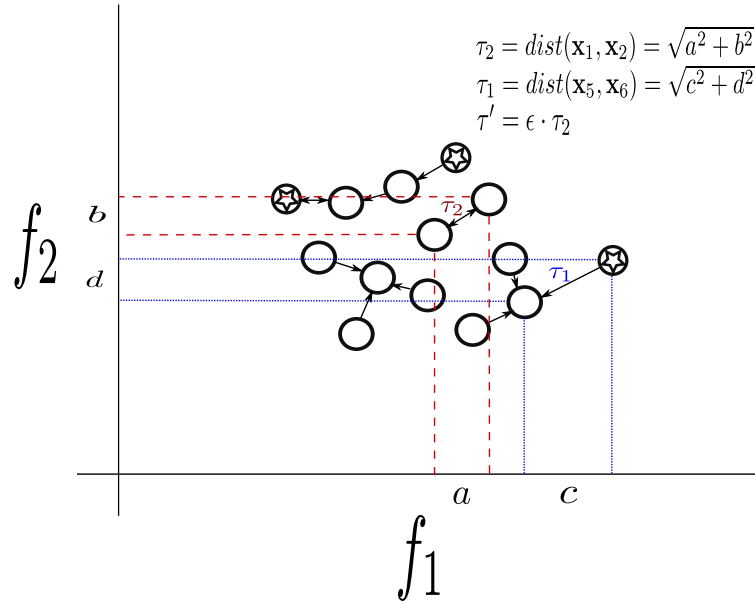


Figure 5.3: This figure demonstrates the calculation of the τ parameter in the ocNN classifiers, and the effect of erroneous instances in the training set on the learned target rejection rate threshold.

the second class are outliers. Thus, they are expected to reside on the periphery of the “real” background distribution. Therefore, with a high probability, the learned parameter, τ , which is intended to record variability in the background class [23], can be expected to represent the distance between a background instance and an erroneously labelled member of the outlier class. A hypothetical learned distance of this sort is illustrated in Figure 5.3 as τ_1 .

Ideally, however, the algorithm should learn the distance that is denoted as τ_2 , because it is the maximum target rejection rate threshold found in the set of pure background instances (represented as empty circles).

Because this scenario creates an unsupervised learning environment, in which we cannot explicitly identify the members of the outlier class during training, we rely on a rejection rate parameter to be engrained in the one-class classification algorithm in order to facilitate the exclusion of these instances. However, while Datta coined τ as the target rejection rate threshold, by definition it does not exclude any instances in

the training set. Indeed, this was not the intention. Thus, in this exceptional domain, it incorporates the erroneous information provided by the mislabelled members of the outlier class into the learned threshold, as is depicted by τ_1 in Figure 5.3.

As a means of accounting for the overestimate, we have added a scaling parameter, ϵ , where $0 < \epsilon \leq 1$, such that

$$\tau' = \epsilon \cdot \tau, \quad (5.22)$$

with the understanding that the optimal value of ϵ will enable the rejection of the majority of the outlier instances, by reducing the magnitude of the learned threshold.

Combined Density and Class Probability Estimation

The Combined Density and Class Probability Estimation (CDCPE) technique for one-class classification is based on the generation of an artificial second class of data. The artificial class is generated according to a user-specified *reference* distribution, such as a multivariate Gaussian, which ideally reflects that of the available class, ω_1 .

In their introductory work, Hempstalk *et al.* [42] utilized Bayes' rule:

$$P(\omega_1|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{p(\mathbf{x})}, \quad (5.23)$$

where ω_1 is the background, or training class, \mathbf{x} is instance drawn from the domain, and $P(\omega_1|\mathbf{x})$ is the true probability mass function, to illustrate how the density function of the reference distribution can be combined with the class probability estimate of a classification model to yield a description of the background class.

They noted that in a binary scenario, the probability of an instance, \mathbf{x} , includes both the chance of it occurring as a member of class ω_1 or class ω_2 , and thus, Bayes' rule can be expanded to:

$$P(\omega_1|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{p(\mathbf{x}|\omega_1)P(\omega_1) + p(\mathbf{x}|\omega_2)P(\omega_2)}. \quad (5.24)$$

Of primary interest is the probability density function $p(\mathbf{x}|\omega_1)$, as it is utilized in the one-class classification process. Through a series of algebraic operations $p(\mathbf{x}|\omega_1)$ can be arrived at:

$$p(\mathbf{x}|\omega_1) = \frac{(1 - P(\omega_1))P(\omega_1|\mathbf{x})}{P(\omega_1)(1 - P(\omega_1|\mathbf{x}))}p(\mathbf{x}|\omega_2). \quad (5.25)$$

Equation (5.25) relates the density of the artificial class $p(\mathbf{x}|\omega_2)$ to the density of the background class $p(\mathbf{x}|\omega_1)$ via the class probability mass function and the *a priori* class probability $P(\omega_1)$.

Autoassociator

The Autoassociator (AA), introduced by Hanson and Kegl, in [40], functions in much the same manner as its binary counterpart, the MLP. However, in the absence of instances drawn from the second class, the AA is trained to reproduce the input vector at the output layer [49]. Thus, it is assumed that the instances of training sets were drawn purely from the background class.

As a result, the general structure of the AA network deviates from that of the MLP. In particular, both the input and output layer of the AA network have l units, where l is the number of features, or dimensions, in the dataset. In addition, the AA has a single hidden layer with an arbitrary number of hidden units. The structure of the AA, in this case with three hidden units, is depicted in Figure 5.4. As described

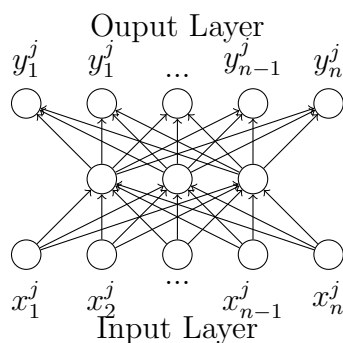


Figure 5.4: This figure demonstrates the neural structure of the autoassociator classifier.

in our presentation of the MLP classifier, the AA is trained using backpropagation to minimize the squared error of the network (see Section 5.3.1 for further details on the training process).

The primary concept that facilitates the application of the AA as a classifier is the assumption that the training process enables it to infer a function, $f(\mathbf{x})$, that

captures the essence of the background class. Thus, when the trained classifier is presented with a novel instance, its corresponding class can be determined according to the AA's ability to reconstruct the components of the instance at the output layer, via the inferred function.

During testing, the quality of the reconstruction is quantified as the sum of the squared error at each of the corresponding input and output units. This value is ultimately required to formulate the class prediction for the novel instance. In particular, if the reconstruction error exceeds the user-defined threshold, the instance is assigned to the outlier class; otherwise, it is classified as a member of the background class.

Support Vector Machines

Schölkopf *et al.*, in [97], modified the standard SVM classification algorithm to produce a similitude that is applicable to one-class classification problems. The one-class Support Vector Machine (ocSVM) utilizes a function that returns +1 in a small region that encompasses the majority of the training instances, and -1 elsewhere.

The proposed algorithm's operations are largely consistent with the inner mechanisms of the binary SVM classifier. In particular, the data is initially mapped onto an extended feature space via a kernel function. Subsequent to the mapping, the process momentarily deviates from its binary counterpart due to the absence of a second class. As a result, the Maximum Margin Hyperplane (MMH) that separates the two classes of the binary problem, is found to divide the majority of the instances in the training set from the origin.

Finally, novel instances are classified in the familiar manner according to their position relative to the MMH, with those residing on the side of the majority of the training data receiving a +1, and the others a -1.

In mathematical terms, the authors presented the algorithm as an initial mapping, ϕ , of the training data $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\} \in \chi$, where $l \in \mathbb{N}$ is the number of training instances, from χ to F .

More specifically, the process, which corresponds to that of the standard binary SVM, applies a mapping on to a dot product feature space F such that the dot product in the image can be computed via a kernel function.

To separate the training instances from the origin, the quadratic program:

$$\min_{\omega \in F, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \left(\frac{1}{2} \|\omega\|^2 + \frac{1}{vl} \sum_{i=1}^l \xi_i - \rho \right),$$

subject to,

$$(\omega \cdot \phi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad i = 1, 2, \dots, l \quad \xi_i \geq 0,$$

is solved. If w and ρ solve the problem, then the discriminant function takes the form:

$$f(\mathbf{x}) = \text{sgn} \left(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho \right), \quad (5.26)$$

where $f(\mathbf{x})$ returns +1 for most instances in χ , while the support vector regulation term $\|\omega\|$ will be small. This trade-off is controlled by the value of v , where $v \in (0, 1]$.

5.4 Experimental Setup

In this section, we present a series of experiments designed to both illustrate the demonstration domain, and to exhibit a first attempt at classifying this sub-category of PR problems.

5.4.1 Application Domain

The CTBT aims to prevent nuclear proliferation through the banning of all nuclear detonations in the environment. A number of verification strategies are currently under study; aimed at ensuring the integrity of the treaty. The primary verification technique being explored relies on the quantity of radioxenon measured at individual receptor sites, distributed throughout the globe, and an ability to distinguish background levels from those affected by clandestine detonations. As a result, verification based on four radioxenon isotopes, i.e., ^{131}Xe , ^{133}Xe , ^{133m}Xe and ^{135}Xe , has previously been formulated as a PR task [106].

For the purpose of this thesis, we note that the presence of these xenon isotopes in the atmosphere can be traced back to one of two scenarios. In general, they are expected to have resulted from the nuclear industry. However, they are also the

byproducts of the detonation of nuclear weapons, thus, presenting a standard binary classification problem.

It must be emphasized, that if a member state were to conduct a test, it would be in breach of international law. Thus, it is logical to assume that should a detonation occur, it would likely be clandestine in nature, and inherently difficult to manually detect within the background noise.

5.4.2 Simulation Aspects: Procuring Data Collection

As a means of acquiring experimental datasets for this research, we utilized the simulation framework presented in Chapter 4. The simulation framework models SE events, such as earthquakes, nuclear explosions, etc., as they propagate through the background noise.

In particular, when generalizing a case-study for a particular environment with a set of means and higher-order moments, the simulation system requires two execution phases. The simulation begins with the user's specification of a set of input variables, which define the nature of the generalized case-study. In particular, the industrial and receptor locations are defined, along with the mean and variance of the pollutant emission rates for each industrial emitter.

The simulation facilitates the definition of one or more seasons. Each user-defined season is accompanied by a specified conceptual, mean wind field over the region of interest, and the corresponding variances. This accounts for the seasonal variability in the background pollutant levels, resulting from board shifts in meteorology, such as those demonstrated in [32]. The variances are applied in the generation of random hourly oscillations in the wind field and emission rates. Finally, the user defines a desired simulation time in terms of hours. One cumulative hourly background result is recorded for each hour of the simulation. Alternatively, the occurrence of an explosion is probabilistic within the specified time period.

Both the background and the explosion execution paths of the simulator can be executed in parallel, as they are independent of each other. The background generation path initiates by generating random Gaussian variations on the user-defined

means, which remain constant for the duration of the hour, in accordance with the steady-state principle for which the Gaussian functions apply.

The result returned from the Gaussian plume model corresponds to the average pollutant volume expected at the receptor site for any one second interval during the simulated hour. Therefore, the returned result is multiplied by 3,600 in order to acquire an estimate of the cumulative pollutant concentration over the course of the hour. Finally, the background prediction is scaled according to the rate of decay for the individual isotope and written to the dataset with the background class label.

The initial stage of the explosion execution path runs a new instance of the background path, with the same environmental variable, in order to derive a new set of background measurements for the region. Should an explosion occur at time i , and subsequently become measurable to the receptor, the effect is added to the background levels at the appropriate time. The Gaussian puff model is iteratively integrated over successive one hour periods, and combined with the background measurements for each time period, until the remnants of the detonation are no longer felt. In the last stage, the results are labelled and written to the dataset.

While it is generally beneficial to develop and study classifiers on “real” data, this is, indeed, impossible within the CTBT verification problem due to the absence of measured detonations, and the limited availability of background instances. Moreover, it has previously been demonstrated that artificial data can be utilized for development in the absence of “real” measurements, and to generate controlled experiments, (generalized case-studies) for the exploration of PR performance [1, 26].

5.4.3 Source Position Recalculation

As a means of accounting for the indirect paths, which may be travelled by the pollutant plume emitted from an industrial source or the pollutant puffs vented from a detonation site, the relative position of the source with respect to the receptor site is re-evaluated each time the mean wind field is updated during the simulation.

The repositioning of the source location according to the prevailing winds is illustrated in Figure 5.5. The source, I , and receptor, R , are illustrated on the map in red.

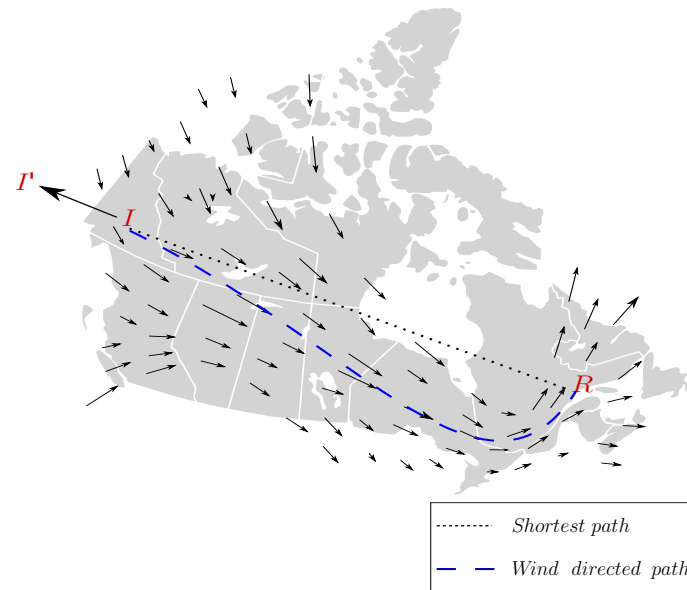


Figure 5.5: This figure illustrates the process of recalculating the relative position of an industrial emitter according to the mean wind field for a particular hour. The relative repositioning of the industrial emitter from I to I' accounts for the additional travel time and distance resulting from the indirect path from the source to the receptor.

A dotted black line connects the source to the receptor highlighting the shortest path. This is the path that the pollutant plume would follow if the atmospheric flow were to travel directly from the source to the receptor. However, this is highly improbable over great distances.

In this figure, the short black arrows form a theoretical wind field, existing for a period of time during the simulation. This wind field transports the plume on an arched path from northern Canada, through eastern Ontario and up the St. Lawrence River, eventually arriving at the receptor site in Quebec city.

Fundamentally, the Gaussian models assume the dispersing pollutant is being advected in a straight trajectory. In compensating for this, the relative position of the source is recalculated each time the atmospheric conditions are recalculated, in order to account for the additional travel time and distance. We can account for

the indirect path by changing the relative position because we are only interested in the concentration at the receptor site, and not the points in between. This process is demonstrated in the figure by shifting the source location to position I' . More specifically, the additional distance induced by the arched wind field is added to the shortest path distance, and the standard Gaussian models are applied with the adjusted wind path blowing directly from the source to the receptor.

5.4.4 Simulation Setup

The simulation framework previously described has been applied in order to demonstrate classifier performance in a simplified, CTBT motivated domain. A similar demonstration domain, is depicted in Figure 5.6, with the receptor site positioned at the centre, and a single industrial emitter located at a radial distance of 3,000 km and at a 180 degree angle from the directed x -axis.

Intuitively, we expect that those detonations occurring nearest to the receptor will be most detectable. In order to test this hypothesis, and to approximate the distance at which classifier performance degrades to unsatisfactory levels, we have formulated our simulation to facilitate the specification of classifier performance as a function of distance. To accomplish this, we have executed a script that iteratively calls the simulator. At each iteration, the atmospheric means and higher-order moments, along with the industrial and receptor parameters are held constant, although, they are subjected to Gaussian fluctuations. The only variable over each iteration, is the radial range of the random explosions. During the first iteration, all detonations occur in a ring around the receptor, between 500 km and 1,000 km. As demonstrated in Figure 5.6, at each successive iteration the upper and lower bounds are incremented by 500 km. For the purpose of this experiment, 23 iterations were utilized, which extended the upper bound of the radial distance to 12,000 km.

As a means of strengthening our results, the iterative process was repeated ten times, creating ten datasets for each radial range, which resulted in a total of 230 independent datasets, and the acquisition of ensemble mean estimates of the performance of each classifier within each radial range.

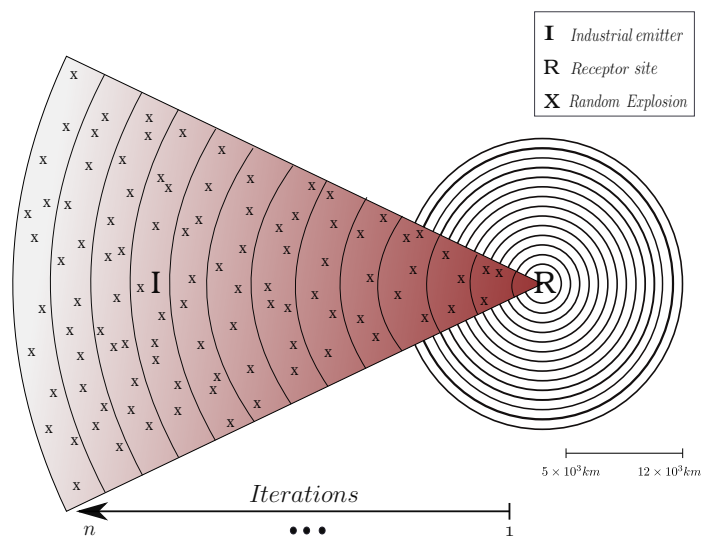


Figure 5.6: This figure demonstrates the iterative composition of the simulated domain. In each iteration of the simulation, a fixed number of explosions are probabilistically generated as uniform, random events in time, space and magnitude, and dispersed according to the prevailing meteorology, which may or may not carry the pollutant cloud past the receptor site. The mean radial distance to the explosion sites is incremented after each successive iteration.

5.4.5 Classification Scenarios

As mentioned in the introductory section, within this challenging domain of classification problems, there exist two conceivable scenarios, which we have denoted as S1 and S2. These scenarios explicitly influence the choice of the classification scheme applied to the task of recognizing the SE events.

Intuitively, the first scenario presents a slightly easier classification problem, because a set, albeit small, of SE events can be extracted from the application domain and applied to train and/or test the PR systems. More specifically, within this scenario, we assume that the outlier class is both identifiable and available in quantities

that facilitate the training of binary classifiers. However, in many ways, the classification problem still presents itself as a so-called one-class classification task, and thus warrants exploration on both fronts.

Alternatively, the second scenario presents itself as a much more difficult PR task, and in many ways more accurately reflects the PR problem suggested by the detection of SE events, in general, and the verification of the CTBT, in particular.

In accordance with the general domain characteristics, as they were originally defined, the data presents itself as a time-series of background measurements that are interwoven with a minute number of SE events. However, unlike the ideal scenario depicted in S1, here we attempt to assume a state-of-nature that is more appropriate for the CTBT task. In particular, we assume that there is a 1% *a priori* probability of a detonation, which, while still an overestimate, is a more accurate depiction, while it still provides insight into the behaviour of PR systems on the SE event class.

Raising the difficulty further, is the recognition that in practice, the clandestine nature of the SE events are such that manually identifying a distant clandestine occurrence in the acquired time-series of readings is extremely difficult, if not impossible. Thus, this prohibits the derivation of a labelled training set, which dictates that practitioners are left to utilize a training set composed largely of background instances, but with a minute number of unidentifiable members of the SE event class.

In the absence of a labelled training set, we propose the application of standard one-class learners as unsupervised classifiers. When applying one-class classifiers to an unlabelled training set, the practitioner must rely on the knowledge of a domain expert to acquire estimates of the individual *a priori* class probabilities.

In particular, estimates of the state-of-nature are required to appropriately specify the parameters of the one-class classifiers, such as the rejection rate, or error rate. This technique aims to prevent the inclusion of the SE event instances in the generalized description of the background class. Our reliance on an error, or rejection rate, presumes that the SE events will reside on the periphery of the background class, and thus, by marginally tightening the generalization of the background class, those instances of the SE event class will no longer be included.

5.4.6 Classifier Assessment Criteria

As discussed in the previous section, this research considers classifier performance within two distinct scenarios. Within each of the scenarios, namely S1 and S2, we consider classifier performance according to a set of criteria. These criteria are subsequently discussed in greater detail.

In particular, we examined the general performance of the classifiers across all of the simulated detonation ranges. Performance in this category is particularly important, as, in practice, the detonation ranges are largely unpredictable. The results of this assessment are presented in Sections 5.5.1 and 5.6.1. In addition, we explored classifier performance within two shorter detonation ranges, the result of which is presented in Sections 5.5.2 and 5.6.2.

The classifier performance as a function of distance was also examined. The results of this comparison are detailed in Sections 5.5.3 and 5.6.3.

Finally, in light of the inherent challenge of distinguishing these two very similar classes according to the four radioxenon isotopes, we were motivated to explore an expanded CTBT feature space. Based on the significant role held by meteorology in affecting the pollutant levels at the receptor site, we surmised that the inclusion of meteorological features would improve the performance of the classifiers. The results of our experiments with an expanded feature space are provided in Sections 5.5.4 and 5.6.4.

5.5 Results: Scenario 1

In this section, we present the results that were obtained according to the four assessment criteria that were motivated in the previous section, on the first classification scenario. We commence our exploration of PR performance by examining the Area Under the ROC Curve (AUC) scores produced by each classifier over the 23 detonation ranges.

5.5.1 General Performance

In this section, we present a general overview of the performance levels of each of the considered classifiers on the simulated CTBT domain. More specifically, we present an assessment of the five binary classifiers and the four one-class classifiers, in terms of their AUC scores averaged over the 230 datasets that spanned the 23 detonation ranges.

In light of the fact that the SE events, which are to be identified, will, in practice, occur at random and unpredictable distances, these results are a particularly insightful overview of the general performance levels.

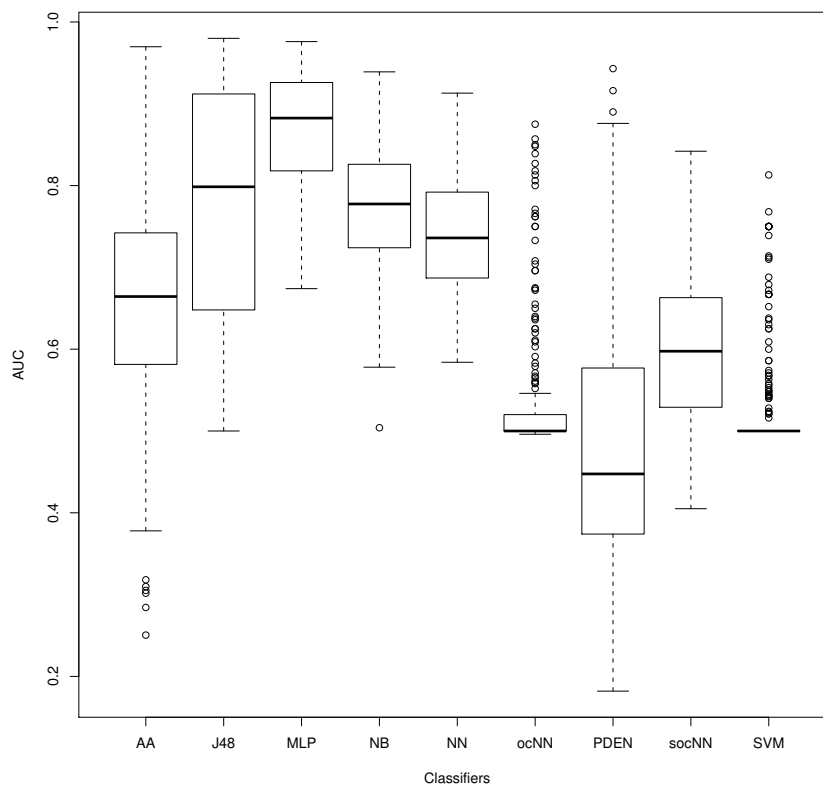


Figure 5.7: This figure displays the performance of the nine classifiers, in terms of their AUC scores on the 230 generated CTBT datasets, in the form of a series of boxplots.

The results depicted in Figure 5.7 were compiled as a series of boxplots; one for each classifier.

The solid lines that bisect the boxes represent the median AUC score produced by the particular classifier. The box itself indicates the distribution of the middle half of the AUC scores produced by the classifier. Thus, it stretches from the 25th percentile (at the lower hinge) to the 75th percentile (at the upper hinge). The boxes that are evenly divided indicate that the classifier's scores are evenly distributed throughout the central region. This is, indeed, the case for AA and NB.

The fact that there is no box around the median indicator for the SVM, suggests that nearly all of the AUC results were equivalent, and in this case, approximately 0.5. The relatively large number of circles extending up from the median, individually identify outliers. This suggests that, in general, the SVM classifier performed poorly, but that it occasionally produced anomalously strong results, which stretched slightly beyond 0.8.

Alternatively, the scenario where the median does not produce an even bisection of the box indicates that the distribution of the inter-quartile range is skewed. This is the case, for example, with PDEN, where the upper-quartile is large, indicating that the points composing the upper-quartile are spread over a larger distance.

The dashed lines, or whiskers, stretch to either the maximum and minimum values, where outliers do not exist, or to 1.5 times the range of the inter-quartile region in scenarios with outliers, such as in the case with the SVM classification results.

The SVM classifier is, surprisingly, by far the worst performing classifier on this data, and in spite of its bias, it is, on average, worse than the one-class classifiers, AA and socNN. This is reiterated in Table 5.2, which contrasts the mean AUC scores of AA and socNN as 0.656 and 0.603, respectively, with the mean value for the SVM classifier of 0.528. Moreover, all four one-class classifiers appear superior to the SVM when considered in terms of their maximum AUC scores.

When assessing the classifiers according to the boxplot, the median value provides a good indication of their performances, in general. However, most interesting are the ranges of the inter- and outer-quartiles along with the presence of the outliers, when combined with a high median value, as these components provide a strong indication

	Mean	Max	Min	STDV
NB	0.772	0.939	0.504	0.074
MLP	0.869	0.976	0.674	0.067
NN	0.741	0.913	0.584	0.071
J48	0.774	0.98	0.500	0.148
SVM	0.528	0.813	0.500	0.065
ocNN	0.540	0.875	0.496	0.087
PDEN	0.487	0.943	0.182	0.156
socNN	0.603	0.842	0.405	0.094
AA	0.656	0.970	0.251	0.140

Table 5.2: This table displays the general classification results, in terms of AUC.

of how likely it is that the classifiers will reproduce the median result.

In these terms, the binary classifier, the MLP, stands out as the superior classifier, with J48, NN, and NB contending for the intermediate positions. The results posted in Table 5.2 confirm that the MLP is the strongest of the classifiers considered here. Furthermore, it indicates that the J48 and NB are very similar, and that the NN is the fourth ranking binary classifier according to the mean and maximum scores. However, the NN is second when ranked according to the minimum AUC scores.

Notably, of the set of one-class classifiers, PDEN produced the most variable range of the AUC scores. It is our suspicion that this variability resulted from the PDEN’s generation of an artificial second class in its training process. However, further exploration of this matter is required.

In general, the AA classifier is identified as the strongest one-class classifier, both with respect to its mean and median values. While the socNN classifier achieved the second highest mean, it is more stable than the AA, and does not produce any anomalous results. Indeed, the socNN has a lower standard deviation and its boxplot spans a smaller range.

5.5.2 Performance on Short- and Long-Range Detonations

In Figure 5.8, we present the AUC results produced over two detonation ranges of particular interest. Boxplot (i) in this figure contains the results for the datasets

that included detonations ranging from 1,000 km and 5,500 km, while Boxplot (ii) has those with detonations between 5,500 km and 10,000 km. Together, these plots contrast the performance of the individual classifiers in the various detonation ranges. This experimental setup demonstrates one technique through which the performance of various receptor network topologies can be examined. For example, if PR within the second range is found to be a considerable challenge, the shorter range may, perhaps, be considered an upper bound on the acceptable distance between receptors.

There are two factors at play when hypothesizing about classifier performance within these ranges. Intuitively, detonations closer to the receptor side will be more visible at the receptor site, provided the meteorological conditions are such that the emissions are advected in the direction of the receptor. Conversely, detonations that occur farther afield are likely to have a smaller influence on the pollutant levels at the receptor site, leading to a more challenging classification problem. On the surface, then, it appears that nearby detonations should be easier to detect. Indeed, the very near detonations are often easily identifiable. However, the scenario is made more complex by the fact that during the simulation, the industrial source was positioned approximately in the middle of the shorter range. Thus, there was, in a sense, a great deal of competing background noise to distort the signal.

Indeed, Figure 5.8 demonstrates that within this scenario it is possible for the performance of the classifiers to improve when detonations occur at greater distances. However, the fact that this only occurred for the binary classifiers, highlights the importance of the second class in the learning process. It turns out that the majority of the binary classifiers are able to, through the training process, utilize the low concentration instances of the detonation class, which resulted from explosions at great distances, to specialize their models to the counter-intuitive point where many of the instances with low concentrations were correctly identified as explosions.

Alternatively, the figure suggests that neither the one-class classifiers, nor the SVM, were able learn a model with this characteristic. Moreover, the SVM exclusively produces AUC scores of 0.5 within the second range, and the ocNN's performance was nearly equivalent. Finally, at greater distances, the PDEN's performance fell even further, with only a minute number of instances exceeding an AUC of 0.5.

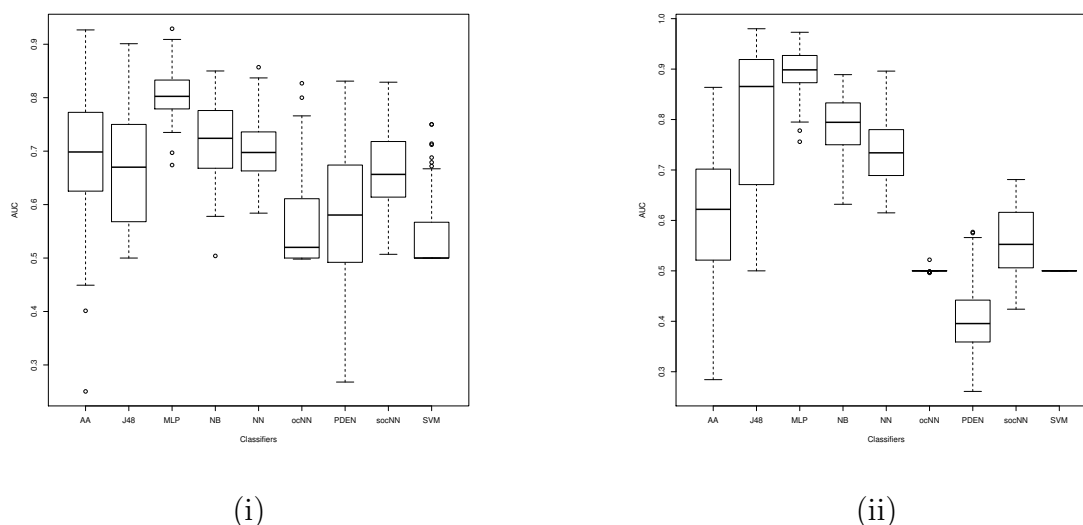


Figure 5.8: In this figure, Boxplot (i) displays the performance of the nine classifiers, in terms of their AUC scores for detonations occurring between the distances of 1,000 km and 5,500 km, and Boxplot (ii) displays their performances for detonations between the distances of 5,500 km and 10,000 km.

Within the shorter range, it is notable that the stronger one-class classifiers, namely the AA and socNN, are very comparable with most of the binary classifiers. However, the distinction in favour of the binary learners is emphasized for the larger detonation range.

5.5.3 Performance as a Function of Distance

In this sub-section, we present the classifier performance as a function of distance. Performance is assessed both according to the AUC and the False Positive Rate (FPR).

A *false positive* occurs when the classifier mislabels a novel instance as a member of the positive class (in this case, a member of the background class), when it is, in fact, a member of the negative class (specifically, a member of the SE event class). Thus, the FPR is the total number of false positives over the total number of negative instances. As a metric, FPR provides insight into whether the model is overly biased towards the positive class, which is a significant risk when the problem is extremely

imbalanced.

These results are particularly interesting, as they provide greater insight into performance trends. Moreover, these suggest a performance scale for successively sparser receptor networks, and enable the interested parties to weigh the cost of receptor stations against the probability of detection.

The performance plots depicted both in Figure 5.9 and Figure 5.10 were produced by calculating the ensemble mean of each classifier's performance at the 23 detonation ranges, and then through the extrapolation of a performance function.

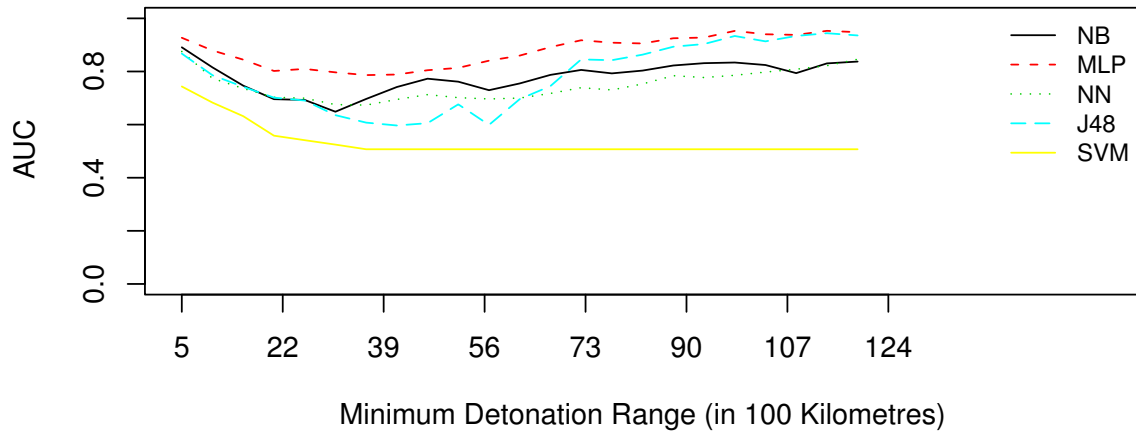
Within Figure 5.9, the MLP classifier is identifiably the superior classifier to the remaining four binary learners in terms of the AUC, across the range of detonation distances. In addition, it is not subject to the abrupt fluctuations that J48, and to a lesser extent NB, incur.

All of the classifiers, with the SVM appearing as the sole exception, have notable hulls in their performance curves that extend over varying distances and to distinct depths. In each case, a slow descent begins immediately, and is subsequently accompanied by a slow ascent. Alternatively, the SVM classifier suffers from a similar initial decline. However, it fails to recover from the degradation at greater distances.

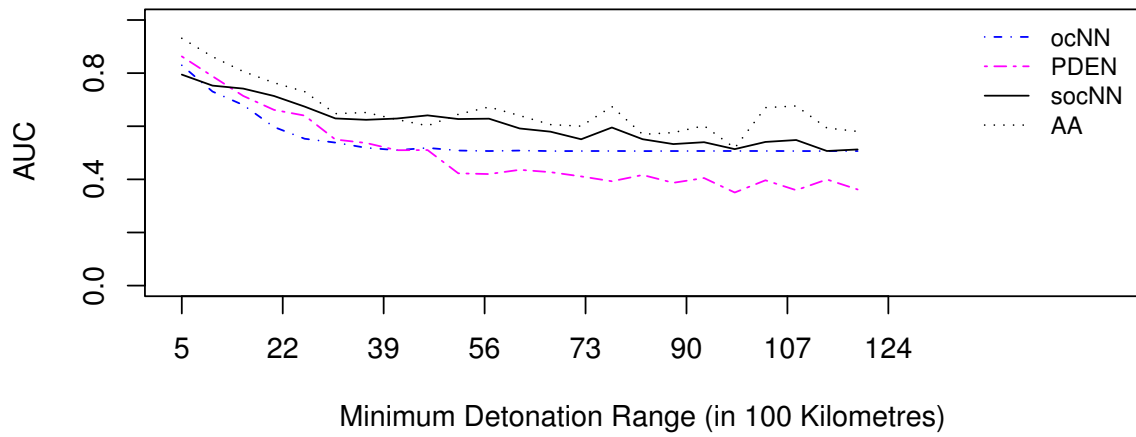
In each case, the position of the performance hull roughly corresponds to the radial distance between the industrial source of radioxenon and the receptor site. Thus, this suggests that detonations occurring at approximately the same radial distance as that of the primary background emitter are a significant challenge for the detection systems.

Plot (i) in Figure 5.9 confirms our previous findings, which identified the MLP as the top classifier in this domain, the SVM as the worst, and the remaining three classifiers as contenders for the inner rankings. Indeed, while there are notable differences in the AUC plots for the J48, the NB, and the NN, the fact that their functions cross at numerous points, prohibits the derivation of a general ranking over the entire range of distances.

Plot (ii) in Figure 5.9 presents the performance of the one-class learners as a function of distance. In general, the plot demonstrates that all of the one-class classifiers follow a similar downward trend from their initial peaks, which occurred between 0.8

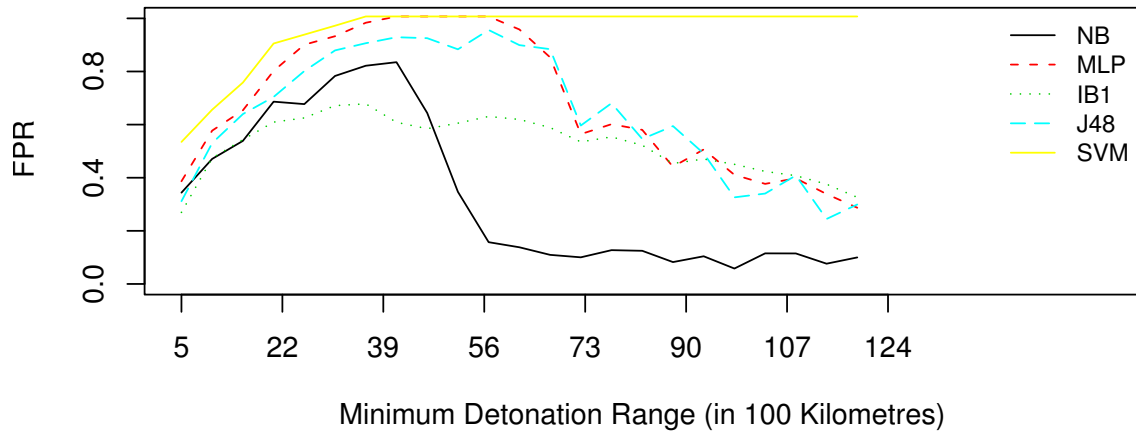


(i)

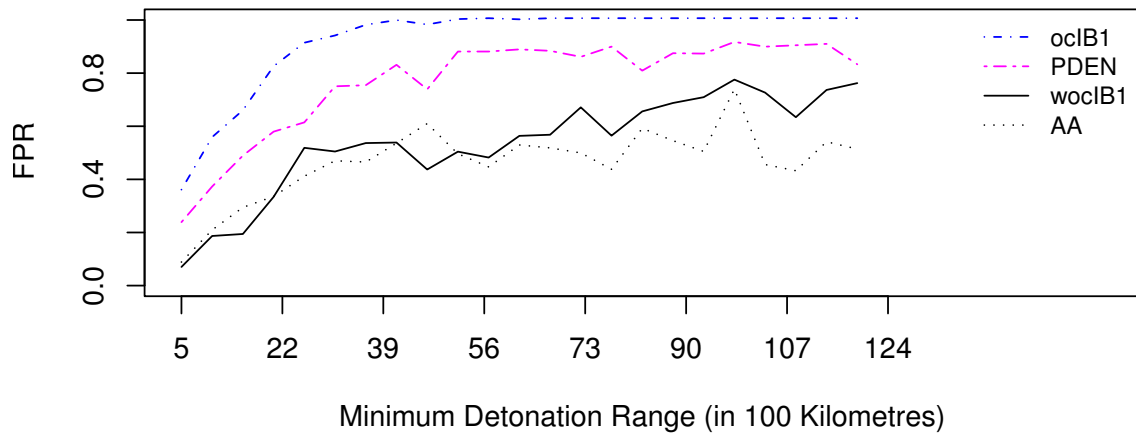


(ii)

Figure 5.9: In this figure, plot (i) displays the performance of the five binary classifiers, in terms of their AUC scores, as a function of distance. Similarly, plot (ii) displays the performances of the four one-class classifiers as a function of distance, according to their AUC scores.



(i)



(ii)

Figure 5.10: In this figure, plot (i) displays the performance of the five binary classifiers, in terms of their FPR scores, as a function of distance. Similarly, plot (ii) displays the performances of the four one-class classifiers as a function of distance, according to their FPR scores.

and 0.9, towards, or beyond in the case of the PDEN, an AUC of 0.5.

Moreover, the performance functions are broadly divisible into two categories. Both the ocNN and the PDEN descend relatively quickly, while the AA and the socNN degrade in a slower, more linear fashion. Therefore, the AA and the socNN are the more suitable of the four one-class learners, with the AA appearing generally superior to the socNN.

The performance of the nine classifiers, measured in terms of the FPR metric, are plotted as a function of distance in Figure 5.10. In this figure, plot (i) emphasizes the significant challenge incurred by the binary learners when the detonations occur at a distance similar to the noise source. Although we previously identified the MLP as the strongest binary classifier on this domain, for a relatively broad range (roughly between 25,000 km and 65,000 km), the vast majority of instances, which are truly of the detonation class, were assigned to the background class. The results are similar for J48. Interestingly, NB has the smallest area under its FPR curve. Thus, it least often identified members of the SE event class as background noise. While we do not consider the FPR results to be individually sufficient for model selection, they do provide some very intriguing insight into the behaviour of the classifiers.

The trends for the one-class classifiers in plot (i) follow much the same trends previously seen in Figure 5.9. In particular, the AA and the socNN are superior to the PDEN and the ocNN. However, the distinction between the AA and the socNN is less clear.

5.5.4 Expanded Feature-Space

Through our exploration of this most interesting of classification problems, we recognized both the inherent challenge presented in the classification of SE events that are interwoven in background noise, and the role of meteorology in effecting the very noise levels that make the task so difficult.

Our extensive consideration of this application domain has led us to identify the particularly strong relationship between the wind direction and pollutant levels at the receptor, which suggests a possibly informative feature.

By expanding the standard CTBT feature space to include wind direction, we have produced a significant increase in the AUC. In particular, the top classifiers (MLP, AA, socNN), now demonstrate the ability to detect detonations that, when considered solely on the basis of the four radioxenon measurements, fit into the background distribution with high probability. This fact is, indeed, depicted for many of the binary and one-class classifiers in Figure 5.11 and Figure 5.12.

In particular, while the depth to the hull in the performance of the MLP decreases only slightly, the J48's hull is entirely removed when the wind direction feature is added. Thus, the J48 classification ceases to be affected by the detonation distance when the new feature is included. In addition, its mean AUC is significantly improved.

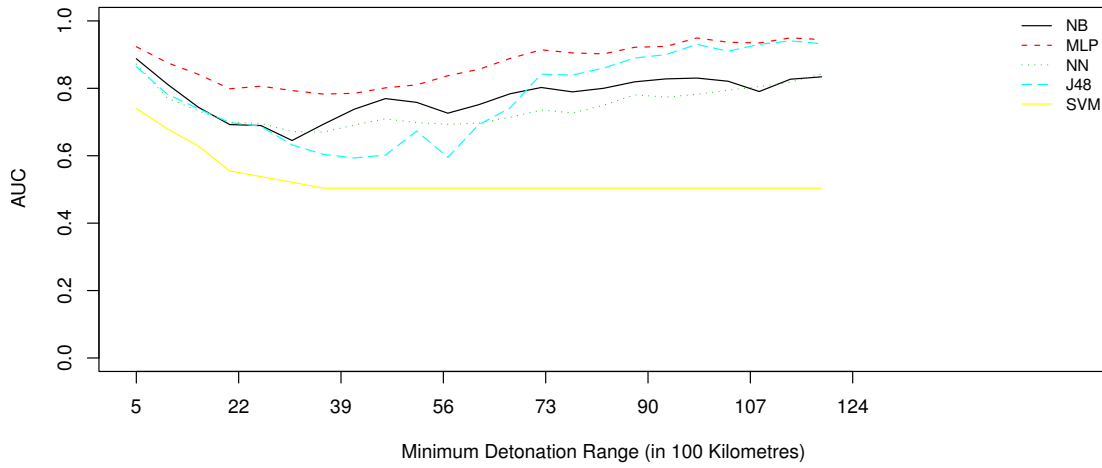
The NN and SVM classifiers also benefit from the inclusion of the wind direction feature. However, the new feature has a slightly negative effect on the NB. It has been noted in the literature, that many of the PR algorithms, including the MLP, SVM and NB may benefit from normalization of the features [28, 126]. Thus, it is conceivable that the performance of these classifier may be improved to some degree. However, these results provide a good baseline from which the individual classifiers can be compared.

By expanding the feature-space to include the wind direction, the one-class learner, socNN, improves significantly, and becomes, in general, the top learner amongst its peers. The classifier, AA, also improves as a result of the new feature. However, its AUC scores do not increase to the same extent as the socNN.

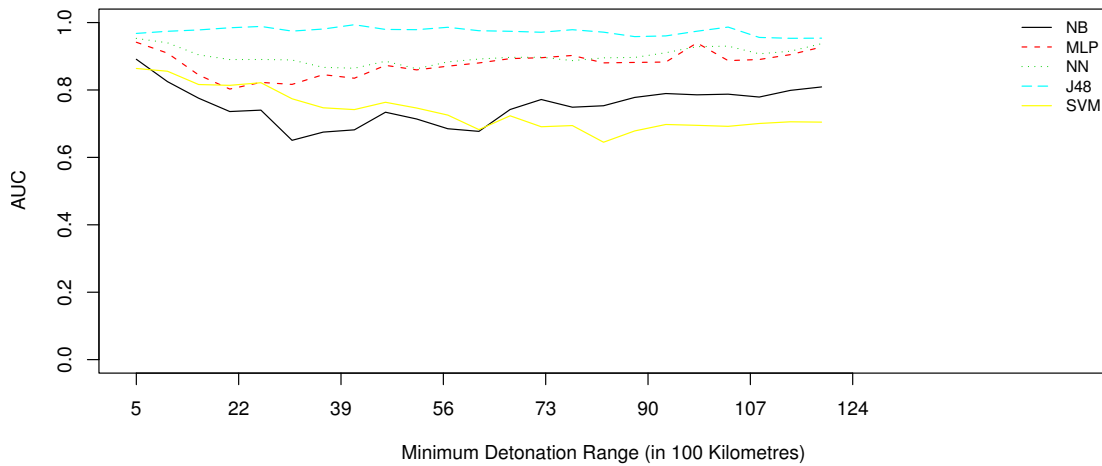
Similar to the socNN, the PDEN's initial performance is lower in the newly expanded feature-space. However, the majority of its performance function is elevated. Finally, the ocNN benefits the least from the new feature, although, its initial performance is improved.

Thus, in the worst case, the wind direction feature produces marginal improvements in the performance of the four one-class learners. However, it significantly improves both the AA and the socNN's ability to perform in scenarios where the detonations occur at distances equivalent to, and beyond the radial distance to the background source.

In Figure 5.13, a series of boxplots are utilized to facilitate the comparison of

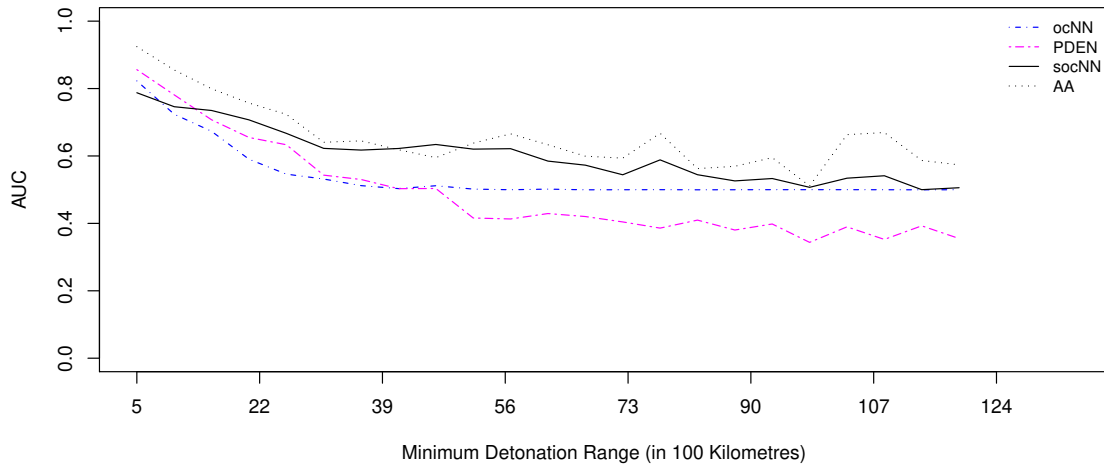


(i)

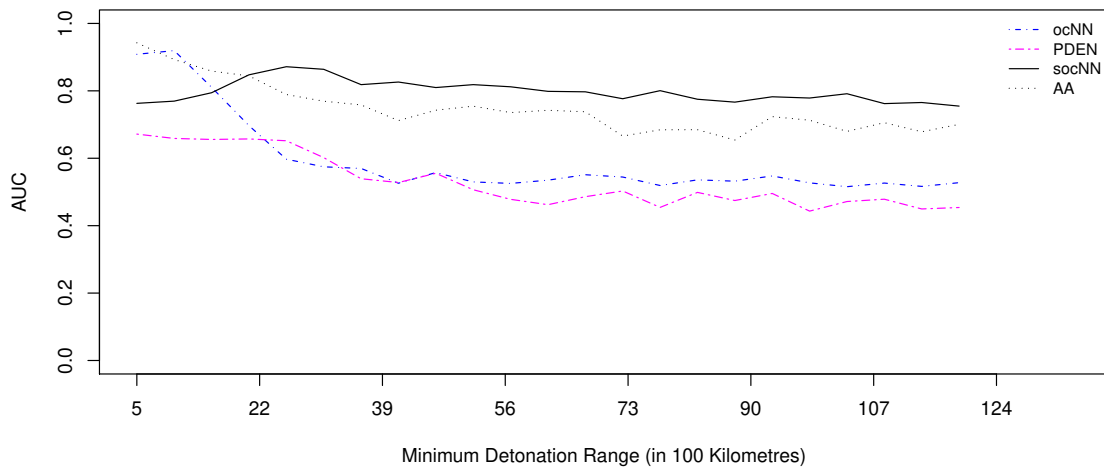


(ii)

Figure 5.11: This figure contrasts the performance of the binary classifiers, in terms of the AUC as a function of distance, on the standard feature-space (see Plot (i)), and when the feature-space is extended to include an assessment of the wind direction (see Plot (ii)).



(i)



(ii)

Figure 5.12: This figure contrasts the performance of the one-class classifiers, in terms of the AUC as a function of distance, on the standard feature-space (see Plot (i)), and when the feature-space is extended to include an assessment of the wind direction (see Plot (ii)).

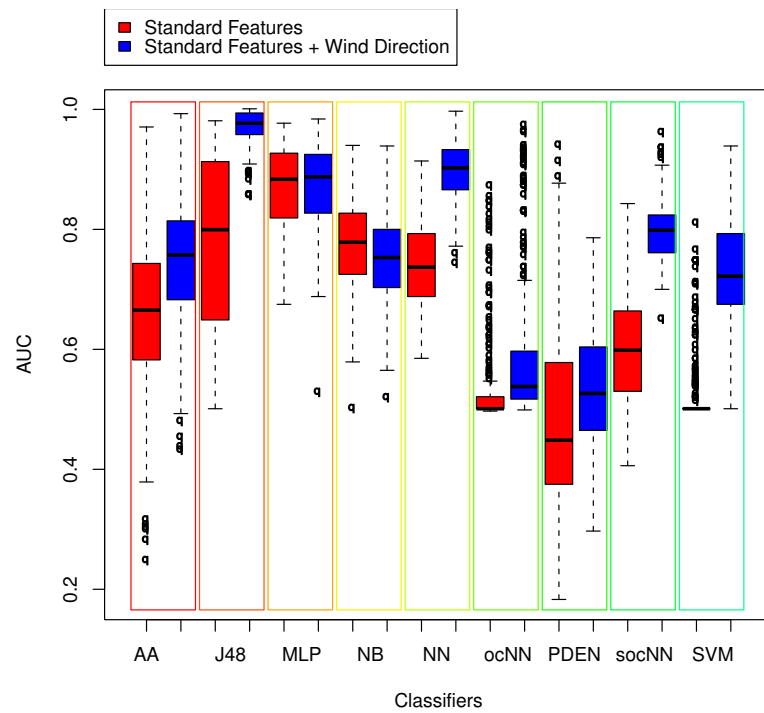


Figure 5.13: This figure utilizes a series of boxplots to compare the performance of the nine classifiers and the standard feature-space, and with the extended feature-space, which is augmented by a wind direction indicator.

classifier performance in the two feature-spaces. Indeed, these results confirm the trends that we have previously identified. Particularly noteworthy is the depiction of J48's performance; this plot emphasizes both the significant increase in the J48's median AUC score, and the impressive stabilization of its classification results when the wind direction feature is added. The benefits to the SVM are also well visualized in this figure.

It is, indeed, well demonstrated in Figure 5.11, Figure 5.12, and Figure 5.13 that the additional information has assisted many of the classifiers to overcome the significant challenges inherent in identifying SE events within the field of background noise.

5.6 Results: Scenario 2

In this section, we present the results that were produced on the four assessment criteria that were motivated, and utilized in the previous sections. In this section, however, as we explore the very intriguing classification scenario, which we previously denoted S1. This exploration follows the same structure that was previously applied in the exploration of the first classification scenario. Thus, we begin by examining the AUC scores produced by each of the one-class classifiers over the 23 detonation ranges; we then proceed to consider the performance over the two successive, smaller distances, the performance as a function of distance, and finally the benefit of expanding the feature-space to include an additional wind direction feature.

5.6.1 General Performance

In this section, we present a general overview of the performance of the set of one-class classifiers on the simulated CTBT domain. More specifically, we present an assessment of the four one-class classifiers, in terms of their AUC scores on the 230 datasets that covered the 23 detonation ranges.

Once again, in light of the fact that the SE event will, in practice, occur at random and unpredictable distances, these results are particularly insightful.

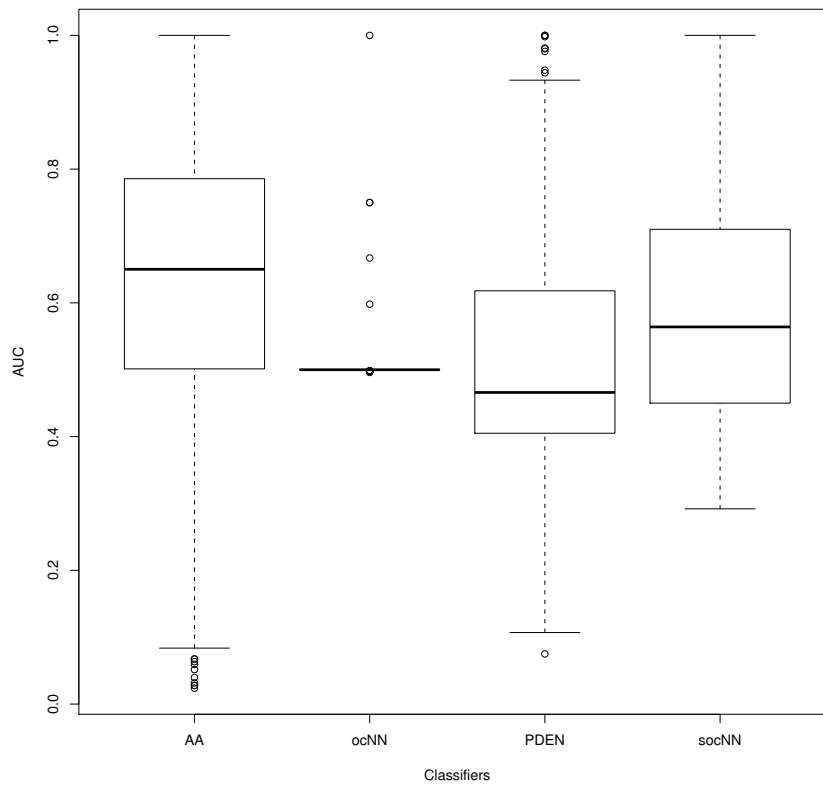


Figure 5.14: This figure displays the performance of the four classifiers, in terms of their AUC scores on the 230 generated CTBT datasets, in the form of a series of boxplots.

	Mean	Max	Min	STDV
ocNN	0.505	1	0.496	0.042
PDEN	0.507	1	0.075	0.185
socNN	0.587	1	0.292	0.171
AA	0.621	1	0.024	0.225

Table 5.3: This table displays the general classification results, in terms of AUC.

The results that are depicted in Figure 5.14 were compiled as a series of boxplots; one for each classifier. In addition, Table 5.3 contains a compilation of the mean, maximum, minimum and standard deviation of the each classifier’s overall results.

Our assessments of both Figure 5.14 and Table 5.3 reveal that, similar to our findings on the S1 scenario, the AA classifier is superior, in terms of its mean, and median scores, to the other one-class classifiers. Indeed, on this, which is a more challenging task, its mean and median values are only slightly lower than in the previous task. However, within this second scenario, it has the lowest minimum AUC scores, which appear as outliers in the boxplot. In addition, it is extremely unstable, with results ranging from perfect to near zero.

The classifier, socNN, ranks second after the AA according to its median and mean, and was considerably more stable, while the ocNN and PDEN classifiers produced values that were near or below 0.5.

5.6.2 Performance on Short- and Long-Range Detonations

In Figure 5.15, we present the results produced over two detonation ranges of particular interest. Specifically, Boxplot (i) in the figure contains the results for the datasets that include detonations between the distances of 1,000 km and 5,500 km, while Boxplot (ii) has those with detonations between 5,500 km and 10,000 km. Together, these plots demonstrate, contrary to the previous results, that there is little change in performance at greater distances.

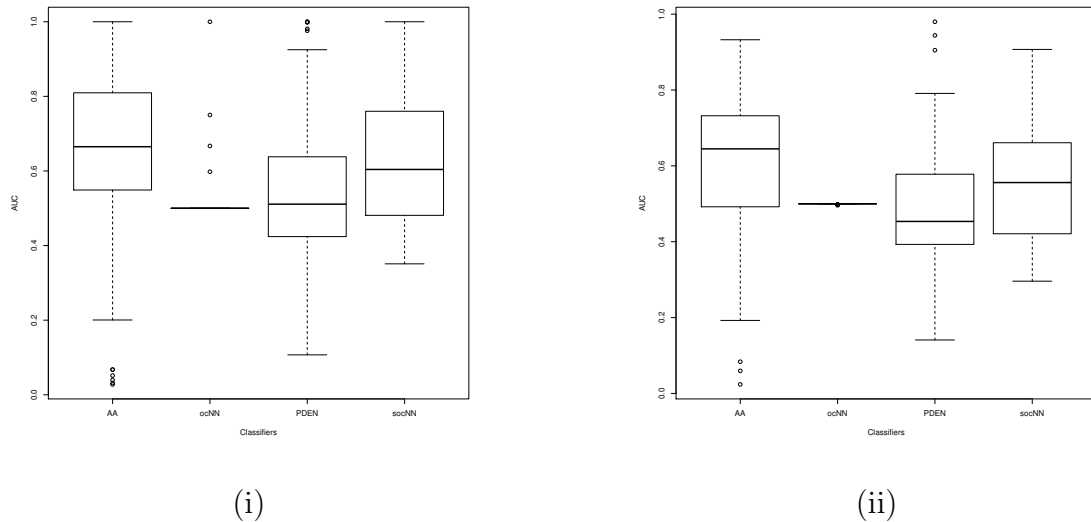


Figure 5.15: In this figure, Boxplot (i) displays the performance of the four classifiers, in terms of their AUC scores for detonations occurring between the distances of 1,000 km and 5,500 km, and Boxplot (ii) displays their performances for detonations between the distances of 5,500 km and 10,000 km.

5.6.3 Performance as a Function of Distance

In this sub-section, we present classifier performance as a function of distance. As in the previous section, performance is assessed both according to the AUC and the FPR.

The AA and socNN are, once again, roughly identifiable as the best of the four classifiers in Figure 5.16 and Figure 5.17. However, all of the classifiers, with the exception of ocNN, which rapidly converges to 0.5, suffer from significant and essentially random fluctuations. These fluctuations in performance suggest that the classifiers' results were as dependent on the nature of the SE events in the 230 datasets, as on the distance at which the events originally occurred.

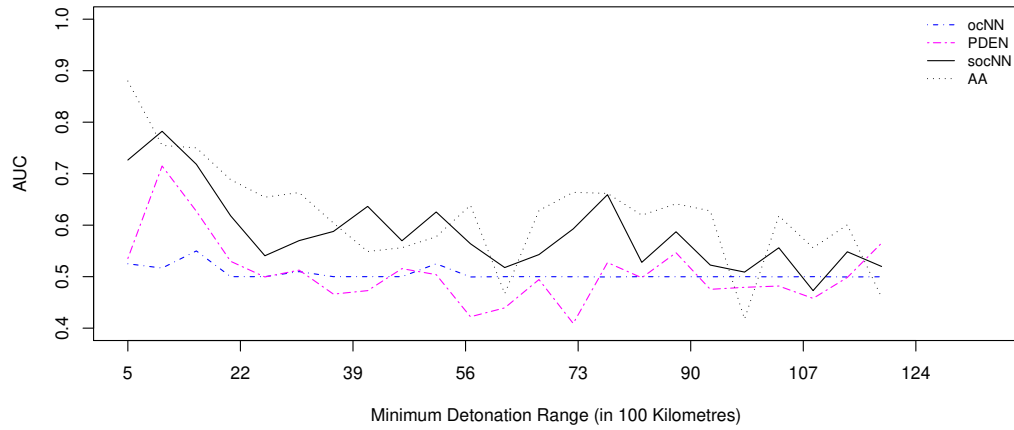


Figure 5.16: This figure displays the performance of the four one-class classifiers as a function of distance, according to their AUC scores.

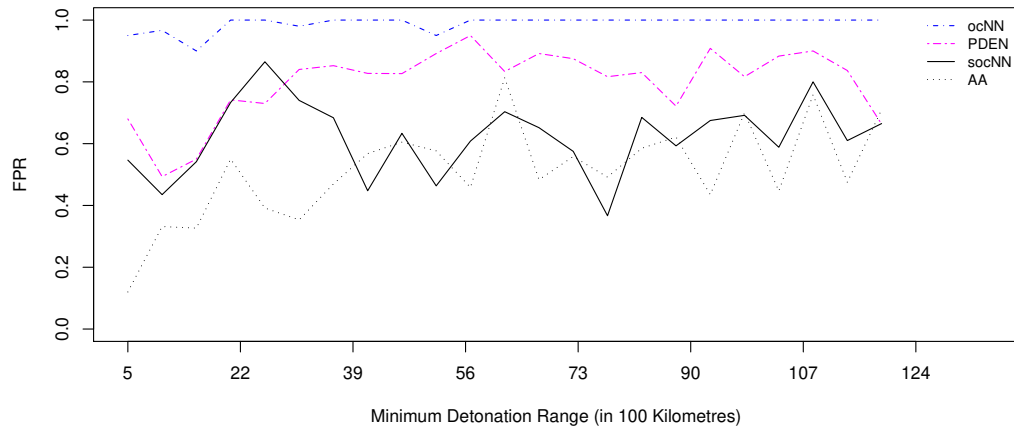


Figure 5.17: This figure displays the performance of the four one-class classifiers as a function of distance, according to their FPR scores.

5.6.4 Expanded Feature-Space

In this final section, we consider the benefits of extending the feature space to include a wind direction indicator. In Figure 5.18, both the original plot of the four classifiers' performances as a function of distance, and their performances on the extended feature-space are plotted. For an alternate view, the comparison is composed of a series of boxplots in Figure 5.19.

These figures illustrate that both the AA and the socNN significantly benefit from the expanded feature-space. Indeed, the socNN benefits the most, as it becomes superior to the AA for the vast majority of distances, and the variability in its results are significantly dampened.

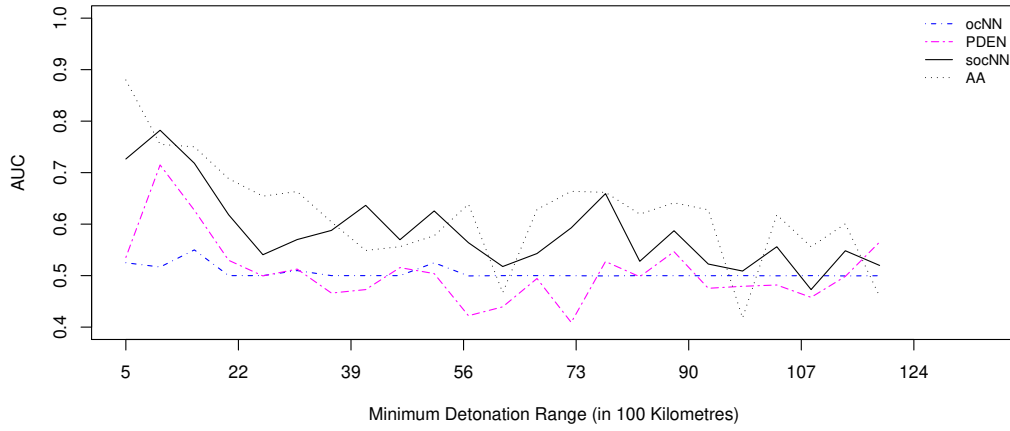
5.7 Discussion

In this section, we consider the results previously reported for the one-class classifiers in comparison to those reported for the binary learners. In particular, Section 5.7.1 compares the two classification strategies within the first scenario, namely S1. Alternatively, the one-class classifiers are considered in comparison to the set of standard binary classifiers on scenario S2 in Section 5.7.2.

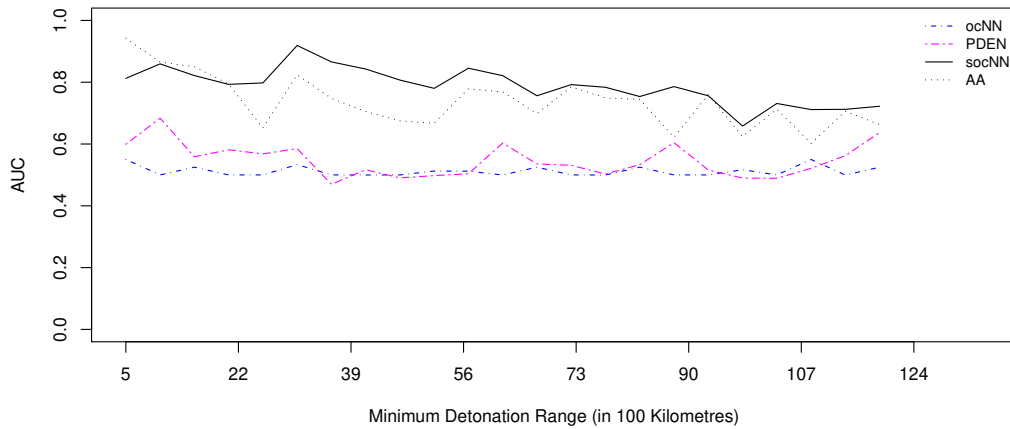
5.7.1 Results: S1

The relatively low mean and median AUC scores produced by the one-class classifiers, combined with the considerable variability in their results on the standard CTBT feature-space, particularly in comparison with the top binary learners, clearly illustrate the many challenges inherent in applying one-class learning to the derivation of a binary classifier. However, Hempstalk *et al.*, in [42], previously identified similar comparisons between binary and one-class learners as “naïve” comparisons, when applied to scenarios that are accurately identifiable as one-class problems.

In particular, in so-called one-class problems, such as the detection of SE events, the second class is inherently ill-understood due to the fact that a characteristic set cannot be drawn from it. Thus, training and testing a binary learner as if one



(i)



(ii)

Figure 5.18: This figure contrasts the performance of the one-class classifiers, in terms of the AUC as a function of distance, on the standard feature-space (see Plot (i)), and when the feature-space is extended to include an assessment of the wind direction (see Plot (ii)).

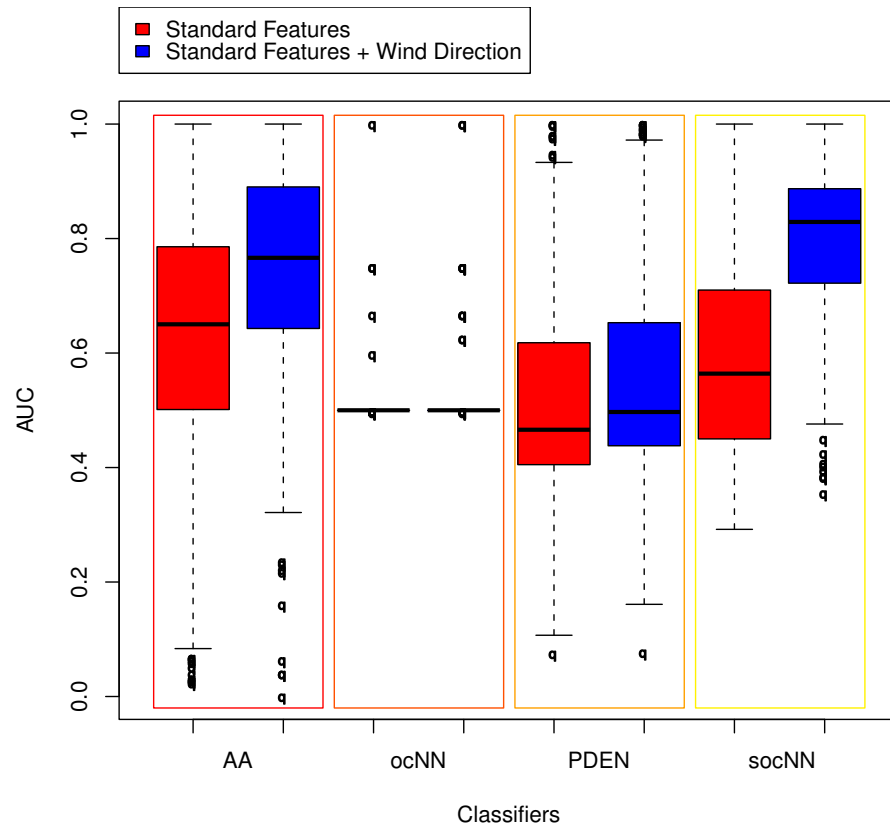


Figure 5.19: This figure utilizes a series of boxplots to compare the performance of the four classifiers and the standard feature-space, and with the extended feature-space, which is augmented by a wind direction indicator.

could draw a representative set from the second class, which is generally assumed when training a binary classifier, provides an upper bound on the classifier's future performance.

The key differences in the performance of the two forms of classifiers is well illustrated in Figures 5.8 and 5.9. While the one-class classifiers are very competitive on the initial radial ranges, when the detonation occurs further afield, their AUC scores drop considerably in comparison to all of the binary classifiers, with the exception of the SVM. The initial success of the one-class classifiers suggests that they are very capable of associating anomalously high levels of radioxenon with the SE event class.

However, the binary learners are not only well adapted to classifying anomalously highly levels as members of the SE event class, through the binary learning process they are also capable of drawing on the anomalously low levels, which commonly result from detonations that occurred well beyond the radial distance to the background source, to specialize their decision boundaries such that similar events are recognized as belonging to the SE event class in the future.

The results of expanding the standard CTBT feature-space to include an indicator of the prevailing wind were, in general, very favourable, and lead to improved AUC scores for most of the classifiers, with the NB being the sole exception.

In its essence, the wind direction feature enabled the classifiers to learn the direction of the background source. As a result, the classifiers were able to identify detonations, which occurred at similar radial distances to the receptor site as the background emissions, and thus, had signatures that were similar to the background levels, but were transported from a different direction. This result is identified very clearly in Figure 5.11, and suggests that the further expansion of the feature-space might additionally improve performance.

5.7.2 Results: S2

A considerable portion of the previous analysis is applicable to this second, more challenging, classification scenario. Most importantly, the benefits of the extended feature-space were witnessed within S2 as well. However, due to the nature of the

problem, only the one-class classifiers were applied to this first attempt at performing PR within this new domain.

As a result of the formulation of the problem, we proposed the use of standard one-class classifiers as unsupervised learners, and relied on inner mechanisms of the individual classifiers to facilitate the derivation of a model that segregated those instances of the training set that were accurately of the background class from the naïvely/erroneously labelled instances of the outlier class.

It is clear that the instability in performance that is depicted with respect to distance, and which is significantly more apparent in S2 than S1, results both from the erroneous instances in the training sets of S2, and the variability in classification challenges presented by the few members of the SE event class in the test sets. Indeed, the generation of random SE events over a domain as vast as the simulated CTBT domain, will inevitably produce both very easy, and nearly impossible classification tasks. Thus, when randomly including only a minute number of these events in the test sets, it is probable that performance on the SE event class will fluctuate significantly. This is, of course, why a large number of receptors are required in the global receptor network.

However, while the ensemble mean performance fluctuates considerably over the successive radial ranges, when considered in terms of the overall means, or medians, the performance of the one-class classifiers on the S2 task is only slightly lower than on the S1 task. In addition, this is true if in Figures 5.9 and 5.16, we were to conduct our analysis according to a series of best-fit lines.

Finally, as is depicted in Figure 5.18, in addition to elevating the performance of the top classifiers, the inclusion of the wind direction in the feature-space significantly dampens the variability in their performance. Moreover, Saey, in an extensive study of background radionuclide concentrations in Europe and North America, found that a few outliers representing significant increases in the background concentrations can be expected [95]. These outliers are attributed to alternate background sources, and can be assumed to have arrived at the receptor site via short-lived, and anomalous alterations in meteorology. Based on the standard CTBT feature space, such events undoubtedly suggest the detonation of a nuclear weapon. However, provided

a sufficient quantity of training data is available, it is conceivable that PR systems functioning with the wind direction feature may appropriately identify outliers of the background class.

5.8 Conclusion

In this research, we extend the frontiers of novelty detection through the introduction of a new field of problems open for analysis. In particular, we note that this new realm deviates from the standard set of one-class problems based on the presence of three characteristics, which ultimately amplify the classification challenge. They involve the *temporal* nature of the appearance of the data, the fact that the data from the classes are “interwoven”, and that a labelling procedure is not merely impractical - it is almost, by definition, impossible.

As a first attempt to tackle these problems, we presented two specialized classification strategies as demonstrated within the exemplary scenario intended for the verification of the CTBT. More specifically, we applied the simulation framework presented in Chapter 4, to generate CTBT inspired datasets, and demonstrated these classification strategies within the most challenging classification domain. More specifically, we have shown that one-class classifiers can be successfully applied to classify SE events, which are unknown, although present, at the time of training.

Finally, we have added a weighting parameter to the one-class nearest neighbour algorithm, thereby significantly increasing its performance on our experimental domain, and demonstrated that the expansion of the CTBT feature space significantly improves classifier performance on our simulated data, thus, motivating further exploration of the expansion of the standard CTBT feature space to include meteorological measurements.

Chapter 6

Conclusion and Future Research

This thesis was motivated by a desire to advance the state-of-the-art in PR by simultaneously promoting the existence of a novel and particularly challenging sub-category of PR problems, and presenting a first attempt at utilizing PR algorithms to classify the concepts within the domain.

Through our research, we recognized, as has previously been identified in other PR domains [26], that in the absence of a sufficient supply of “real” data, appropriately generated artificial data could facilitate insightful research into the classification of SE events. In the following section of this chapter, we provide a synopsis of the key contributions that arose first from our work modelling and simulating SE events, and subsequently from the classification of these events in the generated data. Finally, in Section 6.2, we suggest possible avenues for future work.

6.1 Contribution

In this section, we first discuss our contribution to the field of modelling and simulation. This occurs in Section 6.1.1. Subsequently, we discuss our contribution to PR in Section 6.1.2.

6.1.1 Modelling and Simulation

With this research, we have explored a relatively new field, namely the modelling and simulation of SE events as they propagate through a time-series composed, primarily of noise-like measurements.

The challenge, which inevitably arises when modelling systems such as these, that contain rare and unpredictable components, is that the prerequisite knowledge of the distribution is inaccessible. In overcoming this challenge, we demonstrated how, through the application of a divide-and-conquer strategy, which utilized the accessible domain knowledge, such as emission rates and the details of the propagation medium, along with a series of probabilistic decisions, systems of SE events can be explored in a scenario-specific fashion. In particular, we have provided a framework that can be applied to explore such systems in their own right, or to generate labelled time-series PR datasets.

Indeed, this framework was utilized in Chapter 5 to generate artificial datasets for our exploration of SE event classification, and published in [8].

6.1.2 Pattern Recognition

The primary focus of this research examined how PR systems could be deployed in the classification of SE events. Indeed, Chapter 5 discussed two distinct classification scenarios within the domain of SE event recognition. Both of these scenarios contained the following collective set of characteristics, which we demonstrated to be unique to this sub-category of PR problems:

- The data presents itself as a time sequence;
- The state-of-nature is dominated by a single class;
- The minority class is challenging to identify, even for domain experts, and occurs both rarely and randomly within the data sequence.

However, we formulated the first scenario in such a way that a sufficient, although by no means optimal, set of SE events were available to train binary classifiers. Thus,

we demonstrated how both binary and one-class classifiers could be utilized within this scenario.

Alternatively, and more appropriate for many SE domains, in the second scenario, the practitioners' ability to manually identify the presence of a SE event within the time-series of background noise is extremely limited. Thus, the acquisition of a dataset that is entirely appropriate for training binary and/or one-class classifiers is essentially unachievable.

Our research illustrates, however, that provided there is a mechanism built into the one-class classifier that enables the practitioner to adjust, and in particular, to reduce the portion the training set that is encapsulated by the learned discriminant function, the family of one-class classifiers can be applied to classify this extremely challenging second scenario. In addition, we provided a modified version of the one-class nearest neighbour algorithm [23], which is significantly less susceptible to noise, and is a strong classifier for the most challenging SE events.

Some of the research presented in this chapter has been submitted for publication in [7].

6.2 Future work

This work can be seen to motivate a number of continuation and future endeavours. In this section, we discuss the future work as it relates to modelling and simulation in Section 6.2.1, and with respect to PR in Section 6.2.2.

6.2.1 Modelling and Simulation

With respect to modelling and simulation, there are a number of very interesting extensions to this work. Most notably, and extremely relevant given the numerous possibilities for environmental, and human by extension, catastrophe, which could conceivably result from a host of SE events, this framework warrants the exploration of SE events from a policy and preparedness perspective. In particular, this framework is applicable to the exploration of events such as melt-downs at nuclear power plants

or explosion on deep-sea oil platforms,

In addition, researchers interested in modelling and simulating scenarios with a more acute set of parameters, within the air pollution domain, are directed to the many sophisticated ATM discussed in Chapter 3. These models, in many ways, are more suitable for studies that require increasingly sensitive analyses.

Finally, of significant interest to us is the the exploration of this framework in more complex scenarios, which, for example, involve multiple sources, simultaneously occurring SE events, the involvement of real meteorological datasets and more realistic emission rates. Indeed, with respect to emission rates, researchers in the fields of physics and chemistry continue to make advancements that will ultimately improve the predictions of the model.

6.2.2 Pattern Recognition

Continuing with the theme of disaster preparedness, we view the task of SE event recognition as having significant relevance with respect to *earlier warning systems*. In PR, this can be viewed as time-lapse detection efficiency (i.e., how quickly after a distant event can it be detected at the receptor site, and by extension how “early” is the warning).

We demonstrated in Chapter 4 that certain SE events can be seen to produce an approximately bell-shaped rise and subsequent fall in the features measured at the receptor site. Thus, if PR systems could be derived that detect SE events early in the initial rise, they would be very useful in evacuation scenarios. These systems could conceivably be applied to earthquake, tsunami and pollution earlier warning systems.

More directly related to this thesis, we intend to continue our exploration of one-class classifiers within the second scenario that was proposed in Chapter 5. In particular, the future work will include the examination of additional one-class classifiers and further improvements to socNN. This will be conducted on a greater number of test sets, which include more complex generalized case-studies. In addition, given the nature of the second scenario, we view the state-of-the-art in clustering algorithms as a relevant avenue for exploration.

Finally, we intend to conduct more sophisticated comparisons between the binary and one-class learners, such as those proposed in [\[42\]](#).

Bibliography

- [1] D. W. Aha. Generalizing from case studies: A case study. In *Proceedings of the Ninth International Conference on Machine Learning*, pages 1–10, 1992.
- [2] B. Ainslie and P. L. Jackson. The use of an atmospheric dispersion model to determine influence regions in the Prince George, BC airshed from the burning of open wood waste piles. *Journal of Environmental Management*, 90:2393–2401, 2009.
- [3] T. A. H. Al-Khayat, B. Van Eygen, C. N. Hewitt, and M. Kelly. Modelling and measurement of the dispersion of radioactive emissions from a nuclear fuel fabrication plant in the U.K. *Atmospheric Environment. Part A, General Topics*, 26(17):3079–3087, 1992.
- [4] R. Alaíz-Rodríguez and N. Japkowicz. Assessing the impact of changing environments on classifier performance. In Sabine Bergler, editor, *Canadian Conference on AI*, volume 5032 of *Lecture Notes in Computer Science*, pages 13–24. Springer, 2008.
- [5] S. P. Arya. *Air Pollution Meteorology and Dispersion*. Oxford University Press New York, NY, 1999.
- [6] S. Awasthi, M. Khare, and P. Gargava. General Plume Dispersion Model (GPDM) for point source emission. *Environmental Modeling and Assessment*, 11(3):267–276, 2006.

- [7] C Bellinger and B. J. Oommen. A new frontier in novelty detection: Pattern recognition of stochastically episodic events. In *23rd Australasian Joint Conference on Artificial Intelligence*, Submitted for publication on 20 June, 2010 2010.
- [8] C. Bellinger and B. J. Oommen. On simulating episodic events against a background of noise-like non-episodic events. In *42nd Summer Computer Simulation Conference, SCSC 2010, Ottawa, Canada, July 11-14, 2010. Proceedings*, 2010.
- [9] C. W. Benkley and A. Bass. Development of mesoscale air quality simulation models. volume 2. user's guide to MESOPLÜME (Mesoscale Plume Segment) model. Technical Report EPA-600/7-80-057, US Environmental Protection Agency, Research Triangle Park, NC, 1979.
- [10] M. R. Beychok. *Fundamentals of Stack Gas Dispersion*. Milton R Beychok, Irvine, California, 3rd edition, 1994.
- [11] C. M. Bishop. Novelty detection and neural network validation. *IEE Proceedings-Vision, Image and Signal processing*, 141(4):217–222, 1994.
- [12] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, Walton Street, Oxford OX2 6DP, 1995.
- [13] P. Bonelli, G. Calori, and G. Finzi. A fast long-range transport model for operational use in episode simulation. application to the chernobyl accident. *Atmospheric Environment. Part A. General Topics*, 26(14):2523 – 2535, 1992.
- [14] J. F. Bowers, J. R. Bjorklund, and C. S. Cheney. Industrial Source Complex (ISC) dispersion model user's guide. Technical Report EPA-450/4-79-030, US Environmental Protection Agency, Research Triangle Park, NC, 1979.
- [15] T. W. Bowyer, K. H. Abel, C. W. Hubbard, M. E. Panisko, P. L. Reeder, R. C. Thompson, and R. A. Warner. Field testing of collection and measurement of radioxenon for the comprehensive test ban treaty. *Journal of Radioanalytical and Nuclear Chemistry*, 240(1):109–122, 1999.

- [16] T. W. Bowyer, C. Schlosser, K. H. Abel, M. Auer, J. C. Hayes, T. R. Heim-bigner, J. I. McIntyre, M. E. Panisko, P. L. Reeder, H. Satorius, et al. De-tection and analysis of xenon isotopes for the comprehensive nuclear-test-ban treaty international monitoring system. *Journal of environmental radioactivity*, 59(2):139–151, 2002.
- [17] G. A. Briggs. Momentum and buoyancy effects. In D. H. Slade, editor, *Meteo-rology and Atomic Energy*, pages 189–202, 1968.
- [18] Y. Chen, X. Zhou, and T. S. Huang. One-class svm for learning in image retrieval. In *IEEE International Conference on Image Processing*, pages 34–37, 2001.
- [19] A. J. Cimorelli, S. G. Perry, A. Venkatram, J. C. Weil, R. J. Paine, R. B. Wilson, R. F. Lee, W. D. Peters, and R. W. Brode. AERMOD: a dispersion model for industrial source applications. Part I: general model formulation and boundary layer characterization. *Journal of Applied Meteorology*, 44(5):682–693, 2005.
- [20] R. H. Clarke, D. Harwell, and O. X. Oxon. *A model for short and medium range dispersion of radionuclides released to the atmosphere*. National Radiological Protection Board, 1979.
- [21] J. R. Cooper, K. Randle, and R. S. Sokhi. *Radioactive Releases in the Envi-ronment, Impact and Assessment*. John Wiley & Sons Canada Ltd, 2003.
- [22] S. Corrsin. Limitations of gradient transport models in random walks and in turbulence. In *Turbulent diffusion in environmental pollution; Proceedings of the Second Symposium*, volume 75, pages 25–60, 1974.
- [23] P. Datta. *Characteristic concept representations*. PhD thesis, Irvine, CA, USA, 1997.
- [24] J. P. Dawson, P. N. Racherla, B. H. Lynn, P. J. Adams, and S. N. Pandis. Simu-lating present-day and future air quality as climate changes: Model evaluation. *Atmospheric Environment*, 42(19):4551 – 4566, 2008.

- [25] F. Desiato. A long-range dispersion model evaluation study with Chernobyl data. *Atmospheric Environment. Part A. General Topics*, 26(15):2805–2820, 1992.
- [26] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.
- [27] R. R. Draxler. Determination of atmospheric diffusion parameters. *Atmospheric Environment (1967)*, 10(2):99 – 105, 1976.
- [28] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. New York, Wiley, 2nd ed. edition, 2001.
- [29] W. H. Duewer, M. C. MacCracken, and J. J. Walton. The Livermore regional air quality model. II. verification and sample application in the San Francisco Bay Area. *Journal of Applied Meteorology*, 17:273–311, 1978.
- [30] R. Damours. Modeling the ETEX plume dispersion with the Canadian emergency response model. *Atmospheric Environment*, 32(24):4335–4341, 1998.
- [31] C. Forster, U. Wandinger, G. Wotawa, P. James, I. Mattis, D. Althausen, P. Simmonds, S. O’Doherty, S. G. Jennings, C. Kleefeld, et al. Transport of boreal forest fire emissions from canada to europe. *Journal of Geophysical Research-Atmospheres*, 106(D19):22,887–22,906, 2001.
- [32] C. Gariazzo, V. Papaleo, A. Pelliccioni, G. Calori, P. Radice, and G. Tinarelli. Application of a lagrangian particle model to assess the impact of harbour, industrial and urban activities on air quality in the taranto area, italy. *Atmospheric Environment*, 41(30):6432 – 6444, 2007. Harbours and Air Quality.
- [33] C. Gariazzo, A. Pelliccioni, M. P. Bogliolo, and G. Scalisi. Evaluation of a lagrangian particle model (spray) to assess environmental impact of an industrial facility in complex terrain. *Water, Air, & Soil Pollution*, 155(1):137–158, 2004.

- [34] A. K. Ghosh, A. Schwartzbard, and M. Schatz. Learning program behavior profiles for intrusion detection. *Proceedings of the Workshop on Intrusion Detection and Network Monitoring*, 1:5162, April 1999.
- [35] F. A. Gifford. Turbulent diffusion-typing schemes: A review. *Nuclear Safety*, 17(1):68–86, 1976.
- [36] F. A. Jr Gifford. Use of routine meteorological observations for estimating atmospheric dispersion. *Nuclear Safety*, 2:47–51, 1961.
- [37] C. D. Hall. The simulation of particle motion in the atmosphere by a numerical random-walk model. *Quarterly Journal of the Royal Meteorological Society*, 101(428):235–244, 1975.
- [38] S. R. Hanna. *Review of Atmospheric Diffusion Models for Regulatory Applications*. N-8312662, World Meteorological Organization, Geneva (Switzerland), 1982.
- [39] S. R. Hanna, G. A. Briggs, and R. P. Hosker Jr. *Handbook on Atmospheric Dispersion*. DOE/TIC-11223. U.S. Department of Energy, 1982.
- [40] S. J. Hanson and J. Kegl. PARSNIP: A connectionist network that learns natural language grammar from exposure to natural language sentences. In *Ninth Annual Conference of the Cognitive Science Society*, pages 106–119, 1987.
- [41] D. M. Hargreaves and C. J. Baker. Gaussian puff model of an urban street canyon. *Journal of Wind Engineering and Industrial Aerodynamics*, 69-71:927–939, 1997. Proceedings of the 3rd International Colloquium on Bluff Body Aerodynamics and Applications.
- [42] K. Hempstalk, E. Frank, and I. H. Witten. One-class classification by combining density and class probability estimation. In *In Proceedings of the 19th European Conference on Machine Learning*, volume 5211, pages 505–519. Springer, 2008.
- [43] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. McGraw-Hill, Inc., New York, 1981.

- [44] P. Horton and K. Nakai. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. In *International Conference on Intelligent Systems for Molecular Biology*, volume 5, pages 147–152, 1997.
- [45] L. Hyafil and R. L. Rivest. Constructing optimal binary decision trees is NP-complete. *Information Processing Letters*, 5(1):15–17, 1976.
- [46] Y. A. Izraehl, V. N. Petrov, and D. A. Severov. Modelling of the transport and fallout of radionuclides from the accident at the Chernobyl nuclear power plant. In *International Symposium on Environmental Contamination Following a Major Nuclear Accident*, Proceedings series (International Atomic Energy Agency). IAEA, 1989.
- [47] Y. Izumi and J. S. Caughey. Minnesota 1973 atmospheric boundary layer experiment data report. Technical Report No. 547, AFGL Environmental Research Paper, National Technical Information Center, Springfield, Virginia, 1976.
- [48] M. Z. Jacobson. *Fundamentals of Atmospheric Modeling*. Cambridge University Press, Cambridge, United Kingdom, 2nd edition, 2005.
- [49] N Japkowicz. *Concept-Learning in the Absence of Counter-Examples: An Autoassociation-Based Approach to Classification*. PhD thesis, Rutgers University, 1999.
- [50] F. H. Kamst and T. J. Lyons. The evaluation of diffusivities for a non-uniform site. *Atmospheric Environment (1967)*, 16(3):379 – 390, 1982.
- [51] M. Khare and P. Sharma. *Modelling Urban Vehicle Emissions*, volume 9 of *Advances in Transport*. WIT Press, Computational Mechanics, Inc., 2002.
- [52] A. Kowalczyk and B. Raskutti. One class SVM for yeast regulation prediction. *SIGKDD Explorations Newsletter*, 4(2):99–100, 2002.
- [53] M. Kubat, R. C. Holte, and S. Matwin. Machine learning for the detection of oil spills in satellite radarimages. *Machine Learning*, 30(2):195–215, 1998.

- [54] R. G. Lamb, H. Hogo, and L. E. Reid. A Lagrangian approach to modeling air pollutant dispersion: Development and testing in the vicinity of a roadway. Technical report, EPA Research Report EPA-600/4-79-023, 1979.
- [55] R. Lange. ADPICA three-dimensional particle-in-cell model for the dispersal of atmospheric pollutants and its comparison to regional tracer studies. *Journal of Applied Meteorology*, 17(3):320–329, 1978.
- [56] P. Langley. Selection of relevant features in machine learning. In *Proceedings of the AAAI fall symposium on relevance*, pages 140–144. Institute for the Study of Learning and Expertise, Palo Alto, CA., 1994.
- [57] J. Langner, L. Robertson, C. Persson, and A. Ullerstig. Validation of the operational emergency response model at the Swedish Meteorological and Hydrological Institute using data from ETEX and the Chernobyl accident. *Atmospheric Environment*, 32(24):4325–4333, 1998.
- [58] B. Lauritzen, A. Baklanov, A. Mahura, T. Mikkelsen, and J. H. Sorensen. K-model description of probabilistic long-range atmospheric transport in the northern hemisphere. *Atmospheric Environment*, 40(23):4352 – 4369, 2006.
- [59] B. Lauritzen, A. Baklanov, A. Mahura, T. Mikkelsen, and J. H. Sorensen. Probabilistic risk assessment for long-range atmospheric transport of radionuclides. *Journal of Environmental Radioactivity*, 96(1-3):110 – 115, 2007. Radioactivity in the Environment (including the Arctic and Antarctic) ,The International Conference in Nice, France, 2-6 October 2005.
- [60] B. Lauritzen and T. Mikkelsen. A probabilistic dispersion model applied to the long-range transport of radionuclides from the Chernobyl accident. *Atmospheric environment*, 33(20):3271–3279, 1999.
- [61] D. H. Lenschow, X. S. Li, C. J. Zhu, and B. B. Stankov. The stably stratified boundary layer over the great plains. *Boundary-layer meteorology*, 42(1):95–121, 1988.

- [62] D. H. Lenschow, S. F. Zhang, and B. B. Stankov. The stably stratified boundary layer over the great plains. *Boundary-layer meteorology*, 42(1):123–135, 1988.
- [63] D. Lubinsky. *Bivariate Splits and Consistent Split Criteria in Dichotomous Classification Trees*. PhD thesis, Rutgers University, 1994.
- [64] T. J. Lyons and W. D. Scott. *Principles of Air Pollution Meteorology*. Belhaven Press, London, England, 1990.
- [65] M. C. MacCracken, D. J. Wuebbles, J. J. Walton, W. H. Duewer, and K. E. Grant. The Livermore regional air quality model. I. concept and development. *Journal of Applied Meteorology*, 17:254–272, 1978.
- [66] L. M. Manevitz and M. Yousef. One-class svms for document classification. *Journal Machine Learning Research*, 2:139–154, 2002.
- [67] R. H. Maryon and M. J. Best. Estimating the emissions from a nuclear accident using observations of radioactivity with dispersion model products. *Atmospheric Environment*, 29(15):1853 – 1869, 1995.
- [68] T.M. Mitchell. *Machine learning*. McGraw-Hill, 1997.
- [69] K. M. Mok, A. I. Miranda, K. U. Leong, and C. Borrego. A Gaussian puff model with optimal interpolation for air pollution modelling assessment. *International Journal of Environment and Pollution*, 35:111–137(27), 5 November 2008.
- [70] A. M. Monin, A. S. A. Yaglom. *Statistical fluid mechanics; mechanics of turbulence*. Cambridge, Mass., MIT Press, 1971.
- [71] N. Moussiopoulos, E. Berge, T. Böhler, F. de Leeuw, K. E. Grnskei, S. Mylona, and M. Tombrou. Ambient air quality, pollutant dispersion and transport models. Technical report, European Environment Agency, 1996.
- [72] F. T. M. Nieuwstadt. Some aspects of the turbulent stable boundary layer. *Boundary-Layer Meteorology*, 30(1):31–55, 1984.

- [73] F. T. M. Nieuwstadt. The turbulent structure of the stable, nocturnal boundary layer. *Journal of the Atmospheric Sciences*, 41(14):2202–2216, 1984.
- [74] Office of Technology Assessment. The containment of underground nuclear explosions. Technical report, US Congress, 1989.
- [75] A. P. Oliveira, J. Soares, T. Tirabassi, and U. Rizza. A surface energy-budget model coupled with a skewed puff model for investigating the dispersion of radionuclides in a sub-tropical area of Brazil. *Nuovo Cimento Della Società Italiana Di Fisica. C*, 21(6):631–646, 1998.
- [76] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [77] F. Pasquill. The estimation of the dispersion of windborne material. *Meteorological Magazine*, 90:33–49, 1961.
- [78] F. Pasquill. The dispersion of material in the atmospheric boundary layer: The basis for generalization. *Lectures on air pollution and environmental impact analyses*, pages 1–34, 1975.
- [79] F. Pasquill. Atmospheric dispersion parameters in gaussian plume modeling. part ii: possible, requirements for change in the turner workbook valves. Technical Report No. EPA-600/4-76030b, U.S. Environmental Protection Agency, Research Triangle Park, NC., 1976.
- [80] G. M. Pasquill, F. Smith. *Atmospheric diffusion*. Ellis Horwood, 3 edition, 1983.
- [81] T. Petsche, A. Marcantonio, C. Darken, S. J. Hanson, G. M. Kuhn, and I. Santoso. A neural network autoassociator for induction motor failure prediction. *Advances in neural information processing systems*, pages 924–930, 1996.
- [82] K. Pruess. The TOUGH codes—a family of simulation tools for multiphase flow and transport processes in permeable media. *Vadose Zone J.*, 3(3):738–746, 2004.

- [83] K. Pruess, C. M. Oldenburg, and G. J. Moridis. TOUGH2 user's guide version 2. Technical report, Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory, 1999.
- [84] J. R. Quinlan. The effect of noise on concept learning. *Machine learning: An artificial intelligence approach*, 2:149–166, 1986.
- [85] J. R. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers, 1993.
- [86] J. D. Reid. Markov chain simulations of vertical dispersion in the neutral surface layer for surface and elevated releases. *Boundary-Layer Meteorology*, 16(1):3–22, 1979.
- [87] L. Rendell and H. Cho. The effect of data character on empirical concept learning. In *Artificial Intelligence Applications, 1989. Proceedings., Fifth Conference on*, pages 199–205, 1989.
- [88] G. Ritter and M. T. Gallegos. Outliers in statistical pattern recognition and an application to automatic chromosome classification. *Pattern Recognition Letters*, 18(6):525–539, 1997.
- [89] S. Roberts and L. Tarassenko. A probabilistic resource allocating network for novelty detection. *Neural Computation*, 6(2):270–284, 1994.
- [90] S. J. Roberts. Novelty detection using extreme value statistics. *IEE Proceedings - Vision, Image, and Signal Processing*, 146(3):124–129, 1999.
- [91] H. C Rodean. Stochastic Lagrangian models of turbulent diffusion. *American Meteorological Society*, 26(48), 1996.
- [92] K. R. Rodriquez, D. J Peterson. Simulating the venting of radioactivity from a Soviet nuclear test. *Atmospheric Environment*, 23(5):953–958, 1989.
- [93] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. pages 318–362, 1986.

- [94] P. R. J. Saey, T. W. Bowyer, and A. Ringbom. Isotopic noble gas signatures released from medical isotope production facilities – Simulation and measurements. *Applied Radiation and Isotopes*, 2010.
- [95] Paul R.J. Saey. The influence of radiopharmaceutical isotope production on the global radionuclide background. *Journal of Environmental Radioactivity*, 100(5):396 – 406, 2009.
- [96] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [97] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. Platt. Support vector method for novelty detection. *Advances in neural information processing systems*, 12:582–588, 2000.
- [98] K. T. Schwarz, T. W. Patzek, and D. B. Silin. Dispersion by wind of CO₂ leaking from underground storage: Comparison of analytical solution with simulation. *International Journal of Greenhouse Gas Control*, 3(4):422 – 430, 2009.
- [99] J. S. Scire, D. G. Strimaitis, and R. J. Yamartino. A user’s guide for the CALPUFF dispersion model. Technical report, Earth Tech, Concord, MA, 2000.
- [100] P. Seibert, F. Beyrich, S. E. Gryning, S. Joffre, A. Rasmussen, and P. Tercier. Mixing height determination for dispersion modelling (COST Action 710). In B. E. A. Fisher, J. J., S. Erbrink, P. Finardi, S. Jeannet, M. G. Joffre, U. Morselli, P. Pechinger, Seibert, and D. J. Thomson, editors, *EUR 18195 - COST Action 710 - Final report. Harmonisation of the pre-processing of meteorological data for atmospheric dispersion models*, pages 1–120. Office for Official Publications of the European Communities, 1998.
- [101] C. C. Shir. A preliminary numerical study of atmospheric turbulent flows in the idealized planetary boundary layer. *Journal of the Atmospheric Sciences*, 30(7):1327–1339, 1973.

- [102] C. C. Shir and L. J. Shieh. A generalized urban air pollution model and its application to the study of SO_2 distributions in the St. Louis metropolitan area. *Journal of Applied Meteorology*, 13:186–204, mar 1974.
- [103] D. H. Slade (ed.). Meteorology and atomic energy. Technical report, U.S. Atomic Energy Commission. Division of Reactor Development and Technology, 1968.
- [104] K. Soetaert and P. M. J. Herman. *A Practical Guide to Ecological Modelling: Using R as a Simulation Platform*. Springer Verlag, 2008.
- [105] J. H. Sorensen. Sensitivity of the derma long-range gaussian dispersion model to meteorological input and diffusion parameters. *Atmospheric Environment*, 32(24):4195 – 4206, 1998.
- [106] T. J. Stocki, N. Japkowicz, G. Li, R. K. Ungar, I. Hoffman, and J. Yi. Summary of the data mining contest for the iee international conference on data mining. Pisa, Italy, 2008.
- [107] A. Stohl, M. Hittenberger, and G. Wotawa. Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data. *Atmospheric Environment*, 24:4245–4264, 1998.
- [108] A. Stohl and D. J. Thomson. A density correction for Lagrangian particle dispersion models. *Boundary-Layer Meteorology*, 90:155–167, 1999.
- [109] K. S. Suh, M. H. Han, S. H. Jung, and C. W. Lee. Numerical simulation for a long-range dispersion of a pollutant using Chernobyl data. *Mathematical and Computer Modelling*, 49(1-2):337 – 343, 2009.
- [110] O. G. Sutton. *Micrometeorology*. McGraw-Hill, 1953.
- [111] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady. Novelty detection for the identification of masses in mammograms. *IEE Conference Publications*, 1995(CP409):442–447, 1995.

- [112] D. M. J. Tax. *One-class classification; Concept-learning in the absence of counter-examples*. PhD thesis, Technische Universiteit Delft, Netherlands, 2001.
- [113] G. I. Taylor. Diffusion by continuous movements. *Proceedings of the London Mathematical Society*, 20(2):196–212, 1921.
- [114] Aria Technologies. MINERVE wind field models volume 7.0, general design manual. Technical report, ARIA Tech, 2001.
- [115] H. Tennekes and J. L. Lumley. *A first course in turbulence*. MIT press, Cambridge MA, 1972.
- [116] R. Thompson. Numeric calculation of turbulent diffusion. *Quarterly Journal of the Royal Meteorological Society*, 97(411):93–98, 1971.
- [117] D. J. Thomson. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*, 180:529–556, July 1987.
- [118] G. Tinarelli, D. Anfossi, G. Brusasca, E. Ferrero, U. Giostra, M. G. Morselli, J. Moussafir, F. Trombetti, and F. Tampieri. Lagrangian particle simulation of tracer dispersion in the lee of a schematic two-dimensional hill. *Journal of Applied Meteorology*, 33(6):744–756, 1994.
- [119] T. Tirabassi and U. Rizza. A practical model for the dispersion of skewed puffs. *Journal of Applied Meteorology*, 34(4):989–993, 1995.
- [120] B. D. Turner. *Workbook of Atmospheric Dispersion Estimates: an Introduction to Dispersion Modeling*. CRC Press, Inc., Boca Raton, Florida, second edition edition, 1994.
- [121] D. B. Turner and United States. National Air Pollution Control Administration. *Workbook of atmospheric dispersion estimates*. Number No. AP-26. U.S. EPA: National Air Pollution Control Administration, Research Triangle Park, NC., 1970.

- [122] D. B. Turner and J. H. Novak. User's guide for RAM. volume 1: Algorithm description and use. Technical Report EPA-600/8-78-016A, US Environmental Protection Agency, Research Triangle Park, NC, 1978.
- [123] USEPA. Users guide for the industrial source complex (isc3) dispersion models. Technical Report 2, US EPA: Ofce of Air Quality Planning and Standards Emissions, Monitoring, and Analysis Division, Research Triangle Park, N.C, 1995.
- [124] A. P. Van Ulden. Surface-layer similarity model for the dispersion of a skewed passive puff near the ground. *Atmospheric Environment*, 26(4):681–692, 1992.
- [125] B. L. Wilson, J. D. Brian L. Sawford. Review of lagrangian stochastic models for trajectories in the turbulent atmosphere. *Boundary-Layer Meteorology*, 78(1–2):191–210, 1995.
- [126] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Publishers, second edition, 2005.
- [127] D. Yeung and C. Chow. Parzen-window network intrusion detectors. *Pattern Recognition, International Conference on*, 4:40385, 2002.
- [128] P. Zannetti. *Air Pollution Modeling: Theories, Computational Methods and Available Software*. Van Nostrand Reinhold, New York, 1990.
- [129] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*, pages 49–56. The MIT Press, 2004.
- [130] P. Zib. *Urban Air Pollution Dispersion Models: A Critical Survey*. Department of Geography and Environmental Studies, University ofthe Witwatersrand, Johannesburg, South Africa, 1977.