Applying remote sensing and data science techniques to enhance the monitoring and detection of environmental crime: examples from the NarcoView project

Tatjana Kuznecova [55], Nilay Swarge & Jaap Knotter

https://doi.org/10.3013/cepol-bulletin.envcrime.2024.005-applicatio

Abstract

The Netherlands has gained an international reputation as a centre for the production and export of synthetic drugs such as MDMA and amphetamine. The number of synthetic drug production laboratories and chemical waste dump sites discovered in recent years is a cause for concern. Dumping and discharging synthetic drug waste is a serious environmental crime, since synthetic drug waste contains various harmful chemicals. Such waste is being disposed of unsafely, in a number of illicit ways, causing environmental harm and risking public health and safety. This was highlighted as a growing problem by reports published in 2016 and 2019 by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), now the European Union Drugs Agency (EUDA), and Europol. The main regions in the EU facing the drug waste dumping problem are the southern part of the Netherlands and the northern part of Belgium.

The research presented in this article was conducted in the framework of the project NarcoView, with the main objective being to enhance the monitoring and detection of the environmental crime of illegal dumping of chemical waste from synthetic drug production. This is accomplished by gathering and processing intelligence through the use of RS, data science and ML algorithms. This article thus presents the results from formulating and testing methodologies for selected use cases with the aim of more efficiently and effectively detecting locations used for waste dumping. The main scenarios outlined here include (1) the detection of 'classic' dump sites and (2) the detection of chemically treated crop fields. The methods and resulting analytical products developed for

(55) Tatjana Kuznecova (corresponding author), Saxion University of Applied Sciences (t.kuznecova@saxion.nl); Nilay Swarge, Saxion University of Applied Sciences (n.n.swarge@saxion.nl); Jaap Knotter, Saxion University of Applied Sciences and Police Academy Netherlands (j.c.knotter@saxion.nl).

the two aforementioned scenarios will be further integrated into a web-based platform for end users, such as law enforcement and other state agencies.

Keywords: remote sensing, data science, artificial intelligence, environmental crime, synthetic drug waste.

Introduction

The Netherlands is known to be a producer (and exporter) of cannabis and synthetic drugs and a transit country for cocaine and heroin. The number of synthetic drug production laboratories discovered and dismantled has increased in recent years, with a significant number of synthetic drug waste dump sites having also been discovered (Politie, 2018, 2021; Schoenmakers and Mehlbaum, 2017). Most of the laboratories discovered were involved in the production of amphetamine and MDMA/ecstasy and methamphetamine. The dumping of chemical waste resulting from drug production was highlighted as a large and growing problem in reports published by EMCDDA and Europol in 2016 and 2019 (EMCDDA and Europol, 2016, 2019), supported by the Dutch national police data presented in Figure 5.1 (Politie, 2018, 2021). Figure 5.2 clearly shows that the drug waste dumping problem concerns mostly the southern part of the Netherlands and the northern part of Belgium, adding a transborder element to this issue (Claessens et al., 2019; EMCDDA and Europol, 2019).

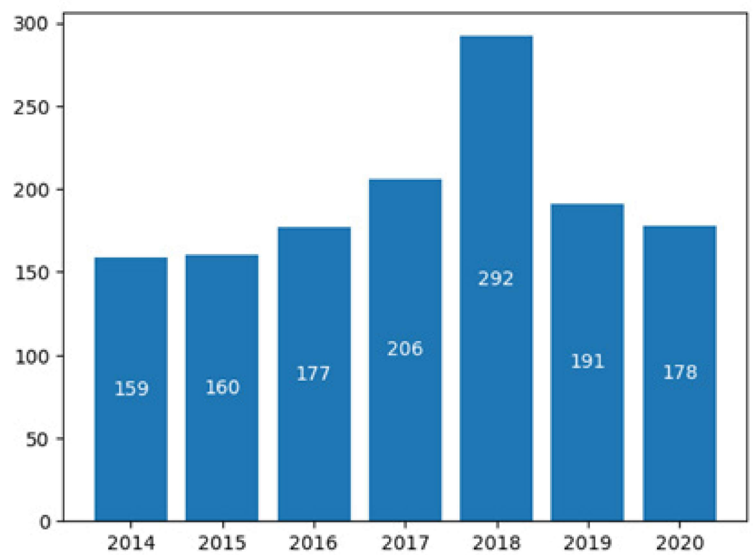

Figure 5.1. Numbers of synthetic drug dump sites discovered in 2014–2020. Source: Based on data from reports of Police Netherlands (Politie, 2018, 2021).

The waste generated from the production of synthetic drugs is usually disposed of unsafely, causing environmental harm and risking public health and safety. It is estimated that the production of 1 kg of MDMA (ecstasy) or amphetamine (speed) results in many more kilograms of waste: 6–10 kg for MDMA and 20–30 kg for amphetamine or methamphetamine (EMCDDA and Europol, 2019).

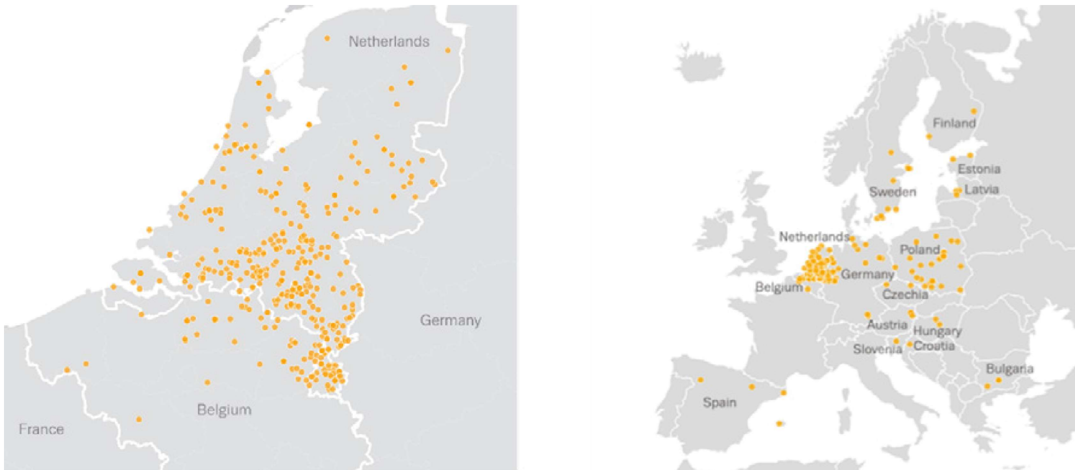

Figure 5.2. Locations of synthetic drug waste dump sites in the Netherlands and Belgium in 2015–2017 (left) and amphetamine production sites in the EU (right). Source: EMCDDA and Europol (2019), under the European Commission reuse policy.

According to Schoenmakers and Mehlbaum (2017), dumping drug waste is a serious environmental crime, with drug waste being considered a hazardous waste on the basis of its composition, as it usually contains acids, solvents and other substances that may be toxic (van Leerdam et al., 2022). Containers containing chemical waste products are often dumped in remote and secluded areas like forests and fields, and sometimes waste is spilled directly into the environment or buried underground. Dumped chemicals penetrate soil, groundwater and sewers, contaminating the environment (EMCDDA and Europol, 2019). Environmental health is strongly linked to human health and well-being, and contamination of the environment can lead to health problems via the consumption of contaminated agricultural produce or water. Contamination of water with chemical waste may lead to the disruption of water treatment plants, which has an impact on (safe) water supply (Ter Laak and Emke, 2023). The cost of cleaning areas where chemicals have been dumped is high, and it poses high safety risks for the personnel involved. Specialised professionals with protective gear are usually needed to clean up affected sites.

Disrupting these forms of organised crime is one of the priorities of the European agenda on security, and the fight against drug waste crime is in line with the European Commission's nine-point action plan on environmental compliance assurance (European Commission: Directorate-General for Environment, 2016). It was also a priority of the 2018–2021 EU policy cycle for organised and serious international crime (EMPACT) (Council of the European Union, 2018).

Geodata science comes under the domain of data science, as it uses statistics, data mining and predictive modelling to study spatial relationships and patterns. It often takes advantage of modern computational techniques and big data technologies (Zuo and Xiong, 2020).

In recent times, Remote Sensing (RS) technology has gained traction in various fields of study and application due to its ability to collect data without the need for direct contact with the subject. To achieve this, it utilises sensors to measure signals, such as electromagnetic radiation, emitted, reflected or scattered by the object under scrutiny (Campbell and Wynne, 2011). Two types of sensors can be used: passive sensors that capture naturally occurring energy like sunlight that is reflected by objects (e.g. multispectral and hyperspectral sensors) and active sensors that emit and then measure reflected signals (e.g. lidar and radar sensors). These sensors enable data collection using satellites for global coverage, aircraft for finer details and drones for localised information (Janga et al., 2023).

RS data are used to calculate spectral indices, which are derived from combining values from multiple spectral bands of a multispectral image. Such indices, for instance the Normalized Difference Vegetation Index (NDVI), are instrumental in the early detection of stress in vegetation (Dash et al., 2017) or other phenomena, depending on the index used.

The integration of Artificial Intelligence (AI) techniques with RS technology presents a groundbreaking opportunity in data analysis across diverse domains (Dash et al., 2017; Hameed, 2024; Rashidiyan and Rahimzadegan, 2024). Traditional Machine Learning (ML) methods have been extensively utilised in RS tasks like classification, object detection and change detection. Deep Learning (DL), a subset of ML, utilises complex neural networks to discern patterns and extract intricate features from large datasets, revolutionising RS applications.

Combining data science or ML techniques with RS approaches presents various opportunities in many application domains and could be instrumental in monitoring and detecting environmental crime. For instance, change detection is a widely used RS task that aims to detect and analyse gradual or abrupt changes occurring in the same area or object over time (Bai et al., 2023; de Bem et al., 2020). Anomaly or outlier detection is considered one of the vital applications of data mining, which deals with identifying data points that are dramatically different from the rest of the observations (Samariya and Thakkar, 2023). Object detection is a common problem in aerial and satellite image analysis, and RS enables the detection of various objects of interest visible from space or air (Cheng and Han, 2016). Such techniques, separately or in combination, may help uncover the signs of an illegal activity or crime by surveying or monitoring a region of interest.

RS technology also has the potential to revolutionise the identification of pollution or chemical stress in soils and vegetation by offering efficient monitoring solutions. By utilising spectral imaging sensors on Unmanned Aerial Vehicles (UAV's) or satellites, RS technology can identify plant distress caused by toxic substances and enable early detection, even before visible symptoms appear. Therefore, it is interesting to explore the applicability of such techniques for detecting or monitoring discharges of synthetic drug waste into the environment.

Several interesting studies conducted so far include those related to the use of herbicides in the context of agricultural crop production, in particular glyphosate (Kouakou et al., 2017; Nehurai et al., 2023; Suarez, 2018; Yao et al., 2012); to the monitoring of brownfield sites and oil-contaminated soils (Adamu et al., 2018; Lassalle et al., 2019); to heavy metal and metalloid pollution (Kooistra et al., 2004; Saha et al., 2022; Sobura et al., 2022); and to a chemical spill where toluene was a pollutant (Kim et al., 2021). Many of the experiments described in these peer-reviewed studies were conducted in the laboratory or under controlled conditions, and most utilised UAV-based or proximity sensing. To date, we are not aware of comparable studies that consider the effects of pollution from synthetic drug waste. A small pilot project was conducted in the Netherlands on chemical spills from drug production as part of an initiative by the national police, but the results were not conclusive.

This research explores the applicability of several data science and ML techniques in combination with RS and other geospatial data for enhancing the detection of synthetic drug waste dump or discharge sites based on formulated scenarios. The article was developed as part of the NarcoView project funded by ISF Police. The NarcoView project aims to develop a platform for monitoring and detecting environmental crime for end users such as law enforcement and other state agencies. The concept is based on combining data from various sources and utilising data-driven algorithms that enable the prioritisation and/or inspection of larger areas in a shorter time.

The objectives of the research presented in this article are as follows:

- to pre-evaluate the feasibility and relevance of using RS technology in combination with data science and ML approaches for selected scenarios;
- to develop approaches and test techniques for enhancing the detection and monitoring of synthetic drug waste dump or discharge sites using (geo)data science and ML for selected scenarios.

The most promising methods and resulting analytical products will be further integrated into the NarcoView platform.

Methodology

At the start of the project, a list of various scenarios was compiled based on a survey and discussions with the relevant experts of the partner organisations (the Dutch national police, Belgian Federal Police and NVWA). The scenarios were later prioritised based on data availability, relevance, technical feasibility potential and other considerations. Several types of synthetic drug waste dumping were considered.

While many of these scenarios carry great relevance in terms of environmental protection and human health impact, the research team was confronted with limited data availability. In view of this limitation, three scenarios were tested for feasibility, of which two were selected for further research.

Various types of data were necessary to conduct the research in accordance with the specified objectives. An overview of the most important datasets for different scenarios is given in Table 5.1.

Table 5.1. Overview of dumping scenarios and datasets used

No	Scenario of waste disposal	RS data	Ground-truth data	Other data
1	'Classic' dumping (use of jerrycans, barrels, intermediate bulk containers (IBC))	Satellite images, aerial images, drone images	Historical locations of dump sites for 2016–2021 in the Netherlands supplied by the Dutch police (approx. 1 000 cases)	Geographical, socioeconomic data (we used public data portals, such as PDOK (Publieke Dienstverlening Op de Kaart) – a platform for geodatasets from the Dutch government (https://www.pdok.nl/))
2	Waste dumped in manure pits / applied to crop fields	Satellite images, drone images	NVWA records: drug waste pollution (1 case) and use of other chemicals (< 10 cases of glyphosate exposure)	—
3	Chemical waste discharged into the soil/vegetation	Satellite images, drone images	Police records (3 cases)	—
4	Waste abandoned in cars/trucks	Excluded (lack of data)		
5	Waste discharged into open waterbodies/ waterways	Excluded (lack of data)		
6	Waste buried underground	Excluded (lack of data)		
7	Waste discharged into a sewer system	Excluded (lack of data)		

The research on the preselected scenarios started with a feasibility evaluation for the detection of dump sites using satellite data. The availability of satellite data (both public and commercial) was assessed, and potential sources/ providers were identified. Initial testing of satellite data processing and analysis was performed to estimate the potential for detection. Where satellite data were proven to be ineffective, aerial or drone data were also considered. After the feasibility assessment, two final scenarios were chosen for more in-depth research, with the aim of developing detection approaches for:

— the detection of 'classic' dump sites, that is, the detection of sites used for the dumping of synthetic drug waste by utilising storage containers such as IBCs, barrels or jerrycans;

— the detection of chemically polluted crop fields, that is, the detection of fields polluted by synthetic drug waste (this study included other chemical stressors due to the shortage of ground-truth data on drug pollution).

Scenario 1–Detection of 'classic' dump sites

For detection of 'classic' dump sites, two components were formulated: (1) a risk map for assessing which areas have a higher or lower risk of being used as a dump site, based on geodata science; and (2) a detection approach based on RS imagery and AI for object recognition. Eventually, the two components could be used in synergy – with the risk map being used to prioritise or narrow down areas for monitoring and inspection and object detection being used to pinpoint suspicious objects from satellite, aerial or UAV data.

Risk map development

Risk map development was based on historical data on synthetic drug dump site locations for 2016–2021 (source: Dutch national police – confidential data) and various complementary spatial data layers, such as roads network data, a land use/land cover map and data on characteristics of neighbourhoods, and other (Figure 5.3). The objective of data modelling was to identify meaningful spatial patterns in the locations of the dump sites and extrapolate this information to other (yet unused) locations. With limited studies available on this type of crime (examples include Schoenmakers et al. (2016) and Schoenmakers and Mehlbaum (2017)), we started by exploring a broader range of potential explanatory variables and then narrowed these down to a smaller subset that was used in the risk modelling.

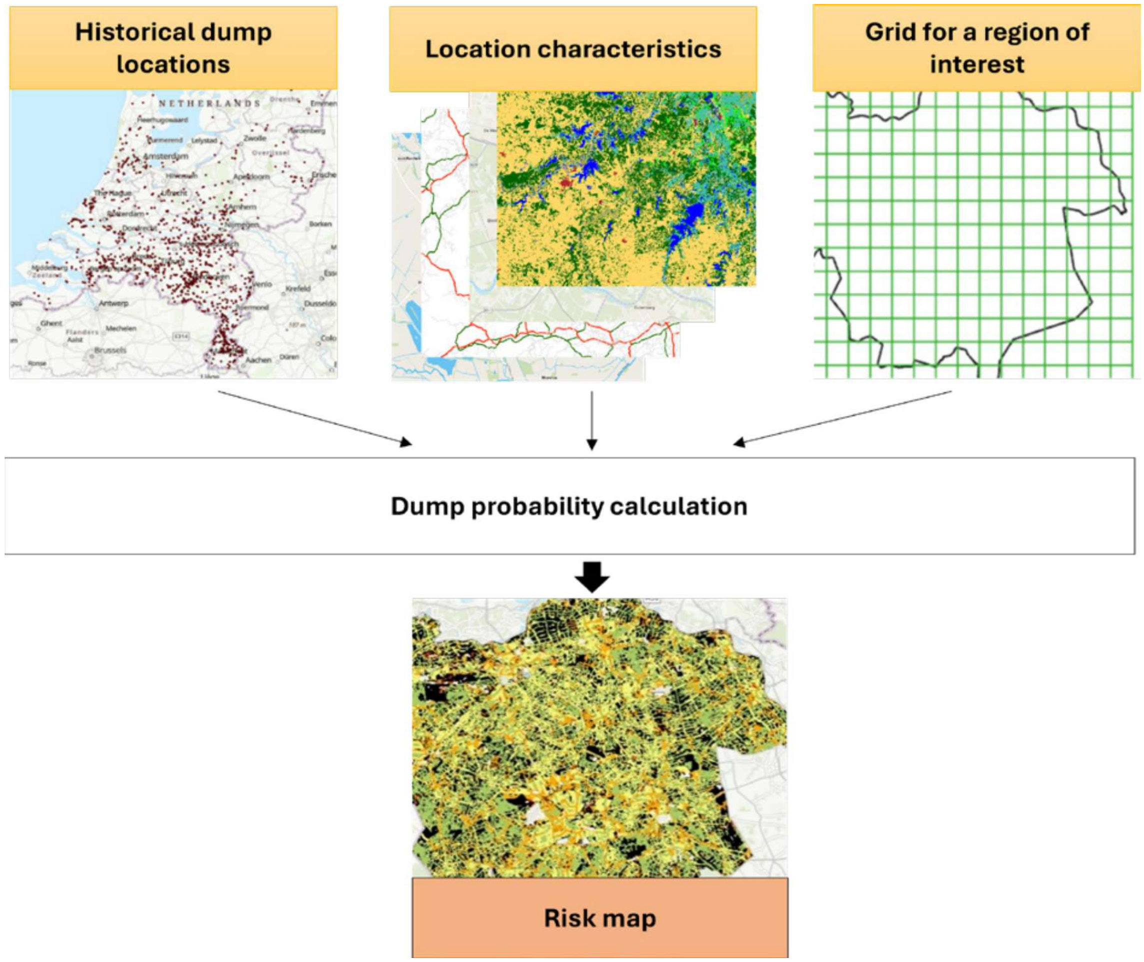

Figure 5.3. Conceptualisation of risk map development. Source: The authors.

The general workflow consisted of the following eight steps.

1. Identify a set of potential explanatory variables, using information from available literature/reports and discussions with experts from the partner organisations.
2. Split a region of interest into a grid of cells (we used 100 × 100 metre squared cells).
3. Preprocess the spatial data and assign variable values to the grid cells (using cell centre location).
4. Test potential explanatory variables for association with the presence of dump sites and select variables with statistically significant association ([56]).
5. Test variables for collinearity (when possible). For continuous variables, we used a maximum collinearity threshold of 0.7 (based on Pearson's coefficient for normal or transformed (e.g. log-normal) variables).
6. Discretise continuous variables.

([56]) We utilised a range of different statistical techniques to analyse the association. Before discretising continuous variables, we conducted analysis of mean and variance of exposed (with dump sites) and unexposed (without dump sites) groups of area cells. For normally distributed variables or normally transformed variables, an analysis of variance (ANOVA) test was used; otherwise, a non-parametric alternative was used, such as the Mann–Whitney *U* test. After discretising variables and with categorical variables, Fisher's exact test and odds ratio estimation were conducted.

7. Calculate dump site probability based on a conditional probability approach, using identified significant explanatory variables (we used a naive assumption that explanatory variables are independent). Conditional probability formula in a general form and in the form of a Bayes theorem is presented in Equation 1 (Bayes and Price, 1763).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \qquad \text{(Equation 1)}$$

8. Visualise computed probabilities on a map.

Before performing the calculations, various geo-processing techniques were applied to the original input datasets (step 3), such as distance calculations, the selection of features and the aggregation of values. The preliminary spatial processing was executed using the ModelBuilder tool of the ESRI software ArcGIS. ArcGIS was also used for final map visualisation. Statistical tests and probability computations were executed in the Python programming environment, using a variety of analytical libraries.

Object detection

We evaluated the potential to use satellite, aerial or drone imagery for detecting objects typically associated with synthetic drug waste dump sites. Publicly available aerial and satellite data and commercial VHR satellite imagery were considered first. Additional drone images were collected by the project team at a training facility of the Belgian Federal Police near Antwerp (Figure 5.4).

Figure 5.4. Drone images of objects tested for detection with DL algorithms: several barrels and jerrycans on the left and two IBCs on the right. Source: The authors in collaboration with Belgian federal Police.

For testing purposes, IBCs, barrels and jerrycans were selected as the most typical objects used for dumping chemical waste. DL object recognition algorithms, such as YOLOv8, were tested by the subcontracted company using the drone imagery collected.

Scenario 2 - Detection of anomalous crop parcels

For analysis and detection of chemical discharge spots and polluted crop fields, vegetation health was used as a proxy. The analysis is based on the use of spectral indices, such as the NDVI, calculated from RS data for the area(s) of interest.

The NDVI, which is calculated using red (*R*) and near-infrared (*NIR*) band reflectance values, quantifies vegetation health, density and seasonal changes (Gandhi et al., 2015; Townshend et al., 1985), with values ranging from – 1.0 to + 1.0 (Equation 2). Values close to 1 correspond to healthier and more dense vegetation, and values close to 0 correspond to little vegetation, early stages of cultivation or bare soil. Negative values are generally associated with water, artificial/built-up surfaces and snow.

$NDVI = (NIR - R) / (NIR + R)$ (Equation 2)

The basic feasibility analysis of chemical discharge spots in soil/vegetation based on the imagery from the Superview and Triplesat satellites showed that automated detection of such cases would not be very likely in practice, at least based on a small number of known cases. This was partly because the affected area was too small. Therefore, we continued with developing an approach for the detection of anomalous crop fields.

Figure 5.5 shows a global conceptual workflow of an approach for the detection of anomalous fields. Due to the shortage of confirmed examples of crop fields polluted by chemical drug waste, the concept was tested on examples of exposure to another type of chemicals - the herbicide glyphosate.

Our test area covered a region of about 80 km^2 south of the city of Zwolle in the Netherlands, partly in the Overijssel province and partly in the Gelderland province. This area was selected because several glyphosate treated crop parcels were confirmed to be located here (exact locations are confidential; data from NVWA). Preprocessing of the data included retrieving available satellite imagery (Sentinel-2 multispectral instrument) in the Google Earth Engine environment. Google Earth Engine combines a multi-petabyte catalogue of satellite imagery and geospatial datasets with planetary-scale analysis capabilities, therefore enabling easy processing of large-scale datasets ([57]). From satellite imagery, NDVI values were computed for the selected region and aggregated at a crop-field level using a crop-parcel dataset for the Netherlands (Basisregistratie Gewaspercelen (BRP)). NDVI values were computed for various points in time spanning 2018–2022 (the temporal interval between data points varied depending on satellite imagery availability).

([57]) https://earthengine.google.com/.

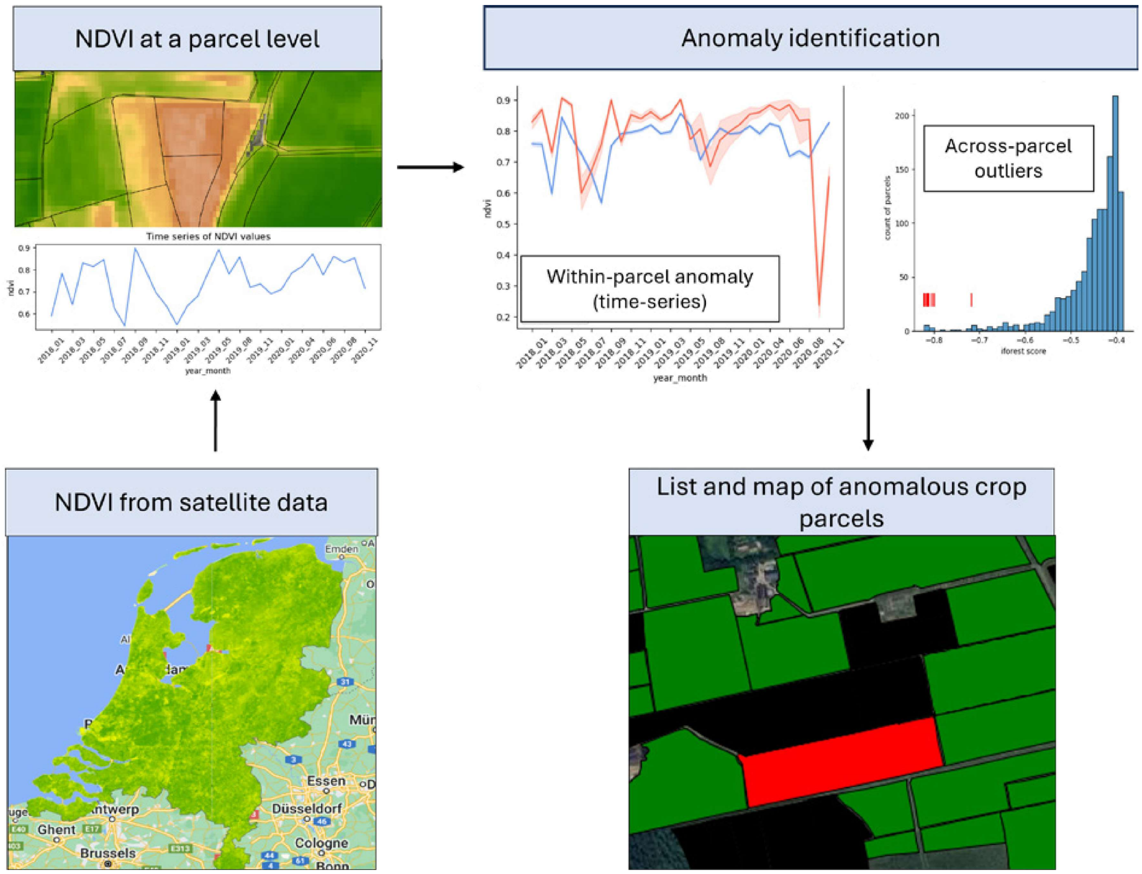

Figure 5.5. Conceptual workflow for detection of anomalous crop fields. Source: The authors.

As spectral reflectance of vegetation may depend on the crop type, as well as various environmental and climatic conditions, certain filters and constraints were applied prior to analysis as follows:

— only parcels within a certain area of interest were included (in our case, we used an area of around 80 km^2, but the sensitivity to different area sizes could be tested in the future) due to possible variations in growing conditions for the crop;

— only parcels with the same crop type were analysed together;

— other filters were used to reduce the noise in the data, such as excluding NDVI values that were uncharacteristic for the vegetation (e.g. negative NDVI values).

For identification of anomalous crop parcels, two approaches were identified:

— outlier parcels were identified in a snapshot in time (e.g. specific month) from the NDVI distribution across parcels in the selected region;

— anomalies were detected in the time series of NDVI values for a specific crop parcel.

For the first approach, two types of techniques were considered: (1) statistical outlier detection (e.g. using z-values) and (2) ML techniques, such as isolation forest. Isolation forest is an unsupervised ML technique that is based on the use of decision trees to isolate outlier data points in a dataset (Liu et al., 2008). After the initial testing, preference was given to the isolation forest technique, as it allows to include multiple variables in the analysis and it does not depend on the input data distribution (unlike the z-value-based technique), thus providing more flexibility in its application.

For the second approach, the techniques for detecting breakpoints or other pattern changes within the time series of values were tested. Work to develop and test the exact methodology based on time series analysis is still in progress.

Results

This section presents the results achieved with the development and testing of the approaches for enhancing detection of synthetic drug dump sites based on the two formulated scenarios.

Scenario 1–Approach for the monitoring and detection of 'classic' dump sites

The results for use case 1 include the proof of concept of the risk map for prioritising monitoring and detection areas, and a preliminary assessment of the feasibility of dump detection using RS data.

Risk map evaluation

We explored the possibility of generating a risk map at the level of a whole province. A proof of concept for the North Brabant province in the Netherlands was developed (Figure 5.6). This province was chosen due to the high concentration of illegal drugs production in the area (close to the Dutch–Belgian border) and, as a result, the highest number of chemical drug waste dump sites among all Dutch provinces.

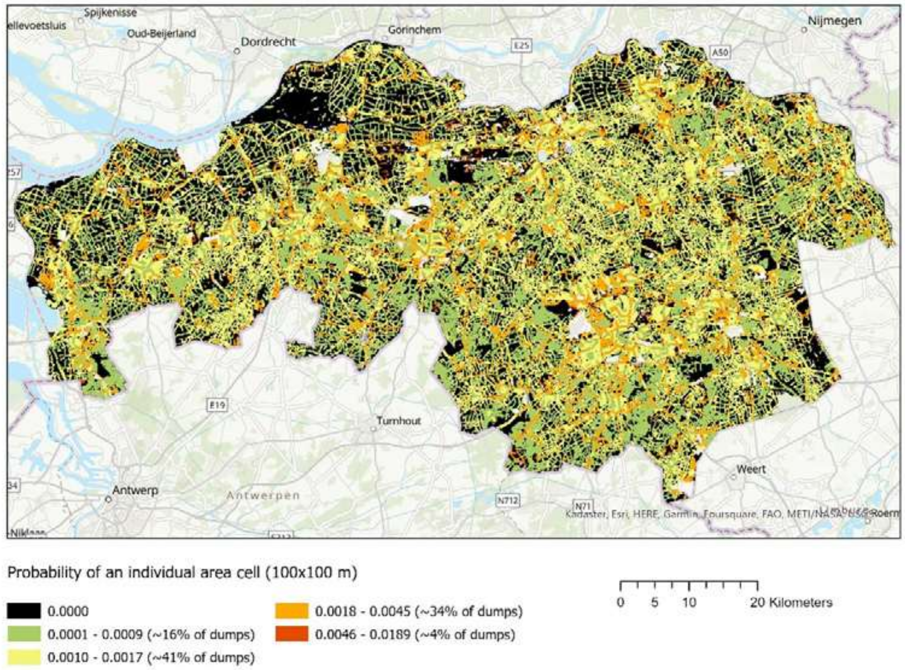

Figure 5.6. Risk map for North Brabant province in the Netherlands based on data for 2016–2021. Source: The authors.

A number of variables that characterise a location were initially explored for the risk modelling, such as variables reflecting accessibility, land use type, urbanisation and various socioeconomic factors, as well as various preprocessing approaches. Based on statistical tests, several variables were selected for inclusion in the model. Some variables were rejected due to a large number of missing values in a dataset. Including many variables in one model also caused a complication: the large number of possible unique combinations of characteristics and relatively small number of dump records could have led to unreliable conditional probability distributions. Therefore, five variables were left in the model: distance to roads, distance to high significance roads (eg. highways or arterial roads), distance to built-up areas, land use type and population density. We also introduced a threshold for the minimum number of cells required with the same characteristics (100 cells ([58])) in order to avoid inflation of the conditional probability range due to a random chance in small groups of cells.

Conditional probabilities were further calculated for each area cell of 100 × 100 metres for the whole province. Resulting values were classified for visualisation using the natural breaks method and displayed with the proportion of the total dump site count covered by the respective area.

([58]) We recommend experimenting with different thresholds to evaluate their influence on the risk map.

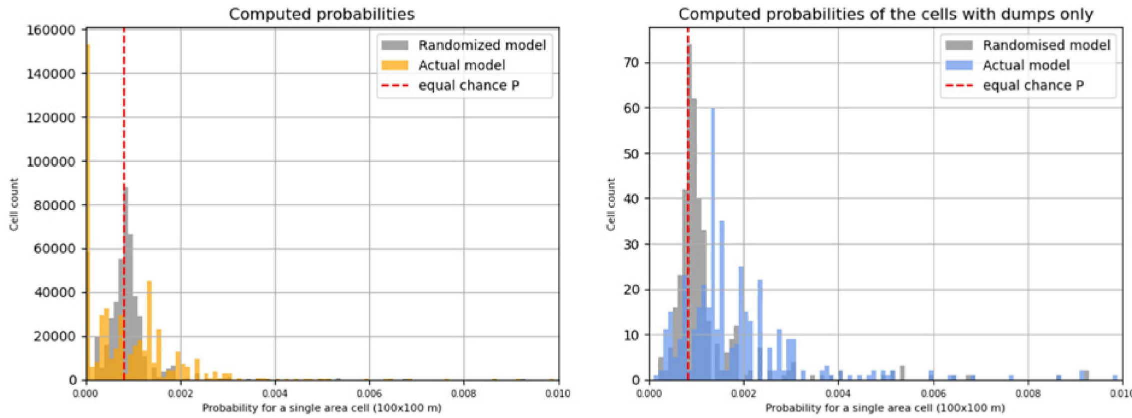

Figure 5.7. Distribution of probabilities of map cells – comparison of the model with the real data and one of randomised experiments (the *x* axis is limited to 0.01 for better visualisation). Source: The authors.

A large portion of cells had an estimated probability of 0 or very close to 0. After splitting the resulting probability range into classes, almost 80 % of the known dump sites occurred in less than 40 % of the total area (yellow, orange and red areas on the map in Figure 5.6), while almost 40 % of all dump sites occurred in only about 12 % of the total area (orange and red areas on the map). In a practical application, such knowledge could enable the exclusion of certain areas from monitoring or inspection routines and narrow down the search areas.

The results of the model using the actual data were compared with a number of experiments using simulated randomised sets of dump locations (Figure 5.7) in order to exclude the possibility that the results we obtained could easily have occurred by chance.

Object detection

Based on the initial feasibility assessment, using satellite data for detecting chemical storage containers was shown to be impractical for several reasons, such as the very narrow window of opportunity (in the Netherlands and Belgium, suspicious objects are quickly found by citizens and reported to the police), difficulty in finding suitable imagery for the specific locations and time / time frames, the resolution of the imagery necessary for detection of small objects and the high costs of VHR satellite imagery. Based on these considerations, it was decided to further test the applicability of aerial and drone-based imagery. Training and testing algorithms for this task is still a work in progress (being undertaken by an external company).

Scenario 2 – Detection of anomalous crop fields

We introduced two approaches to the analysis of potentially chemically treated crop parcels. The first approach investigated patterns at a given point in time across comparable parcels. To control for factors such as soil type and weather circumstances, parcels from the surrounding area of known treated fields were used for comparison. All parcels were filtered down to preserve only one crop type (in this case, permanent grassland) in a specific year and month (e.g. September 2020). The NDVI score was used as an input variable for the isolation forest algorithm, and, later, the difference from the previous month was added as another input. Based on a number of iterations, each value of a variable was assigned a score indicating its 'outlierness' given the full set of values (Figure 5.8). The number of estimators (or number of times the isolation forest algorithm was performed and used for computing the isolation forest scores) was also tested. From 10 iterations and higher, the results stabilised to a constant pattern. However, even with 5 iterations, the distribution of isolation forest scores showed a similar pattern. In each case, known treated fields were classified as outliers.

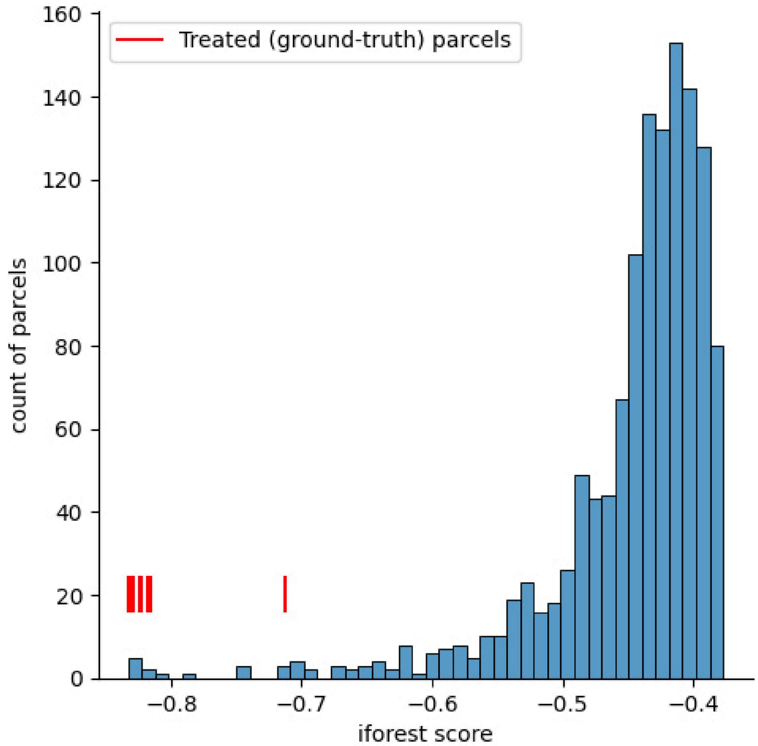

Figure 5.8. Results of outlier detection with isolation forest for September 2020. Source: The authors.

The filter-based approach to finding outliers was shown to be consistent in classifying the known treated fields as such. An example of applying filters to the combined dataset is provided in Table 5.2.

Table 5.2. Example of filtering crop parcels using NDVI values for the month of interest and difference from the previous month as inputs for the isolation forest algorithm

Filter step	Action	Total records remaining	Individual plots remaining
—	Starting state	71 762	2 057
1	Filter crop type (permanent grassland)	43 340	1 318
2	Filter out plots with bare soil (maximum yearly NDVI threshold of 0.2)	43 340	1 318
3	Take out parcels with changed crop type compared with the previous year	43 340	1 318
4	Select the data from September 2020 (slice in time)	1 265	1 265
5	Apply isolation forest algorithm and select the parcels classified as outliers (isolation forest score < – 0.5)	185	185
6	Select parcels classified as stronger outliers (isolation forest score < – 0.7)	21	21
7	Validation	known parcels under exposure (ground-truth) classified as outliers: 8 out of 8	

For the second approach – time series analysis – we first focused on visual inspection. The NDVI values were plotted chronologically. For the treated plots, the months in which the treatment occurred showed drastic differences compared with the non-treated parcels (Figure 5.9).

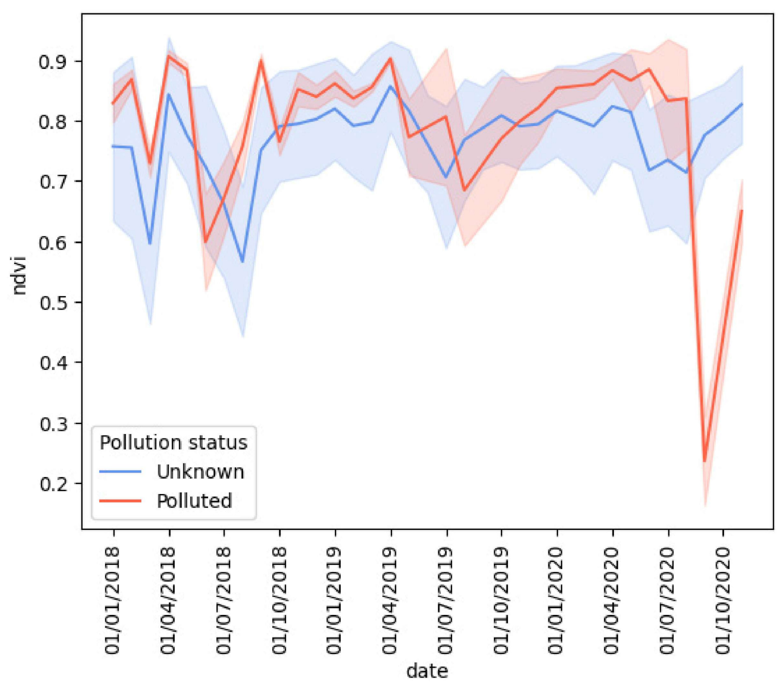

Figure 5.9. Comparison of average NDVI values of normal and affected grass fields. Source: The authors.

For this acute drop, the difference from the previous month was tested as an additional input parameter for applying the isolation forest algorithm (Table 5.2). At this stage, adding the change compared with the previous month did not add a lot to the isolation forest outcome for stronger anomalies, but it slightly reduced the overall number of detected outliers (isolation forest score < – 0.5). By further investigating the possibility of change detection in the time series of spectral data, there is potential to develop a proper detection method suitable for multiple applications in the domain of monitoring crop growth and vegetation stress detection.

Discussion

RS imagery and other types of geospatial data, alongside data science and ML techniques, offer unique possibilities for the monitoring and detection of illegal activities, and assessing the consequences of these. In the context of the illicit dumping of chemical waste from synthetic drug production, the body of knowledge is limited and insufficient data are available. In light of this, our objective was to assess the feasibility of and develop proofs of concept for the two scenarios outlined for detecting synthetic drug waste dump sites, considering the aforementioned constraints. During the process, various limitations and opportunities for future improvement were identified.

Risk map

In the risk map development, several limitations were related to data availability and quality. For example, most socioeconomic variables were available only at a neighbourhood level as the finest level of detail. Uncertainty about the locations of historical drug waste dump sites was also a serious concern. Police records contained specific GPS coordinates for only the year 2021, while other records had only a partial address. Our team used OpenStreetMap ([6]) to convert recorded addresses or postal codes into approximate coordinates; however, the error in location in this case is difficult to estimate. The availability of other potentially interesting information was also a limitation – for instance information on abandoned industrial or agricultural businesses or those no longer operating. Furthermore, it could be beneficial to analyse different subgroups of dump sites separately (e.g. based on the manner of dumping); however, this was not done in this case due to the absence of dump site descriptions in a significant portion of police records.

Further refining the set of input variables would be beneficial. However, including more variables at this stage was not feasible with the currently available dump site records and the method used. Including more variables would lead to an overly large number of possible combinations of characteristics that each area cell could have, and therefore a reliable conditional probability distribution for each of these combinations could not be ensured.

Our current risk map model makes use of the conditional probability concept and could be presented as a very simple Bayesian network in the future. Further risk map development could explore Bayesian network modelling with spatial data in more depth (e.g. the possibilities outlined in Krapu et al., 2023) and more thorough testing of model structure and robustness.

Remote sensing and image recognition

Tasks like object detection are widely used in combination with RS data, based on satellite and/or UAV imagery. Existing algorithms, such as YOLOv8, are well developed and enable the easy implementation of the detection workflows. However, data availability and quality still represent serious limitations. Based on a preliminary analysis, various limitations to the use of satellite data were identified in our application context. Among them were a lack of suitable imagery for the required locations and dates, the specific image resolution requirements for the detection of small objects or features and, consequently, the high costs of VHR imagery that would accommodate such requirements on demand. Our team resorted to the use of aerial and UAV data instead. This approach, however, is not in line with the original idea of a detection workflow, since UAVs cover a smaller area than satellites and require an operator. Aerial imagery, meanwhile, is collected only about once a year by the Dutch government, and therefore would not be suitable for the timely detection of crime.

Availability of labelled examples for training object detection algorithms and examples of ground-truth data for testing and validating the techniques still represent serious limitations. For several interesting scenarios, only a few recorded examples of ground-truth data could be provided by law enforcement partners. This is in line with the observations of Schoenmakers et al. (2016), who pointed out the lack of records, especially on chemical discharge cases. For applications with rather uniform objects (e.g. containers or barrels), the new opportunities that modern generative AI tools (e.g. DALL-E/ChatGPT) can offer should be explored, for example the generation of an artificial training sample for the image classification algorithms.

Detection of anomalous crop fields/parcels

Our approach was based on detecting anomalies in the spectral vegetation index of the crop fields from satellite data. Any observation that is inconsistent with the rest of the dataset would be highlighted, thus reducing the need to account for all possible effects of each specific chemical substance or their combinations. However, anomalies in crops may be caused by various factors, including natural factors (drought, pests, etc.) and anthropogenic factors (mechanical disturbance, use of phytosanitary products, etc.), and, with the approach presented, the distinction between various potential stressors cannot be determined. Overall, the research on sensing the effects of chemical stressors is still limited and does not deliver a full picture of the theoretical body of knowledge (e.g. the response of different crop types to different chemical stressors and the best techniques to sense these responses) or how such techniques can be used in practice in the context of real application scenarios of the stakeholders concerned and in real-world conditions. Other indices are available in addition to the NDVI (e.g. the Optimised Soil-adjusted Vegetation Index (OSAVI)) that may be potentially effective for this application (Rondeaux et al., 1996). Furthermore, with the rapid development of sensing techniques, hyperspectral data may provide additional opportunities for the detection of chemical pollution, using either satellite or UAV platforms.

Another important point of discussion is the availability of ground-truth data. In this research, only a few confirmed cases of affected crop fields were known, with most of these cases being due to glyphosate exposure and only one case being related to drug waste pollution (that could not be used). Identifying and adding more ground-truth data (on both confirmed exposed and unexposed fields) would add robustness to the eventual analysis and hopefully help establish some baselines in the differences between the two groups.

The time series analysis approach showed some promise, but also posed challenges. The cyclical character of seasonal growth patterns is sometimes hard to capture for a few reasons. Firstly, parcels themselves can change in size and be split up or merged with another parcel. Secondly, the type of crop can change over time. Lastly, other factors such as the weather and other environmental factors can vary over the years. Moreover, the analysis may consider other factors, such as the variability of spectral information within the parcel, soil type variations and finer moisture scores.

Conclusions

This article presents the results achieved so far in the context of the NarcoView project, where approaches for improving the detection of synthetic drug waste dump sites were explored using RS and data-driven techniques. Two main use cases were outlined: (1) the detection of 'classic' synthetic drug waste dump sites with the help of a risk map and object detection algorithms; and (2) the detection of polluted crop fields.

For the risk map, we used a small number of variables (five) to define a risk model. It is possible that the model could be further improved through the inclusion of other variables, as well as the use of more and better quality records on cases of dump sites over time. The evaluation of the risk map showed that, even at the proof-of-concept level, it is possible to narrow down or prioritise inspection areas, compared with a purely randomised approach. Eventually, a risk map could be used in synergy with object detection approaches for the more efficient surveying of areas.

The utilisation of VHR satellite imagery was found to be impractical for the detection of containers and small chemical discharge spots due to high costs and limited temporal and spatial coverage and resolution of the imagery. As an alternative, we started testing the applicability of UAV imagery. The initial tests with container detection showed promise; however, employing the latest drone technology equipped with higher resolution cameras is recommended when the detection of smaller objects (e.g. jerrycans) is of interest.

For the detection of anomalous crop fields, two approaches were used. Firstly, crop fields with the same crop type (grass) were compared at a selected point in time (month of interest). Using different statistical and ML methods, as described above, parcels were filtered to obtain a list of parcels labelled 'anomalous'. The use of the isolation forest approach to isolate anomalous crop parcels shows promise for the further development of this method.

As for the time series analysis approach, there are some interesting starting points, such as the acute drops in the NDVI values of some of the affected crop fields. However, further work and more ground-truth data are required to develop a method for the automated detection of time series changes in polluted crop fields. As of now, the testing had to be conducted using data on other chemical stressors (such as glyphosate). Furthermore, to improve the methodological approach, more theoretical knowledge on the response of vegetation to chemical stressors and the best techniques to sense such a response will be crucial before continuing with pattern analysis.

Acknowledgements

We cordially thank the European Commission and ISF Police for funding the NarcoView project, which enabled this research. Our gratitude goes to the consortium partners of the NarcoView project who supported this research with data, knowledge and organisational efforts – the Dutch national police, Belgian Federal Police, NVWA, Space53, LIBRe Foundation, Saxion University of Applied Sciences and Dutch Public Prosecution Service, as well as other partners and collaborators.

References

Adamu, B., Tansey, K. and Ogutu, B. (2018), 'Remote sensing for detection and monitoring of vegetation affected by oil spills', *International Journal of Remote Sensing*, Vol. 39, No 11, pp. 3628–3645, https://doi.org/10.1080/01431161.2018.1448483.

Bai, T., Wang, L., Yin, D., Sun, K., Chen, Y. et al. (2023), 'Deep learning for change detection in remote sensing: A review', *Geo-Spatial Information Science*, Vol. 26, No 3, pp. 262–288, https://doi.org/10.1080/10095020.2022.2085633.

Bayes, M. and Price, M. (1763), 'An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr Bayes, F. R. S. Communicated by Mr Price, in a letter to John Canton, A. M. F. R. S.', *Philosophical Transactions (1683–1775)*, Vol. 53, pp. 370–418, http://www.jstor.org/stable/105741.

Campbell, J. B. and Wynne, R. H. (2011), *Introduction to Remote Sensing*, 5th edition, Guilford Publications, https://books.google.nl/books?id=NkLmDjSS8TsC.

Cheng, G. and Han, J. (2016), 'A survey on object detection in optical remote sensing images', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 117, pp. 11–28, https://doi.org/10.1016/j.isprsjprs.2016.03.014.

Claessens, M., Hardyns, W., Vander Laenen, F. and Verhaeghe, N. (2019), *An analysis of the costs of dismantling and cleaning up synthetic drug production sites in Belgium and the Netherlands*, Ghent University, Belgium.

Council of the European Union (2018), *The EU policy cycle to tackle organised and serious international crime*, Publications Office of the European Union, Luxembourg.

Dash, J. P., Watt, M. S., Pearse, G. D., Heaphy, M. and Dungey, H. S. (2017), 'Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 131, pp. 1–14, https://doi.org/10.1016/j.isprsjprs.2017.07.007.

de Bem, P. P., Abílio de Carvalho, A., Guimarães, R. F. and Gomes, R. A. T. (2020), 'Change detection of deforestation in the Brazilian Amazon using Landsat data and convolutional neural networks', *Remote Sensing*, Vol. 12, No 6, https://doi.org/10.3390/rs12060901.

EMCDDA and Europol (2016), *EU Drug Markets Report 2016*, Publications Office of the European Union, Luxembourg, https://doi.org/10.2810/219411.EMCDDA and Europol (2019), *EU Drug Markets Report 2019*, Publications Office of the European Union, Luxembourg, https://doi.org/10.2810/766585.

European Commission: Directorate-General for Environment (2016), *Environmental Compliance Assurance and Combating Environmental Crime*, Publications Office of the European Union, Luxembourg.

Gandhi, G. M., Parthiban, S., Thummalu, N. and Christy, A. (2015), 'NDVI: Vegetation change detection using remote sensing and GIS – A case study of Vellore district', *Procedia Computer Science*, Vol. 57, pp. 1199–1210, https://doi.org/10.1016/J.PROCS.2015.07.415.

Hameed, A. A. (2024), 'Enhancing hyperspectral remote sensing image classification using robust learning technique', *Journal of King Saud University – Science*, Vol. 36, No 1, 102981, https://doi.org/10.1016/j.jksus.2023.102981.

Janga, B., Asamani, G. P., Sun, Z. and Cristea, N. (2023), 'A review of practical AI for remote sensing in Earth sciences', *Remote Sensing*, Vol. 15, No 16, p. 4112, https://doi.org/10.3390/rs15164112.

Kim, H., Kim, W. and Kim, S. D. (2021), 'Damage assessment of rice crop after toluene exposure based on the vegetation index (VI) and UAV multispectral imagery', *Remote Sensing*, Vol. 13, No 1, p. 25.

Kooistra, L., Salas, E. A. L. and Clevers, J. G. P. W. (2004), 'Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains', Vol. 127, pp. 281–290, https://doi.org/10.1016/S0269-7491(03)00266-5.

Kouakou, A. K., Soro, A. P., Taky, A. K. and Zoueu, J. T. (2017), 'Multi-spectral and fluorescence imaging in prevention of overdose of herbicides: The case of maize', *Spectral Analysis Review*, Vol. 5, No 2, pp. 11–24, https://doi.org/10.4236/sar.2017.52002.

Krapu, C., Stewart, R. and Rose, A. (2023), 'A review of Bayesian networks for spatial data', *ACM Transactions on Spatial Algorithms and Systems*, Vol. 9, No 1, pp. 1–21, https://doi.org/10.1145/3516523.

Lassalle, G., Fabre, S., Credoz, A., Hédacq, R., Borderies, P. et al. (2019), 'Detection and discrimination of various oil-contaminated soils using vegetation reflectance', *Science of the Total Environment*, Vol. 655, pp. 1113–1124, https://doi.org/10.1016/j.scitotenv.2018.11.314.

Liu, F. T., Ting, K. M. and Zhou, Z.-H. (2008), 'Isolation forest', in: *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, https://doi.org/10.1109/ICDM.2008.17.

Nehurai, O., Atsmon, G., Kizel, F., Kamber, E., Bar, N. et al. (2023), 'Early detection of the herbicidal effect of glyphosate and glufosinate by using hyperspectral imaging', *Agronomy Journal*, Vol. 115, No 5, pp. 2558–2569, https://doi.org/10.1002/agj2.21415.

Politie (2018), *ERISSP Landelijk overzicht Synthetische Drugs 2018,* Cluster Synthetische Drugs Intel & Expertise, https://www.politie.nl/binaries/content/assets/politie/nieuws/2019/11-le/erissp-landelijk-overzicht-synthetische-drugs-2018.pdf.

Politie (2021), *ERISSP Landelijk overzicht Synthetische Drugs 1e helft 2021*, Cluster Synthetische Drugs Intel & Expertise, https://www.politie.nl/binaries/content/assets/politie/onderwerpen/drugs/landelijk-overzicht-synthetische-drugs-eerste-helft-2021.pdf.

Rashidiyan, M. and Rahimzadegan, M. (2024), 'Investigation and evaluation of land use–land cover change effects on current and future flood susceptibility', *Natural Hazards Review*, Vol. 25, No 1, p. 1, https://doi.org/10.1061/NHREFO.NHENG-1854.

Rondeaux, G., Steven, M. and Baret, F. (1996), 'Optimization of soil-adjusted vegetation indices', *Remote Sensing of Environment*, Vol. 55, No 2, pp. 95–107, https://doi.org/10.1016/0034-4257(95)00186-7.

Saha, A., Gupta, B. Sen and Patidar, S. (2022), 'Identification of soil arsenic contamination in rice paddy field based on hyperspectral reflectance approach', *Soil Systems*, Vol. 6, No 1, p. 30.

Samariya, D. and Thakkar, A. (2023), 'A comprehensive survey of anomaly detection algorithms', *Annals of Data Science*, Vol. 10, pp. 829–850, https://doi.org/10.1007/s40745-021-00362-9.

Schoenmakers, Y., Mehlbaum, S., Everartz, M. and Poelarends, C. (2016), 'Elke dump is een plaats delict', *Political Science*, Vol. 83, pp. 9, 67, https://www.politieenwetenschap.nl/publicatie/politiekunde/2016/elke-dump-is-een-plaats-delict-278.

Schoenmakers, Y. M. M. and Mehlbaum, S. L. (2017), 'Drugsafval in Brabant', *Justitiële verkenningen*, Vol. 43, No 2, pp. 89–102, https://doi.org/10.5553/jv/016758502017043002006.

Sobura, S., Hejmanowska, B., Widłak, M. and Muszyńska, J. (2022), 'The application of remote sensing techniques and spectral analyzes to assess the content of heavy metals in soil – A case study of Barania Góra Reserve, Poland', *Geomatics and Environmental Engineering*, Vol. 16, No 4, pp. 187–213, https://doi.org/10.7494/geom.2022.16.4.187.

Suarez, L. A. (2018), 'Proximal and remote sensing for early detection and assessment of herbicide drift damage on cotton crops', PhD thesis, University of Southern Queensland, Australia.

Ter Laak, T. L. and Emke, E. (2023), *Environmental Impact of Synthetic Drug Production: Analysis of groundwater samples for contaminants derived from illicit synthetic drug production waste*, European Union Drugs Agency, Lisbon.

Townshend, J. R. G., Goff, T. E. and Tucker, C. J. (1985), 'Multitemporal dimensionality of images of Normalized Difference Vegetation Index at continental scales', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 23, No 6, pp. 888–895, https://doi.org/10.1109/TGRS.1985.289474.

van Leerdam, R. C., van Driezum, I. H. and Broekman, M. H. (2022), *De gevaren van dumpingen en drugsproductieafval voor de kwaliteit van drinkwaterbronnen*, https://doi.org/10.21945/RIVM-2022-0104.

Yao, H., Huang, Y., Hruska, Z., Thomson, S. J. and Reddy, K. N. (2012), 'Using vegetation index and modified derivative for early detection of soybean plant injury from glyphosate', *Computers and Electronics in Agriculture*, Vol. 89, pp. 145–157, https://doi.org/10.1016/j.compag.2012.09.001.

Zuo, R. and Xiong, Y. (2020), 'Geodata science and geochemical mapping', *Journal of Geochemical Exploration*, Vol. 209, https://doi.org/10.1016/j.gexplo.2019.106431.