# Discriminative potential of ion mobility spectrometry for the detection of fentanyl and fentanyl analogues relative to confounding environmental interferents†‡

Thomas P. Forbes*, Jeffrey Lawrence, Jennifer R. Verkouteren, and R. Michael Verkouteren

*National Institute of Standards and Technology, Materials Measurement Science Division, Gaithersburg, MD, USA*

* Corresponding author*: E-mail: thomas.forbes@nist.gov; Tel: 1-301-975-2111*

KEYWORDS: Ion mobility spectrometry; Illicit narcotics; Fentanyl; Synthetic opioids; Receiver operating characteristic curve; Environmental background;

---

## Abstract

The opioid crisis and emergence of fentanyl, fentanyl analogues, and other synthetic opioids has highlighted the need for sensitive and robust detection for interdiction at screening points, notably vehicles at border crossings and packages at postal facilities. This work investigates the discriminative potential, sensitivity and specificity, of ion mobility spectrometry (IMS) for the detection of fentanyl and fifteen (15) fentanyl-related compounds (analogues, other opioids, and metabolites) relative to confounding environmental interferents. The environmental background interferent levels, frequency and intensity, were derived from over 10,000 screening samples collected from delivery vehicles entering a federal site. A receiver operating characteristic (ROC) curve methodology was employed to quantify the relationship between sensitivity and specificity for these target compounds on two instruments/configurations. These instrument configurations

differed in desorption and drift tube temperatures, reactant ion dopant chemistry, and analysis time. This work identified reduced mobility areas of high interference that resulted in increased false positive rates (FPR), effectively reducing sensitivity (true positive rate: TPR) in those regions. Except for a few target compounds on either of the instruments that exhibited elevated FPRs, detection of fentanyl and fentanyl-related species was achieved at single to tens of nanograms with ≥ 90% TPR and ≤ 2% FPR. This work established the importance of systematic environmental background characterization at each specific screening setting in evaluating a platform's true performance.

---

## Introduction

The ongoing opioid epidemic has necessitated the need for rapid and sensitive chemical detection of fentanyl and fentanyl analogues [1, 2]. The detection and identification of these species remains imperative for law enforcement and first responders, as well as customs and border protection, the military, and the postal service. International shipments of relatively pure fentanyl-related species are flooding the U.S. directly through international mail facilities (IMFs), overwhelming traditional screening processes. Approaches and technologies are greatly needed for rapid and high-throughput screening of not only the mail stream, but also vehicles entering at the border. According to the U.S. Bureau of Transportation Statistics, over 6 million trucks and nearly 77 million personal vehicles entered the United States through ports of entry along the southern border [3]. The screening difficulties of these high-throughput environments is further burdened by the more than 25 different fentanyl-related substances reported in seizures since 2015 [4]. In general, the expected levels of contraband contamination on the outside of vehicles or packages containing illicit substances is unknown. However, screening of surfaces in forensic laboratories that perform analyses on seized drug samples has yielded levels up to 100 ng $cm^{-2}$, even on nominally clean surfaces [5].

Ion mobility spectrometry (IMS) is a robust fieldable technology that has been widely deployed worldwide [6, 7] for threat mitigation and the trace detection of contraband materials, specifically, explosives [8-12], illicit narcotics [13-18], and chemical warfare agents [19-21]. IMS has demonstrated the detection of nanograms levels of a wide range of fentanyl and fentanyl analogues [16-18]. However, its limitations in peak resolution [22] often hinders complete discrimination of various analogues from each other [16]. In most traditional studies, neat materials are used to determine the instrument sensitivity, with process blank controls identifying false positives. However, this may not be representative of instrument performance during field deployment in scenarios such as vehicle screening at border crossings or within a mail facility. The confounding effects of environmental background, interferents, and contaminants collected in the wipe-based sampling of dirty target surfaces may limit the effective sensitivity due to overlapping drift times with target species. These interferents may result in diminished specificity and false positive alarms.

Receiver operating characteristic (ROC) curves, introduced during World War II to evaluate radar signals, provide a valuable representation of the trade-offs between sensitivity and specificity of a detection technology [23]. ROC curves are often used in clinical diagnostics [24-27], but have more recently been employed for the characterization of chemical sensors [23, 28-34]. The Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense has defined guidelines for the characterization of chemical sensors using ROC curves [34]. In the analysis of dichotomous systems, such as the use of IMS for the detection of illicit narcotics, ROC curves provide a graphical representation of the sensitivity (true positive rate, TPR) and specificity (1 - false positive rate, FPR). This analysis enables the evaluation of TPRs, FPRs, and appropriate alarm thresholds for a given screening environment and set of instrumental parameters.

The present study employed a ROC curve approach to characterize the discriminative potential of IMS for the detection of fentanyl and fifteen (15) other fentanyl-related compounds against confounding environmental interferents collected during screening of vehicles entering a federal government site. We evaluated instrument sensitivity (TPR) and specificity (1-FPR) under these conditions to supplement traditional controlled environment performance measurements, by characterizing the environmental background from a vehicle screening arena and identifying approximate mass loading necessary to achieve

assigned metrics of ≥ 90% TPR and ≤ 2 % FPR. We investigated two separate IMS instruments, specifically two configurations of the same model instrument, one commercialized for explosives detection and the other narcotics detection, evaluating their detection capabilities for high-throughput vehicle screening. The hardware of the instruments was identical, however there were differences in chemical dopants, drift tube and desorber temperatures, and sampling times. In many situations, IMS instrumentation intended for explosives detection is already deployed and in operation for screening purposes. The comparison of these two instrument configurations served to evaluate the relative performance and discriminative potential of the explosives-based configuration for narcotics detection. Each instrument was deployed for the screening of commercial delivery vehicles across a multi-month period, and then returned to a laboratory setting for additional target compound testing. Though drug detection was not the intent of the established screening activity, signal from each sample relevant to fentanyl-related detection and other narcotics was collected and archived. Signal intensities arising from environmental interferents were evaluated in the channels corresponding to the measured detection windows of these target compounds. The IMS response for a range of mass loadings for each target was directly compared against the background levels derived from over 10,000 archived sample files. ROC curves for all sixteen (16) target compounds were generated for both instruments. We identified drift time and reduced mobility areas of high background, with the potential to confound target detection. These measurements provided useful information for the selection of instrument parameters and implications for their discriminative potential for field applications, such as vehicle screening at ports of entry or package screening at international mail facilities.

## Experimental Methods

**Instrumentation.** Environmental background samples and targeted true positive samples were acquired on two separate instruments, each operated under a different set of conditions and incorporating different

reactant ion chemistries. Both instruments were dual tube ion mobility spectrometers (Ionscan 500 DT, Smiths Detection, Edgewood, MD, USA), with tubes 1 and 2 operated in positive (generally narcotics) and negative (generally explosives) polarity modes, respectively. The analysis conducted here focused on narcotics, detected in positive mode. The first instrument (SN: 52468AE) was operated under "AE" system parameters as demonstrated in Table 1. This instrument included isobutyramide for both the reactant ion chemistry and calibrant ($K_0$ = 1.495 $cm^2/sV$) for positive ion detection. The second instrument (SN: 50152N/E) was operated under "N/E" system parameters (Table 1) and used nicotinamide ($K_0$ = 1.86 $cm^2/sV$) for both the reactant ion and calibrant chemistries in positive mode. Sample introduction was achieved through the thermal desorption of wipe-based collections, followed by ionization via a $^{63}Ni$ radiation source and reactant ion chemistry.

**Table 1**. IMS instrument positive ion detection mode parameters.

| **IMS Parameter** | **AE** | **N/E** |
|---|---|---|
| Desorber temperature (°C) | 205 | 245 |
| Drift tube temperature (°C) | 220 | 244 |
| Inlet temperature (°C) | 285 | 265 |
| Drift flow ($cm^3$ $min^{-1}$) | 300 | 300 |
| Sampling time (s) | 5 | 8 |
| Total Segments (#) | 24 | 30 |
| Scans/Segment (#) (Period 1/Period 2) | 5 / 20 | 5 / 20 |
| Time (s) (Period 1/Period 2) | 2 / 3 | 2 / 6 |
| Calibrant (positive ion detection) | isobutyramide | nicotinamide |

**Sample and Data Collection.** Environmental background was measured through the screening of delivery trucks entering a federal facility. During the deployment of each instrument, no control over environmental or ambient conditions was applied, however, routine preventative maintenance was conducted according the manufacture guidance. Samples were collected on meta-aramid wipes (Nomex, Manual Swab, p/n 6822344-B, Smiths Detection, Edgewood, MD, USA) using standard swipe sampling methods [35] from a range of target locations, for example, door handles, steering wheels, or gear shifters. A total of 9,359 background samples were collected from the AE configured instrument over a year period (December 2016

through November 2017) followed by 1,996 samples from the N/E instrument (December 2017 through August 2018). There was a break in sample collection on the N/E instrument during the spring of 2018 for instrument maintenance.

Following the collection of environmental background and interferent samples, each instrument was relocated to a laboratory setting within the Advanced Measurement Laboratory at NIST. There, the IMS response to true positive samples of fentanyl, fentanyl analogues, select opioids, and metabolites was determined by direct solution deposition of known masses onto collection wipes. True positive IMS response measurements were conducted for a total of sixteen (16) analytes. Standard narcotic samples or solutions were purchased from Cerilliant (Round Rock, TX, USA) or Cayman Chemicals (Ann Arbor, MI, USA) and diluted in methanol to nominal concentrations of (1, 10, or 100) µg/mL depending upon desired mass deposition. For each target analyte and mass loading (Tables S1 and S2), ten to thirty replicates were conducted. Due to safety concerns, fewer replicates were conducted at the elevated mass loading for each (Tables S1 and S2) and fewer individual loading levels were considered for the more potent analogues (*e.g.,* three to five replicates for carfentanil (100× more potent than fentanyl) and *trans*-3-methylfentanyl (10× more potent than fentanyl)) [36].

**Data Processing.** Sample files from each instrument deployment and laboratory-based measurement set were archived and later retrieved for data processing and analysis. Individual raw data files (.pgr) for each sample collected during the deployment period, as well as each laboratory-based true positive test were exported. These files contained the signal intensity data as a function of drift time for each segment during a sampling period. It is important to note that the peak identification and detection/alarm algorithms of the IMS firmware were not used here. A custom MATLAB-based code (MATLAB, R2017a, Mathworks, Inc., Natick, MA, USA) was developed to extract and process the raw signal data for each file. Peaks identified within the target windows (± 0.003 $cm^2/sV$) that appeared in a minimum of two sequential segments were

flagged and the corresponding file number, sample date, drift time, reduced mobility, average intensity (across all segments), maximum intensity, and cumulative intensity logged. Additional details of the data processing can be found in the supporting information.†

**Data Analysis Methodology.** The study methodology and ROC curve analysis employed here was formulated based on guidelines from the DARPA report, "Chemical and Biological Sensor Standards Study"[34] and has been previously implemented for similar investigations [28, 30-32]. The DARPA report supports the acquisition of system response data for target compounds under laboratory conditions and environmental background data from relevant deployment location conditions. These data are then combined (as schematically represented in Scheme S1 - ROC curve creation), and the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) evaluated to determine the detection probability, or true positive rate (TPR=TP/(TP+FN)) and false positive rates (FPR=FP/(FP+TN)) at each alarm threshold (more detailed reviews in the literature [37]). The need to separate the analysis of target analytes from the envisioned deployment environment is founded in the difficulties with the release or deposition of hazardous or toxic chemicals. The hazardous nature of fentanyl and most of the analogues investigated here imposed the need for more rigorous safety protocols than immediately available at the deployed location. Specific safety considerations taken for the true positive experiments under laboratory conditions are described in detail below.

Under this experimental method, the system response (drift times and sensitivity) was assumed to be comparable across both experimental locations. Drift times were continuously calibrated and calibrant information was extracted from each sample file for processing. A cursory set of true positive heroin measurements (10 replicates at 3 mass loadings) were taken on the N/E instrument while deployed. The measured reduced mobility and signal intensities matched the values within the uncertainty (95% confidence interval) measured on that same instrument in the laboratory setting (Figure S1). While ambient

conditions introduced a significant difference between laboratory-based and deployed environments, the environmental background interferent peaks (frequency and intensity) exhibited no clear correlation with the seasons across the year-long deployment (Figures S2-S3). Previous work investigating a number of similar deployed instruments also observed little effect of environmental temperature and humidity on the system response [12]. In addition, for the construction of relevant ROC curves, all environmental background data collected at the deployed location were presumed to be true negatives. More specifically, these samples were defined as a combination of non-specific environmental interferent species, as well as the possibility of unknown environmental baseline levels of the target analyte, similar to the documented baseline levels of cocaine on U.S. currency [38-41]. This enabled the differentiation between fentanyl (or fentanyl-related species) and background detection levels. In addition, the background collection from delivery vehicle screening used in this study was meant to provide an estimation and imitation of port of entry checkpoint screening. However, given the nature of this scenario (i.e., delivery vehicles entering a federal facility), the assumption that fentanyl-related species were not actively smuggled through the checkpoint was believed to be reasonable. Any potential target species were therefore categorized as true negatives and represented a baseline background level.

**Safety Considerations.** Sample storage and preparation were conducted in accordance with best practices. Sample deposition was conducted in a fume hood with HEPA filtration. Nitrile gloves and safety glasses with side shields were worn [42]. Preparation materials, consumables, and used samples were disposed of in a dedicated waste container. Out of an abundance of caution, the IMS instruments were housed in a chemical fume hood for all true positive experiments. Given the hazardous nature of these materials, instruments were thoroughly cleaned and where appropriate, exposed tubes, gaskets, and filters were replaced based on manufacturer procedures, prior to removal from the hood and returning to general use.

## Results and Discussion

**IMS Response.** The true positive IMS response to fentanyl and fifteen (15) fentanyl-related species was measured across a range of mass loadings under laboratory conditions for both the AE (Figure 1 and Table S1) and N/E (Figure 2 and Table S2) instruments. Critical measurements include reduced mobility ($K_0$) (Table 2) and the sensitivity (Figures 1 and 2), which were used to set the location and width of detection windows, as well as to inform the selection of intensity thresholds for positive detection. In this study, we characterized the performance of two instrument configurations, one primarily optimized for explosives (AE) and the other, narcotics (N/E). This comparison provides vital data for numerous screening applications that may already have a fleet of instruments specifically optimized and deployed for explosives detection, generally considered the more critical threat.

Previous studies have demonstrated single to tens of nanogram sensitivities from neat samples for a range of fentanyl analogues and other opioids [16]. In this study, mass loading levels were chosen based on both previous sensitivity studies as well as the observed background levels within the target window during instrument deployment. Figures 1 and 2 provide box plot representation of the resulting IMS signal intensities for the mass loading investigated for each target compound (tabulated in Tables S1 and S2). Histograms of the IMS response for each mass and compound were also generated for both AE (Figure S4) and N/E instruments (Figure S5) for use in the ROC curve analysis below. Differences in sensitivity were observed between the AE and N/E instruments as would be expected given the differences in system parameters (Table 1). In general, the N/E instrument and configuration, which was optimized and marketed for narcotics, provided superior sensitivity from neat samples, in line with previous investigations. The lower desorber temperature and shorter analysis time on the AE instrument led to incomplete desorption of the target analytes (Figure S6). The fentanyl analogues and related species investigated here demonstrated temporal distributions that had just reached, or had yet to reach, peak intensity (Figures S6 and S7) on the AE instrument, a trend that was observed across the range of masses investigated. For most of the target compounds, the peak reached maximum intensity near the end of the 5 s sampling period. The more

complete thermal desorption, from higher temperatures and longer analysis times, aided the sensitivity improvements exhibited by the N/E instrument (Figure S8). At the elevated desorption temperature, the maximum peak intensity was observed between 2 s to 4 s of the 8 s sampling period on the N/E instrument. An evaluation of the reactant ion chemistry differences was not completed here but may also yield differences based on ionization pathways and efficiencies.

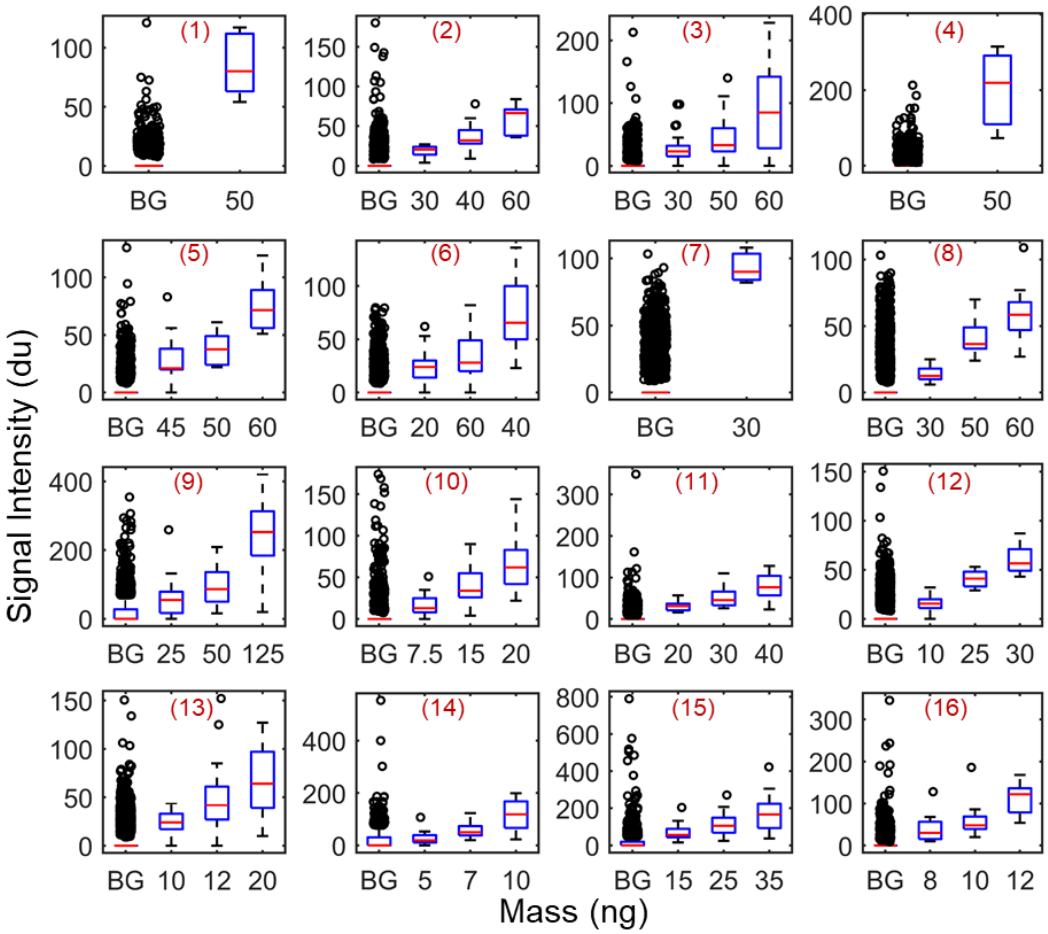


**Figure 1**. Box plots showing signal intensity for nominal mass loadings for each compound under laboratory conditions for the AE instrument. Environmental background (BG) interferent peak intensity levels during their deployments for target analyte drift time windows are displayed for direct comparison. Boxes represent the median and lower and upper quartiles, whiskers represent 1.5× the interquartile range (length of the box), and outliers (o) represent values out of the whisker range. Numerical values provided in Table S1. Analyte labels (#) correspond to identifications in Table 2.

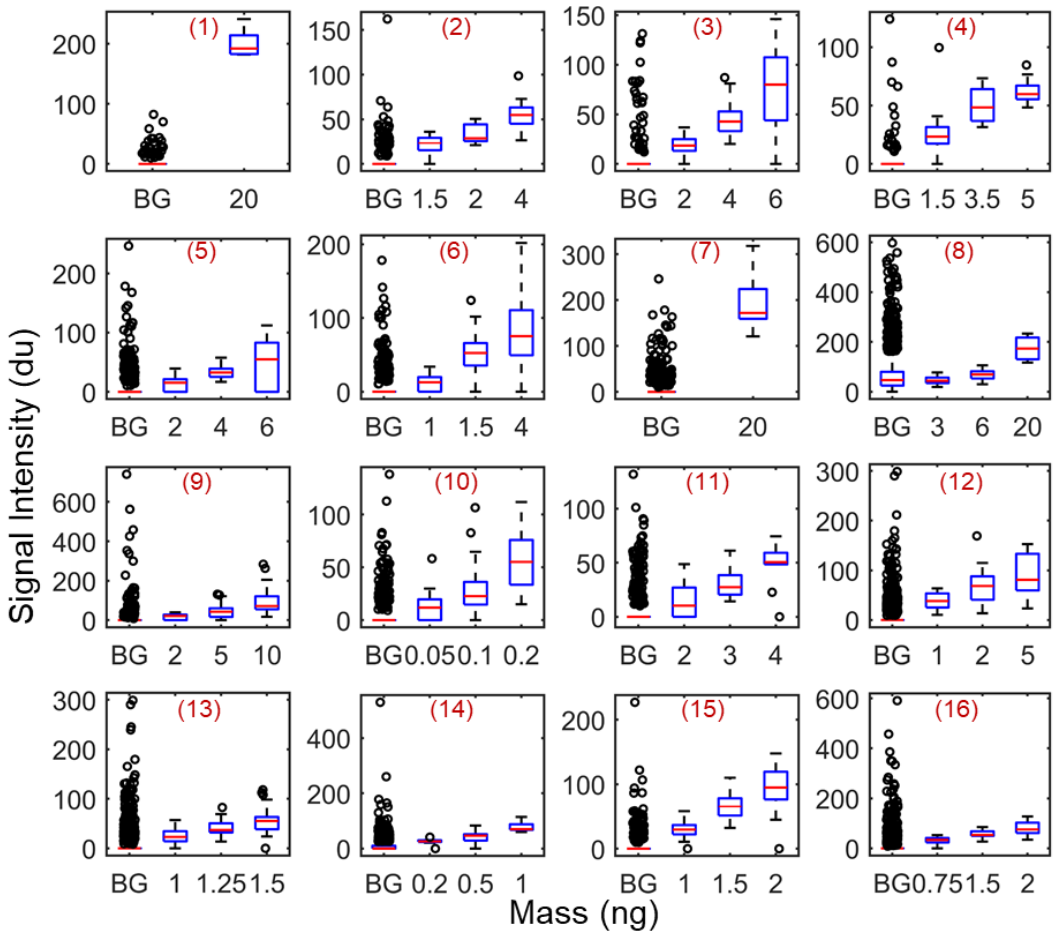


**Figure 2**. Box plots showing signal intensity for nominal mass loadings for each compound under laboratory conditions for the N/E instrument. Environmental background (BG) interferent peak intensity levels during their deployments for target analyte drift time windows are displayed for direct comparison. Boxes represent the median and lower and upper quartiles, whiskers represent 1.5× the interquartile range (length of the box), and outliers (o) represent values out of the whisker range. Numerical values provided in Table S2. Analyte labels (#) correspond to identifications in Table 2.

The reduced mobility ($K_0$) for each target analyte was determined on both the AE and N/E instruments as provided in Table 2. Data processing and analysis of the raw signal was transformed between drift time and reduced mobility based on the extracted calibrant drift time and assigned reduced mobility

using, $K_0 = (K_0 t_d)_{cal}/t_d$, where "*cal*" refers to the calibrant for the specified instrument [43, 44]. The measured reduced mobilities differed slightly between instruments, attributed to the differences in drift tube temperature and gas composition of the two instruments. The reduced mobilities were consistent with those obtained in the literature and in related studies [16]. As inspection of the reduced mobilities in Table 2 indicates, a handful of the fentanyl analogues (*e.g.,* crotonyl and butyryl fentanyl or acetyl and benzyl fentanyl) were not resolvable for the ±0.003 cm$^2$/sV alarm window chosen, and therefore would not yield unique alarms for a library defined by these values. In addition to the sixteen reduced mobility values targeted, a combined heroin/fentanyl mixture peak was also considered. Fentanyl is commonly found in mixtures with heroin, and the presence of both narcotics has been shown to result in a shifted combined peak on these IMS instruments [16]. This combined peak for a 10:1 heroin:fentanyl mixture was observed at $K_0$ of 1.0379 cm$^2$/sV on the AE and 1.0479 cm$^2$/sV on the N/E instruments [16].

**Table 2**. Experimentally determined analyte reduced mobilities and percentage of background samples that exhibited corresponding peaks at that value (± 0.003 cm$^2$/sV) for fentanyl analogues and select related compounds on AE and N/E IMS instruments. Reduced mobility values represent the average and standard deviation across all replicates at all mass loadings for each compound.

| # | Analyte | Mass (g mol$^{-1}$) | $K_0$ measured (cm$^2$ V$^{-1}$ s$^{-1}$) | | Samples with Peak (%) | |
|---|---|---|---|---|---|---|
| | | | AE | N/E | AE | N/E |
| (1) | Carfentanil | 394.2 | 0.9762 ± 0.0003 | 0.9761 ± 0.0004 | 2.5 | 1.7 |
| (2) | Valeryl fentanyl | 364.3 | 0.9887 ± 0.0009 | 0.9900 ± 0.0024 | 3.5 | 2.9 |
| (3) | Furanyl fentanyl | 374.2 | 1.0021 ± 0.0006 | 1.0041 ± 0.0027 | 2.7 | 1.7 |
| (4) | *p*-Fluoroisobutyryl fentanyl (FIBF) | 368.2 | 1.0057 ± 0.0003 | 1.0070 ± 0.0021 | 3.2 | 1.0 |
| (5) | Crotonyl fentanyl | 348.5 | 1.0168 ± 0.0009 | 1.0208 ± 0.0022 | 3.3 | 6.0 |
| (6) | Butyryl fentanyl | 350.2 | 1.0205 ± 0.0007 | 1.0202 ± 0.0019 | 3.8 | 4.5 |
| (7) | *trans*-3-methylfentanyl | 350.2 | 1.0255 ± 0.0006 | 1.0212 ± 0.0012 | 6.6 | 7.1 |
| (8) | Cyclopropyl fentanyl | 348.5 | 1.0275 ± 0.0009 | 1.0284 ± 0.0027 | 9.7 | 79.0 |
| (9) | Heroin | 369.2 | 1.0311 ± 0.0004 | 1.0425 ± 0.0028 | 32.0 | 3.8 |
| (10) | Fentanyl | 336.2 | 1.0506 ± 0.0006 | 1.0531 ± 0.0023 | 2.4 | 5.1 |
| (11) | Acryl fentanyl | 334.2 | 1.0588 ± 0.0006 | 1.0614 ± 0.0035 | 3.3 | 5.7 |
| (12) | Acetyl fentanyl | 322.2 | 1.0807 ± 0.0006 | 1.0851 ± 0.0010 | 9.9 | 21.0 |
| (13) | Benzyl fentanyl | 322.2 | 1.0818 ± 0.0007 | 1.0833 ± 0.0018 | 13.5 | 17.4 |
| (14) | U-47700 | 328.1 | 1.0898 ± 0.0007 | 1.0937 ± 0.0017 | 41.0 | 25.2 |
| (15) | Norfentanyl | 232.3 | 1.2993 ± 0.0006 | 1.3084 ± 0.0025 | 26.7 | 5.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| (16) | Acetyl norfentanyl | 218.3 | 1.3548 ± 0.0012 | 1.3638 ± 0.0010 | 4.2 | 20.3 |

Of additional note, the $K_0$ reproducibility differed between the two instruments, with the standard deviation for the N/E instrument approximately double that of the AE instrument. Though the N/E configuration in general provided superior sensitivity (higher signal intensity for comparable masses), at the lower detectable masses, the observed distributions in reduced mobility were noticeably larger (Figure S9). As loadings approached the limits of detection, the uncertainty in the observed reduced mobility led to sample peaks falling outside the target window, which resulted in missed detections, yielding false negatives (FN). The increase in false negatives caused decreases in the true positive rate (TPR=TP/(TP+FN)) and effective sensitivity. In addition, the standard deviation in reduced mobility measurements may be used to inform the selection of appropriate detection windows for the target analytes, with smaller windows favored for improved selectivity as long as they capture the true peak at the required rate. It is important to note that some of the uncertainty in the IMS response was likely a result of the inherent difficulties in reproducibly pipetting small volumes. This uncertainty could be reduced by employing another deposition technique such as inkjet dispensing [45, 46], however, additional safety considerations related to potential aerosolization of these toxic materials must be considered.

**Environmental Background**. Here, we characterized the environmental interferent background observed for vehicle checkpoint screening at a federal facility. Interferent peaks within a ± 0.003 $cm^2/sV$ window at each of the sixteen reduced mobility locations in Table 2 were identified across 9,359 samples taken on the AE instrument and 1,996 samples taken on the N/E instrument, in the screening of incoming vehicles. The distribution of background peak intensity (signal intensity given in digital units (du)) for each target analyte is displayed in Figure 3. The distribution of background peak intensity was also represented as a function

of the reduced mobility and drift times of the target analytes (Figure S10) and histograms of interferent peak frequency as a function of peak intensity for use in the ROC curve analysis below (Figures S11 and S12). The overall percentage of background samples that exhibited interferent peaks in each of the specified target windows is summarized in Table 2 and correspondingly labeled for each analyte in Figure 3.

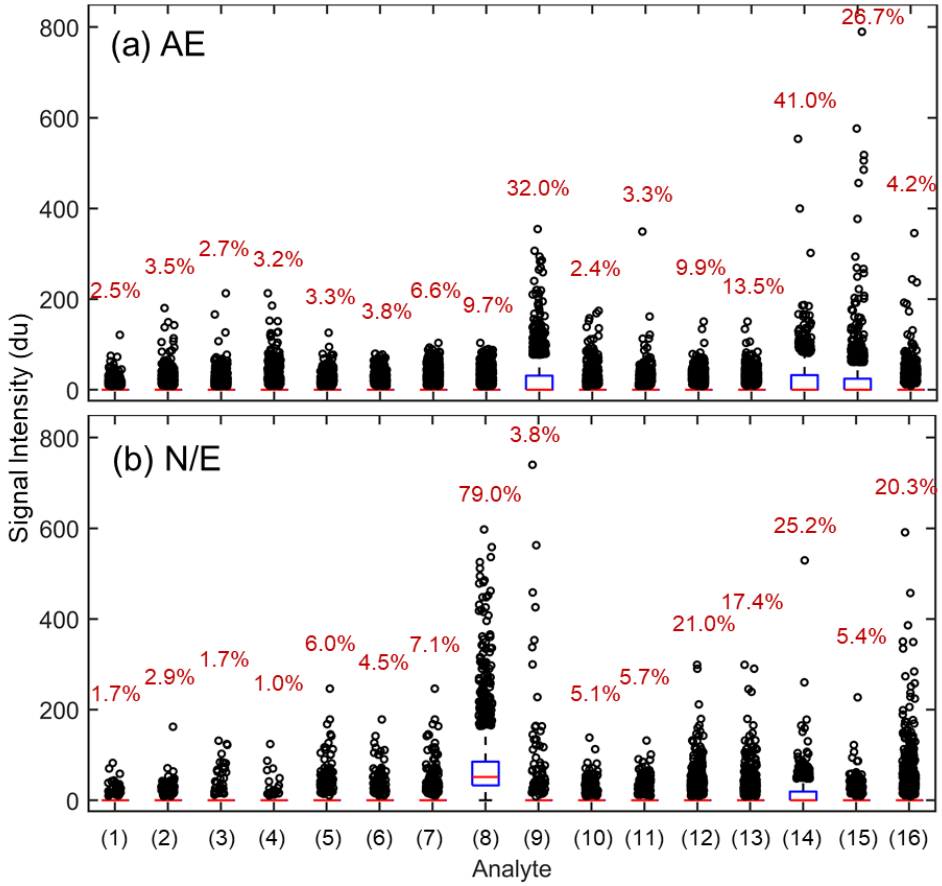


**Figure 3**. Box plots of environmental background interferent peak intensity levels for target analytes from (a) AE and (b) N/E instruments during their deployments. Boxes represent the median and lower and upper quartiles, whiskers represent 1.5× the interquartile range (length of the box), and outliers (o) represent values out of the whisker range. Analyte labels (#) correspond to identifications in Table 2. Labels represent the percentage of background samples that exhibited corresponding interferent peaks at each target reduced mobility value (± 0.003 $cm^2/sV$).

In general, the overall environmental background due to interferent peaks was not instrument dependent and consistent across both instruments, with shifts in drift times of associated interferents resulting from the differences in dopant chemistry and system parameters. However, this background distribution of interferent peaks was likely unique to this specific screening setting. Other screening scenarios, environments, and locations may (and most likely will) result in different background, interferent peaks, and distributions. Many of the fentanyl analogues exhibited similar background intensity distributions on the AE and N/E instruments (Figure 3). On both instruments, those compounds with relatively lower reduced mobilities (< 1.0275 $cm^2/sV$) demonstrated a low number of interferent peaks in background samples (Figure 3 and Table 2). The interferent peaks observed in this range were of relatively low intensity (*i.e.,* < 100 du). This range included carfentanil, valeryl fentanyl, furanyl fentanyl, FIBF, crotonyl fentanyl, butyryl fentanyl, and trans-3-methylfentanyl (Analytes #1 - #7). A similarly low rate of peaks was observed in the reduced mobility range around fentanyl and acryl fentanyl (1.05 $cm^2/sV$ to 1.06 $cm^2/sV$, Analytes #10 and #11). At the high end of the reduced mobilities, specifically, acetyl fentanyl, benzyl fentanyl, and U-47700 (#12 - #14), at around 1.08 $cm^2/sV$ to 1.09 $cm^2/sV$, peaks were identified in approximately 10 % to 40 % of the background samples with some variations between the two instruments. The high reduced mobility (low drift time) range for the two metabolites, norfentanyl and acetyl norfentanyl (#15 and #16), exhibited background peaks in a moderate number of the samples, with noticeable differences between the two instruments (Figure 3), as the interferent drift times shifted slightly (Table 2).

Notable high intensity background levels were observed near the heroin window ((Analyte #9) partial overlap) on the AE instrument and cyclopropyl fentanyl window ((Analyte #8) significant overlap) on the N/E instrument. The specific unidentified background contaminant resulting in these high background frequencies and levels was observed in approximately 89 % of all the vehicle screening samples taken across both instruments (Figures S11 and S12). The contaminant reduced mobility was measured at $K_0 \approx$ 1.0365 $cm^2/sV$ on AE and $K_0 \approx$ 1.0301 $cm^2/sV$ on N/E (Figure S13). The environmental contaminant was not observed in neat sample tests taken under laboratory conditions. However, similar peaks have been

observed in previous studies on dirt and artificial sebum matrices [16]. The role these interferent peaks played in the discriminative potentiation for each IMS instrument was evaluated using a ROC curve approach, as presented in the next section.

**Discriminative Potential – ROC curves.** The IMS response to the array of target analytes under both instrument configurations was directly compared to the environmental background interferents measured for the screening of vehicles. Figures 1 and 2 display not only the box plots of each mass loading, but the corresponding background interferent distributions from Figure 3. These figures provide a quick visual comparison of the expected IMS response for mass loadings and the expected confounding background levels. However, combining these data using a ROC curve methodology enabled a more quantitative comparison. In order to evaluate the discriminative potential of these systems for the detection of fentanyl compounds, which exhibited significant toxicity, the environmental interferent signals and target analyte signals were collected separately. To appropriately combine these datasets, we made the assumption that matrix effects, and specifically competitive ionization, would be minimal in a hypothetical sample of both together. Competitive ionization in a range of spectrometric techniques is widely documented and can play a significant role in signal suppression and other matrix effects [16, 17, 47]. The assumption of minimal interaction employed here was based on the temporal distribution (desorption time) of the interferent signals relative to the target compounds. For nearly all of the target compound reduced mobility windows, the maximum interfering background signal was predominately observed in the first couple seconds of the desorption period (Figures S7 and S8). However, the maximum true positive analyte signals were detected following the interferent signals – close to the end of the 5 s sampling period for the AE instrument (Figure S7) and mostly in the range of 2 s to 4 s of the 8 s sampling period for the N/E instrument (Figure S8), with its elevated desorption temperature. This temporal separation of the majority of interfering and target signals would minimize the gas phase interaction and potential competitive ionization.

The relative intensity distributions of the background and target analyte signals in overlaid histograms graphically represented the ability and confidence of a compound being detected above typical background (Figure 4). As the mass loadings increase, the distribution of target samples shifted to higher signal intensities, demonstrating increased discriminative potential from the background interferent signals. ROC curves were generated for three mass loadings of each analyte by varying the demarcation (or alarm threshold) between target identification and background clutter from 0 du to 600 du in 10 du increments (Scheme S1). At each signal intensity delineation, the TPR and FPR were determined and plotted (Figures S15 and S16).

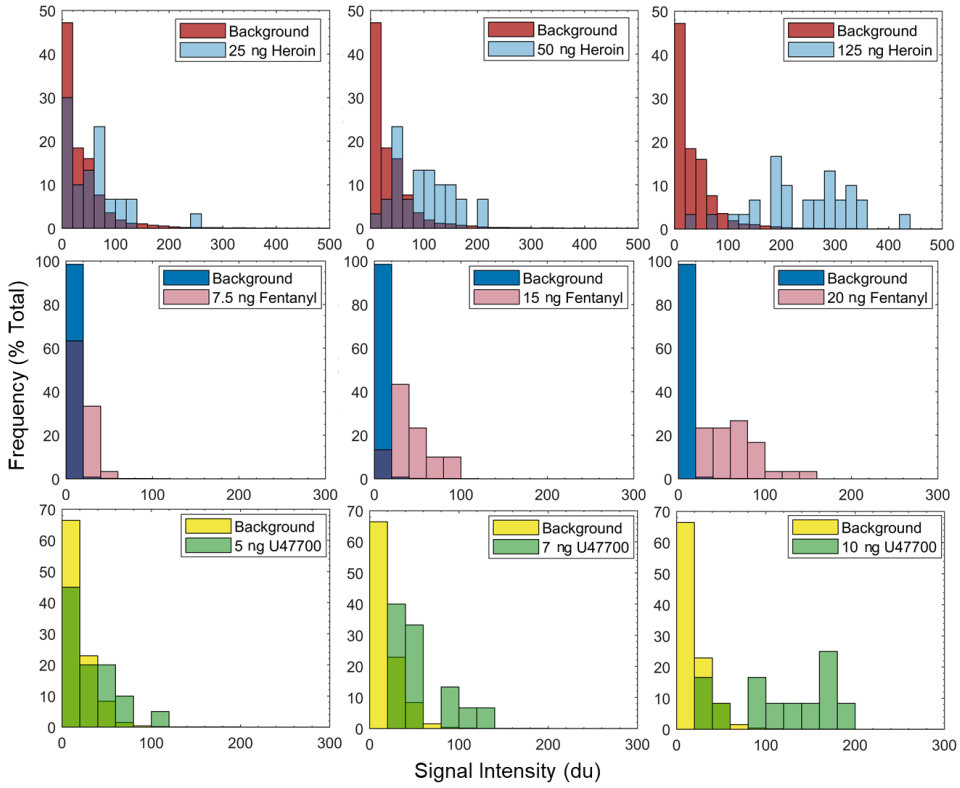


**Figure 4**. Demonstrative overlaid frequency histograms of background and target analyte intensity data for heroin, fentanyl, and U-47700 on the AE IMS instrument. Histograms represent the percent of total samples.

Figure 5(a) demonstrates traditional ROC curves (TPR vs FPR) for heroin, fentanyl, and U-47700 on the AE IMS instrument at masses of (25, 50, and 125) ng, (7.5, 15, and 20) ng, and (5, 7, and 10) ng, respectively. These ROC curves demonstrated the importance of understanding the expected confounding interferents for a deployed location in the evaluation of system performance and sensitivity for the accurate (true positive) detection of target species. In the reduced mobility (drift time) window for fentanyl, significantly fewer background samples exhibited peaks and at noticeably lower signal intensity than those detected in the heroin and U-47700 windows (Figure 3). In evaluating true positive and false positive rates (TPR and FPR), the background distribution significantly affected the discriminative potential. The results demonstrated a high level of discrimination between the TP fentanyl signals and interferent peak levels in the corresponding window. At the lowest level of fentanyl investigated (7.5 ng), an alarm threshold of 10 du yielded a 66% TPR and 2% FPR. Though, doubling the loading of fentanyl to 15 ng, resulted in a 93% TPR and 2% FPR for the same alarm threshold.

Even though the instrument response to heroin was comparable to fentanyl (20 ng fentanyl demonstrated ≈ 62 du (median) response compared to ≈ 55 du (median) for 25 ng heroin, see Table S1), the background in the heroin window restricted the ability to discriminate a true positive signal. As displayed in Figure 5(a), low(er) alarm thresholds yielded significant FPRs that would be a hinderance on throughput in a screening type arena. An alarm threshold of 100 du yielded only 17% TPR and 7% FPR for a heroin mass of 25 ng. However, for the same threshold, five times the loading (125 ng) of heroin achieved 90% TPR with corresponding 5% FPR. Figure 5 also displays the representative ROC curves for the synthetic opioid U-47700. Excluding the metabolites, U-47700 exhibited the highest reduced mobility (fastest drift time) of the opioids and fentanyl analogues investigated here. The sensitive IMS response of U-47700 directly competed with the moderate background observed for vehicle screening. An alarm threshold of 30 du yielded greater than 90% TPR for 10 ng loading, however, this led to a 23% FPR. Given the background distribution in the U-47700 window, a single digit FPR (5%) required a sacrifice in TPR (83% - alarm threshold 50 du) for 10 ng loading. However, pushing the mass loading up to the elevated 50

ng level enabled near complete discrimination from the background: 100 % TPR and 1 % FPR. For the evaluation and initial characterization of chemical sensors in screening applications, it may also be useful to plot the resulting ROC curve data in the form of target loading as a function of FPR for specified TPRs [34]. Figure 5(b) demonstrates the data from Figure 5(a) with additional mass loadings in this form. Here, each curve represents a TPR of 75%, 90%, or 95%. On the AE IMS instrument, for a desired TPR, the mass loading necessary for a target FPR can be directly extracted from the plots. ROC curves for all target species investigated on the AE instrument can be found in the supporting information (Figure S14).

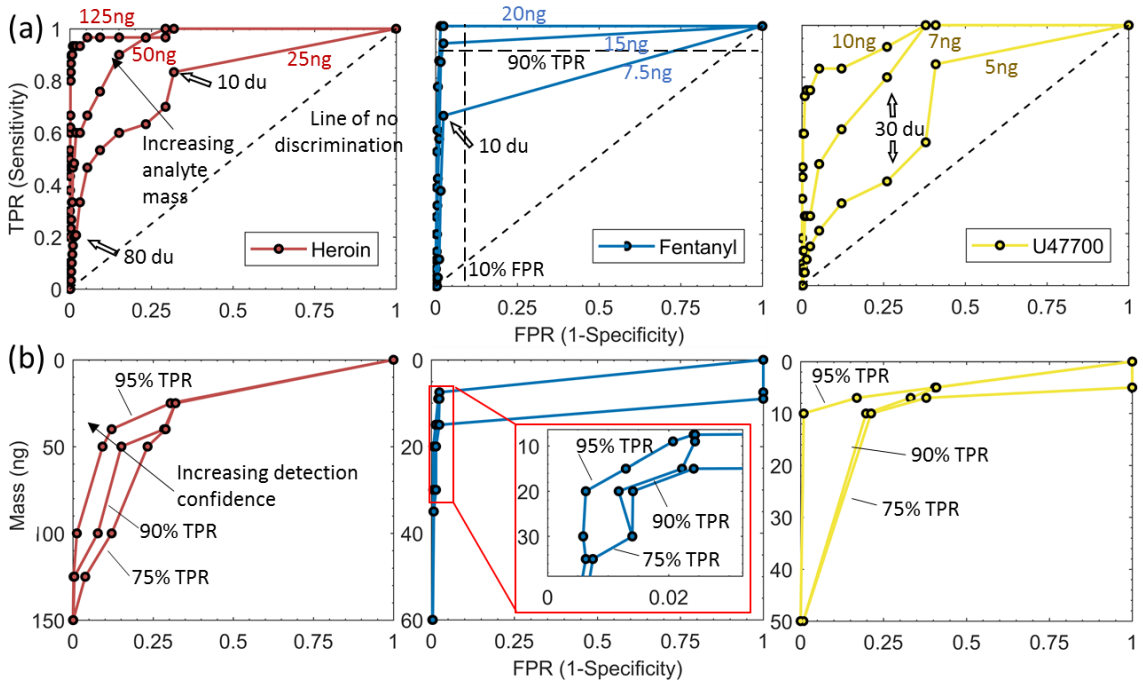


**Figure 5**. (a) ROC curves for heroin, fentanyl, and U-47700 at three mass loading levels (see Figure 2) for the AE IMS instrument. Nuisance/alarm thresholds were varied from 0 du to 600 du in 10 du increments. The line from (0,0) to (1,1) represents the line of no discrimination. Lines of 90% TPR and 10% FPR are labeled on the fentanyl ROC curve, as well as various alarm thresholds for reference. (b) Target mass as a function of FPR for heroin, fentanyl, and U-47700 at TPRs of 75%, 90%, and 95%.

Figure 6 demonstrates representative ROC curves for heroin, fentanyl, cyclopropyl fentanyl, furanyl fentanyl, *p*-fluoroisobutyryl fentanyl (FIBF), and U-47700 on the N/E IMS instrument (ROC curves for all species on the N/E can be found in the supporting information, Figure S15). Proper accounting for potential background interference was critical in evaluating the system performance. For example, previous published work on an N/E configured 500 DT IMS instrument reported single nanogram limits of detection for heroin [16]. Here, 2 ng heroin samples yielded only approximately 70% TPR and 4% FPR (Figure 6). However, a 10 ng sample loading of heroin was required to achieve 93% TPR with only 3% FPR. Similarly, the referenced study reported sub-nanogram limits of detection for the synthetic opioid, U-47700 [16]. However, against the environmental background experienced for the deployed location investigated here, a comparable mass loading resulted in an over 25% FPR (at 90% TPR). Many high throughput screening applications would be severely burdened by such a high rate of false positives. Increasing the U-47700 mass loading to a single nanogram level drastically reduced the FPR to 2%, while maintaining the targeted ≥ 90% TPR (Figure 6). As shown in Figure 6, only a 200 pg loading of fentanyl was necessary to discriminate from background with a greater than 90% TPR and 3% FPR. However, the significant interferent peak that was observed in the environmental background on both instruments (Figures S11 and S12) overlapped significantly with the cyclopropyl fentanyl window on the N/E instrument. This led to significantly higher rates of false positives and poorer discrimination between the fentanyl analogue and background interreference (Figure 6). The 20 ng cyclopropyl fentanyl samples yielded an approximately 15% FPR for assured detection.

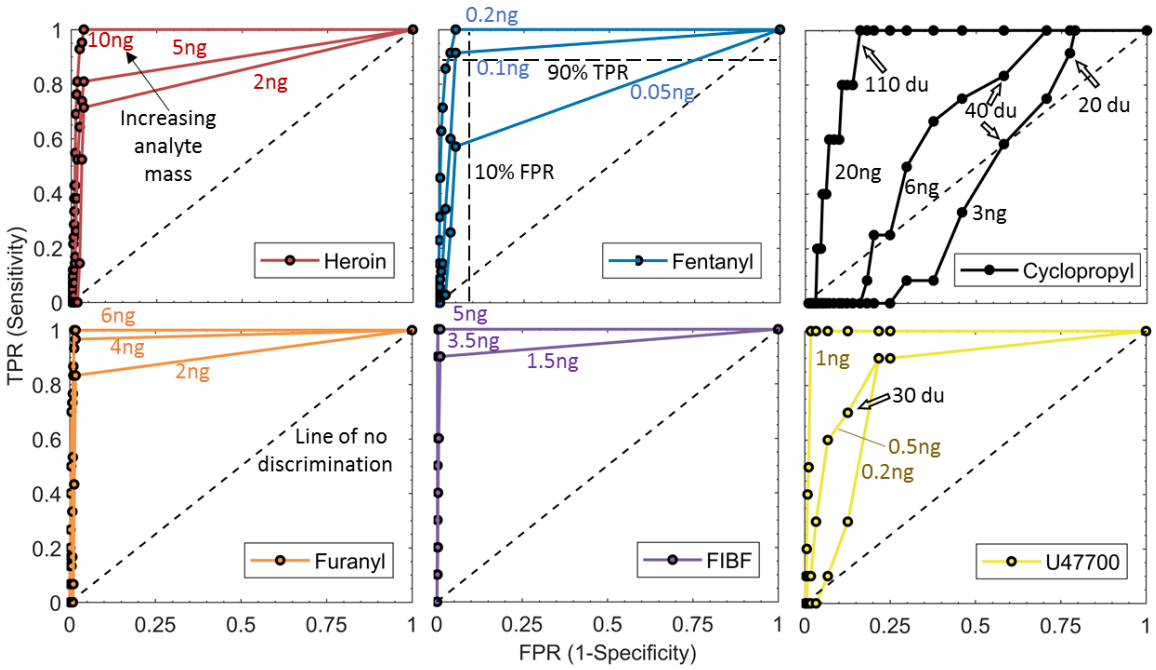


**Figure 6**. ROC curves for heroin, fentanyl, cyclopropyl fentanyl, furanyl fentanyl, *p*-fluoroisobutyryl fentanyl (FIBF), and U-47700 at three mass loading levels for the N/E IMS instrument. Nuisance/alarm thresholds were varied from 0 du to 600 du in 10 du increments. The line from (0,0) to (1,1) represents the line of no discrimination. Lines of 90% TPR and 10% FPR are labeled on the fentanyl ROC curve, as well as various alarm thresholds for reference.

In general, the AE and N/E instruments were able to discriminate the target analytes at the mass loadings investigated from the environmental interferent peaks obtained during vehicle screening. For all the results presented here, further increases in the mass loading and resulting target analyte signal, aided in improving discrimination, reducing the rate of false positives. The intensity of interferent peaks were largely less than 200 du (Figure 3). Typically, mass loadings of a few hundred nanograms yielded signals above this background. Again, ROC curves for all sixteen analytes measured on the AE and N/E instruments, corresponding to the mass loading in Figures 1 and 2, can be found in the supporting information (Figures S14 and S15).

These results and ROC curves evaluated the ability of IMS to discriminate fentanyl (and fentanyl-related compounds) from environmental background interferents experienced in vehicle screening. These

curves provided direct characterization of how the TPR and FPR vary for increasing mass loading as a function of the specific interferent distribution in each target reduced mobility window. From this analysis a critical value was estimated to achieve a specified TPR and FPR. The IUPAC (International Union of Pure and Applied Chemistry) specifies TPR and FPR of 95 % and 5 % for general critical value or minimum detectable value determinations.[48] However, specification of desired or required TPR and FPR is highly dependent on the application and may be based on the risk of missed detection (e.g., chemical warfare agent sensors may target very high TPR and very low FPR) or throughput requirements (e.g., checkpoint screening). For the screening of vehicles at a port of entry or other high throughput checkpoint and based on a number of proposed federal requirements, these IMS instruments were evaluated for ≥ 90% TPR and ≤ 2 % FPR. Table 3 tabulates the mass loading investigated here that achieved or approached the required TPR and FPR for each analyte on both the AE and N/E configured instruments. Apart from the region of elevated interferent signals in the cyclopropyl fentanyl and heroin reduced mobility ranges on each instrument, fentanyl and fentanyl analogues were detected in the single to tens of nanograms range with ≥ 90% TPR and ≤ 2% FPR.

In general, the N/E configured instrument demonstrated superior sensitivity and comparable selectivity, due to its higher desorption temperatures, alternative dopant chemistry, and longer analysis time (Table 3). However, the AE configured instrument still provided sensitive detection (tens of nanograms) of fentanyl and the fentanyl-related species as well as discrimination against the observed background interferences (generally < 2% FPR), with a couple exceptions. This comparison provided important information for a range of screening applications that may already have a fleet of instruments deployed for explosives detection. For example, the concept of operations for a specific setting might dictate instrument optimization for sensitive explosives detection, even at the expense of narcotics sensitivity. In addition, overall background levels of narcotics are expected to be higher than explosives (*e.g.,* cocaine on U.S. currency [39] or narcotics in airborne particles [38, 40, 41]), providing further justification for this trade-off.

**Table 3**. IMS TPR and FPR for fentanyl and related species, targeting ≥ 90% detection sensitivity (TPR) and ≤ 2 % FPR. Target analyte mass, TPR, and FPR are listed for both AE and N/E instruments.

| | Analyte | AE | | | N/E | | |
|---|---|---|---|---|---|---|---|
| | | **Mass (ng)** | **TPR (%)** | **FPR (%)** | **Mass (ng)** | **TPF (%)** | **FPR (%)** |
| 1 | Carfentanil | 50 | 100 | < 1 | 20 | 100 | < 1 |
| 2 | Valeryl fentanyl | 40 | 90 | 2 | 4 | 90 | 1 |
| 3 | Furanyl fentanyl | 60 | 90 | 2 | 6 | 90 | 1 |
| 4 | *p*-Fluoroisobutyryl fentanyl (FIBF) | 50 | 100 | < 1 | 5 | 90 | < 1 |
| 5 | Crotonyl fentanyl | 60 | 90 | < 1 | 20 | 100 | < 1 |
| 6 | Butyryl fentanyl | 40 | 90 | 1 | 20 | 100 | < 1 |
| 7 | *trans*-3-methylfentanyl | 30 | 100 | < 1 | 20 | 100 | < 1 |
| 8 | Cyclopropyl fentanyl | 60 | 90 | 3 | 100 | 100 | 1 |
| 9 | Heroin | 150 | 100 | < 1 | 20 | 90 | < 1 |
| 10 | Fentanyl | 20 | 90 | 1 | 20 | 100 | < 1 |
| 11 | Acryl fentanyl | 40 | 90 | 2 | 20 | 100 | < 1 |
| 12 | Acetyl fentanyl | 30 | 90 | 1 | 20 | 100 | < 1 |
| 13 | Benzyl fentanyl | 50 | 100 | < 1 | 20 | 100 | < 1 |
| 14 | U-47700 | 50 | 100 | 1 | 1 | 90 | 2 |
| 15 | Norfentanyl | 35 | 90 | 2 | 2 | 90 | < 1 |
| 16 | Acetyl norfentanyl | 12 | 93 | 1 | 20 | 100 | < 1 |

## Conclusions

The ROC curve framework employed here identified the reduced mobility areas of high interference that resulted in elevated false positive rates from a suite of sixteen fentanyl-related compounds. Similarly, this study identified the mass loading ranges that resulted in signal intensities (above background interferent signals) to achieve desired FPR (≤ 2 %). Nevertheless, an understanding and characterization of the expected environmental background and interferents of each screening setting (*e.g.*, vehicle screening at a border crossing or mail and package screening at a postal facility) is imperative to evaluating the platform's effective sensitivity and discriminative potential, which can be used to specify appropriate alarm thresholds with known level of detection risk. For the screening applications targeted here (*i.e.,* border crossings and mail facilities) pure materials would likely be the focus, however, in our measurement of true positive data, mixtures of the target narcotic with common excipients were not considered. In general, street samples of fentanyl are often found in the presence of a range of other adulterants, diluents, and cutting

agents [49]. The analysis and detection of fentanyl and related compounds in the presence of complex street sample mixtures, against environmental interferents is an interesting topic, however, beyond the scope of this work. Similarly, the distribution and intensity of interfering signals may vary greatly from setting to setting, further demonstrating the importance of this type analysis for evaluating the effective performance on a location-specific basis.

## Acknowledgments

The U.S. Department of Homeland Security Science and Technology Directorate sponsored a portion of the production of this material under Interagency Agreement IAA HSHQPM-15-T-00050 with the National Institute of Standards and Technology. The authors declare no competing financial interest.

## Notes and References

† Electronic supplementary information (ESI) available: Additional experimental details, IMS spectra, ROC curves, and other figures as noted in the text can be found in the online version.

* Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

‡ Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.

1. *Overdose Death Rates (2018) National Institute of Drug Abuse.* Accessed: March 12, 2019; Available from: https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.
2. *World Drug Report 2018. United Nations Office on Drugs and Crime.* Accessed: March 12, 2019; Available from: https://www.unodc.org/wdr2018/.

3. *Border Crossing/Entry Data*. U.S. Department of Transportation, Bureau of Transportation Statistics Aug. 9, 2019; Available from: https://www.bts.gov/content/border-crossingentry-data.
4. *U.S. Drug Enforcement Administration, Special Testing and Research Laboratory. Emerging Threat Report, multiple years.* Accessed: March 12, 2019; Available from: www.ndews.umd.edu/resources/dea-emerging-threat-reports.
5. E. Sisco, M. Najarro and A. Burns, *Forensic Chem.*, 2018, **11**, 47-57.
6. G. A. Eiceman, Z. Karpas and H. H. Hill, *Ion Mobility Spectrometry*, 3rd edn., CRC Press, Boca Raton, FL, 2013.
7. G. A. Eiceman and J. A. Stone, *Anal. Chem.*, 2004, **76**, 390 A-397 A.
8. R. G. Ewing, D. A. Atkinson, G. A. Eiceman and G. J. Ewing, *Talanta*, 2001, **54**, 515-529.
9. J. Kozole, J. Tomlinson-Phillips, J. R. Stairs, J. D. Harper, S. R. Lukow, R. T. Lareau, H. Boudries, H. Lai and C. S. Brauer, *Talanta*, 2012, **99**, 799-810.
10. M. Najarro, M. E. Davila Morris, M. E. Staymates, R. Fletcher and G. Gillen, *Analyst*, 2012, **137**, 2614-2622.
11. G. Reid Asbury, J. Klasmeier and H. H. Hill Jr, *Talanta*, 2000, **50**, 1291-1298.
12. J. R. Verkouteren, J. Lawrence, G. A. Klouda, M. Najarro, J. Grandner, R. M. Verkouteren and S. J. York, *Analyst*, 2014, **139**, 5488-5498.
13. L. T. Demoranville and J. R. Verkouteren, *Talanta*, 2013, **106**, 375-380.
14. T. Khayamian, M. Tabrizchi and M. T. Jafari, *Talanta*, 2006, **69**, 795-799.
15. A. H. Lawrence, *Forensic Sci. Int.*, 1987, **34**, 73-83.
16. E. Sisco, J. Verkouteren, J. Staymates and J. Lawrence, *Forensic Chem.*, 2017, **4**, 108-115.
17. J. R. Verkouteren and J. L. Staymates, *Forensic Sci. Int.* , 2011, **206**, 190-196.
18. H. Zaknoun, M.-J. Binette and M. Tam, *Int. J. Ion Mobil. Spectrom.*, 2019.
19. G. R. Asbury, C. Wu, W. F. Siems and H. H. Hill Jr, *Anal. Chim. Acta*, 2000, **404**, 273-283.
20. P. Rearden and P. B. Harrington, *Anal. Chim. Acta*, 2005, **545**, 13-20.
21. S. Zimmermann, S. Barth, W. K. M. Baether and J. Ringer, *Anal. Chem.*, 2008, **80**, 6671-6676.
22. G. A. Eiceman, W. Yuan-Feng, L. Garcia-Gonzalez, C. S. Harden and D. B. Shoff, *Anal. Chim. Acta*, 1995, **306**, 21-33.
23. T. Fawcett, *Pattern Recognit. Lett.*, 2006, **27**, 861-874.
24. L. A. Adams, M. Bulsara, E. Rossi, B. DeBoer, D. Speers, J. George, J. Kench, G. Farrell, G. W. McCaughan and G. P. Jeffrey, *Clin. Chem.*, 2005, **51**, 1867-1873.
25. C. Heeschen, B. U. Goldmann, L. Langenbrink, G. Matschuck and C. W. Hamm, *Clin. Chem.*, 1999, **45**, 1789-1796.
26. M. S. Pepe, *The statistical evaluation of medical tests for classification and prediction*, Oxford, New York, NY, 2003.
27. K. H. Zou, W. J. Hall and D. E. Shapiro, *Stat. Med.*, 1997, **16**, 2143-2156.
28. C. G. Fraga, A. M. Melville and B. W. Wright, *Analyst*, 2007, **132**, 230-236.
29. A. Wysoczanski and E. Voigtman, *Spectrochim. Acta B*, 2014, **100**, 70-77.
30. T. P. Forbes and M. Najarro, *Analyst*, 2016, **141**, 4438-4446.
31. I. Cotte-Rodriguez, D. R. Justes, S. C. Nanita, R. J. Noll, C. C. Mulligan, N. L. Sanders and R. G. Cooks, *Analyst*, 2006, **131**, 579-589.
32. C. C. Mulligan, D. R. Justes, R. J. Noll, N. L. Sanders, B. C. Laughlin and R. G. Cooks, *Analyst*, 2006, **131**, 556-567.

33. D.-M. K. Dennis, M. R. Williams and M. E. Sigman, *Forensic Sci. Int.*, 2016, **259**, 179-187.
34. J. Carrano, *Chemical and Biological Sensor Standards Study*, DARPA Microsystems Technology Office, U.S. Department of Defense,Washington, DC, 2004, https://apps.dtic.mil/dtic/tr/fulltext/u2/a458370.pdf.
35. J. R. Verkouteren, J. L. Coleman, R. A. Fletcher, W. J. Smith, G. A. Klouda and G. Gillen, *Meas. Sci. Technol.*, 2008, **19**, 115101.
36. J. R. Riches, R. W. Read, R. M. Black, N. J. Cooper and C. M. Timperley, *J Anal Toxicol*, 2012, **36**, 647-656.
37. K. H. Zou, O'Malley, A.J. and Mauri, L., *Circulation*, 2007, **115**, 654-657.
38. A. Cecinato and C. Balducci, *J. Sep. Sci.*, 2007, **30**, 1930-1935.
39. A. J. Jenkins, *Forensic Sci. Int.*, 2001, **121**, 189-193.
40. C. Postigo, M. J. Lopez de Alda, M. Viana, X. Querol, A. Alastuey, B. Artiñano and D. Barceló, *Anal. Chem.*, 2009, **81**, 4382-4388.
41. M. Viana, X. Querol, A. Alastuey, C. Postigo, M. J. L. de Alda, D. Barceló and B. Artíñano, *Environ. Int.*, 2010, **36**, 527-534.
42. J. Howard and J. Hornsby-Myers, *Am. J. Ind. Med.*, 2018, **61**, 633-639.
43. G. A. Eiceman, E. G. Nazarov and J. A. Stone, *Anal. Chim. Acta*, 2003, **493**, 185-194.
44. G. Kaur-Atwal, G. O'Connor, A. Aksenov, V. Bocos-Bintintan, C. L. Paul Thomas and C. Creaser, *Int. J. Ion Mobil. Spec.*, 2009, **12**, 1-14.
45. R. M. Verkouteren and J. R. Verkouteren, *Anal. Chem.*, 2009, **81**, 8577-8584.
46. R. M. Verkouteren and J. R. Verkouteren, *Langmuir*, 2011, **27**, 9644-9653.
47. L. M. Matz, P. S. Tornatore and H. H. Hill, *Talanta*, 2001, **54**, 171-179.
48. *International Union of Pure and Applied Chemistry, Detection and Quantification Capabilities*. Accessed: August 27, 2019; Available from: http://media.iupac.org/publications/analytical_compendium/Cha18sec437.pdf.
49. C. Cole, L. Jones, J. McVeigh, A. Kicman, Q. Syed and M. Bellis, *Drug Test Analysis*, 2011, **3**, 89-96.