

## Article

# An Incongruence-Based Anomaly Detection Strategy for Analyzing Water Pollution in Images from Remote Sensing

Maurício Araújo Dias <sup>1,\*</sup>, Erivaldo Antônio da Silva <sup>2</sup>, Samara Calçado de Azevedo <sup>3</sup>,  
Wallace Casaca <sup>4</sup>, Thiago Statella <sup>5</sup> and Rogério Galante Negri <sup>6</sup>

<sup>1</sup> Department of Mathematics and Computer Science, School of Sciences and Technology, São Paulo State University (UNESP), Campus Presidente Prudente, SP 19060-900, Brazil

<sup>2</sup> Department of Cartography, School of Sciences and Technology, São Paulo State University (UNESP), Campus Presidente Prudente, SP 19060-900, Brazil; erivaldo.silva@unesp.br

<sup>3</sup> Natural Resources Department, Federal University of Itajubá, Av. BPS 1303, Itajubá, MG 37500-903, Brazil; samara\_calçado@unifei.edu.br

<sup>4</sup> Department of Energy Engineering, São Paulo State University (UNESP), Campus Rosana, SP 19274-000, Brazil; wallace.coc@gmail.com

<sup>5</sup> Federal Institute of Education, Science and Technology of Mato Grosso (IFMT), 95 Zulmira Canavarro, Cuiabá, MT 780025-200, Brazil; t.statella@gmail.com

<sup>6</sup> Department of Environmental Engineering, Sciences and Technology Institute, São Paulo State University (UNESP), Campus São José dos Campos, SP 12247-004, Brazil; rogerio.negri@unesp.br

\* Correspondence: ma.dias@unesp.br

Received: 7 October 2019; Accepted: 18 December 2019; Published: 20 December 2019



**Abstract:** The potential applications of computational tools, such as anomaly detection and incongruence, for analyzing data attract much attention from the scientific research community. However, there remains a need for more studies to determine how anomaly detection and incongruence applied to analyze data of static images from remote sensing will assist in detecting water pollution. In this study, an incongruence-based anomaly detection strategy for analyzing water pollution in images from remote sensing is presented. Our strategy semi-automatically detects occurrences of one type of anomaly based on the divergence between two image classifications (contextual and non-contextual). The results indicate that our strategy accurately analyzes the majority of images. Incongruence as a strategy for detecting anomalies in real-application (non-synthetic) data found in images from remote sensing is relevant for recognizing crude oil close to open water bodies or water pollution caused by the presence of brown mud in large rivers. It can also assist surveillance systems by detecting environmental disasters or performing mappings.

**Keywords:** remote sensing; incongruence; anomaly detection; classification; analysis of images; pattern recognition

## 1. Introduction

The potential applications of computational tools for analyzing remote sensing data taken from environments containing large rivers attract much attention in the scientific literature [1–12]. Two of these tools, anomaly detection [13,14] and incongruence [15–18], are important for studies to find and categorize non-conforming patterns present in images, such as outliers [19–24]. For example, Qi Liu et al. in [25] presented a framework to detect outliers and anomalous events [19–24]. They highlighted the importance of basing anomaly detection studies on taxonomies of anomalies.

Kittler et al. in [13] proposed a taxonomy of anomalies which expanded the concept of anomaly beyond the conventional meaning of outlier. They used sensory data quality assessment [26], contextual [9] and non-contextual [27–30] classifiers [5,15–18], and an incongruence indicator [16–18] to identify each type of anomaly [13]. According to this taxonomy [13] anomalies can be, for example, of the types *unknown object*, *measurement model drift*, *unknown structure*, *unexpected structural component*, *component model drift*, and *unexpected structure and structural components*. This taxonomy [13] is well-known and widely accepted by the scientific community because it has the potential to be applied for solving problems in many different research areas [13]. Therefore, studies related to the application of the taxonomy [13] on synthetic data are common, such as in [16–18]. However, to the knowledge of the authors, studies have not addressed the practical application of the taxonomy [13] to solve real-world problems [2,31–33], because it remains a challenge for all research areas.

Anomaly detection [13] and incongruence [15–18] are two powerful computational tools from pattern recognition (PR) [3,11–18,34,35] and computer vision (CV) [3,36]. PR is a scientific area of study dedicated to analyze patterns and regularities in data [3]. PR provides powerful tools [34] for many different applications and research areas [35] such as scientific research, private and public industries, military activities, etc., [14,21–24,32,37–48]. For example, PR is important for geosciences as its tools are used to analyze geographical features of environments in digital images from remote sensing, i.e., scenes [11,12,49–54]. Additionally, PR also provides powerful tools to help machine perception for CV [3,36]. CV is a scientific area of study committed to developing artificial systems responsible for obtaining information from multidimensional data, such as images [36]. A relevant example of machine perception is the ability of computers to recognize behaviors of geographical features of interest, such as rivers, lakes, fields, plantations, forests, etc. Our interest goes further than a solution based on PR and CV for a human user of a computer to apply in order to solve a specific and already-known problem from remote sensing, such as only detecting water pollution [55–60] as an outlier in an image. In this case, we could use spectral analysis [61] (spectral index [62] or slope ratio [63]) to solve our problem directly. Our motivation has been to find a real application to the taxonomy [13] inspired by the real-world situations captured in images from remote sensing [2,31–33].

When analyzing an image for the first time, a machine could apply each of all known solutions already used in remote sensing (e.g. spectral analysis [61]), one at a time, until detecting a problem. However, this could be considered trial and error learning [35]. The use of the taxonomy [13] can bring to the computational systems a more organized and structured approach of machine learning [28,35,64–74] for detecting problems, i.e., more “intelligence.” In this case, for example, a decision-making system should itself first identify a problem with the minimum interaction with humans before offering an adequate solution that could be based on spectral analysis [61]. A decision-making system [13,16–18] is a software installed in a computer which needs to identify different types of problems in order to automatically select appropriate responses to deal with each type of problem [13]. In other words, the machine should choose the appropriate response to the detected problem based on the type of anomaly identified by the system that could be, for example, *unexpected structure and structural components* [13]. Our research is applied in this context.

The innovation of this work is to introduce the first solution to a practical application of the taxonomy [13] to solve real-world problems [2,31–33]. What makes our study unique is that our study allows computers to categorize real-world problems based on the use of a thorough taxonomy [13] which is a “scaffolding” [75] to increase the machine’s abilities [35]. In other words, beyond identifying problems, this study opens precedents for the machines to recognize the contexts in which the problems are inserted. Each type of anomaly represents a different context [13]. Therefore, the most appropriate response to a real-world problem depends on the context in which that problem is inserted [13]. Humans can learn how to solve problems inserted in different contexts much better than machines [75]. However, our study focuses on machines rather than humans. Providing conditions for a machine to learn how to distinguish and detect different real-world problems taking into consideration the

different contexts and based on a single concept (the taxonomy [13]) make our study unique. This leads to a higher level of meaningful learning, i.e., an evolution in machine learning [28,35,64–74].

Among all relevant areas of research, we have found the first solution to this problem just in remote sensing. It opens up an entirely new perspective and a set of possibilities for future work in remote sensing and many other areas. In remote sensing, for example, the already known solutions for specific problems, such as those based on spectral analysis [61], can be used to spread the taxonomy [13] to be applied on different problems. In this research, the real-world situation captured in images from remote sensing which inspired us to investigate solutions in PR and CV is water pollution, mainly regarding occurrences of disasters [31–33]. If this situation could be revealed by the disagreement between classifiers [13,15–18], then water pollution could be discovered by anomaly detection [13].

Therefore, this paper presents an incongruence-based anomaly detection strategy for analyzing images from remote sensing, with the aim of describing its practical application in detecting water pollution. Therefore, the practical application of the taxonomy [13] is introduced as a strategy for detecting anomalies in real-world application [31–33], i.e., non-synthetic data present in images from remote sensing. We used divergence [16,17] between two image classifications [5,13,15–18] (contextual [9] and non-contextual [8,27–30]) in order to detect incongruence [13,15–18], revealing the presence of anomalies [13]. This research focuses only on the detection of anomalies of the type *unexpected structure and structural components*, which have been described by Kittler et al. in [13].

This paper is organized as follows: Section 2 presents a background; Section 3 comments on related work; Section 4 summarizes the proposed strategy; Section 5 describes the materials and methods used for performing the experiments; Section 6 reviews the achieved results; Section 7 discusses the results and Section 8 presents the conclusions.

## 2. Background

### 2.1. The Differences between Outlier, Anomaly, and Incongruence

In a statistical analysis, an outlier [19–24] is an observation which deviates significantly from other observations [19]. The deviation makes the observation appear to be generated by a different mechanism. Therefore, the observation appears to be inconsistent with the remainder of the data series. For example, consider an image of a large river for which all pixels are blue except one red pixel. In this case, the red pixel is an outlier. Outlier, in the conventional meaning, is a term used to define any type of anomaly [19]. The term “detection of outlier” [19–24] must be used when it is not necessary to categorize the anomaly. The term “detection of anomaly” [13] must be used only when it is necessary to categorize the anomaly. Therefore, it is common to find inadequate exchange of these terms in the scientific literature [14,19].

Anomaly detection is the activity of discovering and categorizing non-conforming patterns and occurrences when there is a failure to relate observed sensory data to information we already know about the image [13]. For example, the detection of absence of water in the part of an image that is correspondent to a large river is an out-of-pattern occurrence and, therefore, an anomaly. Because there are different types of anomalies [13], their detection must distinguish each type of anomaly.

For example, anomalies of the type *unexpected structure and structural components* are those ones detected when the analyses are performed taking into consideration the features which allow classifiers to identify components of samples and how they are structured to compose the samples [13]. The same analyses are applied to detect where differences related to components or their structures can be found in the image, i.e., spatial analyses [13]. Nonetheless, it does not consider when they occurred, i.e., temporal analyses [13]. Therefore, there is no time-series [76–78], i.e., satellite images before and after the occurrence of an event [19] or disaster [31–33], involved in the context. Moreover, detection of anomaly of this type depends on other three conditions to be satisfied: high sensory data quality [26], contextual [9], and non-contextual [8,27–30] classification [5,13,15–18] of the same image, and incongruence [13,15–18]. An example of this type of anomaly is related to detecting brown mud

where there are high levels of turbidity in large rivers. In contrast, for example, incongruent results from contextual and non-contextual classification of images qualified with low values indicate the presence of another type of anomaly [13]. In another example, incongruent results from contextual and non-contextual classification of time-series images [76–78] qualified with high values indicate the presence of a third type of anomaly [13].

Incongruence [13,15–18] is the presence of discordant results and occurs whenever there is contradictory evidence provided by the analyses of sensory data performed by different types of classifiers [13]. For example, taking into consideration the two parallel analyses of the same part of an image that is correspondent to a large river, there is incongruence if water is detected by one analysis and not detected by the other one.

## 2.2. Non-Contextual and Contextual Classifiers

An environment can be modeled by some computational tools, e.g., those based on machine perception, using samples (aggregation of pixels) from the application domain and previous information about the problem [15]. Models trained with different samples are able to be represented by classifiers [16]. In PR, classifiers are experts or learners which perform classifications, i.e., they assign objects to one class or category among predefined others [79]. While predicting classes for a common input, such as a scene, classifiers are expected to present similar probability estimates despite the environment [16]. Consequently, computational tools, which base their decision-making actions on the use of multiple classifiers, usually expect all classifiers to support the same hypothesis [17].

There are two different types of classifiers, non-contextual and contextual [13]. Non-contextual classification is a general-level task that results in a weak learner, but it is less constrained than the contextual one [79]. It is more powerful than a random classifier, despite being much less precise when compared to the contextual one [79]. Contextual classification is a specific-level task which leads to a strong learner, but it is more dependent on specific knowledge such as previous knowledge or training data [9]. It is composed of multiple weaker classifiers that work in synergy to reach a more robust classification [9].

On one hand, when either two non-contextual [27–30] or two contextual [9] classifiers diverge in their classifications [13,15–18], this condition only exposes the limited classification of one of the classifiers. It does not present any relation to incongruence detection [13,15–18] for both classifiers. On the other hand, models of an environment which are represented by non-contextual and contextual classifiers can be used for identifying incongruences if they are used concurrently [15]. After receiving input data, both types of classifiers produce class-posterior probabilities as results [15]. Incongruence happens when the input data presented to both classifiers induces a significantly large discrepancy between posterior probabilities, i.e., incongruence is a conflicting prediction that happens when the probability calculated by some non-contextual classifier is much bigger than the probability calculated by some contextual classifier [15].

Therefore, the concurrent use of non-contextual and contextual classifiers allows systems to recognize the occurrence of conflicting classifications [15]. It enhances the capability of decision-making systems to detect incongruence [15–18]. Consequently, decision-making systems can have their control actions also conditioned by the monitoring and detection of incongruence [15–18]. Decision-making systems can guide computers to give adequate and fast responses after detecting problems [13], e.g., to prevent water pollution disasters or to analyze their extents in the case of their occurrences. For this reason, incongruence is of considerable interest for decision-making systems, since the strategy reveals the occurrence of some type of anomaly [13].

### 3. Related Work

#### 3.1. Research on Outliers Detection

Most current research focuses on outlier detection, such as [1,14,19–25,44,51–54,66,76–78,80–82], however we discuss below only those research which are more relevant to this article. An outlier-based change detection method to detect abnormal points in multi-scale time-series images from remote sensing was proposed by Yin Shoujing et al. in [76]. The method analyzes time-series images from remote sensing and detects temporal and spatial changes. For that research, changes caused by weather condition variation, climate changes, sensor aging, vegetation phenology, human activities, and emergencies (such as drought, fire, pest, insect, etc.,) were considered abnormal. The research focused on land cover changes. These changes are relatively rare over large areas, even for a long period. The research concluded that the land cover changes registered in the images were outliers.

An algorithm capable of identifying outliers in the global data (distributed earth science databases) without moving the whole data to a single database was presented by Bhaduri et al. in [80]. The algorithm analyzes the data to search for extremes or outliers. The algorithm detects outliers which are missed when data is available only at a single database. Their achieved results were compared to the achievements of other algorithms which also detect outliers. An unsupervised outlier detection framework was proposed by Qi Liu et al. in [25]. The framework requires no prior knowledge to detect outliers and anomalous events. The research defined outliers as objects which have either low spatial or temporal coherence with their neighbors, and it defined an anomalous event as a group of outliers that share similar spatial and temporal anomalous behaviors. The research has defined four different categories of outliers or anomalous events and has reinforced the importance of basing anomaly detection studies on taxonomies of anomalies. Nonetheless, outlier detection is not enough for systems to categorize many types of problems. Anomaly detection can overcome this drawback.

A new daily snow cover dataset was developed by Bormann et al. in [81]. The dataset is a satellite-based observational record useful to characterize snow duration and cover extent. These snow cover observations are important to detect outliers such as anticipated melt of snow, declines in snow cover extent, and short season duration. The dataset was assessed using snow detection estimates derived from Landsat Thematic Mapper data and was compared to another snow cover dataset. Although the study focused on satellite data for the alpine region in Australia, the approach can be applied to other regions and other sensors to help assess snow monitoring. Consequently, the study contributes to research on water resource management and snow hydrology.

A method for detecting anomaly regions in each image of satellite image time-series was proposed by Zhou et al. in [77]. The method identifies spatial-temporal dynamic processes of unexpected changes of land cover based on seasonal autocorrelation analysis. The detection of spatial-temporal processes of flooding in satellite data was used to assess the method. A deep-learning approach for change detection applied to satellite time-series images was presented by Sublime and Kalinicheva in [66]. Their approach was compared against other machine-learning methods. The achieved results are better than the results of other existing techniques because of the method's higher performance and relative fast analysis. An online change detection algorithm was proposed by Chandola and Vatsavai in [78]. They used a Gaussian process based non-parametric time-series prediction model in an online mode. The process solves a large system of equations involving the associated covariance matrix which grows with every time step. Their algorithm identifies changes monitoring the difference between the predictions and previous observations within a statistical control chart framework.

An automatic clearance anomaly detection method was proposed by Chen et al. in [82]. The clearance anomaly detection measures the distance between power lines and other objects (e.g., trees and buildings) to evaluate whether the clearance is within the safe range. The clearance measurements were compared with the standard safe threshold to find the clearance anomalies. The results were validated through qualitatively visual inspection, quantitatively manual measurements in raw point



clouds, and on-site field survey. The achieved results show that the proposed method detects the clearance hazards, such as tree encroachment, effectively, and the clearance measurement is accurate.

### 3.2. Research on Anomaly Detection

Anomaly detection can be used by systems to categorize problems based on the use of taxonomies, such as in [13]. Regarding identification of different types of anomalies useful to help decision-making systems, a unified framework for anomaly detection was proposed by Kittler et al. in [13]. According to them, the scientific literature conventionally understands an anomaly as an outlier. However, their framework has expanded the concept of anomaly beyond the conventional meaning of outlier. They claim that anomalies are related to many other occurrences, such as rare events, unexpected events, distribution drift, noise, and novelty detection of an object or an object primitive. The framework presented the multifaceted nature of anomalies and suggested effective mechanisms to identify and distinguish each of them. The research has provided a taxonomy of anomalies that includes, for example, *unknown object*, *measurement model drift*, *unknown structure*, *unexpected structural component*, *component model drift*, and *unexpected structure and structural components*. Because the practical application of the taxonomy is relatively new, and the detection and identification of the type of the anomaly involves complex aspects, it is appropriate to study the practical application of each type of anomaly individually. The research asserts that by identifying different types of anomalies, systems can select appropriate responses to deal with each type of anomaly. For decision-making systems, the application of the anomaly detection is potentially increased when it is based on incongruence.

### 3.3. Research on Incongruence

The use of incongruence for detecting the occurrences of some types of events was deeply discussed by Weinshall et al. in [15]. They identified distinct types of events based on the conflicting predictions given by weak and strong classifiers. Their study presented a framework for the representation and processing of those incongruent events. According to them, an incongruent event is an event for which the probability is divergent when computed based on the use of different types of classifiers. They applied their methodology on pictures of dogs and motorcycles, and indoor videos recorded to simulate common situations, such as people walking and talking. Similar to the other studies we have reviewed, this work also does not present the application of their methodology to solve real-world problems.

The decision cognizant Kullback–Leibler (DC-KL) divergence was proposed by Ponti et al. in [16]. The DC-KL divergence reduces the contribution of the minority classes which obscure the true degree of classifier incongruence. Simulations, i.e., synthetic data, were used to analyze the properties of the novel divergence measure. The measure is more robust to minority class clutter, less sensitive to estimation noise, and achieves much better statistics for discriminating between classifier congruence and incongruence than the classical KL divergence.

The delta divergence measure was proposed by Kittler and Zor in [17]. This measure focuses on the most probable hypotheses. Consequently, it reduces the effect of the probability mass distributed over the non-dominant hypotheses. The measure satisfies symmetry and independence of classifier confidence.

A new classifier incongruence measure was proposed by Kittler and Zor in [18]. The measure overcomes the shortcomings of previous measures and presents relatively low sensitivity to estimation noise, under the assumption of constrained Gaussian distribution. Moreover, the measure determines incongruence thresholds at given levels of statistical significance for different measured values corrupted with different levels of noise.

### 3.4. Research on Real-World Problems, Which Inspired This Study

The studies commented in this subsection inspired us to investigate the potential application of the taxonomy [13] to solve real-world problems. Spatially explicit long-term monitoring frameworks and

priority mitigation measures to cope with acute and chronic risks were proposed by Fernandes et al. in [31]. The research focused on the dam failure in Mariana, Brazil, where a dam breach abruptly discharged between 55 and 62 million m<sup>3</sup> of ore tailings into the Doce River. They claimed that environmental disasters like that of the Doce River would become more frequent in Brazil. The study is based on differential analyses of Landsat 8 scenes.

An automatic approach to enhance and detect river networks was proposed by Kang Yang in [2]. The research characterizes rivers in accordance with their Gaussian-like cross-sections and longitudinal continuity. River cross-sections were enhanced by a Gabor filter. The approach was applied to detect rivers in Landsat 8 images.

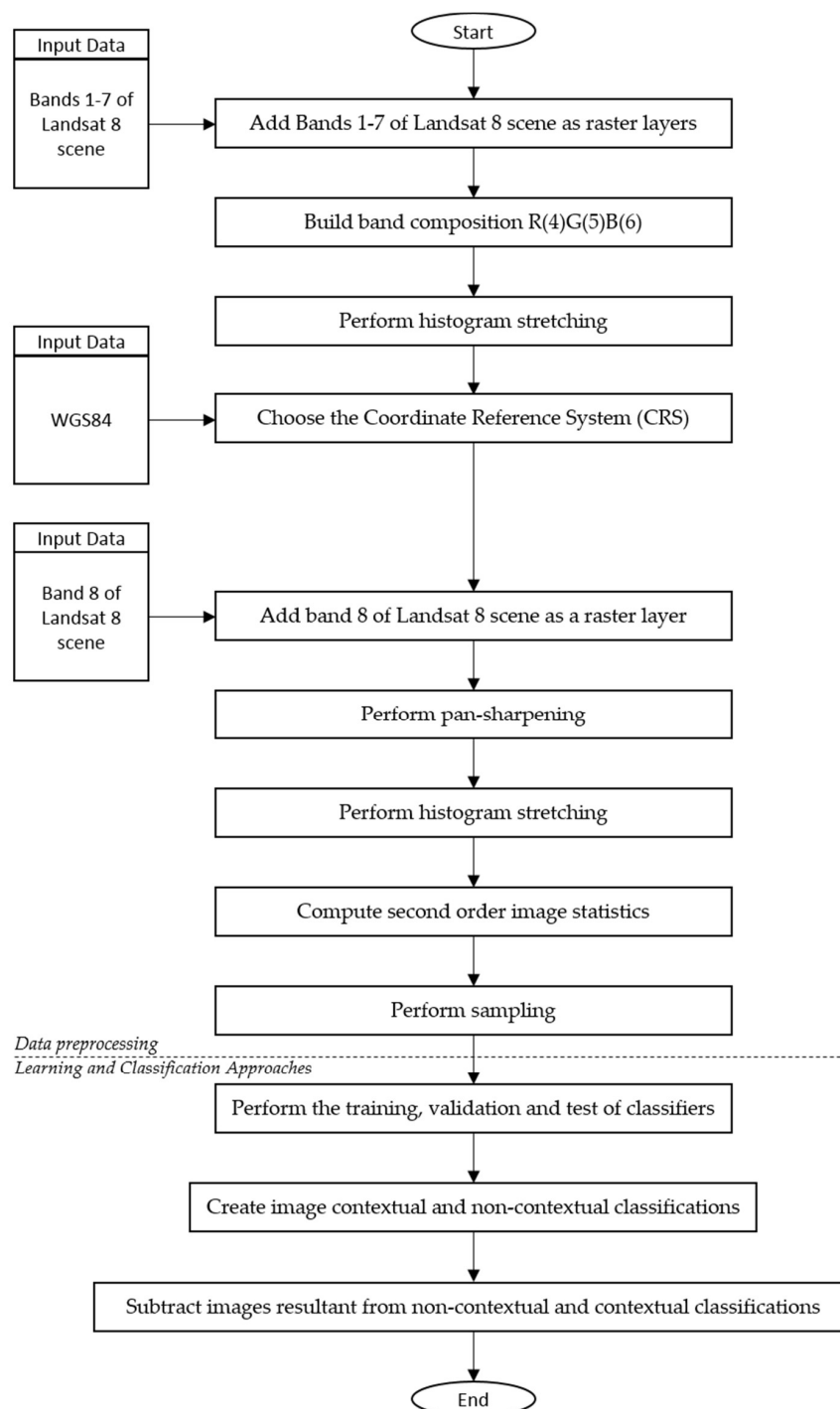
A study which focuses on the mapping and monitoring of the spatial extent of surfaces affected by mine waste was presented by Mielke et al. in [32]. They studied the influence of sensors on the ability to discriminate mine waste surfaces from their surroundings. The research explored the potential use of images from remote sensing to map and monitor the spatial extent of mine waste surface material in areas with mine tailings. They claimed that remote sensing analysis is very important to monitor the extent of mine waste surfaces and that mine waste sites have the potential to contain problematic contaminants, such as chrome, lead, uranium, etc. The research was based on the use of Landsat 8 data.

The impacts of oil extraction on the environment and health of indigenous communities in the Northern Peruvian Amazon (Marañón) were assessed by Rosell-Melé et al. in [33]. They investigated the occurrence of crude oil pollution in soils and river sediments caused by voluntary discharges or accidental spills. Their study is related to a very productive oil area in Peru, which is placed in the headwaters of the Amazon River. Their findings suggest that wildlife and indigenous populations in the Marañón region are exposed to the ingestion of oil, and local spillage of oil in the watercourses could have eventually reached the Amazon River.

## 4. The Proposed Strategy

### 4.1. Introduction to the Proposed Strategy

The flowchart in Figure 1 summarizes the methodology of the proposed strategy. More detailed descriptions of this methodology as well as of each parameter used here can be found in Section 5 and Appendix A. All chosen Landsat 8 images (see the fourth column in Table 1) were thoroughly evaluated by applying the proposed strategy. Each environment was evaluated separately. Many Landsat 8 scenes were previously checked visually to assess their conditions in order to select only those with low overlay contamination by cloud cover. This is critical to making geographical features as visible as possible. Afterwards, each Landsat 8 scene (see the fourth column in Table 1) was chosen according to its date of acquisition. In cases in which disaster had already occurred, the scene with the date of acquisition closest to the date of the disaster (see the fifth column in Table 1) was chosen in order to best evaluate the environment after the disaster. In the cases in which no environmental disaster had occurred, the most recent images with low overlay contamination by cloud cover were chosen (that is the case of the Athabasca, Elbe and Tietê Rivers).



**Figure 1.** Flowchart of the proposed incongruence-based anomaly detection strategy. Landsat 8 band used, wavelength, and resolution: Band 1—coastal aerosol (0.43–0.45  $\mu\text{m}$ ) 30 m; Band 2—blue (0.45–0.51  $\mu\text{m}$ ) 30 m; Band 3—green (0.53–0.59  $\mu\text{m}$ ) 30 m; Band 4—red (0.64–0.67  $\mu\text{m}$ ) 30 m; Band 5—near infrared (NIR) (0.85–0.88  $\mu\text{m}$ ) 30 m; Band 6—SWIR 1 (1.57–1.65  $\mu\text{m}$ ) 30 m; Band 7—SWIR 2 (2.11–2.29  $\mu\text{m}$ ) 30 m, and Band 8—panchromatic (PAN) (0.50–0.68  $\mu\text{m}$ ) 15 m. The coordinate reference system (CRS) adopted uses the WGS84 geodetic reference frame and the universal transverse mercator (UTM) map projection. The samples selected by the user (under input data) can be lakes, rivers, artificial reservoirs, waterways, oceans, fields, forests, clouds, cities, highways, mountains, plantations, and shadows of clouds. The figure also shows the border between the data preprocessing and learning and classification approaches.



**Table 1.** Information about the studied environments.

Environment and Country	UTM Zone and Center: Latitude/Longitude	Disaster	Landsat 8 Image (Scene)	Date of the: Image Acquisition/Environmental Disaster	Number of Samples
Doce River Brazil	23 20°13'48.07"S 42°43'47.24"W	Dam Breach	LC08_L1TP_217074_20151112_20170402_01_T1	12 November 2015 5 November 2015	250
Athabasca River Canada	12 57°18'35.86"N 112°33'13.21"W	—	LC08_L1TP_043020_20170826_20170913_01_T1	26 August 2017 No Disaster	100
Elbe River Germany	32 54°30'31.36"N 9°29'07.80"E	—	LC08_L1TP_196022_20180509_20180517_01_T1	9 May 2018 No Disaster	100
Marañon River Peru	18 4°20'20.54"S 74°45'45.97"W	Pipeline Disruption	LC08_L1TP_007063_20160827_20170321_01_T1	27 August 2016 22 August 2016	100
Tietê River Brazil	23 23°06'46.33"S 46°30'16.96"W	Permanent Sewage Disposal	LC08_L1TP_219076_20170827_20170914_01_T1	27 August 2017 Permanently	100
Arava Valley Israel	36 30°18'22.86"N 35°02'25.51"E	Pipeline Disruption	LC08_L1TP_174039_20141214_20170416_01_T1	14 December 2014 4 December 2014	100

#### 4.2. The Advantages of Our Strategy

The following are advantages of our strategy: (1) It takes into consideration the multifaceted nature of anomalies exposed by [13]. (2) It helps to offer an appropriate response to deal with anomalies of the type *unexpected structure and structural components*. (3) To the knowledge of the authors, this is the first effort of a series of studies dedicated to investigate the practical application of detection of anomalies in accordance with the taxonomy published in [13] to solve real-world problems. (4) This research overcomes the number of categories studied in previous investigations, e.g., [25], which presented four different anomalies, because our strategy is based on that taxonomy [13] which presents ten different domain anomalies. (5) It processes just a single Landsat 8 image (scene), whereas other proposed anomaly detection methods use time-series of Landsat 8 images for each environment to be able to detect outliers. Time-series will only be used to detect anomalies of the types *measurement model drift* and *component model drift* in future papers.

Additionally, this research has established a relationship between the occurrence of anomalies of the type *unexpected structure and structural components* and the presence of brown mud in large rivers. Let us consider the example of a mining company which needs to monitor whether the brown mud of its reservoir is polluting a large river. In this example, the mining company does not need to deal with any other type of pollutant other than brown mud, therefore a system able to detect only brown mud is enough for the company. The fewer the images, the fewer the demands for computational resources, consequently making the system cheaper. Moreover, the fewer the images the less data there is to be processed, consequently making the system faster. Therefore, a system based on the application of the proposed strategy, which deals with only one image per environment, will be advantageous. This system will save more money and time for the mining company than systems based on other proposed anomaly detection methods, because they generally deal with time-series.

For the case studies presented in this paper, the anomalies are related to brown mud or crude oil present as either potential contaminants or contaminators of large rivers. Our strategy identifies incongruences to detect oil spills in some areas where there is water, or brown mud, and where there are high levels of turbidity in the water. These findings are important because an incongruence-based anomaly detection strategy can greatly increase the ability of surveillance systems to detect environmental disasters and to perform mappings.

## 5. Materials and Method

### 5.1. Materials

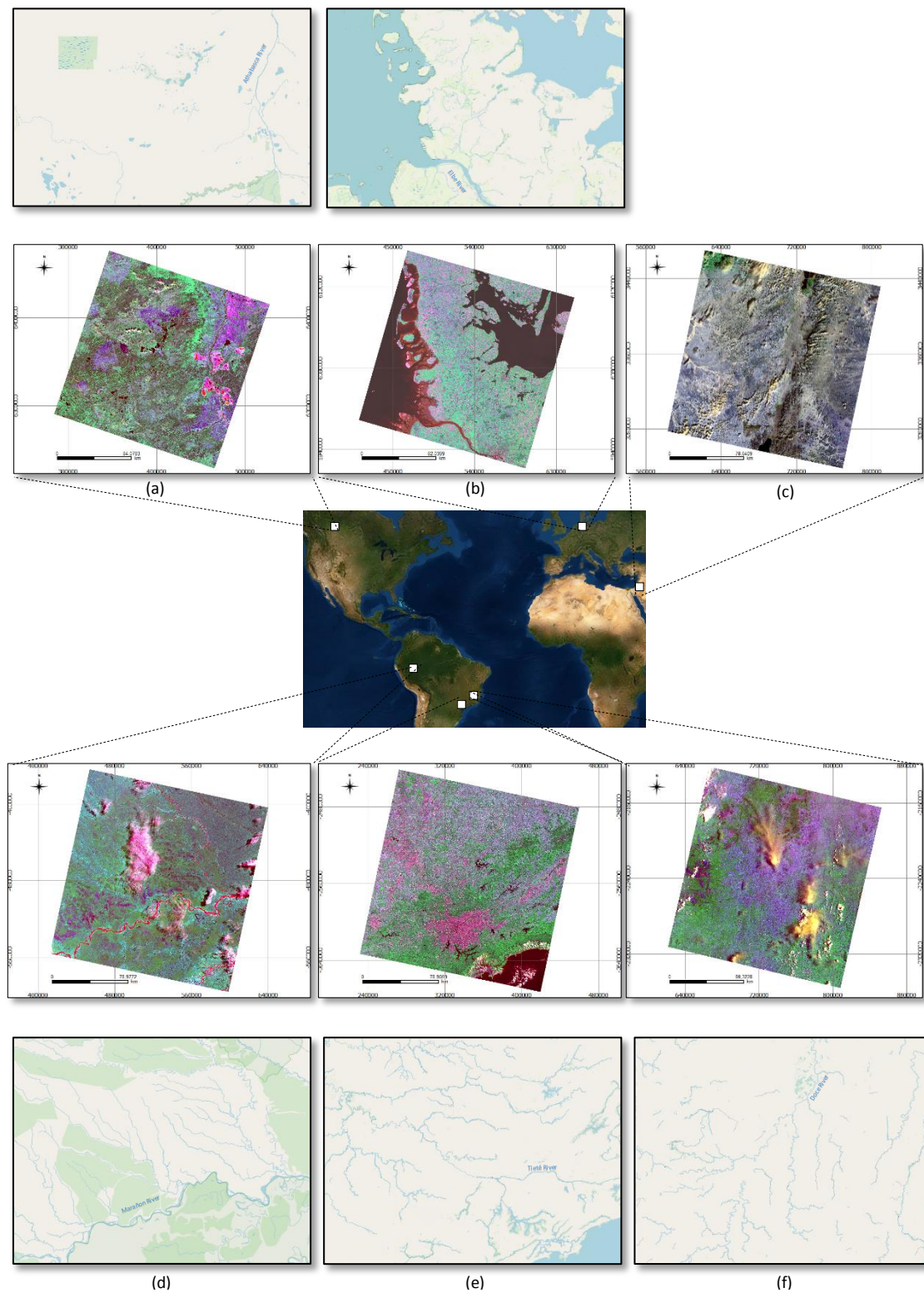
Landsat 8 images (scenes of environments) from the United States Geological Survey (USGS) data sets and benchmarks [83] were used to evaluate the proposed strategy. The USGS consists of Landsat images and many other important data set collections, which are all freely available in [83] for non-commercial purposes. These data sets were chosen to perform the experiments in this study because they provide diversity regarding the images and geographical features for many different environments. Moreover, they have been frequently used to evaluate many recently-published proposals, such as [2,31,32], which allows us to compare the present proposed strategy with other studies. For this study, each scene was provided by the USGS through the Earth Explorer Platform.

All images used in this study are qualified with value “9,” which represents the best quality for Landsat 8 images [26]. Moreover, all Landsat 8 data used in the present experiments were collected by the operational land imager (OLI), which is an instrument onboard the Landsat 8 satellite. The size of each image is approximately 170 km north-south by 183 km east-west, i.e., 106 mi by 114 mi. These images are available in the tagged image file format (TIFF). The three most common types of TIFF images are gray scale, indexed, and RGB.

For this study, each task of our strategy, including all classifications, was systematically performed using the Qgis 2.18.19 (Las Palmas) software with the Orfeo tool box. The software and the tool box are all freely available in [84] for non-commercial purposes. Qgis is a well-known and extensively used software composed of many tool sets, which are useful for analyzing images, mainly from remote sensing, for geosciences.

### 5.2. Study Areas

Figure 2 shows the scenes and the hydrographical maps of the six different areas studied in this research. The information about the studied environments, including the identification of the Landsat 8 images (scenes) used in the experiments, is listed in Table 1. The geographical settings are listed in Table 2. All data are related to areas of large river basins, except the area of the Arava Valley. The image of the Arava Valley in southern Israel, acquired by the Landsat 8 satellite on 14 December 2014, was chosen because its soil received an estimated 3–5 million liters of crude oil on 4 December 2014 (29°40′29.28″N 35°0′25.56″E) [85]. This image was used to contrast the experiments with the Marañon River since there is absence of water in that part of the Arava Valley. This contrast is important because only two classes were defined for all experiments: “water,” which categorizes any open water body, and “no-water,” which categorizes anything different from open water bodies, for example land, trees, cities (this approach is similar to the one found in [86]). The idea is to evaluate if the proposed strategy would detect oil as if it were water. If oil were to be detected as water, then all experiments with images of the Marañon River basin would be invalidated. The image of the Marañon River in Peru, acquired by the Landsat 8 satellite on 27 August 2016, was chosen because its water received high quantity of crude oil on 22 August 2016. This oil spill was the consequence of a Colombian environmental disaster caused by a pipeline disruption (4°48′45.6″S 75°23′56.8″W) [87]. The crude oil is present around almost all extensions of the same pipeline since it was the source of other similar environmental disasters in the preceding three years.



**Figure 2.** The band compositions (R(4)G(5)B(6)) and the hydrographical maps of the six different areas studied in this research. (a) Athabasca River in Canada. (b) Elbe River in Germany. (c) Arava Valley in Israel. (d) Marañon River in Peru. (e) Tietê River in Brazil. (f) Doce River in Brazil. A band composition is an image composed of different bands of a satellite image.

Table 2. Geographical settings.

Environment and Country	Basin Size	Length of the River	Predominant Features in the Environment	Human Activity in the Region	Potential Threat to the Environment	Potential Contaminators or Contaminators
Doce River Brazil	83,400 km <sup>2</sup> (32,201 sq mi)	853 km (530 mi)	Mountains and Valleys	Mining	Ore Tailings Reservoirs	Brown Mud
Athabasca River Canada	95,300 km <sup>2</sup> (36,800 sq mi)	1231 km (765 mi)	Lakes	Mining	Ore Tailings Reservoirs	Brown and Red Mud
Elbe River Germany	148,268 km <sup>2</sup> (57,247 sq mi)	1094 km (680 mi)	Close to the Coastline	Mining	Ore Tailings Reservoirs	Red Mud
Marañon River Peru	358,000 km <sup>2</sup> (138,000 sq mi)	1737 km (1079 mi)	Forest	Transport of Crude Oil	Oil Pipeline	Crude Oil
Tietê River Brazil	150,000 km <sup>2</sup> (58,000 sq mi)	1150 km (710 mi)	Densely Populated Region	Sewage Disposal	Sewer	Sewage
Arava Valley Israel	_____	_____	Arid Valley	Transport of Crude Oil	Oil Pipeline	Crude Oil

The image of the Doce River in Brazil, acquired by the Landsat 8 satellite on 12 November 2015, was chosen because its water received around 55–62 million m<sup>3</sup> of iron ore tailings (brown mud) on 5 November 2015, as a result of Brazil's worst environmental disaster caused by a dam breach (20°12'23.4"S 43°28'01.6"W) [31]. The image of the Elbe River in Germany, acquired by the Landsat 8 satellite on 9 May 2018, was chosen because of the location of a dam close to its water (53°38'20.4"N 9°25'02.0"E). The river could receive many millions of m<sup>3</sup> of bauxite tailings (red mud) in the case of an environmental disaster similar to that which occurred in Brazil. The image of the Athabasca River in Canada, acquired by the Landsat 8 satellite on 26 August 2017, was chosen because there is a set of dams located close to its water (57°03'06.3"N 111°35'18.0"W), which could cause the river to receive many millions of m<sup>3</sup> of brown or red mud in the case of an environmental disaster similar to that which occurred in Brazil. The image of the Tietê River in Brazil, acquired by the Landsat 8 satellite on 27 August 2017, was chosen because its water receives significant quantities of sewage each year (23°23'24.3"S 46°58'36.5"W).

For each Landsat 8 image used in this study, at least 100 samples were located and validated based on the topographic maps of the environments, adopted as ground truth for the purpose of this work. There are samples spread all over the scene. The numbers of samples used as training data to model each environment according to this approach are shown in the sixth column of Table 1, being 50% constituted by open water bodies and 50% by other geographical features. Regarding samples of open water bodies which are expected to be classified as “water” by the classifiers, the chosen areas were carefully selected from lakes, small rivers, artificial reservoirs, waterways, ocean (if present), etc., present in the image. Regarding samples of other geographical features which are expected to be classified as “no-water” by the classifiers, the chosen areas were carefully selected from fields, forests, clouds, cities, highways, mountains, plantations, shadows of clouds, etc., present in the image.

Each image presents different difficulty levels to locate and analyze geographical features, in the environments. For example, in any of these environments it is possible to find parts of the image with an easy to detect, large, and singular geographical feature that differs from other parts of the image with hard to detect, small, numerous, and geographically spread features. By using diversified environments, it was possible to verify the accuracy of the proposed strategy to identify incongruences in order to detect anomalies, when it is applied to different scenes as well as to different types of geographical features present in images.



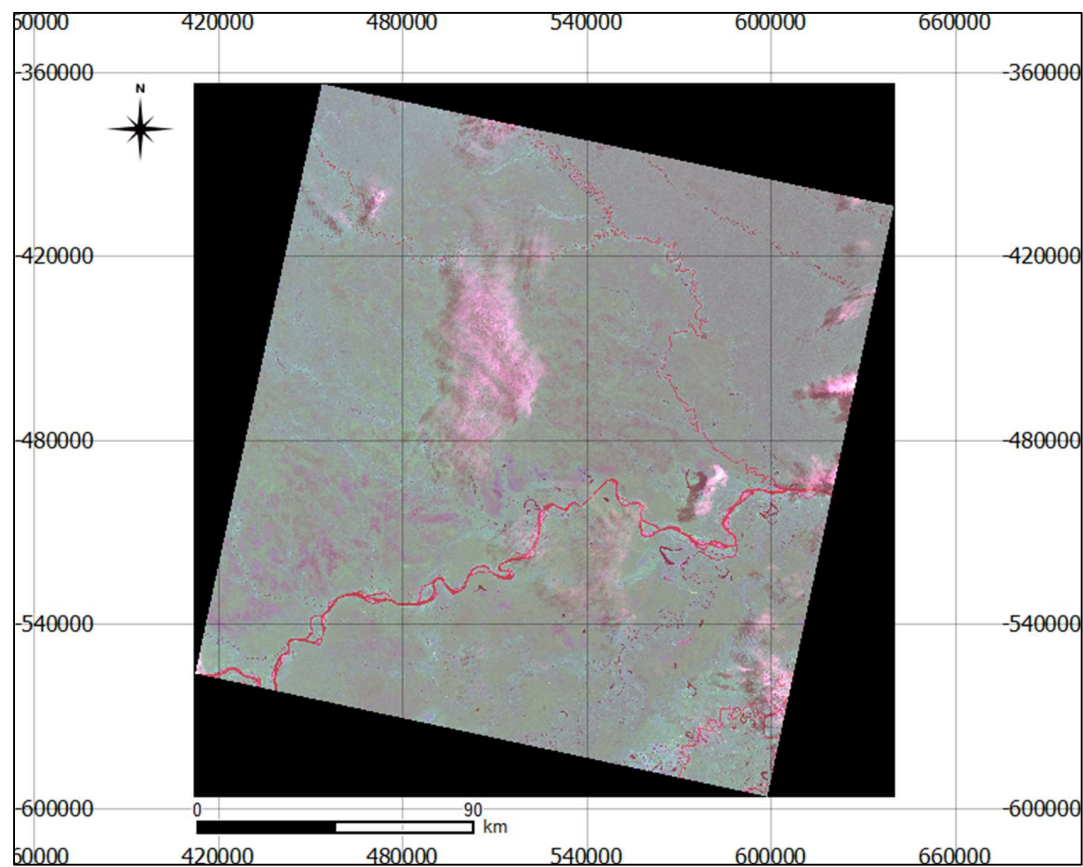
### 5.3. Method

#### 5.3.1. Data Preprocessing

Each image was processed performing tasks to identify incongruences in accordance with Figure 1. Initially, seven bands from a Landsat 8 scene were added as raster layers to a new project of Qgis: Band 1—coastal aerosol (0.43–0.45  $\mu\text{m}$ ) 30 m; Band 2—blue (0.45–0.51  $\mu\text{m}$ ) 30 m; Band 3—green (0.53–0.59  $\mu\text{m}$ ) 30 m; Band 4—red (0.64–0.67  $\mu\text{m}$ ) 30 m; Band 5—near infrared (NIR) (0.85–0.88  $\mu\text{m}$ ) 30 m; Band 6—SWIR 1 (1.57–1.65  $\mu\text{m}$ ) 30 m; and Band 7—SWIR 2 (2.11–2.29  $\mu\text{m}$ ) 30 m. This first task was done in order to organize the seven aforementioned bands into a single multiband raster to allow easy access.

Second, a band composition was built in order to allow the user of the system to analyze the scene visually. A band composition is an image composed of different bands of a satellite image. The band composition uses pixels of different color values in order to represent various properties of objects from the real world. A band composition is represented by Equation (1), for which, considering each Landsat 8 band as a two-dimensional array of pixels,  $x$  and  $y$  are coordinates of the pixels. The composition  $vr$  is a tridimensional array composed of three overlapped elements (bands), being  $b4(x, y)$  = Band 4,  $b5(x, y)$  = Band 5, and  $b6(x, y)$  = Band 6. Figure 3 shows the band composition R(4)G(5)B(6) as an example.

$$vr(x, y) = \begin{bmatrix} b4(x, y) \\ b5(x, y) \\ b6(x, y) \end{bmatrix} \quad (1)$$



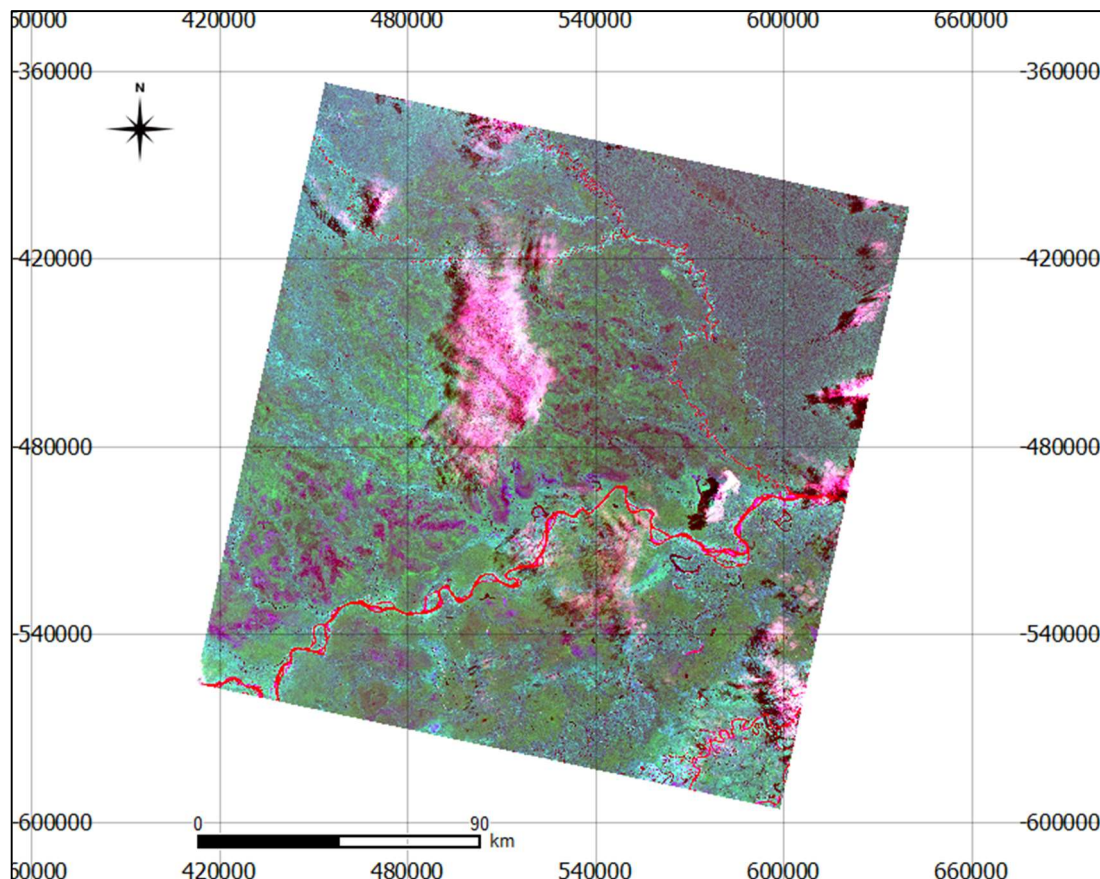
**Figure 3.** Example of a band composition R(4)G(5)B(6) of a Landsat 8 scene related to the Marañon River (in Peru). The Marañon River is the red thick winding line from left to right in the image. In the coordinate grid, the universal transverse mercator (UTM) false northing (10,000 km) was not added to the N coordinate.

Next, we applied a histogram stretching to the composite R(4)G(5)B(6). The procedure was performed by a Qgis tool from the menu properties (style). Because Qgis needs to create a color table in order to render the band composite, the tool calculates the table based on the mean and standard deviation for the three selected bands.

The mean and the standard deviation were calculated respectively according to Equations (2) [88] and (3) [88], for which,  $x$  and  $y$  are coordinates of the pixels,  $g_i(x, y)$  is an image,  $K$  is the total number of images  $g(x, y)$ ,  $i$  is the identifier of each image  $g(x, y)$ , and  $f(x, y)$  is the image formed averaging  $K$  images  $g(x, y)$ . In Equation (3),  $h(x, y)$  is the image formed calculating the standard deviation. Figure 4 shows an example of an enhanced image [89].

$$f(x, y) = \frac{1}{K} \sum_{i=1}^K g_i(x, y) \quad (2)$$

$$h(x, y) = \sqrt{\frac{1}{K} \sum_{i=1}^K (g_i(x, y) - f(x, y))^2} \quad (3)$$



**Figure 4.** Enhanced Landsat 8 scene (R(4)G(5)B(6)) related to the Marañón River (in Peru). The Marañón River is the red thick winding line from left to right in the image.

Shortly after, we created a new project in Qgis to receive the Landsat 8 processed image. The chosen coordinate reference system (CRS) was WGS84 and with the universal transverse mercator (UTM) map projection (see the second column in Table 1). Another band from the Landsat 8 scene was added as a raster layer to the project of Qgis: Band 8 panchromatic (PAN) (0.50–0.68  $\mu\text{m}$ ) 15 m.



Next, a pan-sharpening step based on component substitution (CS) or projection substitution was performed [89]. Equation (4) represents a general formulation of this CS method [90],

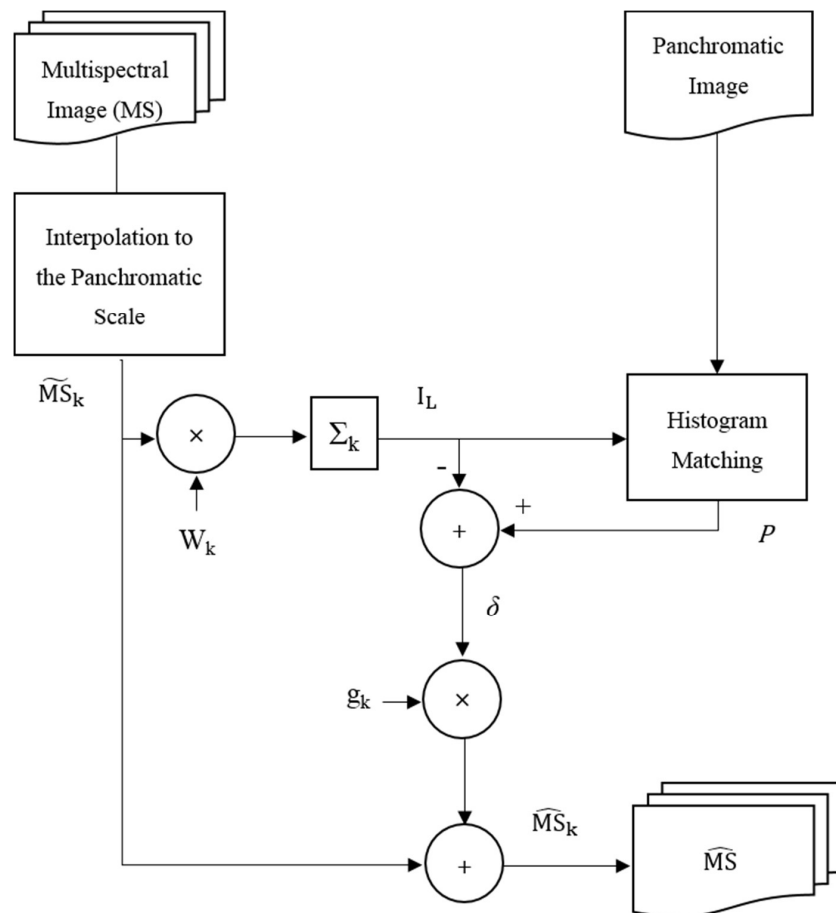
$$\widehat{MS}_k = \widetilde{MS}_k + g_k(P - I_L), \quad k = 1, 2, \dots, N \quad (4)$$

in which  $MS$  is a multispectral image,  $\widehat{MS}$  is a pan-sharpened image,  $\widetilde{MS}$  is a multispectral image interpolated at the scale of the panchromatic image, the subscript  $k$  indicates the  $k^{th}$  spectral band,  $g = [g_1, \dots, g_k, \dots, g_N]$  is the vector of the injection gains,  $P$  is the histogram-matched panchromatic image, and  $I_L$  is defined by the Equation (5) [91],

$$I_L = \sum_{i=1}^N w_i \widetilde{MS}_i \quad (5)$$

in which the weight vector  $w = [w_1, \dots, w_i, \dots, w_N]$  is the first row of the forward transformation matrix and may be chosen, whenever possible, to measure the degrees of spectral overlap among the multispectral and panchromatic channels.

Figure 5 presents a flowchart summarizing how the pan-sharpening method based on this CS method works. The flowchart has blocks responsible to interpolate the multi-scale image for matching the scale of the panchromatic image, calculate the intensity component based on Equation (5), match the histograms of the panchromatic image and the intensity component, and inject the extracted details based on Equation (4).



**Figure 5.** Flowchart of a generic pan-sharpening procedure based on the component substitution method. In this sketch,  $\delta = (P - I_L)$ . Adapted from [91].

For the pan-sharpening method, interpolation must guarantee the overlap of multispectral and panchromatic images at a finer scale. Depending on the acquisition geometry of the imaging instruments, interpolation with conventional zero-phase linear finite-impulse response filters may require a manual realignment, e.g., after a bi-cubic interpolation. Alternatively, linear nonzero-phase filters, having even numbers of coefficients, may be profitably used, according to [91,92].

Although the 30-meter original spectral bands are enough for dealing with large rivers, the pan-sharpening was necessary for dealing with lanes that support oil pipelines. Those lanes, although long, are much narrower than large rivers. Working with 15-meter images made possible the detection of the presence of oil on those narrow lanes. It is important because, for example, there is a very long lane, built in the vicinity of the Marañón River in order to support oil pipelines, that has already received high quantities of crude oil as a consequence of successive oil spills. Whereas those lanes should normally present either dry or humid soil around the pipelines, a significant oil spill turns the soil impermeable causing accumulation of water around the pipelines. The detection of water retained in the lane, therefore, is indicative of the occurrence of an oil spill on the soil of that area.

Next, the pan-sharpening resulting image had its histogram stretched. This step was performed exactly the way it was previously mentioned, including the calculation of Equations (2) [88] and (3) [88].

Then, the second-order image statistics were calculated using the Orfeo toolbox in order to compute the global mean and standard deviation for each band from a set of images. Their applications are related to geometric modeling, i.e., to create a model taking into consideration the spatial distribution of the pixels in the image. The model was created in order to statistically represent the enhanced image, and the statistics are not affected by the enhancing parameters. The second-order image statistics are those for which the slope of the power spectrum tends to be close to negative two. The power spectrum of an  $M$  by  $M$  image is represented by Equation (6) [93], for which,  $F$  is the Fourier transform of the image,  $\varphi$  are directions, and  $u$  and  $v$  are two-dimensional frequencies represented in polar coordinates in accordance with Equations (7) and (8) respectively.

$$S(u, v) = \frac{|F(u, v)|^2}{M^2} \quad (6)$$

$$u = f \cos \varphi \quad (7)$$

$$v = f \sin \varphi \quad (8)$$

According to Reinhard et al. in [93]:

*"(...) averaging over all directions  $\varphi$  and all images in the ensemble, it is found that on log-log scale power as function of frequency  $f$  lies approximately on a straight line (...) This means that spectral power as function of spatial frequency behaves according to a power law function. Moreover, fitting a line through the data points yields a slope  $\alpha$  of approximately negative two for natural images:*

$$S(f) \propto A f^2 = A f^{-2-\eta} \quad (9)$$

*Here,  $\alpha \approx -2$  is the spectral slope,  $\eta$  is its deviation from  $-2$ , and constant  $A$  describes the overall image contrast."*

Contrast can be obtained by image statistics, for example, calculating the standard deviation of all pixel intensities divided by the mean intensity ( $\sigma/\mu$ ) in accordance with Equation (10).

$$\frac{\sigma^2}{\mu^2} = \sum_{(u,v)} S(u, v) \quad (10)$$

Afterward, in the image, at least 100 samples were selected and validated based on maps of the environments (ground truth). The details regarding the chosen samples were described in the

Section 5.2. The selection of samples is important because they are used as training data to model each environment for contextual or non-contextual classifiers. Whenever users select samples for classifiers, their classifications are named supervised or semi-automatic. A supervised classification allows the user to collect samples from the image and to use a classifier trained with these samples to classify the image [90].

The preprocessing is essential to prepare the data to be used for the learning and classification approaches adequately. Our main contribution in the preprocessing is regarding the care for selecting the samples. The samples were not located on the large rivers, ore tailings reservoirs, oil pipelines, or lanes for oil pipelines, in order to avoid the experiments yielding biased results. In other words, no areas located on the large rivers were selected to be samples because these samples could bias the classifiers to classify the large rivers as “water.” No areas located on ore tailings reservoirs, oil pipelines, or lanes for oil pipelines were selected to be samples because these samples could bias the classifiers to classify those structures as “no-water.”

### 5.3.2. Learning and Classification Approaches

The training of classifiers step was performed based on the enhanced 15-meters image, the statistical model, and the selected samples. The training is important in order to create the models which will be used by the contextual or non-contextual classifiers. The classifiers use these models as references based on the selected samples to classify features of interest present in the enhanced 15-meters image [94]. Next, classifications were applied using various classifiers, such as Naïve Bayes [35], support vector machine (SVM) [8,65], decision tree (DT) [29,30], k-nearest-neighbor (kNN) [27,28], and boost [9]. Among the tested classifiers, the best results were achieved by kNN and decision tree (non-contextual classifiers), and boost (contextual classifier). Although the other tested classifiers were used the same way as kNN, DT, and boost, they were excessively time-consuming. Moreover, their results did not overcome those achieved by kNN, DT, and boost for the experiments of this study. Therefore, we decided to consider only kNN, DT, and boost as the classifiers of the experiments in this study.

The kNN classifier is useful to perform discriminant analysis in situations for which it is hard to determine parametric estimates of probabilities [27]. Let us consider that  $X = [x_1, \dots, x_N]$  is the training data with  $N$  points of dimensionality  $D$ ,  $X_i = [x_{i1}, \dots, x_{ik}]$  is the  $k$  nearest neighbors of  $x_i$ ,  $X_t$  is the testing data with  $N_t$  points,  $x_0$  is an arbitrary testing data point,  $X_0 = [x_{01}, \dots, x_{0k}]$  contains its  $k$  nearest neighbors from training data, for which the labels are  $[l_1, \dots, l_k]$ , and  $\Omega = [\Omega_1, \dots, \Omega_C]$  is the set composed of  $C$  classes present in the data. Additionally, the kNN considers that the  $k$  neighbors of a testing point have equal weights [28].

The classification performed based on the kNN finds the nearest neighbors regarding a testing point in the training data. Moreover, the classifier assigns the test point to the most frequently occurring class of its  $k$  neighbors. The kNN classifies  $x_0$  based on the majority voting rule presented by Equation (11) [28], for which  $\delta$  is the Kronecker delta represented by Equation (12).

$$j^* = \arg \max_{j=1, \dots, C} \left( \sum_{i=1}^k \delta(l_i, j) \right) \quad (11)$$

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (12)$$

The decision tree classifier is useful to perform decision analysis by finding the most probable decision for achieving an objective [29]. Many algorithms are able to build decision trees. Among them, the Iterative Dichotomiser 3 (ID3) is a well-known algorithm used to generate decision trees

from data sets. The ID3 is based on information entropy and information gain, which are represented respectively by Equations (13) and (14) [30].

$$Ent(D) = - \sum_{k=1}^k p_k \log_2 p_k \quad (13)$$

$$Gain(D, a) = Ent(D) - \sum_{v=1}^V \frac{|D_v|}{|D|} Ent(D_v) \quad (14)$$

These equations consider that the training sample set is represented by  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , and  $|D|$  stands for the training sample number. The  $p_k$  ( $k = 1, 2, \dots, |D|$ ) is the ratio of every type for every current sample set.  $A = \{a_1, a_2, \dots, a_d\}$  stands for the attribute set of  $D$ , e.g., color, shape, texture, etc., for which  $d = \{1, 2, \dots, k\}$ . For each attribute  $a_i$ ,  $V$  is a set of features composed of  $v$  different values and represented by  $V = \{a_i^1, a_i^2, \dots, a_i^v\}$ , e.g.,  $a_1^1 = \text{green}$ ,  $a_1^2 = \text{blue}$ , etc.  $D_v$  is the sample subset related to the value  $a_i^v$  from  $a_i$  in  $D$ , and  $|D_v|$  stands for the number of the current samples of the subset. The entropy of  $D$ ,  $Ent(D)$ , characterizes the impurity of the collection of samples in  $D$ , i.e., it measures the amount of uncertainty in the data set  $D$ . The gain of  $D$ ,  $Gain(D, a)$ , is the expected reduction in the entropy of  $D$  related to the choice of the attribute  $a$ .

The Boost classifier combines iteratively weak classifiers by taking into consideration a weight distribution on the training samples such that more weight is attributed to samples misclassified by the previous iterations [9]. The final strong classifier is a weighted combination of weak classifiers followed by a threshold. The steps performed by the Boost are described in detail as follows.

(1) Establish training sets.

Let us consider that  $x_i$  and  $y_i$  are respectively the samples and their labels. Moreover,  $D_t(x_i, y_i)$  is the  $t^{th}$  distribution over all these training samples. Let us also consider a training set  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , for which  $y_i \in \{-1, 1\}$ .

(2) Initialize weights.

The initialization of weights is performed according to Equation (15) [9].

$$D_0(x_i, y_i) = \frac{1}{N}, \quad i = 1, 2, \dots, N \quad (15)$$

(3) Training loop.

For  $t = 1, 2, \dots, T$  ( $T$  is the maximum training number):

(a) For each feature  $j$ , train a classifier  $h_j$  which is a simple linear classifier, i.e., a classifier restricted to use a single feature. Equation (16) expresses this type of classifier [9].

$$h(x_i) = \begin{cases} 1, & \text{if } p_j x_{i,j} < p_j \theta_{i,j} \\ -1, & \text{otherwise} \end{cases} \quad (16)$$

In Equation (16),  $x_{i,j}$  is the value of the  $j^{th}$  feature of the sample  $x_i$ ,  $p_j \in \{-1, 1\}$  determines the direction of the inequality sign,  $\theta_{i,j}$  denotes the threshold value of the  $j^{th}$  feature of the sample  $x_i$ . Equation (17) shows that the error  $\varepsilon_t$  is evaluated with respect to  $D_t(x_i, y_i)$  [9].

$$\varepsilon_j = \sum_{i: y_i \neq h_j} D_t(x_i, y_i) \quad (17)$$

(b) Choose the weak classifier  $h_t$ , with the lowest error  $\varepsilon_t$ .

(c) If  $\varepsilon_t \geq 1/2$ , then stop.

(d) If  $\varepsilon_t < 1/2$ , calculate  $\alpha_t$  which is the weight assigned to the classifier  $h_t$ , in accordance with Equation (18).

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (18)$$

(e) Equation (19) shows how to update the weights [9],

$$D_{t+1}(x_i, y_i) = \frac{D_t(x_i, y_i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \quad (19)$$

for which  $Z_t$  is a normalized constant computed to insure that  $D_t(x_i, y_i)$  represents a true distribution, established by Equation (20) [9].

$$\sum_{i=1}^N D_t(x_i, y) = 1 \quad (20)$$

(4) Output the final classifier in accordance with Equation (21) [9].

$$H(x) = \text{sign} \left( \sum_{t=1}^{T_k} \alpha_t h_t(x) \right) \quad \text{sign}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \quad (21)$$

The image classifications are important in order to group features of interest into classes according to the similarities of their characteristics. In the images resulting from the classifications, each class is represented by a different color. In the case of this research, because only two different classes were used for grouping features of interest, “water” and “no-water,” the images resulting from the classification processes are binary [94].

At the end, a subtraction of images resultant from non-contextual and contextual classifications was performed, computing the difference of all pairs of corresponding pixels from both images [88]. The subtraction was calculated according to Equation (22) [88], for which,  $x$  and  $y$  are coordinates of the pixels, and  $f(x, y)$  and  $h(x, y)$  are respectively the images resultant from non-contextual and contextual classifications. Image  $g(x, y)$  results from the subtraction. This subtraction is important because it allows the automatic identification of congruence, in which case the subtraction results in zero, or automatic identification of incongruence, in which it results in at least one pixel equal to one. Incongruences can reveal the presence of anomalies as explained in [16]. When anomalies were identified, they were analyzed and categorized by type in accordance with [13].

$$g(x, y) = f(x, y) - h(x, y) \quad (22)$$

The learning and classification approaches are essential to make feasible the application of the taxonomy [13] to solve real-world problems. Our main contributions in the learning and classification approaches are as follows: (1) The use of a single Landsat 8 image previously assessed as high-quality sensory data. (2) The creation of statistical and classification models adequate enough to make PR tools learn how to classify the features present in the environments. (3) The subtraction of the results from contextual and non-contextual classifiers to indicate the occurrence of incongruence. (4) Finding the combination of the three previous conditions to establish a relation between the occurrence of anomalies of the type *unexpected structure and structural components* and the presence of brown mud in large rivers.

Regarding methodological limitations, it is recognized that the higher the resolution of the images, the longer the computational time to process the images. Therefore, long time of processing is inevitable when large images are evaluated by anomaly detection strategy. This time depends on the computer being used and nowadays can vary from dozens of minutes to many hours. However, this variation,

in the computational time, spent by the proposed strategy to process the images is acceptable since different computers have different computational resources.

## 6. Results

Three evaluation measure mechanisms (contextual and non-contextual classifiers; sensory data quality assessment; incongruence indicator) were used to reliably qualify the anomaly, as described in [13], because they are three successful mechanisms to qualify anomalies. According to Kittler et al. [13], three conditions need to be satisfied to guarantee the occurrence of the anomaly of the type *unexpected structure and structural components*. The three conditions are high sensory data quality, contextual and non-contextual classification of the same image, and incongruence. If any of the three conditions are not present, it is not possible to identify an anomaly of this type. In this study, the high sensory data quality was confirmed because all images are qualified with value “9,” which represents the best quality for Landsat 8 images. The contextual and non-contextual classifications were confirmed by the results achieved applying respectively boost and kNN or DT classifiers. The presence of incongruence results was confirmed performing the last step of the strategy.

All assessments were carefully performed taking into consideration the application of the proposed strategy on Landsat 8 images. Since all those Landsat 8 images are large (170 km north-south by 183 km east-west, i.e., 106 mi by 114 mi, approximately), the image of each environment was precisely cropped in order to create a set of smaller images tiles. This was fitting, as the majority of assessment approaches found in the scientific literature are applied to smaller images. Table 3 shows how the cropping was done for each Landsat 8 scene.

**Table 3.** Information about how cropping was done for each Landsat 8 scene.

Environment Country	Landsat 8 Scene	Scene Resolution (Height and Width in Pixels)	Image Tile Resolution (Height and Width in Pixels)	Total Number of Tiles
Doce River Brazil	LC08_L1TP_217074_ 20151112_20170402_01_T1	15705 × 15440	151 × 193	8400
Athabasca River Canada	LC08_L1TP_043020_ 20170826_20170913_01_T1	16499 × 16277	146 × 155	11836
Elbe River Germany	LC08_L1TP_196022_ 20180509_20180517_01_T1	16099 × 15882	145 × 143	12544
Marañon River Peru	LC08_L1TP_007063_ 20160827_20170321_01_T1	15498 × 15185	126 × 146	12792
Tietê River Brazil	LC08_L1TP_219076_ 20170827_20170914_01_T1	15484 × 15250	139 × 162	14641
Arava Valley Israel	LC08_L1TP_174039_ 20141214_20170416_01_T1	15432 × 15067	129 × 127	10528

In order to quantitatively validate the proposed strategy, we used the metrics accuracy, precision, recall [15], and F-measure [82] for which, *TP* are true positives, i.e., images in which incongruences are truly detected; *TN* are true negatives, i.e., images in which congruences are truly detected; *FN* are false negatives, i.e., images in which congruences are falsely detected; *FP* are false positives, i.e., images in which incongruences are falsely detected. Accuracy measures the efficiency of results and is represented by Equation (23), for which *M* is the number of images. Precision measures the relevancy of results and is represented by Equation (24). Recall measures the quantity of truly relevant results and is represented by Equation (25). F-Measure measures the balance between precision and recall and is represented by Equation (26).

$$Accuracy = \frac{TP + TN}{M} \quad (23)$$

$$Precision = \frac{TP}{TP + FP} \quad (24)$$



$$Recall = \frac{TP}{TP + FN} \quad (25)$$

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (26)$$

### 6.1. Experimental Results

Tables 4–9, which are all contingency tables [95–97], present the quantitative evaluations of incongruence detections on real images for, respectively: Doce River, Athabasca River, Elbe River, Marañon River, Arava Valley, and Tietê River.

**Table 4.** Contingency table helping to detect incongruences in the Landsat 8 scene of Doce River consisting of 8400 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 68	<i>FP</i> = 20
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 8312

**Table 5.** Contingency table helping to detect incongruences in the Landsat 8 scene of Athabasca River consisting of 12,084 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 87	<i>FP</i> = 38
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 11,959

**Table 6.** Contingency table helping to detect incongruences in the Landsat 8 scene of Elbe River consisting of 12,544 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 0	<i>FP</i> = 18
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 12,526

**Table 7.** Contingency table helping to detect incongruences in the Landsat 8 scene of Marañon River consisting of 12,915 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 39	<i>FP</i> = 13
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 12,863

**Table 8.** Contingency table helping to detect incongruences in the Landsat 8 scene of Arava Valley consisting of 10,528 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 0	<i>FP</i> = 25
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 10,503

**Table 9.** Contingency table helping to detect incongruences in the Landsat 8 scene of Tietê River consisting of 14,641 tiles (*TP*—true positive, *FP*—false positive, *FN*—false negative, *TN*—true negative).

	Incongruent Event	Congruent Event
Detection Incongruent	<i>TP</i> = 0	<i>FP</i> = 0
Detection Congruent	<i>FN</i> = 0	<i>TN</i> = 14,641

## 6.2. Interpretation of the Results

Tables 4–9 present encouraging results. As can be seen, there is a significant quantity of true positives associated with the image sets that present brown mud, whereas this quantity is slightly lower for images presenting crude oil and considerably lower for images presenting red mud and sewage. False positives are present in almost all image sets, but their number are of little significance for this study.

The results regarding the accuracy, precision, recall, and F-measure of our strategy are shown in Table 10.

**Table 10.** Accuracy, precision, recall, and F-measure of the proposed strategy for different environments (Landsat 8 scenes).

Landsat 8 Scene	Environment	Accuracy	Precision	Recall	F-Measure
LC08_L1TP_217074_20151112_20170402_01_T1	Doce River Basin	~99.76%	~77.27%	100%	87.18%
LC08_L1TP_043020_20170826_20170913_01_T1	Athabasca River Basin	~99.69%	~69.6%	100%	82.08%
LC08_L1TP_007063_20160827_20170321_01_T1	Marañon River Basin	~99.90%	75%	100%	85.71%

Table 11 shows the comparison of the accuracy, precision, recall, and F-measure of this study, presented in the first row, with results achieved by other studies. It seems that the results are quantitatively consistent with those found in the scientific literature. On average, our strategy achieved the highest values of accuracy and recall among the presented studies. Although our strategy did not achieve the highest values of precision and F-measure compared to the other studies, it is of little significance because the achieved precision and F-measure are high.

**Table 11.** Comparison of the accuracy, precision, recall, and F-measure of this study with others.

Study	Accuracy	Precision	Recall	F-Measure
Ours	~99.78%	~73.96%	100.00%	85.04%
[98]	99.59%	—	—	—
[81]	99.20%	91.85%	53.55%	67.66%
[99]	99.14%	—	—	—
[76]	98.49%	83.84%	83.66%	83.76%
[80]	98.00%	—	—	—
[25]	~91.20%	98.10%	95.7%	96.88%
[77]	88.68%	90.62%	79.62%	84.76%
[66]	84.00%	63.00%	81.00%	70.88%
[100]	81.18%	—	—	—
[78]	78.00%	82.00%	75.00%	78.34%
[82]	—	96.50%	94.8%	95.64%

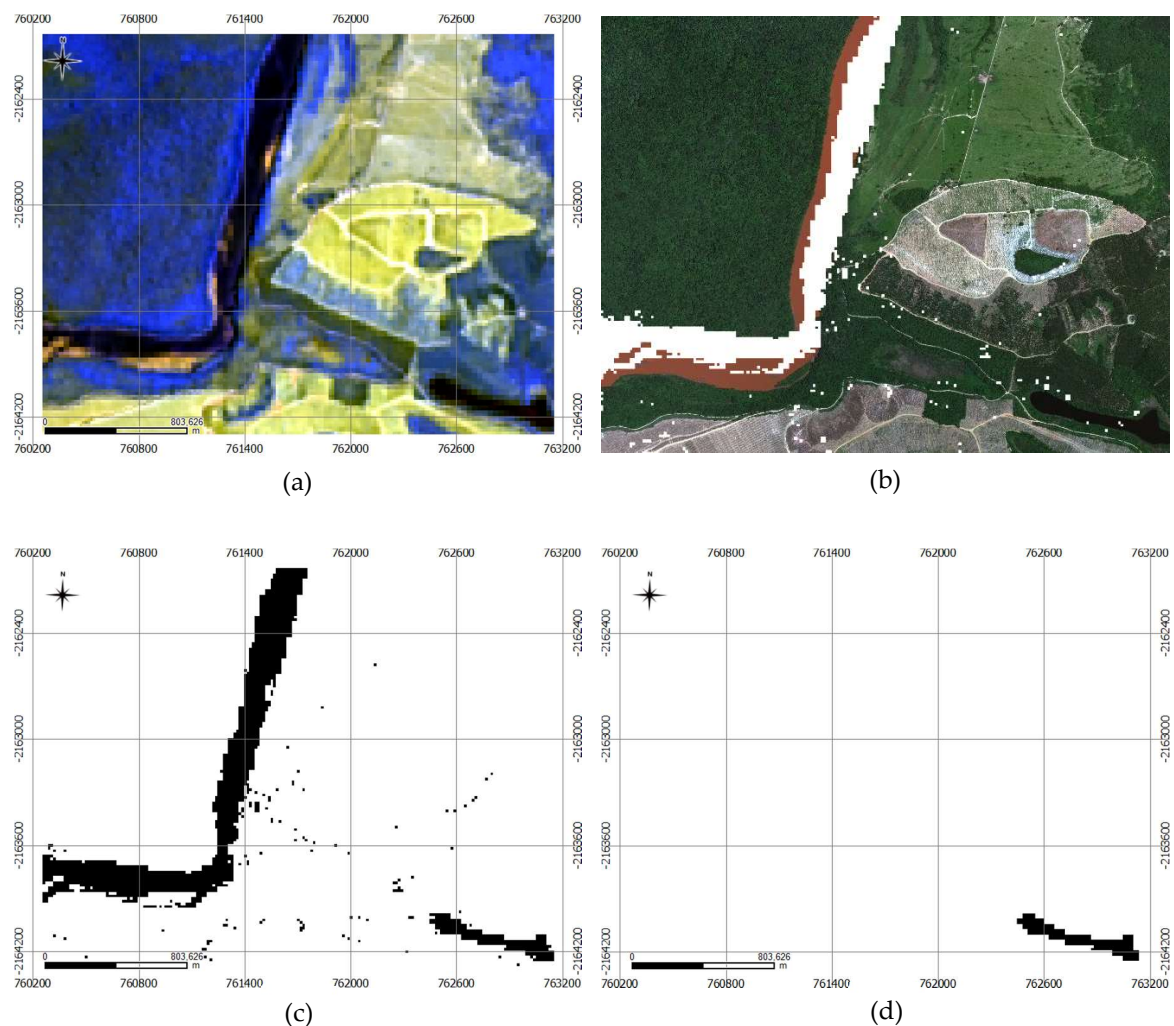
Our strategy achieved the highest value of F-measure among the presented studies, except for [25] and [82]. However, [25] and [82] are very specific case studies. Whereas in this paper and in [66,76–78,80–82,98–100] the areas of study are heterogeneous, in [25], results were presented only related to an ice shelf in the Antarctic region, which is a homogeneous area of study. According to [100], results tend to be higher for homogeneous areas of study compared to heterogeneous ones. Whereas in this paper and in [66,76–78,80–82,98,99] the altitude of the remote sensing is high; in [82], the sensor was very close to the investigated outliers when the images were captured. The unmanned aerial vehicle (UAV) which performed the remote sensing flew in a short distance over the area of study. The higher the proximity of the area of study, the higher tend to be the precision and recall of the detection, consequently making the F-measure higher. Therefore, our strategy achieved the highest value of F-measure among the presented studies, for which the characteristic of the studied area is heterogeneous and the altitude of the remote sensing is high.

Figures 6–8 show the examples of the application of the proposed strategy on three different images (tiles) taken respectively from the Doce River, the Athabasca River, and the Marañon River

scenes: (a) enhanced image, (b) result overlay on the real image, (c) contextual classifier output, and (d) non-contextual classifier output. The incongruences can be recognized comparing the differences between the outputs (c) and (d) of each image.

In the first case, apart from other geographical features, Figure 6a shows the non-polluted water of a narrow and long lake, in black, in the bottom right corner and the water of part of the Doce River polluted with ore tailing waste (brown mud), also in black, starting on the left, making a sharp turn, and ending on the top. Figure 6c shows the capability of the contextual classifier (Boost) for recognizing both polluted and non-polluted water, whereas Figure 6d shows the capability of the non-contextual classifier (decision tree) for recognizing only non-polluted water.

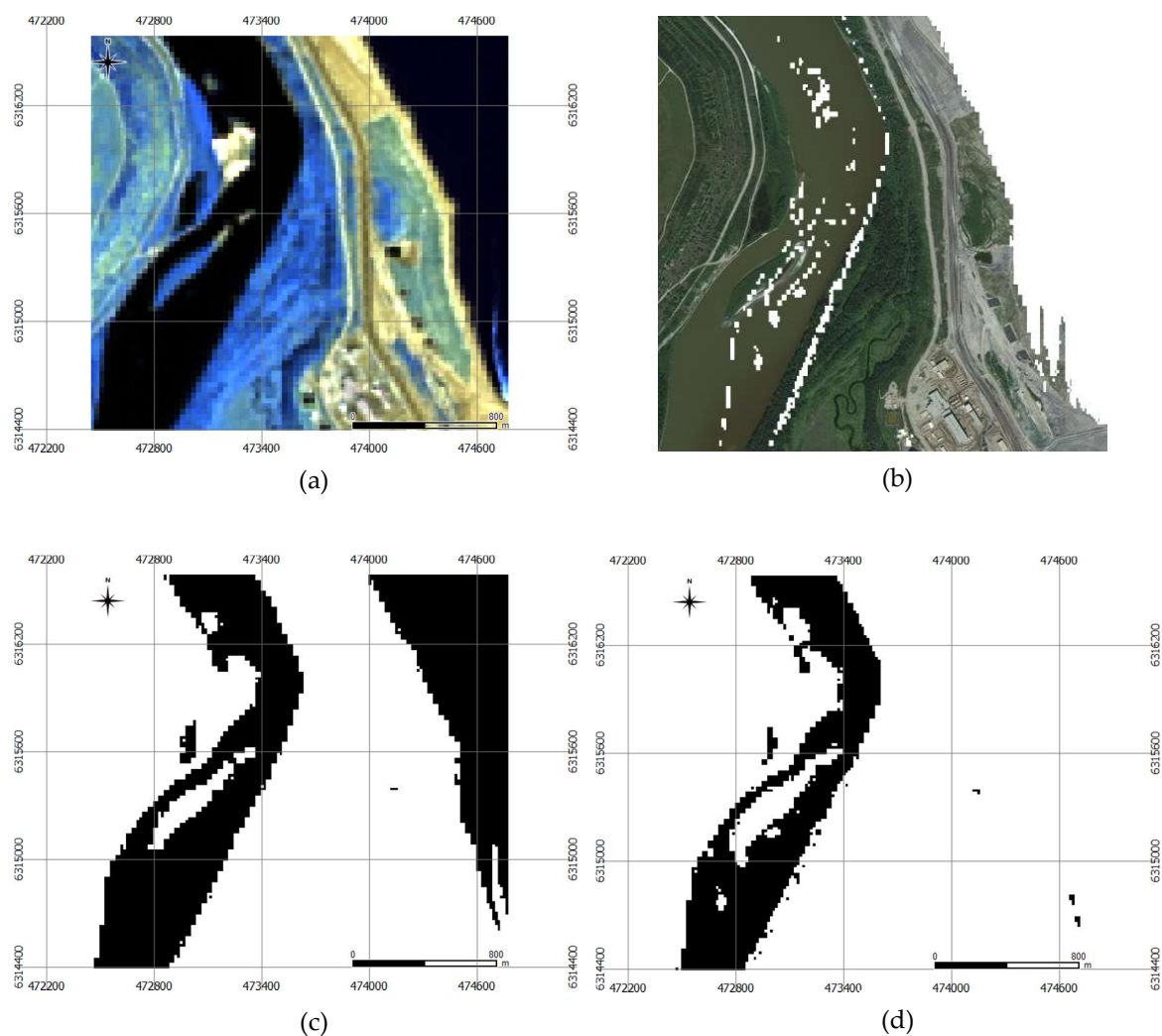
In the second case, apart from other geographical features, Figure 7a shows part of a large reservoir polluted with ore tailing waste (brown mud), in black, in the top right corner and the non-polluted water of part of the Athabasca River, also in black, starting on the bottom left, making a sharp turn, and ending on the top. Figure 7c shows the capability of the contextual classifier (boost) for recognizing both polluted and non-polluted water, whereas Figure 7d shows the capability of the non-contextual classifier (decision tree) for recognizing only non-polluted water.



**Figure 6.** An example of the application of the proposed strategy on an image tile taken from the Doce River scene of Landsat 8: (a) Enhanced image, in which black pixels represent two water bodies, (b) result overlay (white) on the real image, (c) contextual classifier (boost) output, (d) non-contextual classifier (decision tree) output. In (b), white represents pixels detected as water polluted by brown mud. In (c,d), black and white represent pixels classified as water and no-water respectively.

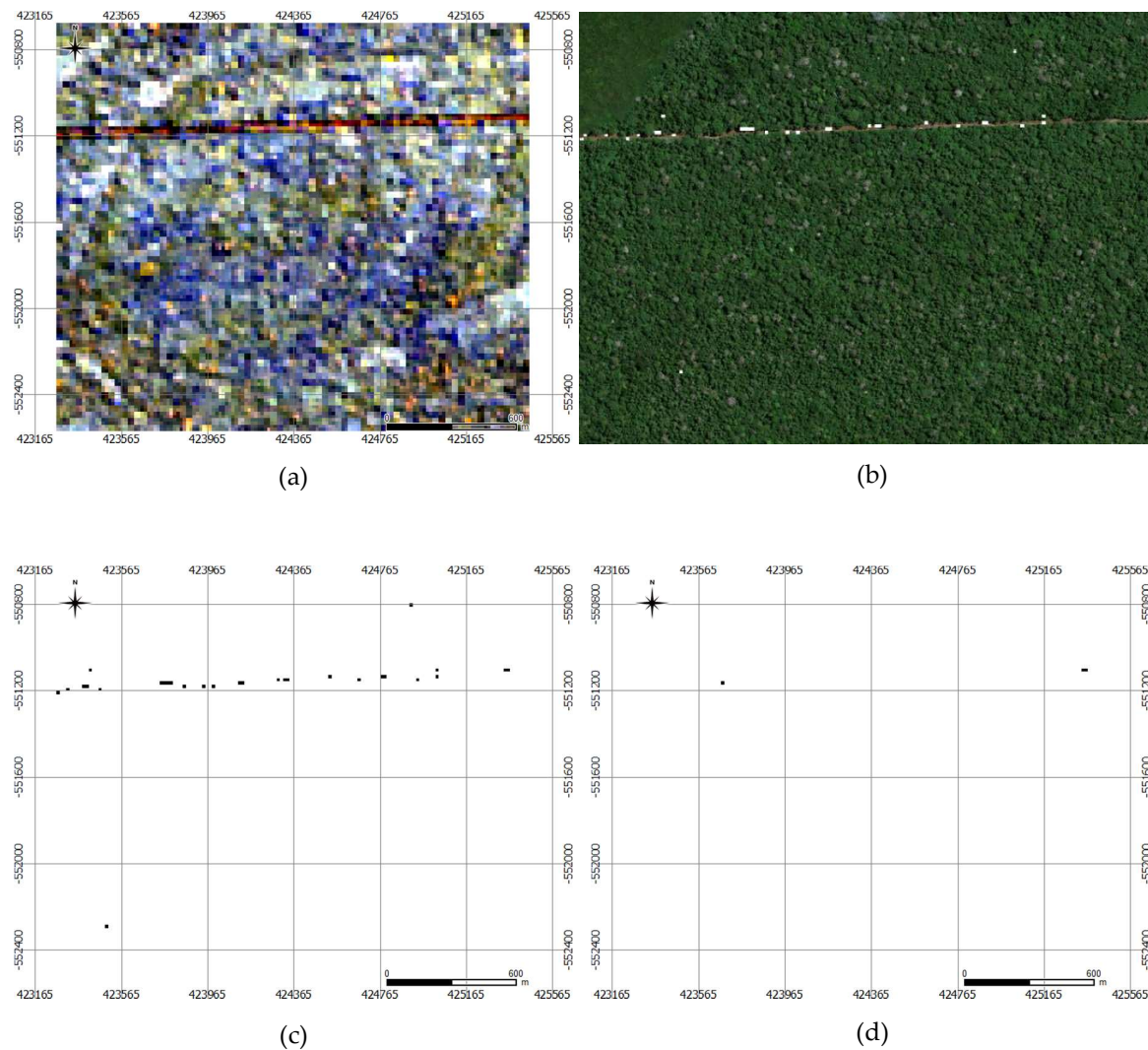
In the third case, apart from other geographical features, Figure 8a shows part of a very long lane built in the vicinity of the Marañon River in order to support oil pipelines, in black and orange, starting on the top left and ending on the top right. Figure 8c shows the capability of the contextual classifier (Boost) for recognizing the polluted water concentrated on the impermeable soil in the lane, whereas Figure 8d shows that the non-contextual classifier (kNN) is not so much capable of recognizing the polluted water in the lane.

In the last step of our method, the image of the Figures 6d, 7d and 8d were subtracted respectively from the images of the Figures 6c, 7c and 8c. Figures 6b, 7b and 8b show the respective results of these subtractions highlighted in white and overlapped on the real images. The small white points spread on Figures 6b and 7b can be easily removed applying a morphological opening filter in the image. However, the same filter would remove the result present in Figure 8b. Therefore, we have decided to not apply the filter.



**Figure 7.** An example of the application of the proposed strategy on an image tile taken from the Athabasca River scene of Landsat 8: (a) Enhanced image, in which black pixels represent two water bodies, (b) result overlay (white) on the real image, (c) contextual classifier (boost) output, (d) non-contextual classifier (decision tree) output. In (b), white represents pixels detected as water polluted by brown mud. In (c,d), black and white represent pixels classified as water and no-water respectively.





**Figure 8.** An example of the application of the proposed strategy on an image tile taken from the Marañon River scene of Landsat 8: (a) Enhanced image, (b) result overlay (white) on the real image, (c) contextual classifier (Boost) output, (d) non-contextual classifier (kNN) output. In (b), white represents pixels detected as water polluted by crude oil. In (c,d), black and white represent pixels classified as water and no-water respectively.

## 7. Discussion

When the classes are very imbalanced, precision and recall measure the success of prediction. If the classifiers return correct results and the majority of all positive results, then precision and recall present high values. We achieved high values of accuracy, precision, and recall with the application of the incongruence based anomaly detection strategy in this study. Therefore, a significant quantity of truly relevant results showed that the anomalies of the type *unexpected structure and structural components* were detected with efficiency by our strategy. The values of precision would be even higher if we had limited the studied areas to the vicinity of the locations with pollutants instead of studying the whole Landsat 8 scene, since many false positives were located far from these locations. Moreover, the achieved results demonstrate that the application of the proposed strategy to detect water pollution is consistent with other studies, although this study has introduced an approach to reach results that is unlike the approaches commonly used.

The results presented in the first row of Table 10 are aligned with the results reached by [31], which reveal high concentration of brown mud in the Doce River. Moreover, the results presented in

the second row of Table 10 compare well with the results reached by [32], which highlight the ability to discriminate mineral mine waste surfaces. Therefore, the present achieved results suggest that the incongruence occurs because, contrary to the non-contextual classifier (decision tree), the contextual one (boost) is able to detect the water, even for cases in which there is high turbidity in the water of the studied river or reservoirs, because of the high concentration of brown mud.

The results presented in the third row of Table 10 are parallel to the results reached by [87], which also correspond to detection of oil spills in optical satellite images. In the case of our study, instead of detecting oil spills on the surface of the water such as in [87], the detection was performed in lanes built in the vicinity of the Marañón River in order to support oil pipelines. A significant oil spill turns the soil impermeable causing accumulation of water around the pipelines, whereas those lanes should normally present either dry or just humid soil around the pipelines. Therefore, the present achieved results suggest that the incongruence occurs because, contrary to the non-contextual classifier (kNN), the contextual one (Boost) presents high sensibility to detect the water around the pipelines (see Figure 9), revealing the presence of oil spill on the soil of that area. Additionally, the results presented in Table 8, which are related to the Arava Valley, have demonstrated that neither the non-contextual nor the contextual classifier classify crude oil as water in a dry environment, which corroborates the causes of the incongruence previously mentioned.



**Figure 9.** Section of an image of a lane built in the vicinity of the Marañón River in order to support an oil pipeline. The pipeline, in the middle-bottom of the image, suffered a severe disruption on 22 August 2016. As a consequence, the lane and the Marañón River received high quantity of crude oil causing an environmental disaster ( $4^{\circ}48'45.6''\text{S}$   $75^{\circ}23'56.8''\text{W}$ ) [87]. The image shows the crude oil around the pipeline. Source: <http://geoportal.regionloreto.gob.pe/mapa-de-derrame-de-petroleo-en-el-tramo-i-del-oleoducto-norperuano/> (accessed on 9 July 2018).

Outcomes from this study align with those presented by Kittler et al. in [13] and Weinshall et al. in [15], because they perform research involving the use of incongruences and recognize their potential to be applied to anomaly detection. Additionally, the current findings expand these prior studies to real-world images from remote sensing, instead of only synthetic data, and apply to real-world applications, such as the detection of water pollution. For example, after detecting anomalies, decision-making systems can guide computers to give adequate and fast responses to prevent water pollution as an environmental disaster or to analyze the extent to which they occur. Furthermore, the proposed strategy broadens a list of studies, such as [55–60,101], engaged to analyze images from remote sensing to offer possible solutions to detect and monitor water pollution.

The results achieved by this study demonstrate higher accuracy and recall than previous studies' results, such as [25,66,76–78,80–82,98–100]. Therefore, the results suggest that our strategy provides



more efficient results and a higher quantity of truly relevant results than the referred studies. The smallest difference regarding accuracy in comparison to our study was reported by Che et al. in [98]. The high level of accuracy was achieved by applying a method based on machine learning. However, Che et al. did not report values in regard to recall.

Another small difference regarding accuracy in comparison to our study was reported by Bormann et al. in [81]. In that study, a method based on detection index allowed them to achieve a high level of accuracy. In contrast, [81] reported the largest difference regarding recall in comparison to our study. Therefore, our method provides much more truly relevant results than the one proposed by them.

Three other studies [76,80,99] reported accuracies, whose values are lower than the one achieved by our study, although they are still close. In [76], a method based on median absolute deviation provided high level of accuracy. However, our method provides more truly relevant results than the one proposed in [76]. Similarly to our study, SVM was a method applied for classification by Bhaduri et al. in [80]. In contrast, we achieved the best results by applying other classifiers in our study compared to the use of SVM. A significant result was provided by a deep-learning method based on convolutional neural network (CNN) in [99]. Although [80,99] achieved high levels of accuracy applying their methods, they did not report results in regard to recall.

Other studies, such as [66,77,78,100], have more significant differences regarding both accuracy and recall in comparison to our study. In these studies, the levels of accuracy and recall were achieved after applying a method based on: (1) seasonal autoregressive integrated moving average (SARIMA) model for autocorrelation analysis in [77]. (2) Gaussian process (GP) based non-parametric time-series prediction in [78]. (3) Maximum likelihood classification in [100]. (4) Joint fully convolutional auto-encoders (FC AE) model in [66]. The results suggest that our method provides significant higher levels of both accuracy and recall in comparison to the methods based on SARIMA, GP, and AE. Unfortunately, [100] did not report values in regard to recall.

Focusing the discussion specifically on the study presented by Sublime and Kalinicheva in [66], they proposed a deep-learning method for change detection (FC AE). Their method is useful for post-disaster damage mapping. Although they also deal with real-world problems, their study differs from ours because their method detects outliers rather than apply the taxonomy [13] (i.e., detecting anomalies) to solve real-world problems. According to Sublime and Kalinicheva, their outlier detection method achieved the best results when it was compared against other machine-learning methods. The levels of accuracy and recall achieved by our strategy are significantly higher than the ones achieved by the deep-learning method proposed in [66]. Additionally, the level of precision achieved by the deep-learning method proposed in [66] is lower than the one achieved by our strategy. Therefore, the results suggest that our method provide results that are more efficient and relevant, and a higher quantity of truly relevant results compared to machine-learning methods applied for post-disaster damage mapping, for example.

As previously mentioned, to the knowledge of the authors this is the first endeavor of a series of studies dedicated to investigating the detection of anomalies in accordance with the taxonomy published in [13] and their potential applications to remote sensing. This study proposed a strategy taking into consideration the multifaceted nature of anomalies [13], whereas other detection methods have been dealing with different types of anomalies as if all of them were outliers. Since the present strategy is based on this taxonomy [13] which presents ten different domain anomalies, this study overcomes the number of categories studied in previous investigations, e.g., [25] which presented four different anomalies. The smallest differences regarding the recall in comparison to our study were reported by Qi Liu et al. in [25] and Chen et al. in [82]. However, Chen et al. did not report their achieved accuracy. Overall, our study provides relevant improvements over other studies and also important support and guidelines for future studies.

Regarding limitations, it should be noted that the accuracy of the proposed strategy varies depending on the environment of application. It is inevitable when many different geographical features are evaluated by some anomaly detection strategy. However, the differences in levels of

accuracy of the proposed strategy are of little significance. Further work is planned to examine the effects of applying filters on the images before applying them to the proposed strategy to determine if the different levels of accuracy can be minimized among different environments. The results indicate that the proposed strategy accurately detects anomalies related to brown mud and crude oil in the majority of evaluated images from remote sensing.

Regarding negative results, it should also be noted that the detection of incongruence, when the water pollution is caused by either red mud or sewage (Tables 6 and 9), was not obtained by the proposed strategy. This limitation is probably caused by the use of seven bands as a pattern for all experiments of this study. However, it is of little significance, since the ability of the proposed strategy detecting incongruence was demonstrated when the water pollution is caused by either brown mud or crude oil. All in all, the achieved results indicate that the proposed strategy detects the presence of anomalies of the type *unexpected structure and structural components*, in accordance with [13].

## 8. Conclusions

This paper presented an incongruence-based anomaly detection strategy for analyzing images from remote sensing, with the aim of describing its practical application in detecting water pollution. Therefore, the practical application of incongruence was introduced as a strategy for detecting anomalies in real-world applications, i.e., non-synthetic data present in images from remote sensing.

The proposed strategy detected anomalies of the type *unexpected structure and structural components*, distinguishing them from outliers. All processing was performed on just a single Landsat 8 image, whereas other methods use time-series for each environment to be able to detect outliers. This means that systems dedicated to observing or monitoring occurrences of anomalies of this type will save time and use few computational resources because they will process a single image. Therefore, systems based on the application of our strategy will help by saving more money than systems based on other proposed detection methods.

This study has established a relationship between the occurrence of anomalies of this type and the presence of brown mud or crude oil as either potential contaminants or contaminants of large rivers. Our strategy identified incongruences to detect oil spills in some areas where there is the presence of water, or brown mud where there are high levels of turbidity in the water. These findings are relevant since an incongruence-based anomaly detection strategy can significantly increase the ability of surveillance systems for detecting environmental disasters and for performing mappings.

Although the achieved results are encouraging, future studies should investigate: (1) The effects of using Gaussian process-based filters on the images before applying them to the proposed strategy in order to know if it would be useful to minimize differences between the levels of the accuracy among different environments. (2) The effects of applying feature extractors on the images as a preprocessing stage before applying the proposed strategy in order to determine if the classification process can be improved for all environments. (3) Whether our strategy will be able to detect other pollutants not using pan-sharpening.

Future work will investigate: (1) Practical applications of other types of anomalies individually, such as *unknown object*, *measurement model drift*, *unknown structure*, *unexpected structural component*, and *component model drift*. (2) A single system to deal with the practical application of all types of anomalies together.

It is expected that this study will open up an entirely new range of applications for computational tools on images from remote sensing in order to model and analyze environments. Moreover, this anomaly detection strategy can be applied to a wide range of studies in geosciences and other scientific areas.

**Author Contributions:** Conceptualization, M.A.D., E.A.d.S., S.C.d.A., W.C., T.S., and R.G.N.; funding acquisition, M.A.D.; investigation, M.A.D.; resources, E.A.d.S., S.C.d.A., W.C., T.S., and R.G.N.; writing—original draft, M.A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by FOUNDATION FOR RESEARCH SUPPORT OF THE STATE OF SÃO PAULO (Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP), grant number 2016/24185-8.

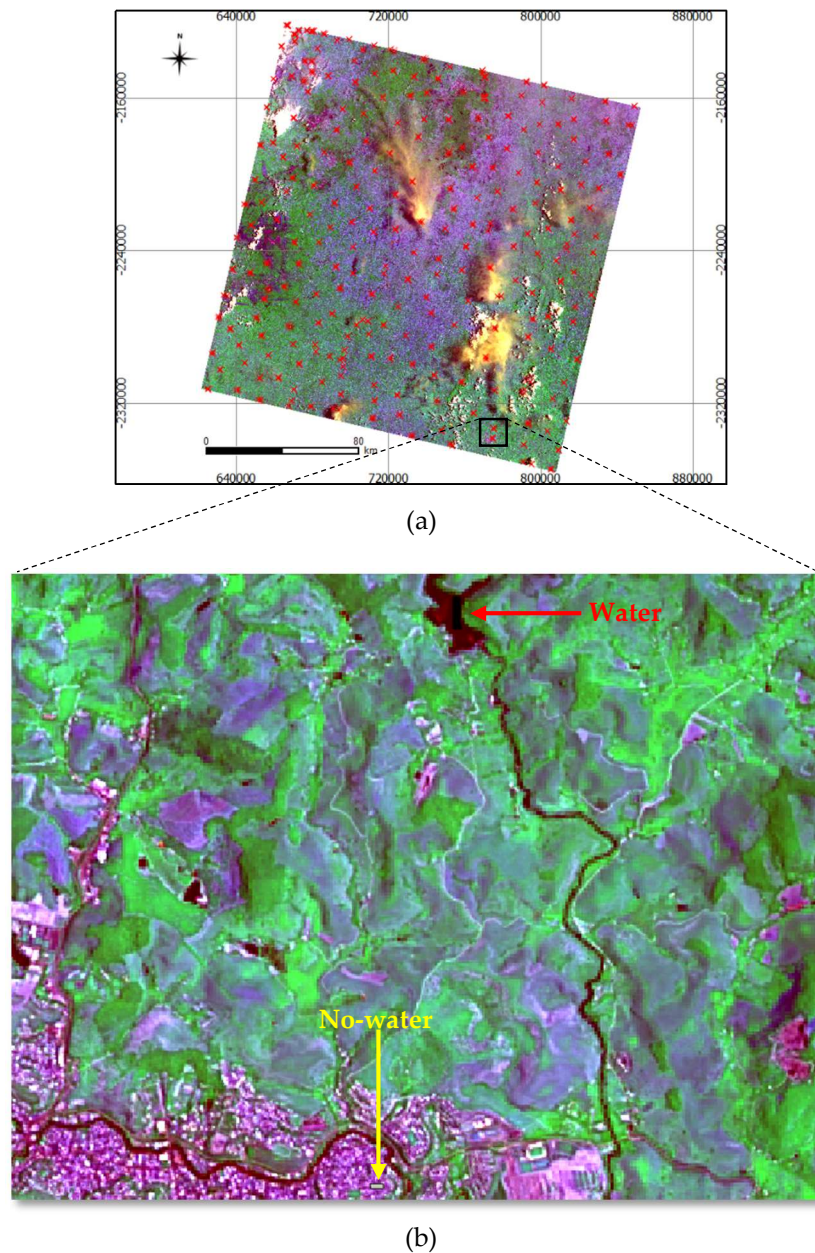
**Acknowledgments:** The Landsat 8 data set was provided by the United States Geological Survey (USGS): <https://earthexplorer.usgs.gov/>. The QGIS software was provided by the QGIS Development Team: <https://www.qgis.org/>. The Orfeo ToolBox was provided by the Open-source Geospatial and the OTB Communities: <https://www.orfeo-toolbox.org/>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Sampling

Figure A1 shows the locations of samples created on an image taken from the Doce River scene of Landsat 8 and highlights two examples of samples: from water and no-water land cover.



**Figure A1.** Samples created on an image taken from the Doce River scene of Landsat 8: (a) the red marks show the locations of the samples, (b) two examples of samples: the red arrow shows a black polygon for the water sample, the yellow arrow shows a white polygon for the no-water sample.

### Appendix A.2. Rendering

We used the default parameters of Qgis 2.18.19 for rendering, which means: *multiband color* to Render type, *Band 4* to Red Band, *Band 5* to Green Band and *Band 6* to Blue Band. The other parameters used were *stretched to MinMax* to Contrast enhancement, *2%–98%* to Cumulative count cut, *2* to mean +/- standard deviation, *normal* to blending mode, and *off* to Greyscale. More detailed descriptions of this tool as well as of each parameter used here can be found in [89].

### Appendix A.3. UTM Projection

In the UTM projection the terrestrial surface is mapped into 60 zones. Each zone spans 6° in longitude and ranges from 84°N to 80°S. The origin of the UTM coordinates (E, for east-west direction and N, for north-south direction) is where the equator crosses the zone center meridian. In order to avoid working with negative values, a false easting of 500,000 m is added to the coordinate E, both east and west of the central meridian. As for N, when in the south hemisphere, a false northing of 10,000,000 m is adopted [102].

### Appendix A.4. Pan-Sharpening

First CS provides a sensor modelling to zooming and register [90] the multi-spectral image on the panchromatic image, i.e., to perform the projection of an image into the geometry of another one. Second, CS fuses the co-registered pixels [90] of the multispectral image with the pixels of the panchromatic one by the application of a pixel-by-pixel fusion operator. This pan-sharpening method operates in the same way on the whole image.

The advantages of this pan-sharpening method are the high fidelity in rendering the spatial details in the final product and the fast, easy implementation. The limitation of this pan-sharpening method is the inability to account for local dissimilarities between the panchromatic and multispectral images originated by the spectral mismatch between the panchromatic and multispectral channels of the instruments, which may produce significant spectral distortions. Another formalization of the CS method can solve this limitation [91]. It takes into consideration that the fusion process can be obtained through a proper injection scheme without the explicit calculation of the forward and backward transformations, if the substitution of a single component and the hypothesis of a linear transformation are considered. This CS method fuses the multispectral image with the panchromatic one allowing the resultant image to show a combination of the high spectral resolution of the multispectral image with the high spatial resolution of the panchromatic one.

The steps used to perform the pan-sharpening method based on this CS method are described in detail as follows. At the start, the method projects the multispectral image into another space. It is assumed that this transformation separates the spatial structure from the spectral information in different components. Then, it replaces the component containing the spatial structure with the panchromatic image to enhance the transformed multispectral image. Meanwhile, it performs histogram matching of the panchromatic image to the selected component. The histogram matching is performed before substitution, because greater correlation between the panchromatic image and the replaced component corresponds to lower levels of distortion introduced by the pan-sharpening method. At the end, the data is brought back to the original space through the inverse transformation.

In Qgis, two tools from the Orfeo tool box, superimpose sensor and pansharpening (RCS), were used to perform respectively the projection and the fusion. Default values were chosen as parameters for performing the projection, i.e., the superimpose sensor tool used the values *zero* for default elevation, *four* for spacing of the deformation field, and *two* for radius for bi-cubic interpolation, and *Nearest Neighbor* interpolation. The projection was necessary in order to prepare the image for the pan-sharpening. Pan-sharpening was then performed in accordance with the algorithm of component substitution (RCS) in order to increase the spatial resolution of the image based on the Band 8. More detailed descriptions of these tools as well as of each parameter used here can be easily found in [92].



### Appendix A.5. Learning and Classification Approaches

This step was performed by an Orfeo tool named TrainImagesClassifiers, which is itself completely responsible for performing automatically the training, validation, and testing of each classifier. Because the samples set is statistically significant to represent the whole scene, the tool uses hold-out validation for which the tool defines separately the percentage of each data set: training data set, validation data set, and test data set.

Some default parameters were used for training all classifiers: *zero* for default elevation, *1000* for maximum training sample size per class, *1000* for maximum validation sample size per class, *one* for bound sample number by minimum, *0.5* for training and validation sample ratio, *class* for the name of the discrimination field, and *zero* for set user defined seed. Other default parameters, which were used for training a specific classifier, are discriminated as follows. For the kNN classifier: *32* for number of neighbors, *knn* for classifier to use for training, and set *off* edge pixel inclusion. For the DT classifier: *65535* for maximum depth of the tree, *10* for minimum number of samples in each node, *0.01* for termination criteria for regression tree, *10* for cluster possible values of a categorical variable into  $K \leq \text{cat}$  clusters to find a suboptimal split, *10* for K-fold cross-validations, set use *1seRule* flag to *false*, set *TruncatePrunedTree* flag to *false*, *dt* for classifier to use for training, and set *off* edge pixel inclusion. For the Boost classifier: *one* for maximum depth of the tree, *100* for weak count, *0.95* for weight trim rate, *real* for boost type, *boost* for classifier to use for training, and set *off* edge pixel inclusion. More detailed description of the TrainImagesClassifiers tool as well as of each parameter used here can be easily found in [94].

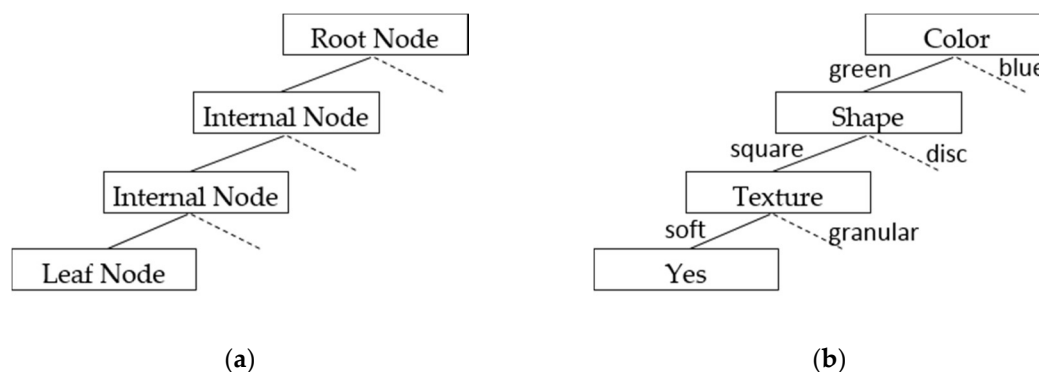
The advantage of the kNN is the fast and easy implementation. The limitation of the kNN is that the commonly employed Euclidian distance metric, which considers the data as isotropic or homogeneous, is generally not suitable for many real-world data sets. Nevertheless, learning a new distance metric can solve this problem. Therefore, it is possible to improve this classification if the kNN learns a distance metric derived from the training data, such as the one represented by Equation (27) [28], for which  $T$  represents a linear transformation.

$$\text{dis}(x_i, x_j) = \|T(x_i - x_j)\|^2 \quad (27)$$

The ID3 can be used to build the block diagram of the decision tree model presented in Figure A2a [30], starting from the root node, after generating internal nodes, and finishing each branch with a leaf node. For the ID3, the maximum information gain is used as a heuristic approach to choose the optimal decision attribute  $a_i$  from  $A$  that will be assigned to the next internal node. The root and internal nodes correspond to the different test attributes, i.e., the root or each internal node tests the assigned attribute  $a_i$ , as shown in Figure A2b. A branch for node is generated for every value  $a_i^v$  of  $a_i$ . Sample sets  $D_v$  included in every node are classified to respective child node based on the heuristic approach. In the extreme of each branch, each leaf node shows a decision result that designates a classification regarding the class  $c_t$ , for which  $t = \{1, 2, \dots, n\}$ . The class  $c_t$ ,  $c_t \in C = \{c_1, c_2, \dots, c_n\}$ , is determined by a specific combination of features attributed to  $c_t$ , e.g.,  $c_t = \{\text{green, square, rough}\}$ , that will be assigned to objects during the training and test procedures. The root node includes the whole sample set  $D$ . The path which goes from the root node to every leaf node represents a test rule.

The advantages of the ID3 are its obvious intuitive characteristic and its easy decomposability. The limitations of the ID3 are the difficult to control tree sizes, the time lost to calculate the information entropy expression by using the logarithmic algorithm, and the tendency to select the attribute which has more values in the procedure to get the optimal attribute. Because these limitations can reduce performance, the use of this version of the ID3 is not appropriate for real-time systems. Real-time systems are not addressed in our study. Using an improved version of the ID3 algorithm based on the simplification of the information entropy can solve these limitations, such as in [30]. However, the description of this other version is of little significance here, because this improvement causes no impact on the quality or quantity of the results of this study.

The advantages of the boost are the ability to integrate different classifiers that focus on different aspects of a problem and assign more weight on features that can train more accurate weak classifiers. The limitation of the boost is the slower and harder implementation compared to weak classifiers. This limitation was addressed by using the Boost as coded to run in our tests.



**Figure A2.** Block diagram of the decision tree: (a) general model, (b) example with decision attributes assigned to nodes. Adapted from [30].

The contextual and non-contextual classifications were performed using the tool Image Classification from Orfeo toolbox. More detailed description of this tool as well as of each parameter used here can be easily found in [94].

## References

- Li, W.; Du, Q. A survey on representation-based classification and detection in hyperspectral remote sensing imagery. *Pattern Recognit. Lett.* **2016**, *83*, 115–123. [\[CrossRef\]](#)
- Yang, K.; Li, M.; Liu, Y.; Cheng, L.; Huang, Q.; Chen, Y. River detection in remotely sensed imagery using Gabor Filtering and path opening. *Remote Sens.* **2015**, *7*, 8779–8802. [\[CrossRef\]](#)
- Chen, C.H. *Handbook of Pattern Recognition and Computer Vision*, 5th ed.; World Scientific: Singapore, 2015.
- Xie, X.; Li, B.; Chai, X. Kernel-based Nonparametric Fisher Classifier for Hyperspectral Remote Sensing Imagery. *J. Inf. Hiding Multimed. Signal Process.* **2015**, *6*, 591–599.
- Li, M.; Zang, S.; Zhang, B.; Li, S.; Wu, C. A review of remote sensing image classification techniques: The role of spatio-contextual information. *Eur. J. Remote Sens.* **2014**, *47*, 389–411. [\[CrossRef\]](#)
- Perveen, N.; Kumar, D.; Bhardwaj, I. An overview on template matching methodologies and its applications. *Int. J. Res. Comput. Commun. Technol.* **2013**, *2*, 988–995.
- Jiang, D.; Zhuang, D.; Huang, Y.; Fu, J. Survey of multispectral image fusion techniques in remote sensing applications. In *Image Fusion and its Applications*; Chinese Academy of Sciences: Beijing, China, 2011; pp. 1–23.
- Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [\[CrossRef\]](#)
- Shen, L.; Li, C. Water Body Extraction from Landsat ETM+ Imagery Using Adaboost Algorithm. In *Proceedings of the IEEE 2010 18th International Conference on Geoinformatics*, Beijing, China, 18–20 June 2010. [\[CrossRef\]](#)
- Dey, V.; Zhang, Y.; Zhong, M. A Review on Image Segmentation Techniques with Remote Sensing Perspective. In *Proceedings of the ISPRS TC VII Symposium—100 Years ISPRS*, Vienna, Austria, 5–7 July 2010; Wagner, W., Székely, B., Eds.; IAPRS: Paris, France, 2010; Volume XXXVIII, Part 7A.
- Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Elsevier: London, UK, 2006.
- Richards, J.A.; Jia, X. *Remote Sensing Digital Image Analysis*, 1st ed.; Springer: Berlin, Germany, 1999.
- Kittler, J.; Christmas, W.; de Campos, T.; Windridge, D.; Yan, F.; Illingworth, J.; Osman, M. Domain anomaly detection in machine perception: A system architecture and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 845–859. [\[CrossRef\]](#)



14. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–72. [\[CrossRef\]](#)
15. Weinshall, D.; Zweig, A.; Hermansky, H.; Kombrink, S.; Ohl, F.W.; Anemuller, J.; Bach, J.-H.; Gool, L.V.; Nater, F.; Pajdla, T.; et al. Beyond novelty detection: Incongruent events, when general and specific classifiers disagree. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1886–1901. [\[CrossRef\]](#)
16. Ponti, M.; Kittler, J.; Riva, M.; de Campos, T.; Zor, C. A decision cognizant Kullback–Leibler divergence. *Pattern Recognit.* **2017**, *61*, 470–478. [\[CrossRef\]](#)
17. Kittler, J.; Zor, C. Delta divergence: A novel decision cognizant measure of classifier incongruence. *IEEE Trans. Cybern.* **2018**, *99*, 1–13. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Kittler, J.; Zor, C. A Measure of Surprise for Incongruence Detection. In Proceedings of the 2nd International Conference on Intelligent Signal Processing (ISP), London, UK, 1–2 December 2015; IET: London, UK, 2015; pp. 1–6. [\[CrossRef\]](#)
19. Chandola, V.; Banerjee, A.; Kumar, V. Outlier detection: A survey. *ACM Comput. Surv. (CSUR)* **2007**, *14*, 1–83.
20. Gupta, M.; Gao, J.; Aggarwal, C.C.; Han, J. Outlier detection for temporal data: A survey. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 2250–2267. [\[CrossRef\]](#)
21. Zimek, A.; Schubert, E.; Kriegel, H.P. A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat. Anal. Data Min. ASA Data Sci. J.* **2012**, *5*, 363–387. [\[CrossRef\]](#)
22. Gogoi, P.; Bhattacharyya, D.K.; Borah, B.; Kalita, J.K. A survey of outlier detection methods in network anomaly identification. *Comput. J.* **2011**, *54*, 570–588. [\[CrossRef\]](#)
23. Niu, Z.; Shi, S.; Sun, J.; He, X. A Survey of Outlier Detection Methodologies and Their Applications. In Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, Taiyuan, China, 24–25 October 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 380–387.
24. Hodge, V.; Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [\[CrossRef\]](#)
25. Liu, Q.; Klucik, R.; Chen, C.; Grant, G.; Gallaher, D.; Lv, Q.; Shang, L. Unsupervised detection of contextual anomaly in remotely sensed data. *Remote Sens. Environ.* **2017**, *202*, 75–87. [\[CrossRef\]](#)
26. Data Management and Information Distribution (DMID)—Landsat Data Dictionary. Available online: [https://lta.cr.usgs.gov/DD/landsat\\_dictionary.html#image\\_quality\\_landsat\\_8](https://lta.cr.usgs.gov/DD/landsat_dictionary.html#image_quality_landsat_8) (accessed on 29 August 2019).
27. Blanzieri, E.; Melgani, F. Nearest Neighbor Classification of remote sensing images with the maximal margin principle. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1804–1811. [\[CrossRef\]](#)
28. Ma, L.; Crawford, M.M.; Tian, J. Local manifold learning-based k-Nearest-Neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [\[CrossRef\]](#)
29. Swain, P.H.; Hauska, H. The decision tree classifier: Design and potential. *IEEE Trans. Geosci. Electron.* **1977**, *15*, 142–147. [\[CrossRef\]](#)
30. Yi-Bin, L.; Ying-Ying, W.; Xue-Wen, R. Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; IEEE: Toulouse, France, 2017; pp. 1526–1530. [\[CrossRef\]](#)
31. Fernandes, G.W.; Goulart, F.F.; Ranieri, B.D.; Coelho, M.S.; Dales, K.; Boesche, N.; Bustamante, M.; Carvalho, F.A.; Carvalho, D.C.; Dirzo, R.; et al. Deep into the mud: Ecological and socio-economic impacts of the dam breach in Mariana, Brazil. *Braz. J. Nat. Conserv.* **2016**, *14*, 35–45. [\[CrossRef\]](#)
32. Mielke, C.; Boesche, N.K.; Rogass, C.; Kaufmann, H.; Gauert, C.; de Wit, M. Spaceborne mine waste mineralogy monitoring in South Africa, applications for modern push-broom missions: Hyperion/OLI and EnMAP/Sentinel-2. *Remote Sens.* **2014**, *6*, 6790–6816. [\[CrossRef\]](#)
33. Rosell-Melé, A.; Moraleda-Cibrián, N.; Cartró-Sabaté, M.; Colomer-Ventura, F.; Mayor, P.; Orta-Martínez, M. Oil pollution in soils and sediments from the Northern Peruvian Amazon. *Sci. Total Environ.* **2018**, *610*, 1010–1019. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Asht, S.; Dass, R. Pattern recognition techniques: A review. *Int. J. Comput. Sci. Telecommun.* **2012**, *3*, 25–29.
35. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
36. Jahne, B. *Computer Vision and Applications: A guide for Students and Practitioners*, 1st ed.; Elsevier: London, UK, 2000.
37. Ye, Q.-Z.; Wu, P.; Zhang, M.-L. Research on Automatic Highway Extraction Technology Based on Spectral Information of Remote Sensing Images. *J. Inf. Hiding Multimed. Signal Proc.* **2017**, *8*, 368–380.

38. Chaker, R.; Al Aghbari, Z.; Junejo, I.N. Social network model for crowd anomaly detection and localization. *Pattern Recognit.* **2017**, *61*, 266–281. [\[CrossRef\]](#)
39. Ahmed, M.; Mahmood, A.N.; Islam, M.R. A survey of anomaly detection techniques in financial domain. *Future Gener. Comput. Syst.* **2016**, *55*, 278–288. [\[CrossRef\]](#)
40. Ahmed, M.; Mahmood, A.N.; Hu, J. A survey of network anomaly detection techniques. *J. Netw. Comput. Appl.* **2016**, *60*, 19–31. [\[CrossRef\]](#)
41. Agrawal, S.; Agrawal, J. Survey on anomaly detection using data mining techniques. *Procedia Comput. Sci.* **2015**, *60*, 708–713. [\[CrossRef\]](#)
42. Akoglu, L.; Tong, H.; Koutra, D. Graph based anomaly detection and description: A survey. *Data Min. Knowl. Discov.* **2015**, *29*, 626–688. [\[CrossRef\]](#)
43. Pawar, A.D.; Kalavadekar, P.N.; Tambe, S.N. A survey on outlier detection techniques for credit card fraud detection. *IOSR J. Comput. Eng.* **2014**, *16*, 44–48. [\[CrossRef\]](#)
44. Zhang, J. Advancements of outlier detection: A survey. *ICST Trans. Scalable Inf. Syst.* **2013**, *13*, 1–26. [\[CrossRef\]](#)
45. Lin, A.Y.M.; Novo, A.; Har-Noy, S.; Ricklin, N.D.; Stamatiou, K. Combining GeoEye-1 satellite remote sensing, UAV aerial imaging, and geophysical surveys in anomaly detection applied to archaeology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 870–876. [\[CrossRef\]](#)
46. Xie, M.; Han, S.; Tian, B.; Parvin, S. Anomaly detection in wireless sensor networks: A survey. *J. Netw. Comput. Appl.* **2011**, *34*, 1302–1325. [\[CrossRef\]](#)
47. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection for discrete sequences: A survey. *IEEE Trans. Knowl. Data Eng.* **2010**, *24*, 823–839. [\[CrossRef\]](#)
48. Zhang, W.; Yang, Q.; Geng, Y. A Survey of Anomaly Detection Methods in Networks. In Proceedings of the International Symposium on Computer Network and Multimedia Technology, Wuhan, China, 18–20 January 2009; IEEE: Toulouse, France, 2009; pp. 1–3. [\[CrossRef\]](#)
49. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008.
50. Transon, J.; d’Andrimont, R.; Maignard, A.; Defourny, P. Survey of hyperspectral earth observation applications from space in the sentinel-2 context. *Remote Sens.* **2018**, *10*, 157. [\[CrossRef\]](#)
51. Frontera-Pons, J.; Veganzones, M.A.; Pascal, F.; Ovarlez, J.P. Hyperspectral anomaly detectors using robust estimators. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 720–731. [\[CrossRef\]](#)
52. Matteoli, S.; Diani, M.; Theiler, J. An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2317–2336. [\[CrossRef\]](#)
53. Matteoli, S.; Veracini, T.; Diani, M.; Corsini, G. Models and methods for automated background density estimation in hyperspectral anomaly detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *51*, 2837–2852. [\[CrossRef\]](#)
54. Matteoli, S.; Diani, M.; Corsini, G. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerosp. Electron. Syst. Mag.* **2010**, *25*, 5–28. [\[CrossRef\]](#)
55. Nazeer, M.; Nichol, J.E. Combining Landsat TM/ETM+ and HJ-1 A/B CCD sensors for monitoring coastal water quality in Hong Kong. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1898–1902. [\[CrossRef\]](#)
56. Ha, N.T.T.; Koike, K.; Nhuan, M.T.; Canh, B.D.; Thao, N.T.P.; Parsons, M. Landsat 8/OLI two bands ratio algorithm for Chlorophyll-A concentration mapping in hypertrophic waters: An application to West Lake in Hanoi (Vietnam). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4919–4929. [\[CrossRef\]](#)
57. Chen, J.; Zhu, W.-N.; Tian, Y.Q.; Yu, Q. Estimation of colored dissolved organic matter from Landsat-8 imagery for complex inland water: Case study of Lake Huron. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2201–2212. [\[CrossRef\]](#)
58. Kotchi, S.O.; Brazeau, S.; Turgeon, P.; Pelcat, Y.; Légaré, J.; Lavigne, M.-P.; Essono, F.N.; Fournier, R.A.; Michel, P. Evaluation of Earth observation systems for estimating environmental determinants of microbial contamination in recreational waters. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3730–3741. [\[CrossRef\]](#)
59. Chang, N.-B.; Vannah, B.; Yang, Y.J. Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2426–2442. [\[CrossRef\]](#)

60. Li, Z.; Liu, H.; Luo, C.; Li, P.; Li, H.; Xiong, Z. Industrial wastewater discharge retrieval based on stable nighttime light imagery in China from 1992 to 2010. *Remote Sens.* **2014**, *6*, 7566–7579. [[CrossRef](#)]
61. Mustard, J.F.; Sunshine, J.M. Spectral analysis for earth science: Investigations using remote sensing data. *Remote Sens. Earth Sci. Man. Remote Sens.* **1999**, *3*, 251–307.
62. Verstraete, M.M.; Pinty, B. Designing optimal spectral indexes for remote sensing applications. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1254–1265. [[CrossRef](#)]
63. Brezonik, P.L.; Olmanson, L.G.; Finlay, J.C.; Bauer, M.E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sens. Environ.* **2015**, *157*, 199–215. [[CrossRef](#)]
64. Xie, X.; Li, B. A Unified Framework of Multiple Kernels Learning for Hyperspectral Remote Sensing Big Data. *J. Inf. Hiding Multimed. Signal Process* **2016**, *7*, 296–303.
65. Li, Y.; Li, J.; Pan, J.-S. Hyperspectral Image Recognition Using SVM Combined Deep Learning. *J. Internet Technol.* **2019**, *20*, 851–859.
66. Sublime, J.; Kalinicheva, E. Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami. *Remote Sens.* **2019**, *11*, 1123. [[CrossRef](#)]
67. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the International Conference on Artificial Neural Network, Rhodes, Greece, 4–7 October 2018; Springer: Cham, Switzerland, 2018; pp. 270–279.
68. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl. Based Syst.* **2015**, *80*, 14–23. [[CrossRef](#)]
69. Shao, L.; Zhu, F.; Li, X. Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *26*, 1019–1034. [[CrossRef](#)] [[PubMed](#)]
70. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [[CrossRef](#)] [[PubMed](#)]
71. Sun, S. A survey of multi-view machine learning. *Neural Comput. Appl.* **2013**, *23*, 2031–2038. [[CrossRef](#)]
72. Xu, Q.; Yang, Q. A survey of transfer and multitask learning in bioinformatics. *J. Comput. Sci. Eng.* **2011**, *5*, 257–268. [[CrossRef](#)]
73. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
74. Taylor, M.E.; Stone, P. Transfer learning for reinforcement learning domains: A survey. *J. Mach. Learn. Res.* **2009**, *10*, 1633–1685.
75. Committee on Developments in the Science of Learning, Committee on Learning Research and Educational Practice & National Research Council. In *How People Learn: Brain, Mind, Experience, and School*, Expanded ed.; National Academy Press: Washington, DC, USA, 2000.
76. Shoujing, Y.; Qiao, W.; Chuanqing, W.; Xiaoling, C.; Wandong, M.; Huiqin, M. A robust anomaly based change detection method for time-series remote sensing images. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing Ltd.: Bristol, UK, 2014; Volume 17, p. 012059.
77. Zhou, Z.-G.; Tang, P.; Zhou, M. Detecting Anomaly Regions in Satellite Image Time Series Based on Seasonal Autocorrelation Analysis. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing Spatial Information Science (XXIII ISPRS Congress), Prague, Czech Republic, 12–19 July 2016; pp. 303–310.
78. Chandola, V.; Vatsavai, R.R. A Gaussian Process Based Online Change Detection Algorithm for Monitoring Periodic Time Series. In Proceedings of the 2011 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, Mesa, AZ, USA, 28–30 April 2011; pp. 95–106. [[CrossRef](#)]
79. Tan, P.-N.; Steinbach, M.; Karpatne, A.; Kumar, V. *Introduction to Data Mining*, 2nd ed.; Pearson Education: Noida, India, 2018.
80. Bhaduri, K.; Das, K.; Votava, P. Distributed Anomaly Detection Using Satellite Data from Multiple Modalities. In Proceedings of the 2010 Conference on Intelligent Data Understanding, CIDU 2010, Mountain View, CA, USA, 5–6 October 2010.
81. Bormann, K.J.; McCabe, M.F.; Evans, J.P. Satellite based observations for seasonal snow cover detection and characterization in Australia. *Remote Sens. Environ.* **2012**, *123*, 57–71. [[CrossRef](#)]
82. Chen, C.; Yang, B.; Song, S.; Peng, X.; Huang, R. Automatic clearance anomaly detection for transmission line corridors utilizing UAV-Borne LIDAR Data. *Remote Sens.* **2018**, *10*, 613. [[CrossRef](#)]
83. USGS—The United States Geological Survey, “Earth Explorer”. Available online: <https://earthexplorer.usgs.gov/> (accessed on 25 April 2018).
84. QGIS Development Team. Available online: <https://www.qgis.org/> (accessed on 25 April 2018).

85. Gordon, G.; Stavi, I.; Shavit, U.; Rosenzweig, R. Oil spill effects on soil hydrophobicity and related properties in a hyper-arid region. *Geoderma* **2018**, *312*, 114–120. [CrossRef]
86. Pisano, A.; Bignami, F.; Santoleri, R. Oil spill detection in glint-contaminated near-infrared MODIS imagery. *Remote Sens.* **2015**, *7*, 1112–1134. [CrossRef]
87. Reátegui-Zirena, E.G.; Stewart, P.M.; Whatley, A.; Chu-Koo, F.; Sotero-Solis, V.E.; Merino-Zegarra, C.; Vela-Paima, E. Polycyclic aromatic hydrocarbon concentrations, mutagenicity, and Microtox® acute toxicity testing of Peruvian crude oil and oil-contaminated water and sediment. *Environ. Monit. Assess.* **2014**, *186*, 2171–2184. [CrossRef] [PubMed]
88. Gonzales, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
89. QGIS Project. QGIS User Guide Release 2.14. 2017. Available online: <https://docs.qgis.org/2.14/pdf/en/QGIS-2.14-UserGuide-en.pdf> (accessed on 25 April 2018).
90. OTB Development Team. The Orfeo ToolBox Cookbook, a Guide for Non-Developers Updated for OTB-3.10. 2011. Available online: <https://www.orfeo-toolbox.org/packages/archives/Doc/CookBook-310.pdf> (accessed on 25 April 2018).
91. Vivone, G.; Alparone, L.; Chanussot, J.; Mura, M.D.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [CrossRef]
92. OTB Development Team. The On-Line Orfeo ToolBox Cookbook, a Guide for Non-Developers Updated for OTB-3.10. 2011. Available online: [https://www.orfeo-toolbox.org/packages/doc/tests-rfc-52/cookbook-3b41671/Applications/app\\_Superimpose.html](https://www.orfeo-toolbox.org/packages/doc/tests-rfc-52/cookbook-3b41671/Applications/app_Superimpose.html); Available online: <https://www.qgis.org/> (accessed on 25 April 2018).
93. Reinhard, E.; Shirley, P.; Ashikhmin, M.; Troscianko, T. Second Order Image Statistics in Computer Graphics. In Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (APGV'04), Los Angeles, CA, USA, 7–8 August 2004; pp. 99–106.
94. OTB Development Team. The On-Line Orfeo ToolBox Cookbook, a Guide for Non-Developers Updated for OTB-3.10. 2011. Available online: [https://www.orfeo-toolbox.org/packages/doc/tests-rfc-52/cookbook-3b41671/Applications/app\\_TrainImagesClassifier.html](https://www.orfeo-toolbox.org/packages/doc/tests-rfc-52/cookbook-3b41671/Applications/app_TrainImagesClassifier.html) (accessed on 25 April 2018).
95. Congalton, R.G.; Mead, R.A. A review of three discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE-24*, 169–174. [CrossRef]
96. Marzano, F.S.; Scaranari, D.; Montopoli, M.; Vulpiani, G. Supervised classification and estimation of hydrometeors from C-Band dual-polarized radars: A Bayesian approach. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 85–98. [CrossRef]
97. Indu, J.; Kumar, D.N. Evaluation of precipitation retrievals from orbital data products of TRMM over a subtropical basin in India. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6429–6442. [CrossRef]
98. Che, X.; Feng, M.; Sexton, J.; Channan, S.; Sun, Q.; Ying, Q.; Liu, J.; Wang, Y. Landsat-based estimation of seasonal water cover and change in arid and semi-arid Central Asia (2000–2015). *Remote Sens.* **2019**, *11*, 1323. [CrossRef]
99. Chen, Y.; Fan, R.; Yang, X.; Wang, J.; Latif, A. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water* **2018**, *10*, 585. [CrossRef]
100. Natesan, S.; Armenakis, C.; Benari, G.; Lee, R. Use of UAV-borne spectrometer for land cover classification. *Drones* **2018**, *2*, 16. [CrossRef]
101. Bernardo, N.; do Carmo, A.; Park, E.; Alcântara, E. Retrieval of suspended particulate matter in inland waters with widely differing optical properties using a semi-analytical scheme. *Remote Sens.* **2019**, *11*, 2283. [CrossRef]
102. Snyder, J.P. Map projections: A working manual. In *US Geological Survey Professional Paper 1395*; United States Government Printing Office: Washington, DC, USA, 1987.

