# On Simulating Episodic Events Against a Background of Noise-like Non-episodic Events

**Colin Bellinger**
**School of Computer Science, Carleton University, Ottawa, Canada**
e-mail:`cbelling@scs.carleton.ca`

**B. John Oommen**
***Chancellor's Professor*; *Fellow: IEEE* and *Fellow: IAPR***
**School of Computer Science, Carleton University, Ottawa, Canada**
Also an *Adjunct Professor* with the University of Agder in Grimstad, Norway.
e-mail:`oommen@scs.carleton.ca`

**Keywords**: Episodic events; propagation; Gaussian dispersion; pattern recognition; data generation

## Abstract

Simulation, as an art and a science, deals with the issue of allowing the practitioner to model events using their respective probability distributions. Thus, it is customary for simulations to model the behaviour of accidents, telephone calls, network failures etc. In this paper, we consider a relatively new field, namely that of modelling *episodic* 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 episodic event does occur, its magnitude and features far overshadow the background, as one observes after a seismic event. In this paper, we demonstrate how the effect of a particular form of episodic event can be modelled as it propagates through the underlying background noise. Furthermore, we illustrate how the subsequent decay of the event can also be modelled and simulated. In demonstrating this concept, we utilize the exemplar scenario posed by the Comprehensive Nuclear-Test-Ban Treaty (CTBT), and model the propagation and decay of radionuclides, emitted from clandestine, subterranean nuclear detonations, through the background levels resulting from the global nuclear industry.

## 1 INTRODUCTION

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 of farmers to predict annual planting and harvesting cycles, in addition to aiding travellers with 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. Phenomena of the latter classes are the focus of this paper. Indeed, as we shall argue, a certain subset of these possesses 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 recreate 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 reactions 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 Pattern Recognition (PR), and for example, Disease Contagion Prediction (DCP) systems, as was fundamental to the study of the SARS crisis.

This paper concentrates on the simulation and subsequent analysis of episodic events. Our interest in this particular aspect of simulation resides in a desire to generate labelled datasets for the training, testing and validation of PR systems. In particular, the simulation strategy, as described, provides a controlled means by which to explore the classification of rare and/or episodic events within a well-defined background distribution. To the best of our knowledge, the results presented here represent both a novel and pioneering step in this venture.

The remainder of this paper is formatted as follows: Section 2 provides a motivating scenario. In Section 3, we provide some background on pattern classification. Section 4 describes the frameworks utilized in the simulation of the background and episodic events. In particular, this section articulates the Gaussian equations applied to the demonstration of plume and puff dispersion in the atmosphere. In Section 5, we describe the motivating factors behind the demonstration scenario. In addition, this section also discusses the parameter selection and generation issues. The results produced by the modelling and simulation system are presented in Section 6. Section 7 includes a discussion of the results, and Section 8 contains our concluding thoughts.

## 2 MOTIVATION

The modelling framework presented in this paper is motivated by two distinct application domains. From a regulatory perspective, the modelling of rare and/or episodic events provides insight, and advances our general understanding of the physical and natural fall-out from rare and/or episodic events, such as a failure at an industrial plant, or within other critical systems.

The primary motivation for this work, however, is the generation of background and episodic events, the effects of which may be measured and included in domain-specific machine learning datasets. Subsequently, the derived datasets may be applied to explore existing and new PR systems on target domains where data is particularly elusive, or altogether unavailable.

Related work, by Dietterich *et al.*, in [7], extended the generalizing strategy presented by Aha in [1], 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. This study, however, relied on a characteristic understanding of the class data being generated. Moreover, it assumed that the available data was, indeed, representative of the broader class, and thus, justified the generation procedure. The accuracy of such a conclusion is typically elusive in domains having characteristically episodic elements.

Alternatively, our approach relies not on an understanding of the distribution of, and statistical relationships within, the episodic data desired, but on knowledge of the environment that affects it. In the case of pollutant dispersion, for example, this implies knowledge of the propagation medium, the physical properties of the chemical being dispersed, and the ability to simulate a source's affect on the measurement site.

## 3 BACKGROUND ON PATTERN RECOGNITION

PR systems and their practitioners specifically rely on data from both the background and episodic classes for training and validation. For further details on PR, readers are directed to the work of Duda *et al.*, in [8]. However, it is sufficient to understand that standard PR systems utilize training instances drawn from the background and episodic classes in order to learn a discriminating function, which can be used to distinguish between the two classes during deployment.

Further training instances are subsequently utilized in testing and validation, which are processes essential to PR model selection, and our interpretation of how the system will perform when deployed. In standard PR applications, the availability of training instances is generally not an issue. However, if the problem is episodic in nature or imbalanced in general, the issue of insufficient data points must be overcome in order to build sufficient confidence in the PR system being deployed.

The standard approach to the testing and validation of PR systems is to set a certain number of the training instances aside, or held-out, for use during each of these important phases. The standard holdout approach was demonstrated by Lubinsky in [14], in an imbalanced scenario. The use of novel instances is essential in these phases in order to assess the performance in a manner that is independent of the training process. When the holdout technique is applied to imbalanced classification problems, the results depend on extremely small testing and validation sets, thus, limiting the confidence which can be placed on the future performance of the PR system.

Alternative approaches train one-class classifiers, in which the training process utilizes the background instances alone [11, 13]. As a result, the few instances from the episodic class can be entirely committed to the testing and validation process. However, the availability of instances of the episodic class may still be insufficient to build confidence in this approach as well. This suggests that alternative mean validation are required, such as the generation of artificial data.

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 episodic 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 episodic data. The background class is considered to be relatively well understood, and in particular, strong estimates of its distribution can be assumed to be known. Alternatively, the episodic 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.

In the spirit of the verification of the Comprehensive Test Ban Treaty (CTBT), this work demonstrates the process of simulating the dispersion of radioxenon emitted from industrial sources, and calculates the effect on the radioxenon levels at points of interest. The initial simulation process is employed to develop the background data, which is characteristically well-defined. More specifically, the background distribution can be determined and is stable over time as a result of the static nature of the sources and the consistency of the atmospheric dispersion when considered over the long-term. In addition, the result of radioxenon emitting, clandestine, detonations of nuclear weapons, whose spacial and temporal locations along with the magnitude of the detonations are random, is simulated to form the episodic events.

# 4 MODELLING SYSTEM

As previously mentioned, this work considers a relatively new field, namely that of modelling episodic events, which are characteristically random in space, time and magnitude. It also explores the relationship between these episodic 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 episodic 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 episodic events on, and propagate their effects through, the background noise.

In doing this, we divide the modelling process into two phases. The details of these stages, and in particular, the modelling of the background and the episodic events, are discussed in detail in the following subsections. Before preceding, 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 episodic interludes into a well-defined background distribution.

We can summarize our hypothesis by the following:

1. The simulation of background noise-like non-episodic pollutants is best modelled by the Gaussian plume model;

2. The simulation of episodic contaminants is best modelled by the Gaussian puff model.

These issues are clarified in the following sections.

## 4.1 Modelling the Background

In this particular application of the theoretical model described above, the background data is modelled after pollutant observations have been made at a receptor site. More specifically, we assume the existence of industrial emitters positioned at static locations and characterized by emission rates that are subjected to Gaussian fluctuations. The effect of the various sources on the receptor site is calculated based on the widely applied Gaussian plume model.

While the Gaussian plume model has seen considerable application, and has been verified to reproduce plume dispersion with relative accuracy (see [3, 4, 17]), it is 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 [19]. The latter justifies its application in the current task.

The Gaussian plume dispersion equation has its foundation in the basic advection-diffusion equation, which through a series of assumptions can be solved analytically to produce the Gaussian puff equation described below. The Gaussian puff equation, Eq. (2), 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, such as neglecting dispersion along the *x*-axis in order to simplify the integration of Eq. (2), the Gaussian plume equations takes the following form, which was articulated by Lyons in [15] as:

$$\begin{aligned}\chi(x,y,z,t) =& \frac{Q}{2\pi\sigma_y\sigma_z\overline{u}} exp\left(-\frac{y^2}{2\sigma_y^2}\right) \\ & \left[exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right],\end{aligned} \tag{1}$$

and describes the air pollutant concentration, $\chi$, 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 $\overline{u}$ 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 center-line.

Turner, in [21], conveniently articulates 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, $\frac{1}{u}$, specifies that parallel to the *x*-axis, the concentrations are inversely proportional to the wind speed. Parallel to the *y*-axis, the crosswind factor,

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

indicates that the concentrations are inversely proportional to the crosswind spreading, $\sigma_y$, 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 $\sigma_y$ 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,

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

specifies that the concentrations are inversely proportional to the vertical spreading, $\sigma_z$, 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 Eq. (1) 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.

## 4.2 Modelling Episodic Events

The episodic events modelled in this simulation are representative of short-term massive releases of pollutants into the environment from a random point in space and time. Subsequent to the random episodic event, a pollutant cloud is instantaneously vented, and takes a Gaussian form in the environment, where it is transported by the host of atmospheric forces. This Gaussian assumption applies to the pollutant cloud as it advects through the atmosphere, and is, once again, supported by Batchelor's supposition, which was originally referred to in the previous subsection. Its influence 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 measurements.

While the Gaussian 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. 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. Indeed, models based on the Gaussian puff framework, as explained in [2, 10, 12], have been demonstrated to produce strong concentration predictions in comparison with physical measurements. 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 an instantaneous point source

The derivation of the Gaussian puff equations is illustrated by Lyons, in [15], and takes the following form,

$$\begin{aligned}\overline{\chi}(x,y,z,t) =& \frac{Q}{(4\pi t)^{\frac{3}{2}}(\sigma_x\sigma_y\sigma_z)^{\frac{1}{2}}} \\ & exp\Big[-\frac{(x-\overline{u}t)^2}{2\sigma_x}+\frac{y^2}{2\sigma_y}+\frac{z^2}{2\sigma_z}\Big],\end{aligned} \tag{2}$$

which is unsurprisingly reminiscent of the Gaussian plume equation displayed in Eq. (1). As in the Gaussian plume equations, $\chi$ indicates the air pollutant concentration at the receptor position, $(x, y, z)$, in *mass units* $m^{-3}$. The parameter $Q$ in the above equation represents the instantaneous point source in *mass units* $s^{-1}$, $t$, is time in seconds, and the mean wind speed, which is assumed to travel along the *x*-axis, is represented by $\overline{u}$, in $m\ s^{-1}$. Finally, the $\sigma$ values represent the downwind, crosswind and vertical dispersion as a function of downwind distance, in meters.

# 5 EXPERIMENTAL SETUP

In this study, we propose a modelling technique designed to facilitate the simulation of episodic 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.

## 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 techniques being explored utilize PR systems trained on quantities of radioxenon measured at sampling stations, otherwise referred to as "receptors", distributed throughout the globe [20].

In general, it can be assumed that radioxenon is present within the atmosphere for one of two reasons, the primary being emissions from the nuclear industry. Alternatively, the detonation of nuclear weapons is known to release massive 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 of nuclear weapons will be characteristic of an episodic event. Moreover, representative instances of the class are unavailable for development of a PR system. 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 episodic explosions propagating through the system.

## 5.2 Modelled System

For the purpose of this demonstration, we assume a simplified environment. In particular, we apply simplifying assumptions to the process of atmospheric dispersion, industrial emissions and the episodic emissions, in order to illustrate the general effect of the episodic event at the receptor site.

### 5.2.1 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. Theorist and practitioners have attempted to classify the diffusive effect that results from these processes in a variety of ways. Traditionally, the Pasquill stability classes and the Pasquill-Gifford dispersion parameters [9, 16] have experienced considerable favour within Gaussian dispersion models, and are applied for the determination of the $\sigma$ terms in Eq. (1) and Eq. (2). More recently, new techniques have been proposed, which are subjected to fewer restrictions. However, for the purpose of this study, we choose to view diffusion as a purely statistical phenomenon. Indeed, we apply this assumption to the complete set of atmospheric parameters required by the Gaussian methods. Thus, we assume that over the course of an experiment, the wind speed, wind direction, and diffusion parameters, $\sigma_{x,y,z}$, take independent Gaussian forms with user-defined means and variances.

### 5.2.2 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 emission cycles, which may result from refuelling, planned maintenance, safety inspections, etc., as a means of concisely achieving our overall objective. Furthermore, the emission 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.

### 5.2.3 Episodic Emissions

The episodic events in this simulation take the form of clandestine nuclear explosions. 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 [18]. 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 [6].

Based on the evidence presented above, and consistent with [18], 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 episodic in nature, the occurrence of an explosion at any particular point in space and time, is best estimated as 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.

# 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 radioxenon levels at a set of regional receptor sites. The subsequent subsection, Subsection 6.2, illustrates how the episodic events propagate through the otherwise consistent background distribution.

## 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[1] 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 Eq. (1) and Eq. (2) assume a specified mean value of 15 with a standard deviation of 5. However, we use the term "approximately isotropic" because, each $\sigma$ 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. 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 1. By calculating the

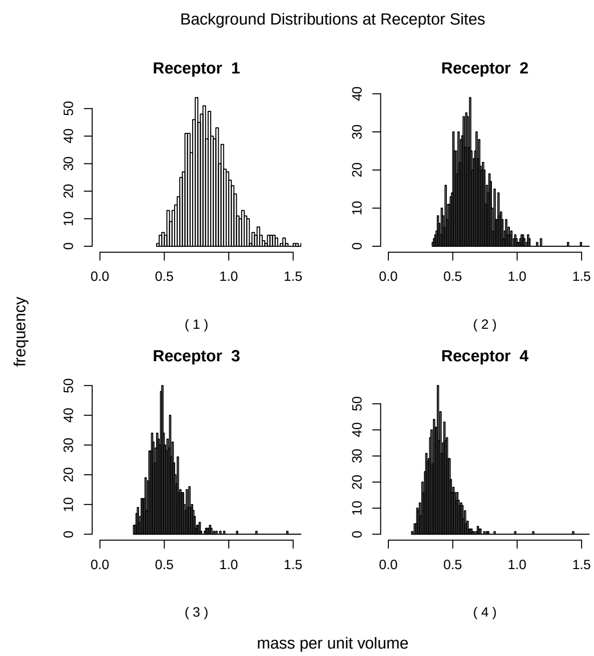


**Figure 1**. The calculated probability distribution for pollutant levels resulting from a single regional industrial emitter at each of the four receptor sites located 100, 130, 160 and 190 meters downwind. These results were calculated over a forty-one day period.

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 2. 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.

## 6.2 Episodic Events

Unlike episodic 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

[1] For demonstration purposes, the distances used in this paper are "small". A corresponding scenario in which the distances involve hundreds of thousands of kilometres is described in [5].

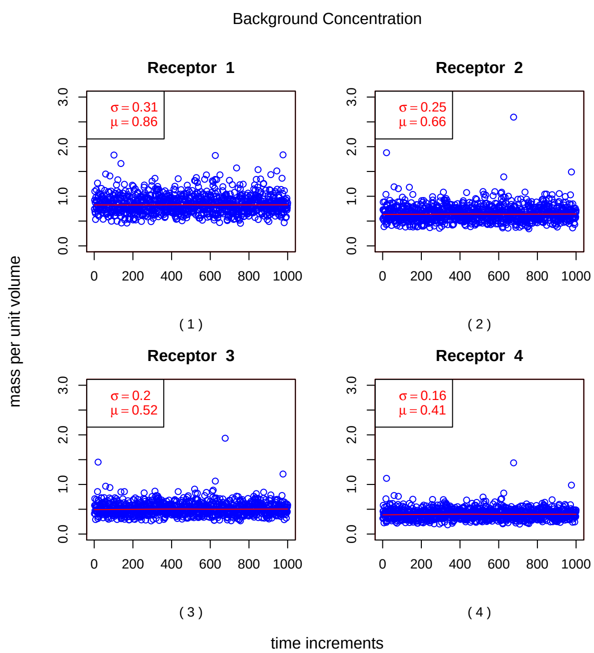


**Figure 2**. The time-series scatter plot illustrating the mean hourly background pollutant concentrations resulting from an upwind industrial emitter, for each of the four receptor sites located 100, 130, 160 and 190 meters downwind. These results were calculated over a forty-one day period.

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 2. 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. In Figure 3, the solid blue circles indicate the background levels discussed earlier. The four peaks in 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 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. The downwind propagation of an instantaneous vented cloud of radioxenon, resulting from a subterranean detonation is depicted in Figure 4. In particular, this figure demonstrates the effects of both dispersion and radioactive decay. By comparing the episodic spike in sub-figure 4.1 with the third episodic spike in Figure 3 (which corresponds to a similar detonation), it becomes apparent that the radioactive decay plays a fundamental role in the removal of radioxenon from the atmosphere. In particular, the episode peaks slightly above a concentration of 2 in sub-figure 4.1, 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.

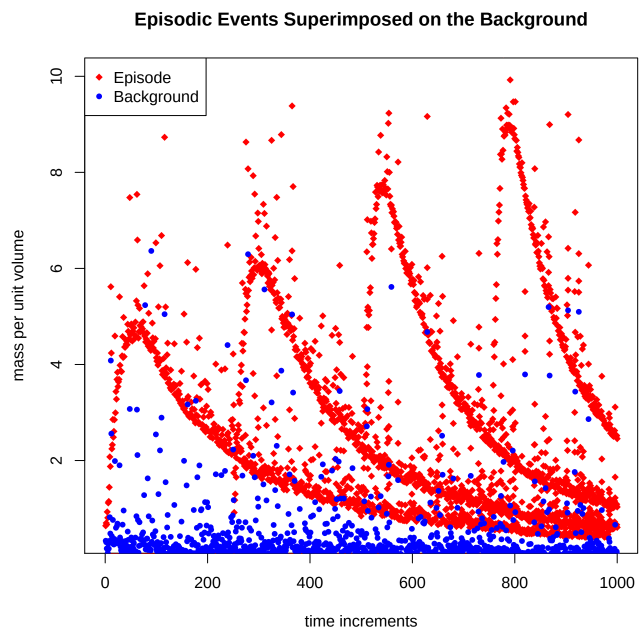


**Figure 3**. The time-series scatter plot illustrating the effect of four episodic events on pollutant levels at the receptor site nearest to the industrial emitter. The details about the figure and its legend are found in the body of the paper.

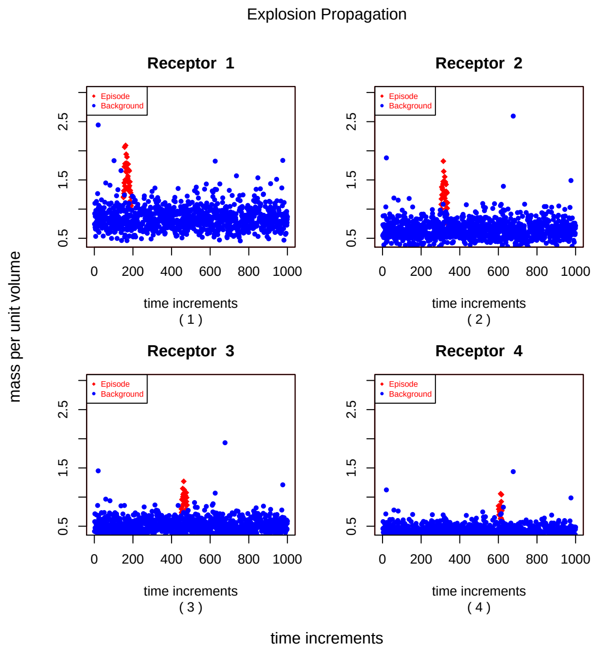


**Figure 4**. This figure illustrates the downwind propagation of a cloud of vented radioxenon. In particular, the effects of diffusion and radioactive decay are demonstrate through the incrementally decreasing concentrations of radioxenon (depicted by red diamonds, which collectively form small spikes) measured at each successive receptor site.

## 7 DISCUSSION

An analysis of the accuracy of the Gaussian dispersion models presented above is beyond the scope of this paper. 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 episodic events, and in particular, how this relationship can be modelled. We theorize that in many instances, over the long-term 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 1 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 features of the system under study. In particular, the simulation articulates 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 2 through the sequentially decreasing mean values and standard deviations. Larger-scale and longer experiments, which include seasonal variations in the atmosphere, and for more emitters at greater distances, will help to further validate this in the future. However, we believe that this simple model presents a solid foundation for any future study.

Episodic 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 properly tune most models in order 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 episodic events. Moreover, we are interested in how the rare episodic events, which possesses a largely unknown distribution in time, space and magnitude, can be simulated, and their effects subsequently propagated through the model. Figure 3 and Figure 4, provided excellent depictions of how a series of probabilistic choices can be applied to generate episodic 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 the individual downwind receptor affected by the episode. In particular, the slumping spikes in Figure 3 illustrate the relationship between the dispersion and distance travelled, as does Figure 4, 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 of analysis within 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. We have, indeed, used these schemes to design such a system, the results of which are presently being compiled for publication [5].

## 8 CONCLUSION

In this paper, we have considered a relatively new field, namely that of modelling *episodic* 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 episodic 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 particular propagation medium and a general description of the background sources that are under study (in this case, the nuclear industry), the major features can be modelled sufficiently for the purpose of exploring the effects of episodic events. Given the largely random and sporadic nature of these episodic 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.

The results obtained and knowledge gained through the application of the state-of-the-art in pattern recognition of labelled (background/episodic) data produced by this simulation system, are discussed elsewhere [5].

## REFERENCES


[1] D. W. Aha. Generalizing from case studies: A case study. In D. H. Sleeman and P. Edwards, editors, *ML*, pages 1–10. Morgan Kaufmann, 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] 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.

[5] C. Bellinger. Simulation and pattern classification for rare and episodic events. Master's thesis, Carleton University, Ottawa, Ontario, 2010.

[6] 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.

[7] 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.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. New York, Wiley, 2nd ed. edition, 2001.

[9] F. A. J. Gifford. Atmospheric dispersion calculations using the generalized gaussian plume model. *Nuclear Saftey*, 2:56–59, 67–68, 1960.

[10] 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 Colloqium on Bluff Body Aerodynamics and Applications.

[11] 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.

[12] R. Jain, J. Vaughan, K. Heitkamp, C. Ramos, c. Claiborn, M. Schreuder, M. Schaaf, and M. Lamb. Development of the clearsky smoke dispersion forecast system for agricultural field burning in the pacific northwest. *Atmospheric Environment*, 41(32):6745 – 6761, 2007.

[13] N. Japkowicz. *Concept-Learning in the Absence of Counter-Examples: An Autoassociation-Based Approach to Classication*. PhD thesis, Rutgers University, 1999.

[14] D. Lubinsky. *Bivariate Splits and Consistent Split Criteria in Dichotomous Classi cation Trees*. PhD thesis, Rutgers University, 1994.

[15] T. J. Lyons and W. D. Scott. *Principles of Air Pollution Meterorology*. Belhaven Press, London, England, 1990.

[16] F. Pasquill. The estimation of the dispersion of windborne material. *Meteorological Magazine*, 90:33–49, 1961.

[17] S. G. Perry, A. J. Cimorelli, R. J. Paine, R. W. Brode, J. C. Weil, A. Venkatram, R. B. Wilson, R. F. Lee, and W. D. Peters. Aermod: A dispersion model for industrial source applications. part ii: Model performance against 17 field study databases. *Journal of Applied Meteorology*, 44(5):694–708, 2005.

[18] K. R. Rodriquez, D. J Peterson. Simulating the venting of radioactivity from a Soviet nuclear test. *Atmospheric Environment*, 23(5):953–958, 1989.

[19] D. H. Slade (ed.). Meteorology and atomic energy. Technical report, U.S. Atomic Energy Commission. Division of Reactor Development and Technology, 1968.

[20] T. J. Stocki, N. Japkowicz, G. Li, R. K. Ungar, I. Hoffman, and J. Yi. Summary of the data mining contest for the ieee international conference on data mining. Pisa, Italy, 2008.

[21] B. D. Turner. *Workbook of Atmospheric Dispersion Estimates: an Introduction to Dispersion Modeling*. CRC Press,Inc., Boca Raton, Florida, second edition edition, 1994.