The Application of Satellite Image Analysis in Oil Spill Detection

Paweł Tysiąc *, Tatiana Strelets and Weronika Tuszyńska

EKO-TECH Center, Faculty of Civil and Environmental Engineering, Gdansk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdansk, Poland; t_strelets@me.com (T.S.); w.tuszynskaa@gmail.com (W.T.) * Correspondence: pawtysia@pg.edu.pl

Abstract: In recent years, there has been an increasing use of satellite sensors to detect and track oil spills. The satellite bands, namely visible, short, medium infrared, and microwave radar bands, are used for this purpose. The use of satellite images is extremely valuable for oil spill analysis. With satellite images, we can identify the source of leakage and assess the extent of potential damage. However, it is not yet clear how to approach a specific leakage case methodologically. The aim of this study is the remote sensing analysis of environmental changes with the development of oil spill detection processing methods. Innovative elements of the work, in addition to methodological proposals, include the long-term analysis of surface water changes. This is very important because oil is very likely to enter the soil when water levels change. The classification result was satisfactory and accurate by 85%. The study was carried out using images from Landsat 5, Landsat 7, Landsat 8, Sentinel-1, and Sentinel-2 satellites. The results of the classification of the oil stains in active and passive technologies differ. This difference affects the methodology for selecting processing methods in similar fields. In the case of this article, the oil spill that occurred on 29 May 2020 in Norilsk was investigated and compared with data from other years to determine the extent of biodegradation. Due to the tank failure that occurred at the Norilnick power plant on that day, a large amount of crude oil leaked into the environment, contaminating the waters and soil of local areas. Research shows that oil spills may be caused by human error or may be the effect of climate change, particularly global warming.

Keywords: oil spill detection; satellite imagery; spatial data; environmental analysis

1. Introduction

Scientists believe that recent, more frequent oil spills may result from climate change in addition to human error. Permafrost is a phenomenon in which parts of the Earth’s crust remain below the freezing point of water [1]. Since the early 20th century, the severity of glacier freezing has been recorded. In the high-latitude regions of Eurasia, temperatures are rising by around 0.12 °C per year, which is seen as a much greater increase compared to other parts of the world [2]. In addition, along with permafrost thaw, large sinkholes open, releasing small amounts of greenhouse gases into the environment. Released carbon dioxide and methane are factors in environmental biodegradation [3].

Satellite imagery plays an important role in the study of the modern world. It allows us to better understand the changes taking place on our planet. Harvey [4] claims that “satellite technologies have led to one of the most productive periods in the history of cartography.” In turn, another author, Hall [5], claims that “Landsat revealed entirely new worlds hidden in the folds of a familiar world that we thought I knew so well.” A review of the literature on the use of remote sensing technology for oil spill detection and mapping confirms the high efficiency of remote sensing technology [6–12]. Research shows that active and passive sensors can be used to respond more quickly and precisely to such disasters [13–15]. To understand this phenomenon, Sentinel-2 [16] and Sentinel-1 [17]...
data are used, as well as images from Landsat 5, Landsat 7, and Landsat 8 [18]. Based on satellite images, many researchers have analysed oil spills that occurred in water reservoirs. Sun et al. [19] studied the spread of oil in the East China Sea as a result of a collision between two ships, a tanker and a freighter, using multisensory satellite images. They showed that by combining remote sensing and modelling techniques used in post-processing, they could provide an effective means of monitoring sea accidents. The same author [18] focused her subsequent research on the northern part of the Gulf of Mexico, which was destroyed in September 2004 by Hurricane Ivan. The hurricane caused oil leaks from nearby plants. The Landsat 5, Landsat 7, and Sentinel-2A satellites have been used for the first time to estimate damage caused by the above-mentioned natural disaster. Consequently, we found that many data can be used to understand the movement of oil residues.

On the other hand, Biermann [20], using the innovative floating debris index algorithm, separated microplastics and other elements, not being the natural environment of water, from algae and marine plants with an overall accuracy of 86%. The study was carried out at four different locations: Vietnam, Canada, Great Britain, and Ghana. Biermann [20] prepared action strategies for the following years. Australian researchers [21] assessed the sensitivity of the type and thickness of hydrocarbon oil on the remote sensing signal. Using the results of the collected data, they calculated the theoretical detection limit of each oil for two sensors: HYMAP hyperspectral and Quickbird multispectral. In their study, they found that the type of oil was critical for the application of optical remote sensing to detect and identify oil leaks. Currently, there are two methods available for remote oil thickness measurements, namely passive microwave radiometry and acoustic travel time with laser measurements [17]. Merv Fingas, together with Carl Brown, carried out a review of remote sensing achievements in oil spill detection [22]. They found that satellite radar technology has dominated the detection and mapping of sea backwaters.

Furthermore, the development of leakage detection on rivers is less than the development of maritime泄漏 detection technology. It is important to note that oil films have a reflective contrast between the surface of the sea and background water, ranging from 400 to 700 nanometres (nm) [23]. According to a 2008 study by Chinese scientists [24], the spectral characteristics of oil points are clear at wavelengths of 550 nm and 645 nm, and these wavelengths are the best wavelengths to monitor marine oil and estimate oil thickness. They also developed hyperspectral data that found that very fine oil spots change reflection at 440 nm and are useful for leakage tests. Klemas [25] in his paper described the colour of the oil as silvery and stated that it reflects light in a wide spectral range. Furthermore, he found that brown heavy oil is most preferably detected at a wavelength of about 650 nm and red-brown impurities at a wavelength of 700 nm. Sun et al. [26] conducted research on the hyperspectral characteristics of the polarized reflection of an oil spot on the sea surface. They found that seawater polarization was higher than oil spectra and observed changes in oil spectra thickness in near-infrared 785 and 880 nm. However, Greek scientists [27] observed that wavelengths ranging from 660 to 760 nm of multispectral sensors are best suited for oil spill detection in coastal areas. The longer the wavelength, the more difficult it is to identify the water and the image becomes unreadable. Another study of 2021 [28] was devoted to the creation of an index to map oil spills using various combinations of the spectral bands of the Sentinel-2 mission and decorrelation stretching methods. Furthermore, scientists were able to distinguish between different thicknesses of the oil. Most satellite sensors have bands that extend over the area that allow oil spills to be observed.

This paper focuses on analysing the environmental characteristics associated with oil spills using satellite data from various missions. Snow, ice, water, and vegetation were examined immediately before and after the accident. The results have been extended to longer-term analysis, helping to understand the origin of the problem and facilitate the choice of appropriate strategies for subsequent actions. Such practical approaches have not been seen in the literature. Therefore, the purpose of this article is to articulate the following research objectives clearly:

- conduct an analysis of oil spills in rivers,
group data for long- and short-term analysis,
- select the appropriate classes according to the remote sensing device,
- propose a practical method for monitoring oil spills.

The above research tasks will enable the practical aspects of evaluating natural disasters such as oil spills in rivers. In the past, they were primarily studied in the sea and oceans. The method of grouping long- and short-term analyses enables us to understand the origins of problems and select appropriate satellite products for analysis. The analysis of satellite images from the point of view of changes in water levels proposed in this paper will provide a practical indication of potential observation sites where oil can enter the soil directly, resulting in soil pollution.

2. Materials and Methods
2.1. Study Site and Problem Description

The river Ambarnaya near Norilsk is the area studied. It has a length of 60 km and a total area of 428 km². The river is supplied by rainwater and melting water [29]. It is shallow and the bottom is made up of gravel and glacier moraine limestone [30]. The river begins with its source, the Ambarnaya River (69°16'17" N 87°45'21" E), then flows northward into Lake Pyasino (69°28'44" N 8°55'27" E) (Figure 1). As a result of an accident, about 20,000 tons of diesel fuel flowed into the Ambarnaya River on 29 May 2020. The river is polluted in an area of approximately 7 km, or 20% of its total length. Dams have been installed in the river to prevent further pollution, but they have not completely prevented the spill of oil products into Lake Pyasino. Freshwater lakes cover an area of 735 km².

![Figure 1. The study location.](image)

The Norilsk region is rich in natural resources, such as nickel, copper, platinum, or high value hard coal [31]. According to 2004 data [21], the region’s mineral resources have reached nearly 480 billion tonnes. The city is mainly known as Nornickel, the company that manufactures heavy metals. It produces approximately 500,000 tonnes of copper annually, 500,000 tonnes of nickel annually, and 2 million tonnes of sulphur dioxide annually [32].
This is more than the annual production of France as a whole [23]. The National Aeronautics and Space Administration reports that heavy metal contamination in the areas near the Norilsk complex is so large and extensive that it has become economically viable to extract contaminated soil containing the elements platinum and palladium [23]. The list includes air pollution caused by solid particles, radioactive isotopes of strontium 90 and caesium 137. These are characterized by a long half-life. In the summer, the temperature difference in the atmosphere prevents smoke from escaping, leading to smoke accumulating as smog. The location of the factory makes the situation even worse, because it is adjacent to the river, which means that regardless of the direction, the wind blows smoke in the city. A 1999 study showed that copper and nickel concentrations in the soil increased up to 60 km from the city [33]. Most of the surrounding areas of Norilsk are Tundra, Lasundra, and the deciduous forest of the boreal, Taiga. The vegetation period is from June to August. Subarctic climates are characterized mainly by small vegetation such as marshes, lichens, and dwarf trees. There are also pines, junipers, and hemisphere plants [23]. In a 2003 study, it was found that the trees with the highest concentrations of sulphur in the needles are found in the most damaged areas of the forest [34].

The disaster was described as the second-largest oil spill in Russian history. On 29 May 2020, a support under the fuel tank of the Norilsk power plant of the Nadezhda metal factory collapsed, ejecting about 21,000 tonnes into the surrounding river. The oil contaminated the Ambarnaya River and surrounding areas, and then entered Lake Piasino, which flows into the Arctic Kara Sea. Total oil pollution has contaminated 350 square kilometres of land. The 1989 Exxon Valdez disaster was similar, when oil tankers descended in Prince William Bay, Alaska, spilling 37,000 tons of oil into the surrounding waters. Svetlana Radionova, chief executive of Rosprirodnadzor (the Federal Russian Natural Resource Surveillance Service), said that water samples tested exceed the permitted concentrations of oil products by thousands [35]. Norilsk-Taimyr Energy Company stated that the incident was caused by global climate change and global permafrost defrost, in particular the 2020 heat wave.

2.2. Methodology Description

In our study, we strongly emphasize the applicability of the solution obtained and its potential for replication in similar research. Therefore, Figure 2 shows the flow of activities to achieve the objectives set in practice.

As shown in Figure 2, Landsat data were used for long-term analyses using the Google Earth Engine. Sentinel missions were used to analyze oil spill to better spatially define the results and divide them into passive and active techniques. In addition to the processes presented to detect changes in water and oil leaks, spatial indices have been used to correctly understand the situation in the field. We used, respectively, NDVI (Normalized Difference Vegetation Index) to study vegetation, NDWI2 (Second Normalized Difference Water Index), and SWM (Sentinel Water Mask) to monitor water in water reservoirs and NDSII (Normalized Difference Snow and Ice Index) to assess the impact of snow cover on the results. The abbreviations are listed in alphabetical order in Abbreviations.

2.2.1. Satellite Missions

Satellite images from the Sentinel-1, Sentinel-2 and Landsat 5, Landsat 7, and Landsat 8 systems were used for the analysis. Landsat mission products were selected for their long-term data coverage, Sentinel-2 images for their good spatial resolution, and Sentinel-1 due to the characteristics of the data that will allow to determine the thickness of the oil spot.
The Sentinel-1 mission was designed as a constellation of two satellites: Sentinel-1A launched on 3 April 2014 and Sentinel-1B, launched on 25 April 2016. The satellites are in synchronous orbit with the Sun at an altitude of 693 km, with an inclination of 98.18°. Each satellite is equipped with an advanced C-SAR radar, regardless of weather conditions. The Sentinel-1 mission was developed to ensure the continuity of Earth’s radar mapping. It is characterized by a short revisit time, which is 12 days. This makes the mission suitable for monitoring land and ocean surface changes, as well as for security and emergency responses [36]. Sentinel-2 is a European mission for multi- and multi-spectral imaging with wide-range data acquisition and high resolution. It consists of two satellites, rotating in the same orbit synchronized with the movement of the Sun, 180° apart from each other. The position of each satellite is measured by a dual-frequency Global Navigation Satellite System (GNSS) receiver. The full specification is designed to ensure a high reimaging frequency of 5 days at the equator. This guarantees the continuity of data collection and delivery.

Landsat 5 is the longest-running Earth observation satellite in history. It was launched on 1 March 1984, with a Delta 3920 rocket launched from the Western Test Range at Vandenberg Air Force Base, California, and spent 29 years collecting and sending information about our planet. It was designed and built at the same time as Landsat 4 and had the same hardware specifications: a multispectral scanner system (MSS) and thematic mapping (TM) instruments. Landsat 7 was launched on 15 April 1999, from the same base as Landsat 5, on a Delta II launch vehicle. It is in a polar orbit, i.e., at a tilt of 98°, synchronized with the azimuthal position of the Sun, which means that it scans 100% of the Earth’s surface, which takes 16 days. The main instrument on board the Landsat 7 satellite is the Enhanced Thematic Mapper Plus (ETM+), a “whisk broom scanner” image sensor. Landsat 8 is another American Earth observation satellite that began its mission on 11 February 2013. It was launched into space with an Atlas V rocket from the same base as its predecessors, Vandenberg, California. The mission ensures continuous acquisition and sharing of Earth data thanks to the use of two sensors: Operational Land Imager (OLI) and Thermal InfraRed Sensor (TIRS) [37].
2.2.2. Indices and Classification Methods

For the purposes of this work, the following indices were used:

- **NDVI**

\[
NDVI = \frac{NIR - RED}{NIR + RED}
\]

NDVI is an indicator for determining the state of vegetation [38–40]. NDVI is based on the contrast between the greatest near infrared reflection and the red absorption. The range of NDVI values is \(-1\) to 1. Negative NDVI values (values close to \(-1\)) correspond to water. Values close to zero (\(-0.1–0.1\)) correspond to barren regions of rock, sand, or snow. Low, positive values are for shrubs and grasslands (values from about 0.2 to 0.4), while high values are for moderate and tropical rainforests (values close to 1).

- **NDWI2**

\[
NDWI2 = \frac{GREEN - NIR}{GREEN + NIR}
\]

NDWI2 is used to monitor changes in water content in water reservoirs. Since water reservoirs strongly absorb light in the visible and infrared electromagnetic spectrum, green and near infrared bands are used to illuminate water reservoirs. Index values greater than 0.3 usually relate to water bodies. The higher the value, the greater the content of healthy water in a given tank. On the other hand, results below 0.3 indicate no water [41].

- **NDSII**

\[
NDSII = \frac{RED - SWIR}{RED + SWIR}
\]

NDSII relies on the ratio of visible and shortwave infrared wavelengths to separate snow/ice pixels from cloud cover and other undesirable elements. It is calculated by dividing the red and short infrared bands by their sum, because snow indicates the lowest and the highest reflectance in these bands, respectively. The index values range from \(-1\) to 1. When surveying areas covered with a significant amount of snow, including Alaska, Minnesota, and Iceland, the NDSI mapped the snow cover well and distinguished it from clouds for a value of about 0.4 [42].

- **SWM**

\[
SWM = \frac{BLUE + GREEN}{NIR + SWIR}
\]

SWM provides fast and effective water detection. It is dedicated to Sentinel-2 products. It is calculated as the sum of the blue and green bands divided by the sum of the near and short infrared bands. The range of index values is from 0–12, and the optimal threshold is values in the range 1.4–1.6, when the index reaches an accuracy of 96–99% depending on the test area [43].

2.2.3. Classification

In order to create the classification, the Sentinel-2 image from 31 May 2020 (2 days after the oil leak) was used. For the purposes of this study, the supervised SVM (support vector machine) classification was selected from Geomatica Banff software (version 2018 SP2), and the object-based SVM classifier was used in Python environment [44]. The support vector machine classifies the data by searching for a hyperplane that maximizes the margin between classes in the training data. The algorithm is named after a relatively small number of observations, called support vectors. The classification parameters of this study are mainly useful for high resolution data [45]. We decided to use Geomatica Banff software as it allows us to carry out such classification on lower resolution images. The first steps in
the workflow are to segment, calculate attributes, and import ground truth points for the initial image. Once that is done, the attributes fields are exported to a text file. Then, the training model is generated and SVM classification is run on that initial image. Figure 3 shows the basic workflow.

<table>
<thead>
<tr>
<th>Initial image</th>
<th>SVM Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Step 1</td>
</tr>
<tr>
<td>A hierarchical region-growing segmentation</td>
<td>Vector Layer Import</td>
</tr>
<tr>
<td>Step 2</td>
<td>Step 2</td>
</tr>
<tr>
<td>Polygons attribute calculation</td>
<td>SVM kernel function choice</td>
</tr>
<tr>
<td>Step 3</td>
<td>Step 3</td>
</tr>
<tr>
<td>Ground Truth Points Import</td>
<td>Optimization and cross-validation</td>
</tr>
<tr>
<td>Step 4</td>
<td>Step 4</td>
</tr>
<tr>
<td>Attributes Names Export</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** Basic SVM classification workflow.

The SVM classification approach, which is used by Geomatica Banff software, is called one against all. This approach generates \( n \) classifiers, where \( n \) is the number of classes. The output is the class that corresponds to the SVM with the largest margin. With multiclass, SVM has to interpret \( n \) hyperplanes. This requires \( n \) quadratic programming (QP) optimization problems, each of which separates one class from the remaining classes. Applying each classifier to the test-data vectors gives one vote to the winning class. The data are assigned to the class label of the class with the most votes.

First, the hierarchical segmentation elements (step 1 of Figure 3) have been defined. The input layer is the initial image. As an output, segmentation data with parameters, i.e., scale parameters, is expected. In this study, we established set scale = 35 to create larger segments. The next step (step 2 of Figure 3) adds attributes of objects to the attribute of the vector layer. In this study, bands 2–5 from Sentinel-2 image were used, which are: Blue, Green, Red and Near-Infrared. Then, the statistical mean values were calculated: Mean_Blue, Mean_Green, Mean_Red, Mean_NIR. Moreover, we calculated geometric values which are: elongation, roundness, compactness, squarness, shape factor, convexity, solidity, main axis length, and minor axis length. Additionally, an NDVI index was calculated which helped to classify vegetation, meaning: the mean of the vegetation indices across all pixels contained within each object. All these values are assigned to the segmentation layer calculated in Step 1.

The segmentation result is shown in Figure 4. Furthermore, training site collection information was obtained and imported into Step 3 for compilation. This information is used to conduct quantitative accuracy assessments after the evaluation program is started. The number of training and test fields is shown in Table 1. The last stage of the initial image processing is to export the obtained data to text files as variables of the Geomatica Banff (version 2019) software classifier.

If two classes cannot be distinguished linearly in two-dimensional spaces, they might be separated in high-dimensional spaces (hyperplanes). The SVM classification uses the kernel function to map support vectors extracted from training data to higher dimensions. In this article, we have used the radial-basis function (RBF) kernel, respectively. Typically, it provides the best results. In Geomatica Banff software, it is recommended to normalize data so that each attribute can be treated equally to discriminate classes. Attributes are normalized by using linear scaling to produce a range from 0 to 1. The minimum value is mapped to zero and the maximum is mapped to one. All of the values in between are scaled linearly.
In Support Vector Machines (C) is a hyperparameter that determines the penalty for misclassification of observations. In addition, there is a parameter gamma (γ) for the Radial-Basis Function kernel. As a rule, the Grid Algorithm is used to optimize these values. The truth is that this method is ineffective and time-consuming. To simultaneously optimize the feature subset and the SVM kernel parameters, the Geomatica Banff software uses attempts to increase the classification accuracy rate by employing two computing optimization-based approaches: genetic algorithm (GA) and particle swarm optimization (PSO) [46]. The calculated parameter achieve values generally with the best accuracy for the training samples while reducing the possibility of model over-fitting. With these rules and parameters, the algorithm built a model that assigns new examples to classes [47,48]. The Confusion Matrix and Accuracy statistics for SVM classification as results are presented in Table 2.

**Table 1. Number of Training and Test Samples.**

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Training</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>88</td>
<td>46</td>
<td>134</td>
</tr>
<tr>
<td>Oil</td>
<td>28</td>
<td>11</td>
<td>39</td>
</tr>
<tr>
<td>Ground</td>
<td>81</td>
<td>54</td>
<td>135</td>
</tr>
<tr>
<td>Vegetation</td>
<td>34</td>
<td>19</td>
<td>53</td>
</tr>
<tr>
<td>Snow</td>
<td>38</td>
<td>19</td>
<td>57</td>
</tr>
<tr>
<td>Buildings</td>
<td>18</td>
<td>9</td>
<td>27</td>
</tr>
</tbody>
</table>
Table 2. Confusion Matrix and Accuracy statistics for SVM Classification.

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Water</th>
<th>Vegetation</th>
<th>Ground</th>
<th>Snow/Ice</th>
<th>Oil</th>
<th>Buildings</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>Vegetation</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Ground</td>
<td>0</td>
<td>10</td>
<td>51</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>Snow/Ice</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Oil</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Buildings</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>19</td>
<td>54</td>
<td>19</td>
<td>11</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Overall accuracy: 87.975%
Overall kappa statistics: 0.839
95% Confidence Interval (82.587–93.363%)

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Producer’s Accuracy</th>
<th>95% Confidence Interval</th>
<th>User’s Accuracy</th>
<th>95% Confidence Interval</th>
<th>Kappa Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>100.000%</td>
<td>(98.913–100.000%)</td>
<td>100.000%</td>
<td>(98.913–100.000%)</td>
<td>1.0000</td>
</tr>
<tr>
<td>Vegetation</td>
<td>42.105%</td>
<td>(17.273–66.938%)</td>
<td>80.000%</td>
<td>(50.208–100.000%)</td>
<td>0.7727</td>
</tr>
<tr>
<td>Ground</td>
<td>94.444%</td>
<td>(87.409–100.000%)</td>
<td>78.462%</td>
<td>(67.698–89.225%)</td>
<td>0.6728</td>
</tr>
<tr>
<td>Snow/Ice</td>
<td>89.474%</td>
<td>(73.043–100.000%)</td>
<td>100.000%</td>
<td>(97.059–100.000%)</td>
<td>1.0000</td>
</tr>
<tr>
<td>Oil</td>
<td>100.000%</td>
<td>(95.455–100.000%)</td>
<td>84.615%</td>
<td>(61.156–100.000%)</td>
<td>0.8346</td>
</tr>
<tr>
<td>Buildings</td>
<td>66.667%</td>
<td>(30.313–100.000%)</td>
<td>85.714%</td>
<td>(52.648–100.000%)</td>
<td>0.8485</td>
</tr>
</tbody>
</table>

Quantity Disagreement: 8.228%
Allocation Disagreement: 3.797%

2.2.4. Short and Long-Term Analysis

In order to detect changes in surface waters, two Sentinel-2B images were used, one before the disaster on 23 May 2020 (25.81% cloudiness) and another after the initial water treatment on 10 July 2020 (1.20% cloudiness). The spatial resolution of the pixel has been reduced to 10 m in order to simplify and speed up the work of the program. Subsequently, a subset was established, cropping the image to the designated survey area. The bands for the analysis were selected: blue, green, red, near, and mid-infrared. Then, the NDWI2 and SWM indices were calculated to detect surface water in wetlands and measure their extent. In the next step, a formula of actions for both images was introduced (if NDWI2 > 0.3 or SWM > 1, then 1 else 0). Based on calculating water values, information on water level changes between time acquisition is calculated. The difference of the results gives information about the change in the water level.

Oil spill mapping technology is successfully used to identify and map oil spills. As mentioned in Section 1, the best wavelength for detecting leaks is 400–700 nm. One of the most effective methods of detecting oil leaks is synthetic aperture radar (SAR). The Sentinel-1A image, obtained on 4 May 2020, was used to visualize oil spills. The image was cropped to the analysed area of study and the image noise was removed with a speckle filter tool. Oil leakage was detected automatically with the Oil Spill Detection tool available in Snap (Sentinel Application Platform) Desktop software. It involves two pre-processing steps: masking of inland areas and radiometric calibration so that the pixel values actually represent the radar backscatter from the surface. After these pre-processing steps, an algorithm detects oil spots and assembles them into a single group. The output is a binary mask. Results were compare with those, obtained from Sentinel-2 image from 31 May 2020. The sequence of steps was the same as during the detection of surface water changes. Bands 1, 9, and 10 that represent aerosols, water vapor, and cirrus clouds, respectively, have been removed [49]. New combinations of bands were created in the range of visible bands to
short infrared. The bands representing edges of the absorption characteristics for the oil were summed as a numerator and the band closest to this absorption as a denominator.

In order to check the water level in the studied area, the JRC (Joint Research Centre) Yearly Water Classification History v1.3 data set was used [50], on the basis of which the script in Google Earth Engine was written. The script needs a few additional commands that define, among others, the name for the loaded bands, the minimum and maximum pixel value, and the colour palette. JRC Yearly Water Classification History is a database containing maps of location and maps of temporal distribution of surface waters from 1984 to 2020. Thus, it provides statistics on the range and changes of these water surfaces based on over 4 million images from Landsat 5, 7, and 8. In addition, this dataset includes an annual classification of the seasonal waters based on annual occurrence values, so the detection of water is classified as either still water or seasonal.

To determine the change in the water content of the reservoir around the Norilsk factory, 74 images were used from 23 March 1985 to 14 May 1998, from the Landsat 5 mission, 24 images of Landsat 7 from the period of 15 March 1999 until 11 April 2013, and 50 Landsat 8 images from 5 March 2014 to 15 April 2020. Each image has a total cloudiness of no more than 15% and a spatial resolution of 30 m. Furthermore, the image properties have been assessed with Google Earth Engine at scales from 0 to 9, where 0 means very unfavourable imaging (meaning: the image is covered by clouds), and 9 excellent (meaning: the image has no cloud coverage or the % is irrelevant) in terms of further analysis. Moreover, image scenes in Google Earth Engine contain imagery with digital numbers (DNs) that represent scaled radiances. The conversion of DNs to at-sensor radiances is a linear transformation using coefficients stored in scene metadata [51]. Conversion to TOA (or at-sensor) reflectance is a linear transformation that accounts for solar elevation and seasonally variable Earth-Sun distance. This means that the data used in the analysis are relatively high-quality and low-cloudy, resulting in more reliable results. Then, the scripts were written in the JavaScript programming language. The first one was used to load the data set. After loading the required dataset, the algorithm calculated the NDWI2 and NDVI indices. The indicators were calculated separately for each satellite. Next, the NDWI2 and NDVI data series were combined into a single table.

Visual data analysis involves uncomplicated data preparation. These types of interpretations are often easier to understand and read for recipients than complex activities that produce unambiguous results at first glance. Such analysis is particularly used in intense or extensive changes. The most suitable image from Sentinel-2 satellite was selected. The criteria used were cloud coverage that could not cover more than 15% of images from 25 May to 10 July 2020, which is before and after the oil spill. Then, using specific commands and selecting appropriate spectral bands, the NDVI and NDWI2 indices were calculated to simulate the actual and distortion colour images. In addition, the data for the analysis of 2016 and 2021 from the same growth period were used to provide a comprehensive overview of the changes in the Arctic region. The only difference from the previous script was the change of the year in the second line of the formula for 2016 and 2020, respectively.

To analyse the cover of snow and ice in the research area, Google Earth Engine was used. This script is used to load a data set (Sentinel-2), as well as to select the most favourable images from a given period of time by the filter (based on cloud coverage). In addition, the script computes the NDSII index. The script was made for the imagery before and 2 days after the leak. The selected dates are: 14 May 2020 and 31 May 2020, respectively.

3. Results

To present test results, Sentinel-2 data in visual analysis were shown in RGB colours composition. The satellite image taken on 10 May 2020 (two weeks before the disaster), was used as a reference image (Figure 5a). The colours of the Ambarnaya River water do not attract objections and the surrounding vegetation is green. The image taken on 31 May 2020 (Figure 5b), i.e., two days after the accident, differs. The earth is barren, and the water colour is not blue, but grey and green. The image clearly shows the red spots of
oil that seeped into the river. The image from 8 June 2020 (Figure 5c) shows an area that has undergone initial cleaning. The image still shows the pollution on the river (red oil stains appear in the surrounding troughs and streams). A satellite image from 13 June 2020 (Figure 5d) shows the Ambarnaya River devoid of visible pollution. Unfortunately, the red pixels in the channel and stream near the river can be seen as red spots of residual oil. This means that despite the response of the services, it was impossible to completely stop and clean up toxins from the environment.

Figure 5. Visual Oil Spill Analysis (a) 10 May 2020, (b) 31 May 2020, (c) 8 June 2020, (d) 13 June 2020.

The exact location of the red leakage point is determined by classification. The results of the SVM classification are shown in Figure 6. Using the Sentinel-2 image to analyse the classification, there is a large number of red pixels representing oil spill-related substances. These are mainly concentrated around water reservoirs, i.e., rivers, beds, and streams. There is a tendency to subside at the edges of the reservoirs, where it can easily enter the soil from shallow waters.

The results of surface water change detection in the area studied are alarming. The data (from Sentinel 2) taken for analysis cover the dates: 23 May 2020 (before the leak) and 10 July 2020 (after the leak). Figure 7 shows the changes that took place during this period. Red pixels indicate areas where the amount of water has decreased during this time. The white colour of the pixel means no change in the water level, and the blue pixel, which is not present on the map, would mean an increase in the water level. By analysing the image, it can be seen that the Ambarnaya River lost a lot of water during this time. This could be due to works related to area cleaning, which removes large amounts of contaminated water from the river bed. This significant decrease in water level could be linked to serious oil pollution.
Figure 6. SVM Classification results.

Figure 7. Water change Detection.
The results of mapping the oil spill with satellite images from Sentinel-1 missions are interesting (Figure 8a). The radar data have revealed more oil contamination in the area studied than by analysing Sentinel-2 images (Figure 8b). Not only is the Ambarnaya River contaminated, but also small reservoirs, basins, channels, and nearby streams. The leak detection results using Sentinel-2 data (Figure 8b) are slightly less accurate than the above-mentioned SAR methodology. However, they can visually evaluate the thickness of the oil detected. In Figure 8b, blue-colour denotes thick oil spills and green-blue spots represent thinner oil layers. Using Sentinel-1 radar data, the actual contamination state can be observed. The appropriate Sentinel-2 bands tell us about oil density and thickness. The combination of both methods gives more accurate information about the leak and allows us a better understanding of its scale.

![Figure 8](image-url)

**Figure 8.** Oil Detection Results (a) from Sentinel 1 imagery (detected oil marked in red), (b) from Sentinel 2 imager (detected oil marked in blue and green-blue).
The results of the analysis of water quality in the Norilsk region using the JRC Annual Water Classification History v1.3 data set are shown in Figure 9. Solid waters are blue, and the areas in which they do not exist are red. Seasonal water was marked with white pixels. The image analysis shows that water losses have increased significantly since 1984. The red pixel covers a large part of the resulting map. They are largely concentrated in the river basins, i.e., streams and channels, as well as around Lake Pyasino (upper left corner). This could mean that the toxins that were draining into the water began to accumulate around the larger reservoirs, slowly degrading the natural environment. Nevertheless, the climate has changed significantly since 1984. Siberia’s regions are warming faster, resulting in lake drying.

The average annual NDWI2 index from 1985 to 2021 monitors changes related to water content in water bodies over a given period (Figure 10). The results greater than or equal to the value of 0.3 are synonymous with the good water quality of the studied areas. Such results in the study area are visible only between 2001 and 2002, and in 2011 and 2013. The remaining data, obtained from 36 years of research, are below the water detectability.

The average annual NDVI index from 1985 to 2021 remains in the range of values from about \(-0.45\) to 0.05 for the analysed period of time (Figure 11). The index is based on the contrast between the near-infrared reflection and the visible red absorption. Values in the
range of $-0.1$–$0.1$ correspond to barren areas, sand, or snow. Index results below this value are water, and above are vegetations. The analysis of the results shows a practical lack of healthy vegetation in the studied area. For 2020, the result fluctuates towards $-0.05$. This means that the test area was detected as sterile, with no clearly visible signs of healthy water, which is confirmed by the NDWI2 indicator described above.

![NDWI2 Values from 1985 to 2021](image)

**Figure 10.** NDWI2 Values from 1985 to 2021.

![NDVI Values from 1985 to 2021](image)

**Figure 11.** NDVI Values from 1985 to 2021.

Data from 14 May 2020 (Figure 12a) and 13 June 2020 (Figure 12b) were used to monitor and study the snow cover of the studied area. The NDSII (Normalized Difference Snow and Ice Index) was used. When analysing data before leakage, it is possible to see large amounts of snow cover, particularly on the left side of the map, showing white pixels. In June, the snowfall declined dramatically. The snowfall in Lake Pyasino is also markedly reduced in the upper and central parts of the two maps. In the months of May and June, Norilsk has normal snow. The sharp decline is largely due to climatic anomalies, making it one of the warmest sources in Siberia’s history. Snow defrosting between May and June resulted in a greater amount of water in the soil, which confirms the increase in vegetation.
4. Discussion

A rapid response from rescue personnel is essential during oil spills and other natural disasters [52]. Oil spills are detectable using the spatial and spectral resolutions provided by the sensors used in this analysis. The proposed methodology is based on a long-term analysis to better understand climate change in the area studied. Analyses have been carried out to identify and analyze the possible causes of the collapse of the foundations under the fuel tank. Based on the results obtained, it can be concluded that May 2020 was a very warm period for Norilsk. This was confirmed by analysis using NDWI2, NDVI, SWM, and NDSII indices. Large water reservoir anomalies can be found by detecting water changes and multi-time analysis.

The leakage tests were carried out with Sentinel-1 and Sentinel-2 missions to determine the thickness of the leak and actual water contamination. Moreover, Norilsk, due to its geographic location, is a complicated area to monitor by remote sensing methods due to year-round cloudiness and precipitation. For this reason, the first data after the leak are from 31 May 2020. According to the official statement of Vladimir Potanin, CEO of Norilsk Nickel, the main cause of the fuel tank failure was the heatwave in May 2020, which caused the ground under the plant to thaw [16]. Permafrost covers about 60% of the total area of Russia and 90% of Norilsk and its vicinity. According to reports by climatologists, since the mid-1970s the rate of temperature increase in northern parts of Russia has been more than five times greater than the value calculated for global warming, which is 0.18 °C. Moreover, since the 1980s, there has been a drastically rapid thawing of sea ice in the Arctic Ocean due to climate change. The volume of sea ice in 2019 decreased by more than 50% compared to the 1979–2019 average. Contemporary climate models confirm a more intense warming of Siberia, which will result in a significant loss and change in the extent of permafrost [49]. The May 2018 decree of the President of Russia, entitled “On the Foundations of the State Policy of the Russian Federation in the Field of Industrial Safety”, states that “60–70% of equipment used in hazardous production plants has reached the standard service life of the equipment. Under such conditions, the socio-economic damage caused by accidents can be estimated at 600–700 billion rubles annually, which will adversely affect the economic stability of the Russian Federation” [53].

According to the official website of Nornickel [54], by October 2020, the leak was: “fully located and most of the fuel-water mixture collected”. By 28 September 2020, a total of 34,451 m³ of mixture had been collected, and by 28 October 2020, “more than 25,000 m³ of mixture was pumped through newly installed pipelines from temporary tanks to storage for subsequent separation.” A study published in Nature [49] found that almost half of all oil and gas fields in Siberia are located in areas with significant thawing and subsidence potential. Global warming can have a devastating effect not only on local nature and people living there, but also on the entire global climate system, e.g., by melting permafrost, reducing snow cover, and melting ice [55]. The Norilsk disaster shows how climate change puts a huge strain on the Arctic, its indigenous peoples, and local flora and fauna.
5. Conclusions

The aim of the study was to analyze the data obtained from the Sentinel-1, Sentinel-2, Landsat 5, Landsat 7, and Landsat 8 missions in Norilsk, Ambarnaya River, and Pyasino Lake. A long-term analysis covered the data from 1984 to 2021. Short-term analysis focused on images obtained immediately before, during, and after the leakage on 29 May 2020. This shows the scale of changes. The changing climate, heat waves, human interference with the natural environment, and pollution produced by factories cause changes in the ecosystem. For years, trace amounts of toxins have been released into the reservoir, causing environmental degradation. Snow thawing and the growth of vegetation in May 2020 may be related to the record-high temperatures for Norilsk. The analysis in this manuscript suggests that the failure of the reservoir that caused the oil spill near Norilsk could have been caused by climate change that melted the permafrost. In this manuscript, classification of the studied area, detection of surface water changes, visual data analysis, and long-term analysis from 1984 were processed, and the change of vegetation, snow cover, and water level before, during, and after the oil spill was examined. The analysis confirms the effectiveness of the use of remote sensing methods for mapping leaks, as presented in the introduction. Moreover, a practical interpretation of data usage is presented with the additional advantage of re-running the experiment in line with our assumptions. Oil spill detection studies are performed worldwide. Thanks to our experience, the use of several analysis systems, their joint analysis, and the formulation of appropriate conclusions can be planned. The methodological innovation presented in the paper uses the most modern tools for remote sensing processing, i.e., Snap Desktop, Geomatica Banff, and Google Earth Engine.

The leakage of crude oil into the environment is extremely dangerous, especially for the delicate natural areas of Siberia. Accidents happen every day and we cannot predict them, but we are able to reduce the risk by changing the energy sources we use. Climate scientists around the world urge a transition to a new type of economy that would use green technologies to generate energy [56–58]. The information provided in this paper is alarming and should not be taken lightly.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12084016/s1.

Author Contributions: Conceptualization, P.T., T.S. and W.T.; methodology, P.T., T.S. and W.T.; software, T.S. and W.T.; validation, P.T.; formal analysis, P.T., T.S. and W.T.; investigation, P.T., T.S. and W.T.; resources, T.S. and W.T.; data curation, T.S. and W.T.; writing—original draft preparation, P.T.; writing—review and editing, P.T.; visualization, P.T., T.S. and W.T.; supervision, P.T.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: To improve the readability of the article, we provide Supplementary Materials that provide information about data sets and analytical methods.

Acknowledgments: Support of these studies from Gdansk University of Technology by the DEC-26/2021/IDUB/L3.3 grant under the Argentum Triggering Research Grants—‘Excellence Initiative—Research University’ program is gratefully acknowledged.

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

ETM+ the Enhanced Thematic Mapper Plus
GNSS Global Navigation Satellite System
JRC Joint Research Centre
MSS Multispectral Scanner System
NDSS Normalized Difference Snow and Ice Index
NDVI Normalized Difference Vegetation Index
NDVI2 Second Normalized Difference Water Index
OLI Operational Land Imager
Rosprirondnadzor The Federal Russian Natural Resource Surveillance
SAR Synthetic Aperture Radar
SNAP Sentinel Application Platform
SWM Sentinel Water Mask
TIRS Thermal InfraRed Sensor
TM Thematic Mapping

References


52. Liu, Y.; Macfadyen, A.; Zhen-Gang, J.; Weisberg, R. Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 195. [CrossRef]