Motivating the Inclusion of Meteorological Indicators in the CTBT Feature-Space

Colin Bellinger
School of Information Technology and Engineering,
University of Ottawa,
Ottawa, Canada
e-mail: `cbell052@uottawa.ca`

Nathalie Japkowicz
School of Information Technology and Engineering,
University of Ottawa,
Ottawa, Canada
e-mail: `nat@site.uottawa.ca`

Abstract—**Verification of the Comprehensive Test-Ban-Treaty (CTBT), as a Pattern Recognition (PR) problem, has been proposed based on four radioxenon features. It has been noted, however, that in many cases this limited feature set is insufficient to distinguish radioxenon levels effected by an explosion from those that are solely products of industrial activities. As a means of improving the detectability of low-yield clandestine nuclear explosions, this paper motivates the inclusion of meteorological indicators in the CTBT feature-space, promotes further research into which meteorological indicators are most informative, and how they may be acquired. In doing so, we present classification results from four simulated scenarios. These results demonstrate that the inclusion of a simple wind direction feature can significantly increase the prospect of classifying challenging detonation events, and suggests the predictive power of meteorological features in general.**

I. Introduction

The verification of the Comprehensive Test-Ban-Treaty (CTBT) poses both an extremely interesting and a challenging Pattern Recognition (PR) problem. When viewed in its standard form, and as a "black box", the PR system is trained to classify vectors quantifying four radioxenon isotopes to either the explosion or the non-explosion class.

Feature vectors belonging to the non-explosion class are assumed to be the results of the combined impact of one or more regional, and to a lesser extent global, industrial nuclear activities, such as those at Medical Isotope Production Facilities (MIPF) and Nuclear Power Plants (NPP). Alternatively, the feature vectors associated with the explosion class, are assumed to be the combined results of a low-yield clandestine nuclear test, which occurs randomly in time and space, and a set of industrial emitters.

By definition, the initiator of a clandestine test goes to great lengths to conceal the nuclear test. Thus, tests are likely to be contained in underground or underwater facilities. These containment facilities are expected to subdue the majority of the isotopes produced in the detonation. Therefore, the occurrence of a significant increase in atmospheric levels of radioxenon cannot be assumed after a test. Moreover, it has been observed that due to radioactive decay and chaining, the indicative "fingerprint" formed by the four radioxenon isotopes, ^{131}Xe, ^{133}Xe, ^{133m}Xe and ^{135}Xe, after the detonation, degrades and becomes indistinguishable from the MIPF emissions, which are the most influential [1]. Together, these factors work to conceal the source (industrial or detonation) of the measured radioxenon, when considered within the standard feature-space. This fact was specifically demonstrated in [2], where a significant number of detonations were shown to reside well within the background distribution. As a result, these instances could only be classified as explosions at the expense of a higher *False Negative Rate* (FNR).

There exists, however, a promising opportunity in this otherwise challenging task, in particular, the possibility to expand the CTBT feature-space, thereby improving the accuracy of the derived PR systems. More specifically, the fundamental role of meteorology in determining the portion of the released radioxenon that is ultimately measured in the atmosphere at the International Monitoring Station (IMS) indicates that meteorology likely contains information that can aid in the discrimination process. Indeed, when knowledge, or estimates, of the atmospheric processes that transport and disperse pollutants are combined with a strong understanding of the emissions source (specifically global positioning, emitted species, and rates of decay), a great deal of information becomes available about the sources of the background radioxenon levels at particular points. Importantly, this includes under which atmospheric conditions the background concentrations are at their highest and lowest.

The remainder of this paper is organized as follows. In the subsequent section, Section II, we motivate the expansion of the CTBT feature-space. Section III demonstrates our hypothesis, and why it is expected to increase the classification results. The details about our data procurement process are included in Section IV. In Section V, issues pertinent to the experimental design are considered. The experimental results, and a subsequent discussion, are contained in Section VI and Section VII, respectively. Finally, our concluding remarks are organized in Section VIII.

II. Motivation

As a result of the considerable challenge of classifying detonations based on the four radioxenon measurements, we propose that, in general, expanding the feature-space will improve the classification results. More specifically, we recognize the significant role held by meteorology in effecting

the amount of radioxenon that is witnessed at each IMS. Atmospheric flows are specifically responsible for transporting the emitted radioxenon from the source to the IMS. Together, the set of atmospheric flows that effect the radioxenon levels at a point, over a particular period of time can be thought of as an *airshed*.

Similar to a watershed, which specifies the area over which melt- and rain-waters migrate downhill into increasingly large river systems (effecting the downstream content of minerals, pollutants, etc.), an airshed can be viewed as a set of air flows that merge to effect the pollution levels at a particular downwind site (see [3] for a further discussion of airsheds). Therefore, by inferring the airshed that effected a particular measurement, a considerable amount of information about the source(s) of radioxenon at the receptor is obtained. This is valuable information that can assist in the classification, particularly if the measured levels are anomalous for the detected airshed.

The value in an airshed estimate resides in the fact that measurements that appear normal in the standard feature-space, may be exceptional when considered with respect to an airshed with only marginal, or no, industrial emitters. Thus, when radioxenon levels are considered in conjunction with the history of the particular airshed that transported them, stronger predictions can be made. Indeed, similar techniques utilizing background radioxenon statistics and *inverse modelling* are applied to manually assess the likelihood of a detonation and its probable location [4].
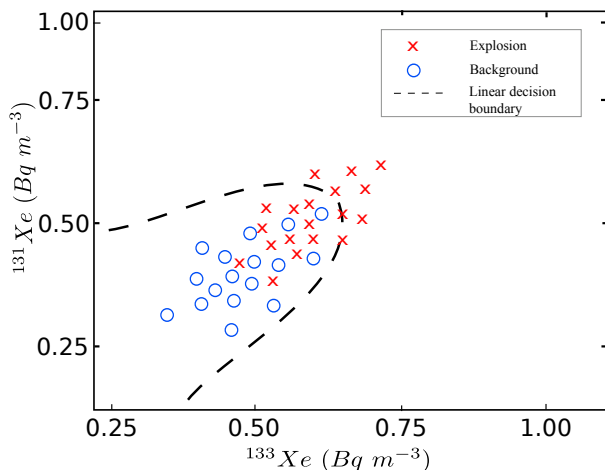

Fig. 1. A hypothetical decision boundary inferred by a simple one-class learner on the background data.

While [2] demonstrated that a wind direction feature indicating the direction to the emission source could greatly improve the classification of clandestine nuclear explosions, it did not provide an in-depth exploration of how such estimates might be acquired, or the type of error that may be involved in "real world" estimates. Based on a series of simulations, this work comprehensively demonstrates that knowledge of the source airshed can improve classification results, thus serving

to motivate further research into *a)* the estimation of airsheds within the CTBT domain of problems, *b)* how such estimates can be quantified, and *c)* their ideal form within the feature-space.

III. JUSTIFICATION FOR AN EXPANDED CTBT FEATURE-SPACE

In this section, we assume that airsheds can be associated with each IMS measurement, and proceed to demonstrate how its inclusion in the CTBT feature-space aids in creating, or increasing, the separability of the explosion data from the background data.

Conceptually, adding an extra dimension to the CTBT feature-space, specifically an airshed or wind direction indicator, facilitates the separation of radioxenon measurements that would otherwise be indistinguishable to a trained classifier.
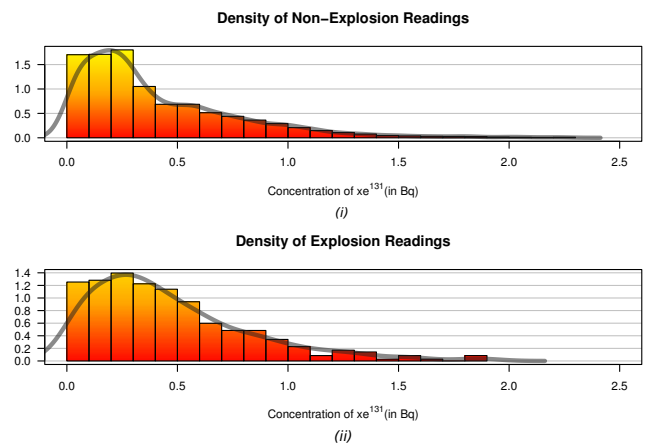

Fig. 2. This figure illustrates the strong similarity between the background and explosion distributions when considered with respect to the radioxenon concentration.

Take, for example, the simple two-dimensional case presented in Figure 1. This plot depicts a hypothetical scenario in which a one-class (OC) classifier has been trained on the background data (specified by the blue circles), and inferred the discriminant function represented by the dashed line. During testing, a significant number of explosions (marked as red crosses) are erroneously classified according to the decision boundary. A similarly challenging scenario was, indeed, discussed based on simulated CTBT data in [2].

In Plot (i) and Plot (ii) of Figure 2, the distributions of the background and explosion data from a single experiment in [2] are respectively displayed. These figures illustrate that the majority of background and explosion instances (specifically the ^{131}Xe measurements) were quantified as being between $0\ Bq\ m^{-3}$ and $0.5\ Bq\ m^{-3}$. Similar scenarios were found for the three remaining radioxenon isotopes. Thus, there is a strong similarity in the multi-variate background and explosion distributions. As a result, classification, and particularly one-class classification (OCC)[1] is extremely challenging. This is

---

[1]From the standpoint of PR, verification of the CTBT is conceptually an OCC problem (for more details see [2])

demonstrated by the relatively poor Area Under the ROC Curve (AUC) results produced on the data, which ranged from 0.505 to 0.621. By adding an estimate of the direction to the airshed, however, many explosion instances, which were previously indistinguishable from the background class, can be correctly classified.

The benefit of the expanded feature-space is emphasized by the two-dimensional histogram plotted in Figure 3. In this figure, the wind direction is plotted in radians along the x-axis, and the radioxenon concentration is plotted along the y-axis. The yellow areas indicate a high frequency of instances, and the dark red areas specify lower frequencies of instances. Finally, the white implies that no such instance occurred in the dataset. No instances, for example, occurred with a wind direction of 3 $radians$ and a concentration around $0.75^{131}Xe$ Bq m^{-3}. Indeed, the wind was only rarely between 2 $radians$ and 4 $radians$. Similarly, the radioxenon concentration was rarely above 0.5 Bq m^{-3}.

If the two-dimensional histogram is assumed to represent the training set (in accordance with OCC, this data is drawn entirely from the background class), it is clear that during application, nearly any radioxenon concentration should be considered suspect when the wind is between 2 $radians$ and 4 $radians$. Similarly, concentrations roughly above 0.5 Bq m^{-3} are suspicious when the wind is between 0.9 $radians$ and 5.5 $radians$. Without the wind direction feature, these instances fit neatly into the background distribution, and thus, could only be classified as explosions at the expense of a higher FPR.
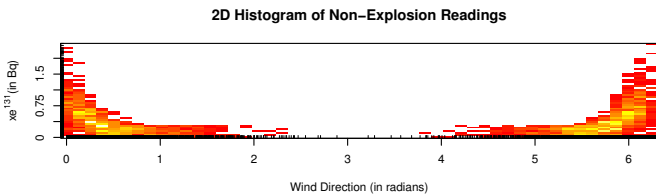

Fig. 3. This figure contains a two-dimensional histogram with the background concentration of ^{131}Xe plotted on the y-axis, and the direction to the airshed plotted on the x-axis. The areas of lighter yellow specify regions of high density.

IV. DATA PROCUREMENT

Due to the absence of a completely satisfactory set of background and explosion data drawn from the CTBT domain, this work relies on the simulation framework of Bellinger and Oommen [5] for the acquisition of experimental datasets. In addition, the utilization of simulated data facilitates our consideration of particular industrial and detonation scenarios in our exploration of the benefits of the expanded feature-space.

In our initial exploration of this PR task, four simulation scenarios were considered. Each scenario considered the affect of one, or more, industrial emitters on a single IMS.

The simulations framework applied is composed of two modules; the first module calculates the combined hourly impact of the set of industrial emission sources on the IMS over a user-defined time period. In accordance with the CTBT domain, the degree to which each source influences the total radioxenon concentration during a particular hour is dependent upon the atmospheric conditions that advected pollutants from the source. These conditions fluctuate over time based on user-defined statistics. Subsequently, the instances of the background class (labelled 0), measured over hour i, are written to a dataset in the following form:

$$\mathbf{x}_{i,0} = {}^{131}Xe_{i,0}, \; {}^{133}Xe_{i,0}, \; {}^{133m}Xe_{i,0}, \; {}^{135}Xe_{i,0}, \; 0. \quad (1)$$

The second phase generates the data for the detonation class (labelled 1). This is done by generating random (in time, space and magnitude) low yield explosions and measuring their impact on the IMS. The affect of the detonation is combined with that of the background source over the appropriate period of time, and written to the dataset with the detonation label. Therefore, a detonation instance measured over hour j, takes the following form:

$$\mathbf{x}_{j,1} = \mathbf{x}_{j,0} + \left\{ {}^{131}Xe_{j,1}, {}^{133}Xe_{j,1}, {}^{133m}Xe_{j,1}, {}^{135}Xe_{j,1}, 1 \right\}. \quad (2)$$

Similar to the Experiments in [2], an incremental detonation scenario was applied. This process highlights the affect of the distance between the IMS and the detonation on the overall classification results. Twenty-three incremental detonation ranges were defined with explosions beginning at a radial distance of $500km$ from the IMS, and extending out an additional $500km$. For each incremental experiment, where $0 > i > 24$, the radial range about the IMS is defined as $500 \times i : 500 \times i + 500$. Thus, during the second iteration, all detonations occurred in a radial range between $1,000km$ and $1,500km$.

V. EXPERIMENTAL DESIGN

In this section, we discuss both the details of our four simulated environments, and some issues pertinent to the classification experiments. The former topics are covered in Section V-A, and the latter are included in Section V-B.

A. Simulation Scenarios

Each of the four simulation scenarios were designed to explore the benefits of the expanded feature-space under a diverse set of atmospheric conditions. We hypothesize that the fundamental strength of the airshed indicator on classification is an increased ability to detect explosions. In particular, we surmise that a detonation's radioxenon signature, which appears to have resulted from an industrial source, but arrived at the IMS via an alternate airshed, becomes more easily identifiable.

As a result, the experiments focus on four wind environments, with the fundamental difference being the distribution of the wind directions. The four wind roses in Figure 4 illustrate the various wind environments. Experiment 1 (E1) and Experiment 2 (E2) describe simple wind environments with uni-modal Gaussian distributions, in which the mean winds
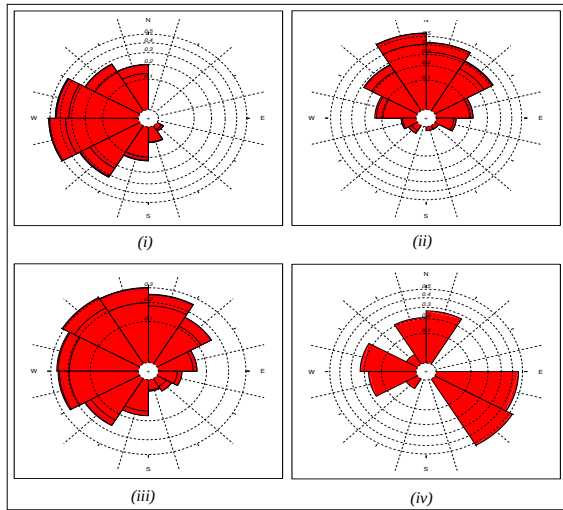
Fig. 4. This figure displays the wind roses for each of the four simulations.
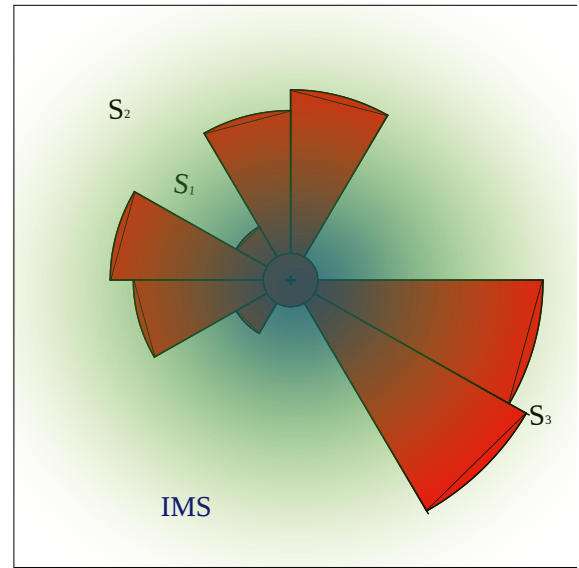

Fig. 5. This figure corresponds to the fourth experiment. It depicts the wind rose superimposed on the relative positions of the IMS and the three industrial sources.

are from the West and North, respectively. In both of these experiments the single industrial emitter is positioned due North of the IMS. Therefore, the background concentrations are generally much higher in E2, where the predominant wind was from the North, than in E1, where it was from the West.

Experiment 3 (E3) is, in essence, a combination of E1 and E2. In particular, E3 has two periods. During the first period, the mean wind is from the West, and in the second period, the mean wind is from the North. This explores the benefit of the airshed feature in a slightly more complex wind environment. The complexity is increased as a direct result of the affect of the wind on the background distribution. Specifically, the greater the number of possible wind events, the greater the range of radioxenon concentrations measured over time. Thus, the set of low-yield nuclear tests is more likely to be concealed by the background distribution.

Finally, the fourth experiment (E4) contains a single period, however, the distribution of the wind takes a multi-modal Gaussian form. In addition, this experiment includes three industrial sources (S_1, S_2, and S_3). The simulation scenario for E4 is depicted in Figure 5. This figure illustrates that the fourth experiment includes a significantly more complex environment, one in which both very high and low radioxenon concentrations can be found. In particular, a large portion of the wind is from the South-East, and a smaller portion is from the West. Under these circumstances the majority of the radioxenon emitted from the three industrial sources is dispersed away from the IMS. However, the simulation also includes a northerly wind, which transmits the emissions from S_1 and S_2, towards the IMS, thus, causing elevated concentrations.

B. Classification Details

From the standpoint of PR, verification of the CTBT is conceptually an OCC problem (for more details see [2]). However, in light of the fact that the verification of the treaty has generally been explored as a standard binary problem,

this paper explores the affect of the expanded feature-space in terms of both binary and OCC. Moreover, this work does not aim to provide a comprehensive analysis of all the possible PR strategies. Indeed, such an exhaustive attempt would be overly burdensome. Alternatively, we have selected five binary classifiers, which encompass the wide range of possibilities, and three OCC approaches in order to provide a concise, yet general, assessment.

Each classification algorithm included in this research has been implemented in the Weka machine learning environment [6]. In the interest of brevity, and due to the fact that the details of the binary learners are widely available in the literature, we merely mention the binary classifiers. Specifically, the Multi-layer Perceptron (MLP), the Support Vector Machine (SVM), the Nearest Neighbour (NN), the Naïve Bayes (NB) and the Decision Tree (J48) algorithms were utilized.

OCC is dissimilar to binary classification, as it does not assume the availability of a representative set of data drawn from both the positive and negative class. More specifically, it is assumed that a representative set can only be acquired from the positive class. Thus, in OCC, the learner must rely on instances drawn from a single class in the derivation of a discriminant function. Due to the fact that the OCC techniques applied in this study have received less consideration, we provide a few details in the following paragraphs.

Hempstalk *et al.*, in [7], converted the one-class classification problem into binary tasks by estimating the distribution of the concept class and generating instances of the non-concept, accordingly. Finally, a standard binary classifier is trained. This process has been denoted the Combined Probability and Density Estimator (PDEN).

Alternatively, the one-class Nearest Neighbour (ocNN) algorithm [8] learns a *target rejection rate*, τ, where τ is

the distance between the two nearest neighbours with the greatest separation in the training data. Subsequently, all novel instances whose nearest neighbours are at greater distances than τ are classified as outliers. We have additionally utilized a modified version of the ocNN in Weka, and denoted it as the scaled ocNN (socNN) [2]. Contrary to the ocNN, the socNN classifier is capable of learning a model that accounts for noise in the training set, based on a user-defined parameter, ϵ.

In each of the following experiments, we rely on the AUC in our performance assessments. A total of 230 datasets were produced for each experiment, which are sub-categorized into ten sets for each detonation range. This sub-categorization enables the consideration of performance as a function of distance, which resulted in Figure 6, for example. In addition, we consider the overall performance levels of each classifier according to the ensemble mean AUC over the twenty-three detonation ranges.

VI. EXPERIMENTAL RESULTS

A. Overview

The results included in this paper can be divided into four categories. Each category relates to a specific wind scenario, as described in Section V-A. The combined objective of these experiments is to *a)* motivate an expansion of the standard CTBT feature-space to include atmospheric indicators, such as a source airshed estimate, *b)* witness the affect of increasingly complex environments on classification in the expanded feature-space, and *c)* explore some possible forms of the new feature, with a consideration of how noise might influence the results.

In the following sub-sections, we, therefore, present similar results for each of the four experiments. In particular, each section includes a results table that specifies the ensemble mean AUC for each classifier over the twenty-three radial detonation ranges, based on the standard feature-space and the expanded feature space.

In terms of the expanded feature-space, we explore three scenarios. In the first scenario, the new feature can take any value between 0 and 2π. This continuous range of values specifies the direction, in radians, to the source airshed. In addition, the experiments were executed with the new feature taking on one of eight discrete values. Therefore, each discrete value specifies a continuous range of possible directions to the source airshed. The discretization process applied to transform the continuous airshed feature to a discrete form is commonly referred to as *Equal weight interval* (EWI) binning. Finally, we recognized the risk of error when estimating the source airshed. Therefore, we report the classification results based on training sets in which thirty percent of the instance in the new feature were randomized.

Our assessment of these results is later discussed in Section VII.

B. Experiment 1

As previously indicated, the first experiment was characterized by a low background concentration of radioxenon. In

TABLE I
THIS TABLE DISPLAYS THE ENSEMBLE MEAN AUC RESULTS FOR THE FIRST EXPERIMENT. THE RESULTS IN THE FIRST COLUMN CORRESPOND TO CLASSIFIER PERFORMANCE IN THE STANDARD CTBT FEATURE-SPACE. FURTHER DETAILS ABOUT THE ENTRIES IS INCLUDED IN THE TEXT.

	Standard	Continuous	Discrete	Noisy
NB	0.885	0.880	0.878	0.867
MLP	0.948	0.936	0.878	0.840
NN	0.896	0.858	0.924	0.931
SVM	0.553	0.691	0.682	0.562
J48	0.839	0.934	0.924	0.896
PDEN	0.702	0.837	0.827	0.822
ocNN	0.539	0.549	0.596	0.544
socNN	0.792	0.891	0.884	0.848

particular, the background concentration of ^{131}Xe at the IMS was $0.049 Bqm^{-3}$, with a standard deviations of 0.178. This was specific due to the fact that the mean wind direction was perpendicular to the industrial source and IMS. As a result, the wind was seldomly oriented such that any of the emitted radioxenon dispersed passed the IMS. Consequently, most low-yield detonations that affected the IMS were distinguishable from the background levels.

This is confirmed by the results reported in Table VI-B and Figure 6. Table VI-B demonstrates that, with the exception of SVM and ocNN, all of the classifiers achieved strong AUC results within the standard four-dimensional feature-space.

Even with the strong results produced within the standard feature-space, however, six of the eight classifiers improved with the expanded feature-space. Furthermore, each of SVM, PDEN, J48 and socNN classifiers improved by at least 0.1. Finally, when noise was added to the airshed feature in the training set, the overall results still represented an improvement.
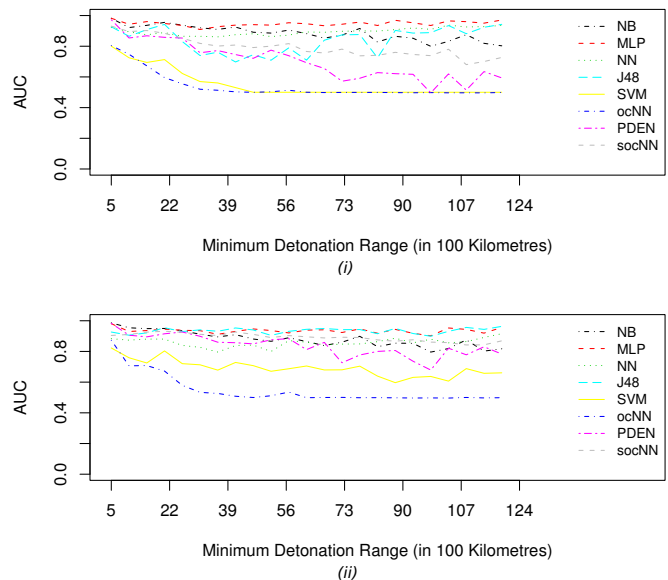

Fig. 6. Classifier performance as a function of Distance, on the first experiment, with the standard (i), and expanded (ii) CTBT feature-space.

The functions in Plot (i) of Figure 6 can be subdivided

	Standard	Continuous	Discrete	Noisy
NB	0.753	0.779	0.797	0.780
MLP	0.822	0.861	0.819	0.807
NN	0.714	0.806	0.820	0.751
SVM	0.521	0.583	0.601	0.523
J48	0.617	0.879	0.797	0.611
PDEN	0.431	0.432	0.656	0.621
ocNN	0.532	0.556	0.603	0.531
socNN	0.614	0.716	0.722	0.677

into two categories of varying clarity. The NB, PDEN, SVM, ocNN and socNN classifiers have strong AUC scores for detonations that occur relatively close to the IMS. However, the results for PDEN, SVM, ocNN and socNN degrade sharply with distance. Alternatively, in Plot (ii), where the continuous airshed estimate feature was included, the degradations of PDEN, SVM and socNN are noticeably less, NB remains consistent, and ocNN improves slightly.

The plotted functions for the remaining classifiers all display a hull in their performances for detonations occurring approximately $4,500km$ away, with J48's hull being extremely accentuated. In the expanded feature-space, however, J48 demonstrates a significant improvement. The MLP classifier improves as well, however, the increase is only slight due to its admirable performance in the standard feature-space. Alternatively, NN appears slightly worse. However, based on the results reported in Table VI-B, it would likely improve if we were to plot the results for the discretized airshed feature as a function of distance.

C. Experiment 2

Although the industrial source is positioned at a significant distance from the IMS, the fact that in E2 the mean wind blows directly from it to the IMS causes elevated radioxenon concentrations. In particular, the mean ^{131}Xe at the IMS was $0.264Bqm^{-3}$, with a standard deviation of 0.368.

The high mean and larger standard deviation suggest that when considered based on the radioxenon levels alone, more simulated detonations will be misclassified. This is, indeed, depicted by the lower results in Table VI-C, and the deeper performance hulls in Plot (i) of Figure 7.

In addition, Table VI-C illustrates the benefits of the continuous airshed feature when utilized in more challenging environments. Specifically, all eight classifiers are improved within the expanded feature-space, with NN, J48, PDEN, ocNN and socNN improving by at least 0.1.

Based on the noisy training set, five of the eight classifiers maintain their superiority in the expanded feature-space, and the remaining three degrade only slightly. It is, of course, worth recalling that thirty percent of the airshed estimates in the noisy training sets were randomly reassigned. This is, indeed, a high error rate, thus, the performance gains are impressive.
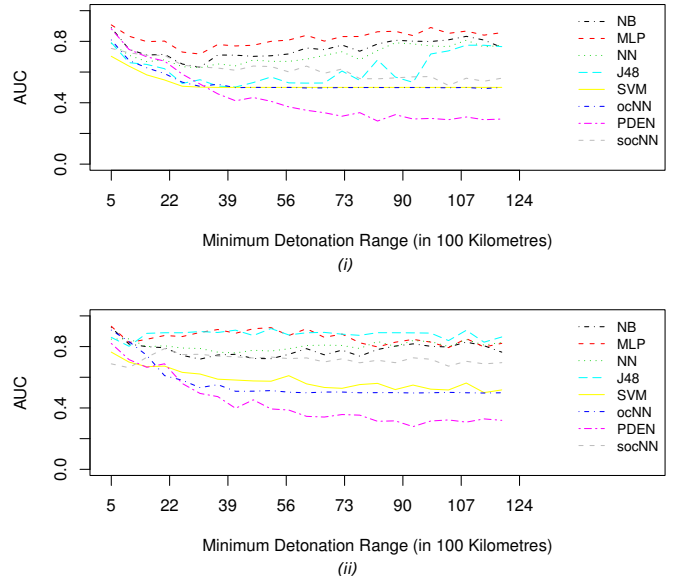

Fig. 7. Classifier performance as a function of Distance, on the second experiment, with the standard (i), and expanded (ii) CTBT feature-space.

	Standard	Continuous	Discrete	Noisy
NB	0.792	0.786	0.786	0.781
MLP	0.881	0.891	0.805	0.789
NN	0.791	0.809	0.843	0.833
SVM	0.527	0.572	0.569	0.528
J48	0.682	0.892	0.886	0.723
PDEN	0.465	0.564	0.651	0.630
ocNN	0.534	0.556	0.576	0.534
socNN	0.679	0.774	0.770	0.735

Similar to E1, and according to Figure 7, J48 and socNN benefit most from the new feature, but with significantly more impressive increases. Indeed, both of these classifiers become comparable with MLP, the top classifier. Alternatively, PDEN, ocNN and SVM remain well below the other classifiers.

D. Experiment 3

In order to explore the affect of a more diverse set of wind directions on the classification results, E3 simulated an environment with two periods. During the first period, the meteorological conditions were similar to those in E1. Alternatively, the conditions during the second period were similar to those in E2. Thus, half of the time the wind travelled perpendicular to the industrial source and IMS, and during the other half, it travelled from the source to the IMS.

These simulation parameters produced a mean ^{131}Xe concentration of $0.172Bqm^{-3}$, with a standard deviation only slightly lower than that in E2 of 0.329.

The benefits of the extended feature-space are similar in this experiment to those in E2. With the exception that NB stands out as it does not improve as a result of the airshed feature.

In addition, fewer of the classifiers perform better with the discretized airshed feature, and six out of the eight maintain their superiority to the standard feature-space when trained on the noisy data.
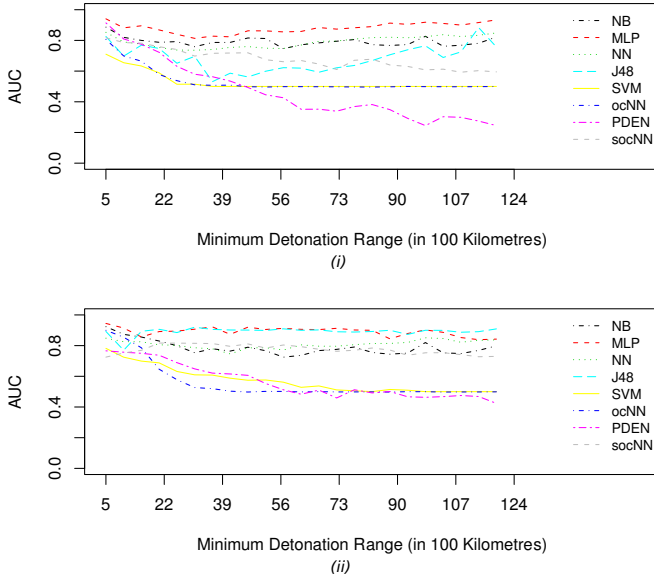

Fig. 8. Classifier performance as a function of Distance, on the third experiment, with the standard (i), and expanded (ii) CTBT feature-space.

As in E2, J48 demonstrates impressive performance when the feature-space is expanded. Likewise, MLP improves significantly. However, their performance curves are not as flat, nor as close to an AUC of 1.0 in the previous experiment. Similarly, socNN follows previous trends and improves from the expanded feature-space, and NB and NN classifiers remain strong, however, their plots increase only slightly.

E. Experiment 4

This fourth and final experiment represents the most diverse of the set. While the simulation only involves a single period, its winds were characterized by a multi-modal Gaussian distribution, creating a diverse set of background scenarios at the IMS. During the most extreme of cases, emissions from two of the three industrial sources were advected towards the IMS. However, through much of the simulation, the majority of the emissions were dispersed away from the IMS. As a result, the mean and standard deviations are, respectively, quite high. Specifically, the mean concentration of ^{131}Xe was $1.871 Bqm^{-3}$, with a standard deviation of 0.352.

The ensemble mean AUC results for this experiment are listed in Table VI-E. These results indicate that the classifiers, when operating within the standard feature-space, are significantly weakened by the challenging wind scenario detailed in this experiment.

However, substantial improvements in the ensemble mean AUC metric are obtained when the airshed estimate feature is included. Indeed, all eight classifiers achieve superior results in the expanded feature-space, with NB, SVM and J48 improving by nearly 0.2, and the remainder improving by at least 0.1.

TABLE IV
THIS TABLE DISPLAYS THE ENSEMBLE MEAN AUC RESULTS FOR THE FOURTH EXPERIMENT. THE RESULTS IN THE FIRST COLUMN CORRESPOND TO CLASSIFIER PERFORMANCE IN THE STANDARD CTBT FEATURE-SPACE. FURTHER DETAILS ABOUT THE ENTRIES IS INCLUDED IN THE TEXT.

	Standard	Continuous	Discrete	Noisy
NB	0.588	0.731	0.775	0.746
MLP	0.690	0.798	0.856	0.807
NN	0.660	0.831	0.831	0.704
SVM	0.527	0.727	0.725	0.531
J48	0.577	0.865	0.809	0.543
PDEN	0.579	0.576	0.747	0.707
ocNN	0.557	0.600	0.696	0.549
socNN	0.607	0.676	0.708	0.668

Moreover, six of the eight classifiers maintain their superiority when learning is performed on the noisy training set. The degradation of SVM and ocNN, the two classifiers that fail to maintain their superiority, is only slight in comparison to their original results. Performance in this category is, once again, of significant interest, given the exceptional degree to which the airshed feature has been randomized in the noisy training set.
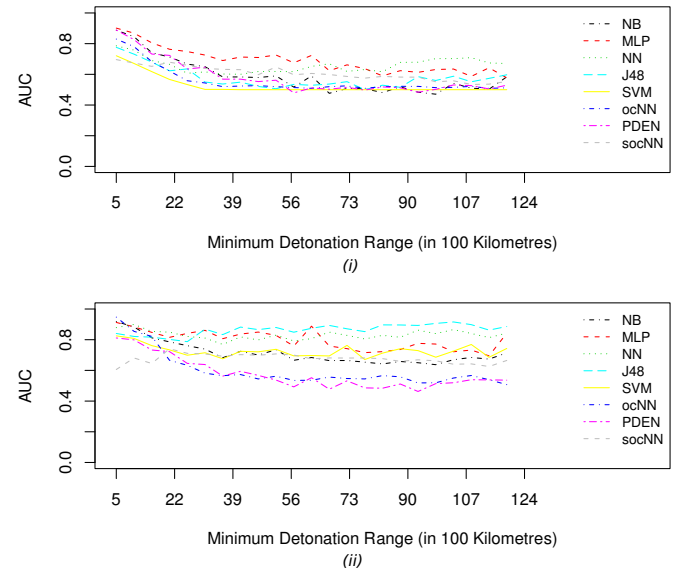

Fig. 9. Classifier performance as a function of Distance, on the fourth experiment, with the standard (i), and expanded (ii) CTBT feature-space.

When performance is plotted as a function of distance in Figure 9, the trend of the top classifiers is contrary to the results we have previously seen. In particular, all of the classifiers have similar performance curves, which are initially strong, but degrade quickly as the radial detonation range is pushed farther from the IMS. Beyond the $4,000km$ mark, performance generally flattens out. However, J48 and NN regain some of their losses.

Consistent with our previous results, adding the airshed estimate leads to significant improvements in the AUC. Indeed, MLP, J48 and NN maintain an AUC at or above 0.8. In general, J48 and NN are respectively the top classifiers, surpassing MLP in the expanded feature-space. Furthermore, their AUC

scores improve slightly with distance. Finally, socNN, NB and SVM improve to near 0.7.

VII. Discussion

Based on the above experiments, the possible improvements offered by the expansion of the CTBT feature-space to include an estimate of the source airshed looks promising. A general examination of these results reveals that the AUC metric was increased in 29 of the 32 classification experiments when the feature-space was extended. Moreover, the benefits of the new feature persisted even when a considerable amount of error was infused into the training set. In particular, 23 of the 32 classifications results maintained their superiority when trained on the noisy data.

When the results are explored as a function of distance, it becomes clear that the J48 decision tree classifier benefited the most from the additional feature. Figure 7, for example, illustrates that the new feature facilitated the mitigation of the performance hull present in the results produced on the standard feature-space. In general, the figures demonstrate that the most significant improvements are achieved on the more challenging experiments, E2 and E4. This is a likely result as there is more space for improvement on these tasks.

Interestingly, unlike the other classifiers, J48 consistently improves the most when the continuous airshed feature is added. In considering this point, we note that being a decision tree, J48 has a discretization procedure incorporated in the algorithm [9], thus suggesting that the discretization process applied by J48 is superior to the simple EIW binning that we have applied. Indeed, it has been observed that the EIW process is vulnerable to outliers that skew the range [10]. Moreover, it is probable that valuable information is lost when putting instances that are strongly related to different classes into the same bin [11]. Therefore, further consideration of the most appropriate discretization process is required in order to maximize the classification results.

Alternatively, ocNN, NN and NB achieve more success with the discretized airshed feature. This raises further questions about how significantly they may be improved with a sophisticated discretization process.

Nonetheless, the overall results of this experiment are very favourable. Indeed, they indicated that further, and more empirical, research on the incorporation of meteorological indicators into the CTBT domain, may produce further improvements in the detection of low-yield nuclear tests.

VIII. Conclusion

Verification of the Comprehensive Test-Ban-Treaty (CTBT), as a Pattern Recognition (PR) problem, has been proposed based on four radioxenon features. It has been noted, however, that in many cases this limited feature set is insufficient to distinguish radioxenon levels affected by an explosion from those that are exclusively the byproducts of industrial activities. As a means of improving the detectability of low-yield clandestine nuclear explosions, this paper has motivated the inclusion of meteorological indicators in the CTBT feature-space, and promoted further research into which meteorological indicators are the most informative, and how they may be acquired.

In doing so, we have contrasted classification results produced in the standard feature-space with those produced in an expanded feature-space, based on four simulated scenarios. Our results demonstrated that the inclusion of an airshed estimate feature significantly increases the prospect of classifying challenging detonation events, thus, suggesting the predictive power of meteorological features in general. More specifically, in 26 of the 32 classification experiments, better ensemble mean AUC results were produced in the expanded feature-space. Moreover, even in scenarios where a significant amount of error was added to the new feature, its inclusion was highly favourable.

While questions remaining about the exact form that an airshed estimate feature should take, and how best to acquire the estimate, this paper clearly demonstrates the rich possibility for the CTBT verification, offered by meteorological features, in general, and airshed estimates in particular.

References

[1] 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 Isotpes*, 2010.

[2] C. Bellinger, "Modelling and classifying stochastically episodic events," Master's thesis, Carleton University, Ottawa, Ontario, 2010.

[3] 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*, vol. 90, pp. 2393–2401, 2009.

[4] G. Wotawa, L. De Geer, P. Denier, M. Kalinowski, H. Toivonen, R. D'Amours, F. Desiato, J. Issartel, M. Langer, P. Seibert *et al.*, "Atmospheric transport modelling in support of CTBT verification–overview and basic concepts," *Atmospheric Environment*, vol. 37, no. 18, pp. 2529–2537, 2003.

[5] 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.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[7] 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*, vol. 5211. Springer, 2008, pp. 505–519.

[8] P. Datta, "Characteristic concept representations," Ph.D. dissertation, Irvine, CA, USA, 1997.

[9] J. R. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann Publishers, 1993.

[10] J. Catlett, "Megainduction: Machine learning on very large databases," Ph.D. dissertation, University of Sydney, 1991.

[11] R. Kerber, "Chimerge: Discretization of numeric attributes," in *Proceedings of the Tenth National Conference on Artificial Intelligence*. AAAI Press, 1992, pp. 123–128.