VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Viktor Nareiko

# FIRE RISK ASSESSMENT AND FORECASTING USING REMOTE SURVEY DATA AND GEOGRAPHICAL INFORMATION SYSTEMS

Doctoral dissertation
Physical sciences, environmental engineering (04T)

Vilnius, 2020

The dissertation work was carried out at Vilnius Gediminas Technical University from 2017 to 2020.

Scientific supervisor:

prof. dr. Jūratė Sužiedelytė VISOCKIENĖ (Vilnius Gediminas Technical University, environmental engineering – 04T).

# Acknowledgements

Acknowledgements go here.

Viktor Nareiko
Vilnius
27th August 2020

# Table of Contents

# Notation

...

# Chapter 1

# Introduction

## 1.1 Research problem

For the last decade in the systems related to geographic information (for the sake of simplicity let call such system as Geographic Information System) start increasingly using publicly and openly available data that can be used free of charge for a variety of purposes and tasks. Satellite data from the Sentinel program run by the European Space Agency (ESA) can be used to monitor environmental pollution parameters, soil, vegetation, forest and water resources, can be used for natural consequences management and risk identification.

The Copernicus program, developed by ESA, provides high-quality and freely available Sentinel satellite data that can be adapted to meet a variety of Geographic Information System (GIS) challenges. One example of such a task is the identification of fire damage to a forest or the classification of arid areas to prevent a potential fire.

Due to global warming, increased and prolonged heat waves, there is an increasing risk of uncontrolled fires. Uncontrolled fires are the cause of air, water pollution, and the destruction of flora and fauna.

Using multi-spectral, hyper-spectral and radio spectral cameras, chemical, thermal and biological contamination can be captured rapidly. Numerous software programs have been developed for the processing of satellite images.

However, many software programs typically consist of separate units used to solve specific tasks or work with specific data formats and do not provide a single information system that can be used to quickly and effectively assess chemical, biological or other sources of pollution and natural disaster risk. The biggest problem when working with satellite imagery is the large amount of data that takes a long time to process due to the different data formats, unclear data processing methodology, consistent data processing, and limited job automation.

A thematic data processing methodology for the analysis of multi-spectral spatial data images with respect to their fast and efficient processing of satellite images is proposed. The methodology is designed to identify environmental parameters to assess the risk of fire or other natural disasters, based on various results of electromagnetic wave reflection analysis in different ranges and selection of optimal electromagnetic wave bands. ESA's Sentinel-2 satellite imagery, which is increasingly used in European countries, is used as the main data source for the analysis.

## 1.2 Relevance of the work

In recent years, there has been an increasing need to effectively assess the risks (large-scale fires, oil spills, etc.) that can be caused by both natural disasters and human activities, and to take all preventive measures where possible. Large uncontrolled fires or oil spills affect climate change, soil acidification, water pollution. Increasingly, remote sensing systems are in place that can cover large areas and perform automated monitoring of the area in question.

Such an analysis requires access to constantly updated and high-resolution remote sensing data. Such data are provided by ESO (European Space Agency) under the Copernicus program. Data from remote sensing data from Sentinel-1 and Sentinel-2 satellites are widely used to address risk management challenges such as fire damage detection or prevention. The images provided by Sentinel-1 and Sentinel-2 satellites are and will be widely used for various environmental tasks, mapping of natural resources, modeling of natural processes in a large area due to their accuracy, free provision and frequent updating. The data provided by the Sentinel-1 and Sentinel-2 satellites open up new opportunities for research.

Many applications of remote sensing to the monitoring of environmental parameters require frequent and dense coverage of spatial data in the study area. Much of the satellite imagery is free and publicly available online. Sentinel satellite imagery supplemented with data from other satellite systems can form the basis for an information system for monitoring a large area, the analysis of which can be used by public authorities, for more effective risk prevention and control.

The leading countries in the field of remote sensing in the world are the USA, Russia, France, Germany, China. For example, U.S. Forest Service has done most of its research into forest fire risk identification and management. The U.S. Forest Service also provides guidelines and various indices for assessing forest status, drought levels, and so on. The paper proposes a methodology for forest fire risk assessment by processing Sentinel satellite images with data from other remote sensing systems, using multi-spectral image classification algorithms as a basis in conjunction with learning computer neural networks. To develop a methodology for the prevention and prediction of wild fire risk. The expected result of the work is the development of a methodology for assessing the condition of forests, assessment of fire risk and prediction it spread, and the provision of recommendations on how to improve environmental monitoring.

## 1.3 Object of research

The object of the research is an algorithm that combines satellite data of different formats to model the risk of uncontrolled forest fires and revise fire indices that are used to assess fire risk in [insert area of concern].

## 1.4 Study objective

To develop a reasonable methodology of classification algorithms combining satellite image data in different formats, allowing to identify the risk of forest fires and present it in a constantly updated thematic map as a graphical representation of fire indices. A map of forest fire risk is the graphical representation of a collection of levels of risk obtained by an index of risk. An index of risk is made up by different variables combined in such a way to obtain particular values; which are referred to be as levels of risk. [1]

An index of fire risk have different nature and methods, so index performances comparison is required in order to establish which index is best suitable to be adopted at European level.

Consequently, the European Commission has incessantly introduced proper regulations and schemes of prevention in which a production of maps of forest fire risk is required. [1]

Thus, this work presents results obtained by a retrospective analysis of long term time series of remote sensing and meteorological data. [1]

## 1.5 Work tasks

In order to achieve the goal of the work, the following tasks need to be solved at work:

- Selection of remote sensing (passive and active) methods and criteria for identification, investigation, monitoring and modeling of fire risk areas.
- Carry out experimental research using optical satellite images, identify the necessary indices for the identification of fire sites.
- Carry out experimental research using radar satellite images, identify the necessary indices for the identification of fire sites.
- To select criteria suitable for monitoring the fire risk in the Baltic States.

- To develop an algorithm combining satellite data of different formats for the identification and modeling of fire sites.
- Create a thematic real-time renewable fire risk map.

## 1.6 Research methodology

The dissertation uses theoretical and experimental research using satellite imagery, geographic information technologies, digital modeling and statistical modeling methods.

## 1.7 Scientific novelty of the work and its significance

During the preparation of the dissertation, the following new results important for environmental engineering science and for environmental protection:

- A methodology has been developed that allows the analysis of large areas using not only passive but also active remote sensing systems.
- The revised and selected indices are intended to determine the forest fire risk or a new methodology for calculating the indices has been developed based on experimental studies.
- An algorithm has been developed for processing satellite images of different formats by modeling fire risk.

## 1.8 Practical significance of the work results

The results of the research can be used to develop, prepare or use satellite imagery to assess and prevent the risk of uncontrolled forest fires or to assess the consequences of forest fires, to select data processing software and to adapt the data to interactive digital thematic maps.

## 1.9 Defensive statements

Assessing the risk of large-scale uncontrolled forest fires can help avoid dire consequences for the environment. Identification of fire risks will avoid the costs associated with fire suppression and its consequences.

# Chapter 2

# Application of remote sensing systems for assessment and prediction of fire risks

The remote sensing methods are reviewed in this chapter that are used to obtain indirect quantitative and qualitative properties about objects. Passive and active research methods are reviewed. The problems related to the large-scale processing of data for the operative retrieval of results about the object under study are reviewed. Methods for obtaining data from optical and radiometric satellite images are considered.

## 2.1 Application of remote sensing data to the environment

Spatial data is one of the fastest growing scientific data, which poses significant computational difficulties for scientists who need to deal with the processing and analysis of large spatial data sets. With the continuous development of sensor technology, the volume of remote sensing images has increased rapidly in recent years and is expected to continue to do so (Yan Ma et al., 2015).

Passive remote sensing data is often used for a variety of environmental challenges:

- Wildfire prevention.
- Wildfire damage assessment.
- Etc.

For the prevention and detection of forest fire damage, a relative normalized deployment index can be applied (N. V. Rodinov. Et al., 2016), distinguishing cloud cover and hydrology from the total data array.

The so-called change detection methodology (CD) can be used to determine forest fire damage. Using the CD methodology, images of different time sections of the same area are usually compared (N. V. Rodinov. Et al., 2016). In order to reduce the images present in the triggers (cloudiness, hydrology), filters are applied to distinguish such triggers using pixel algebra. In this case, the normalized deployment index is calculated for images of different time sections of the same area by making a relative estimate - a relative normalized deployment index. Based on the obtained research results, a data model distinguishing the places where the forest may be affected or most affected is presented.

Currently, more than 100 spectral index variants used for different purposes have been described in different studies using passive remote sensing methods, but only a few of them find wide application in environmental research, monitoring (Vladimir Aleksandrovich Hamedov et al.). In many cases, the described indices do not provide an integrated approach to their complex use in observing a physical phenomenon, which is related to their applications in localized areas using empirical formulas in different ecosystems (Vladimir Aleksandrovis Hamedov et al.). It is for this reason that existing indices used to assess, for example, forest fire risk need to be adjusted for a specific area.

In addition to passive remote sensing, active satellite surveys are increasingly being used, which, for example, make it possible to determine quite accurately:

- Water surface contamination by oil products.

- Water surface contamination with plastic (plastic).
- Fire damage assessment.
- Improving fire risk indices.
- Etc.

There are a lot of example of successful applying of remote sensing for monitoring and assessment various risk for different environmental purposes:

**Example 2.1.1** (Oil detection). *Petroleum products form a film on the surface of the water that is impermeable to sunlight. Without sunlight, the oxidation of bacteria and the multiplication and life of small organisms cease. The death of animal organisms damages the food web. The sea is often polluted by oil products during tanker accidents, but mainly due to sewage from tankers and refineries. Oil products pollute beaches, killing birds and fish.*

*Large spills of oil products at sea can have significant biological and economic impacts. Public and media surveillance is usually intensive after a spill. Remote sensing is playing an increasingly important role in oil monitoring. Active satellite sensors SAR are commonly used to monitor marine oil spills. It is a more effective tool that can penetrate through clouds, rain and snow. The sensor emits microwave radiation that is reflected from the object and the received signal can be determined by the reversible scattering function (Merv Fing et al., 2014).*

*Active satellite remote sensing has been successfully applied to simulate the 2010 oil spill in Dalian, Japan. In this study, a support vector machine was adapted to monitor oil spills based on high radiometric resolution SAR images (Jianchao Fana et al., 2015).*

*Under different weather conditions in different places, the reflection and scattering properties of radio waves from the oil differ. On this basis, a vector machine can be constructed that can automatically classify oil spill zones (Jianchao Fana et al., 2015). In recent years, water has been increasingly contaminated with plastic products, which need to be monitored and identified to determine the extent of the contamination.*

**Example 2.1.2** (Plastic detection). *Plastic pollution in the ocean has been identified as a threat to various coastal areas. Many marine organisms can swallow or become* *entangled in plastic and this can pose a fatal risk to them. Although high concentrations of floating plastic debris are observed as they travel from inland waters to the open ocean, a detailed analysis of the spatial scale and abundance of waste is lacking and monitoring tools are not well developed to determine the distribution of pollution. Remote-sensing images with medium and high temporal, spectral, and spatial resolutions would be a great additional tool to quantify the distribution of floating marine plastic debris (Shungudzemwoyo P. Garabaa et al., 2017).*

*Identifying plastics at sea is difficult because there are many different plastics in the marine environment. The size of the plastic can range from microplastics (less than 5 mm) to large plastic parts such as "ghost nets" (lost or discarded fishing nets). In the first case, the plastic material can be toxic through the absorption of contaminants into the plastic, and in the second case, the plastic contaminants can injure animals and endanger seafarers. Plastics can be made from granules used in manufacturing, from certain cosmetic and personal care products, from textile fibers (Lonneke Goddijn-Murphya et al., 2017).*

*The rays of the sun falling on the surface of the water are partly reflected and the other partially penetrates through the surface. In water, light photons are absorbed and scattered in all directions. Due to the scattering and even distribution of repetitive light rays in all directions, it is possible to identify objects on the water surface. If the water is optically deep (the bottom is invisible), the fraction of light that is scattered and penetrates through the water into the air medium provides information on optically active (e.g., plankton) constituents of the water. Optically active components determine the apparent color of water, and their concentration can be calculated from spectral reflection measurements. In this case, the concentration of plastics can be identified in places where optically active water components cannot be identified (Lonneke Goddijn-Murphya et al., 2017).*

*Satellites that have sensors that can detect the color of the ocean also provide instruments to identify the plastic, such as Sentinel-3 working in conjunction with Sentinel-3 SLSTR comparing nine bands VIS-SWIR 0.55–12 $\mu m$ (Lonneke Goddijn-Murphya et al., 2017).*

**Example 2.1.3** (Nitrogen of plants detection ). *Optical satellite imagery is by far the most suitable for determining the nitrogen content of plants to reduce the use of fertilizers in crop production.*

**Example 2.1.4** (Fire risk assessment). *The estimation of forest fire risk involves the integration of meteorological and other fuel-related variables leading to an index that assesses the different levels of risk. Two indices that are frequently used to estimate the level of fire risk are the Fire Weather Index (FWI) and the Normalized Difference Vegetation Index (NDVI) (ASSESSMENT OF FOREST FIRE RISK IN EUROPEAN MEDITERRANEAN REGION: Comparison of satellite-derived and meteorological indices).*

## 2.2 Open remote sensing data source

ESO (European Space Agency) is developing a Sentinel satellite system under the Copernicus program and there is an increasing need for operative analysis in the territory in question to solve various environmental, landscaping and risk management tasks. Such an analysis requires access to constantly updated and high-resolution remote sensing data. Such data are provided by ESO (European Space Agency) under the Copernicus program. The paper examines the application of open Sentinel-1 and Sentinel-2 satellite remote sensing data to address a risk management task, such as fire damage detection.

Sentinel remote sensing data are being developed by ESA under the Copernicus program. This data is free and freely available. In recent years, the need for access to open, free, and frequently updated time-lapse remote sensing data has been growing. Sentinel is divided into 6 groups of satellites, each of which has 2 satellites that can be used to solve various environmental, landscaping and risk management tasks.

One of the most common tasks is risk management. Sentinel data, due to their accuracy and frequent updating in the time section, can be used, for example, to determine fire damage.

ESA has launched two satellites from each Sentinel family into orbit under the ongoing Copernicus program:

- Sentinel-1 - satellites using active remote sensing sensors. Sentinel-1A launched into orbit in 2014 and Sentinel-1B in 2016.

- Sentinel-2 - satellites using multi-spectral high-resolution passive remote sensing sensors. Sentinel-2A was launched into orbit in 2015, and Sentinel-2B in 2017.
- Sentinel-3 - multi-functional satellites are used for seabed topography, sea and land surface temperature determination. Used to monitor the environment and climate change. Sentinel-3A was released in 2016 and Sentinel-3B was released in 2018.
- Sentinel-4 - geostationary orbit satellites for atmospheric observation.
- Sentinel-5 - polar orbit satellites for atmospheric monitoring.
- Sentinel-6 is a satellite using an altimeter radar that measures the world's sea level. Used for oceanography and climate research.

Sentinel satellite data is free and publicly available. To receive them, it is necessary to register on the Copernicus Hub website and download them according to the selected search criteria (time, place, etc.). Figure 2.6 shows a graphical user interface that can be used to download data.

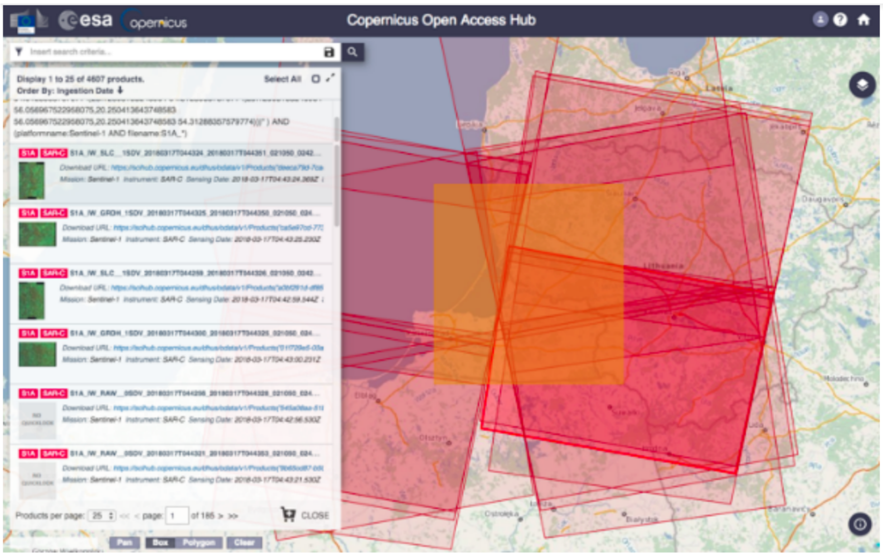


Figure 2.1: Sentinel data download via the Copernicus Hub website

Also the Copernicus Hub programming interface can be used to automate data downloads by specified parameters.

Sentinel data is provided in SAFE (Standard Archive Format for Europe) format. This format stores not only raster information but also textual information that can be used to correct raster data.

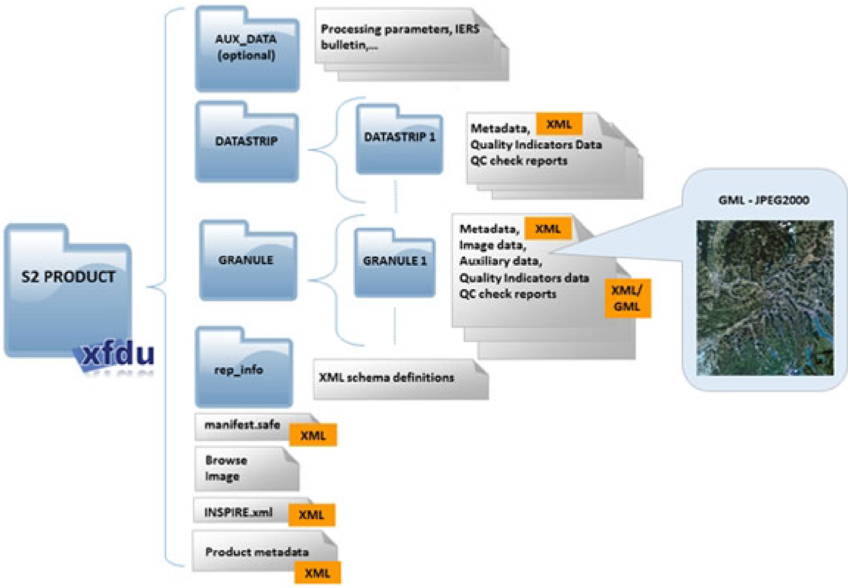


Figure 2.2: Example of Sentinel-2 data format

The Copernicus program, developed by ESA, provides high-quality and freely available Sentinel satellite data that can be adapted to meet a variety of GIS challenges. One example of such tasks is the identification of fire damage to the forest, the detection of oil spills at sea, the identification of plastic sites at sea. ESA provides all the tools needed for data retrieval (Coperinicus Hub data download portal) and GIS analysis (SNAP software). Sentinel data is high quality, frequently updated, and SNAP free software provides all the tools to perform GIS analysis for both scientific and commercial purposes.

## 2.3 Google earth engine as a platform for accessing open remote sensing data

At the same time, petabyte-scale archives of remote sensing data have become freely available from multiple agencies including NASA (Woodcock et al., 2008;

Loveland and Dwyer, 2012; Nemani et al., 2011), as well as the ESA (Copernicus Data Access Policy, 2016). A wide variety of tools have been developed. Despite that we have a lot of resources and tools - it still requires considerable technical expertise and effort. Google Earth Engine is a cloud-based platform that makes it easy to access high-performance computing resources for processing very large geospatial datasets, without having to suffer the IT pains currently surrounding either (N. Gorelick, 2017).

*Google Earth Engine* is a cloud-based platform that allows users to have an easy access to a petabyte-scale archive of remote sensing data and run geospatial analysis on Google's infrastructure. Currently, Google offers support only for *Python*, *JavaScript* or *R*.

Image preprocessing (i.e. download, reprojection, mosaicking, resize, bad pixels control and composite) has always been a time-consuming activity. Although *R* offers incredible open-source "API packages" to easily get geospatial resources (modistsp, getSpatialData, elevation, landsat and so on), these still require that users count with competent computers (and high-end for users that want to analysis large areas). *Google Earth Engine* democratizes access to high performance computing, enabling image preprocessing and analysis in the worst circumstances (e.g. in the middle of the Amazon with a Pentium IV laptop and weak internet access conditions).

One of the way to use Google earth engine is to use python or javascript, but it is also possible to use R programming language with *rgee* package. Such combination let to use python or javascript for quick prototyping and R programming language for statistical analysis and classification.

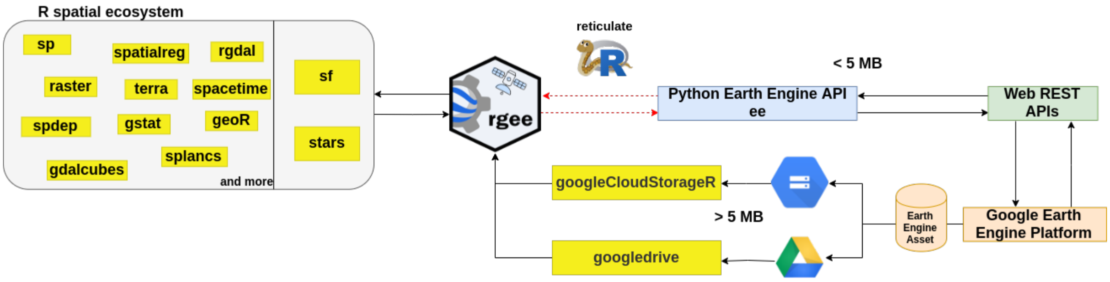


Figure 2.3: Architecture of accessing mutiple sources of remote data by using *Google Earth Engine*

Below is provided an example of code in R programming language for getting sentinel data of interest area:

```
ee_roi <- st_read(system.file("shape/nc.shp",
    package="sf")) %>%
st_geometry() %>%
sf_as_ee()

ee_search_dataset() %>%
ee_search_tagstitle(
   "sentinel", "sr",logical_operator = "AND"
) %>%
'['(4,) %>%
ee_search_display()
```

The Earth Engine public data catalog is a multi-petabyte curated collection of widely used geospatial datasets. The bulk of the catalog is made up of Earth-observing remote sensing imagery, including the entire Landsat archive as well as complete archives of data from Sentinel-1 and Sentinel-2, but it also includes climate forecasts, land cover data and many other environmental, geophysical and socio-economic datasets (table 2.3). The catalog is continuously updated at a rate of nearly 6000 scenes per day from active missions, with a typical latency of about 24 h from scene acquisition time. Users can request the addition of new datasets to the public catalog, or they can upload their own private data via a REST interface using either browser-based or command-line tools and share with other users or groups as desired (N. Gorelick 2017).

Table 2.1: Frequently used datasets in the earth engine data catalog. (Gorelick 2017)

| **Dataset** | **Nominal resolution** | **Temporal granularity** | **Temporal coverage** | **Spatial coverage** |
|---|---|---|---|---|
| Landsat 8 OLI/TIRS | 30m | 16 day | 2013–Now | Global |
| Landsat 7 ETM+ | 30m | 16 day | 2000–Now | Global |
| Landsat 5 TM | 30m | 16 day | 1984–2012 | Global |
| Landsat 4–8 surface reflectance | 30m | 16 day | 1984–Now | Global |
| Sentinel 1 A/B ground range detected | 10 m | 6 day | 2014–Now | Global |
| Sentinel 2A MSI | 10/20 m | 10 day | 2015–Now | Global |
| MOD08 atmosphere | 1° | Daily | 2000–Now | Global |
| MOD09 surface reflectance | 500 m | 1 day/8 day | 2000–Now | Global |
| MOD10 snow cover | 500 m | 1 day | 2000–Now | Global |
| MOD11 temperature and emissivity | 1000 m | 1 day/8 day | 2000–Now | Global |
| MCD12 Land cover | 500 m | Annual | 2000–Now | Global |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| **Dataset** | **Nominal resolution** | **Temporal granularity** | **Temporal coverage** | **Spatial coverage** |
|---|---|---|---|---|
| MOD13 Vegetation indices | 500/250 m | 16 day | 2000–Now | Global |
| MOD14 Thermal anomalies & fire | 1000 m | 8 day | 2000–Now | Global |
| MCD15 Leaf area index/FPAR | 500 m | 4 day | 2000–Now | Global |
| MOD17 Gross primary productivity | 500 m | 8 day | 2000–Now | Global |
| MCD43 BRDF-adjusted reflectance | 1000/500 m | 8 day/16 day | 2000–Now | Global |
| MOD44 veg. cover conversion | 250 m | Annual | 2000–Now | Global |
| MCD45 thermal anomalies and fire | 500 m | 30 day | 2000–Now | Global |
| L1 T radiance | 15/30/90 m | 1 day | 2000–Now | Global |
| Global emissivity | 100 m | Once | 2000–2010 | Global |
| PROBA-V top of canopy reflectance | 100/300 m | 2 day | 2013–Now | Global |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| **Dataset** | **Nominal resolution** | **Temporal granularity** | **Temporal coverage** | **Spatial coverage** |
|---|---|---|---|---|
| EO-1 hyperion hyperspectral radiance | 30 m | Targeted | 2001–Now | Global |
| DMSP-OLS nighttime lights | 1 km | Annual | 1992–2013 | Global |
| USDA NAIP aerial imagery | 1 m | Sub-annual | 1992-2013 | CONUS |
| Shuttle Radar Topography Mission | 30 m | Single | 2000 | 60°N–54°S |
| USGS National Elevation Dataset | 10 m | Single | Multiple | United States |
| USGS GMTED2010 | 7.5″ | Single | Multiple | 83°N–57°S |
| GTOPO30 | 30″ | Single | Multiple | Global |
| ETOPO1 | 1′ | Single | Multiple | Global |
| GlobCover | 300 m | Non-periodic | 2009 | 90°N–65°S |
| USGS National Landcover Database | 30 m | Non-periodic | 1992–2011 | CONUS |
| UMD global forest change | 30 m | Annual | 2000–2014 | 80°N–57°S |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| **Dataset** | **Nominal resolution** | **Temporal granularity** | **Temporal coverage** | **Spatial coverage** |
|---|---|---|---|---|
| JRC global surface water | 30 m | Monthly | 1984–2015 | 78°N–60°S |
| GLCF tree cover | 30 m | 5 year | 2000–2010 | Global |
| USDA NASS cropland data layer | 30 m | Annual | 1997–2015 | CONUS |
| Global precipitation measurement | 6′ | 3h | 2014–Now | Global |
| TRMM 3B42 precipitation | 15′ | 3h | 1998–2015 | 50°N–50°S |
| CHIRPS precipitation | 3′ | 5 day | 1981–Now | 50°N–50°S |
| NLDAS-2 | 7.5′ | 1h | 1979–Now | North America |
| GLDAS-2 | 15′ | 3h | 1948–2010 | Global |
| NCEP reanalysis | 2.5° | 6h | 1948–Now | Global |
| ORNL DAYMET weather | 1 km | Annual | 1980–Now | North America |
| GRIDMET | 4 km | 1 day | 1979–Now | CONUS |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| **Dataset** | **Nominal resolution** | **Temporal granularity** | **Temporal coverage** | **Spatial coverage** |
|---|---|---|---|---|
| NCEP global forecast system | 15′ | 6h | 2015–Now | Global |
| NCEP climate forecast system | 12′ | 6h | 1979–Now | Global |
| WorldClim | 30″ | 12 images | 1960–1990 | Global |
| NEX downscaled climate projections | 1 km | 1 day | 1950–2099 | North America |
| WorldPop | 100 m | 5 year | Multiple | 2010–2015 |
| GPWv4 | 30″ | 5 year | 2000–2020 | 85°N–60°S |

## 2.4 Sentinel passive and active remote sensing systems

Remote sensing is the process of obtaining qualitative and quantitative data about a particular object without physical contact. Remote sensing systems can be divided into passive and active systems using electromagnetic radiation:

- Passive remote sensing systems use an external source of electromagnetic radiation, usually the sun.

- Active remote sensing systems have their own light source, which is an electromagnetic wave generator that emits waves from a satellite to the earth's surface. The reflected waves from the earth's surface are captured by a satellite.

ESA has launched two satellites from each Sentinel family into orbit under the ongoing Copernicus program:

- Sentinel-1 - satellites using active remote sensing sensors. Sentinel-1A launched into orbit in 2014 and Sentinel-1B in 2016.
- Sentinel-2 - satellites using multi-spectral high-resolution passive remote sensing sensors. Seintinel-2A was launched into orbit in 2015, and Sentinel-2B in 2017.

Sentinel data is presented in SAFE (Standard Archive Format for Europe) format, which stores not only raster information but also text that can be used to correct raster data, such as cloud cover, so ESA recommends the use of SNAP (Figure 4) open source software for the processing and analysis of their data (`http://www.eo4sd-eastern.eu/sites/default/files/publications/snap_workbook_english.pdf`).

Since 2014, radar data (SAR) from the European satellite Sentinel-1A have been freely available, opening up new opportunities for research. The Copernicus portal provides many examples of where they can be applied, ranging from the detection of forest fire damage to the detection of oil spills in water bodies. The main parameters, for example, of Sentinel-1A radar photos (E.A. Baldina et al., 2016):

Table 2.2: Sentinel-1A satellite data parameters (E.A. Baldina et al., 2016)

| **Mode** | **Coverage area, km** | **Spatial resolution, m** | **Polarization (H - horizontal; V - vertical)** |
|---|---|---|---|
| Stripmap (SM) | 80 | 5x5 | HH, VV, HH+VV, VV+VH |
| Interferometric Wide Swath (IW) | 250 | 5x20 | HHH, VV, HH+HV, VV+VH |
| Extra-Wide Swath (EW) | 400 | 20x40 | HH, VV, HH+HV, VV+VH |
| Wave (WV) | 20x20 | 5x5 | HH, VV |

The resolution of the data received by the Sentinel-2A satellite depends on the color wave captured by the sensor and ranges from 10 m to 60 m (table 2.3)).

Table 2.3: Sentinel-2A satellite data resolution (Du, Y et al., 2016)

| **Physical band** | **Pixel resolution, m** | **Wave length, mm** |
|---|---|---|
| B1 | 60 | 443 |
| B2 | 10 | 490 |
| B3 | 10 | 560 |
| B4 | 10 | 665 |
| B5 | 20 | 705 |
| B6 | 20 | 740 |
| B7 | 20 | 783 |
| B8 | 10 | 842 |
| B8A | 20 | 865 |
| B9 | 60 | 945 |
| B10 | 60 | 1375 |
| B11 | 20 | 1610 |
| B12 | 20 | 2190 |

## 2.5 Fire, wildfire and fire risk descriptions

Before start moving forward and doing any research we should describe what fire is and what main factors condition it.

Forest fire is a major natural disaster for the European territory. Also known as wildfires, vegetation fires or grass fires, it represents an uncontrolled fire in wild land. It is often caused by human careless or arson and, thus, it is one of the natural disasters more disposable to be prevented by a good preventive scheme.

Forest can be described as an ecosystem characterized by a more or less dense and extensive tree cover, often consisting of stands varying in characteristics such as species composition, structure, age class, and associated processes, and commonly including meadows, streams, fish, and wildlife – note forests include special kinds such as industrial forests, non industrial private forests, plantations, public forests, protection forests, and urban forests, as well as parks and wilderness (Helms 1998).

Forest also can be described as Forest: land with tree crown cover (or equivalent stocking level) of more than 10 % and area of more than 0.5 ha. The trees should be able to reach a minimum height of 5 m. It may consist either of closed forest formations where trees of various stores and undergrowth cover a high proportion of the ground, or of open forest formations with a continuous vegetation cover in which tree crown cover exceeds 10 %. Young natural stands and all plantations established for forestry purposes which have yet to reach a crown density of 10 % or tree height of 5m are included under forest, as are areas normally forming part of the forest area which are temporarily unstocked as a result of human intervention or natural causes but which are expected to revert to forest. The definition of 'forest' includes: forest nurseries and seed orchards that constitute an integral part of the forest; forest roads, cleared tracts, firebreaks and other small open areas within the forest; forest in national parks, nature reserves and other protected areas such as those of special environmental, scientific, historical, cultural or spiritual interest; windbreaks and shelter belts of trees with an area of more than 0.5 ha and a width of more than 20 m. Rubber wood plantations and cork oak stands are included. However, the definition of 'forest' excludes: land predominantly used for agricultural practices (European Commission 2003).

Fire: rapid burning of combustible material with the evolution of heat and usually accompanied by flame (Encyclopedia Britannica 2005). We can boldly say that forest together with air is a main fuel for a wild fire.

Forest fire is a fire which breaks out and spreads on forest and other wooded land or which breaks out on other land and spreads to forest and other wooded land. The definition of 'forest fire' excludes: prescribed or controlled burning, usually with the aim of reducing or eliminating the quantity of accumulated fuel on the ground (European Commission 2003).

A forest fire could occur and evolve assuming different characteristics. Consequently, different types of forest fire have been classified (fig. 2.4). The most familiar fire type is the surface fire. It represents the most common propagation regime and consists in rapidly burning fire that sweep quickly over an area, consuming litter and the above ground portions of herbs, shrubs, grasses and lower branches of trees. If conditions are favorable a surface fire may extend to the upper layers of the crown foliage. A fire affecting mainly the crowns of the woody vegetation is called crown fire. Frequently, it leaves most of the steam and the forest floor relatively untouched and is difficult to control since strictly dependent to wind conditions. Moreover, a fire could evolve below the terrain. Referred to as ground fire, it consists principally in largely flame less fire that burn slowly through thick surface accumulation of organic matter, duff and roots and it is very difficult to detect and control. In some particular conditions a ground fire can became a flaming surface fire if not adequately treated (Viegas 2002).

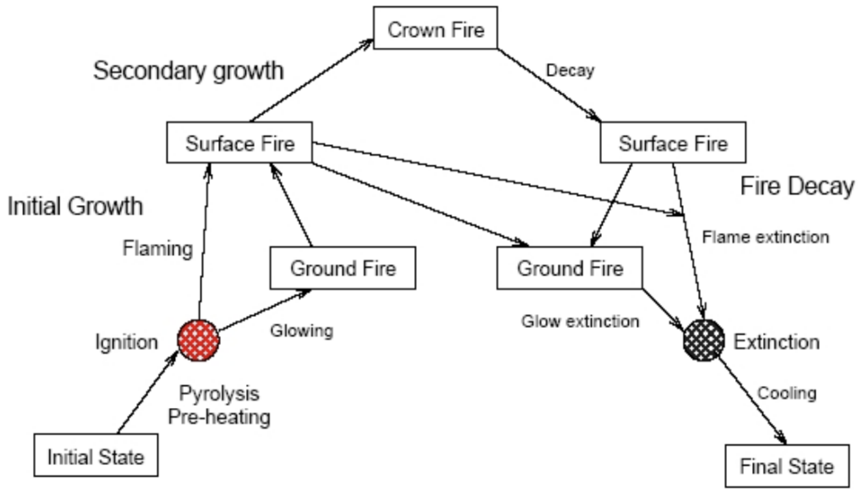


Figure 2.4: Fire growth, spread and decay from (Viegas 2002)

## 2.6 Forest Fire Risk Assessment

The evaluation of the effects of a fire a priors is not always possible because the consequences of a forest fire depend by several aspects. Climate conditions, terrain topography, intensity and permanence of the fire are the prevailing elements. Wind condition influences the fire behavior and is really hard to predict. It depends by topography, vegetation and local heating and cooling. Besides, topography may cause dramatic changes in fire behavior as a fire progress over the terrain. In addition, the fire itself may influence the environment and thus the fire behavior; heating from the fire can modify or produce local winds contributing to atmospheric instability and causing cloud development. The effects of fire on soil vary with the proprieties of fuel, fire and soil itself. The consequences are physical, biochemical and biological as well as economic. Fire is able to influence soil temperature, soil structure and the ability of the soil to absorb and store water (Pyne et al. 1996). Forest fires produce gaseous and particle emissions that impact the composition and functioning of the global atmosphere. They are a source of carbon emitted into the atmosphere which influences climate change but are also an irreplaceable sink of carbon. For this reason, the Kyoto protocol, on article 2.ii, suggests the improvement of sustainable forest management practices, afforestation and reforestation.

Fire danger: "the resultant, often expressed as an index, of both constant and variable factors affecting the inception, spread, and difficulty of control of fires and the damage they cause" Fire hazard: "a measure of that part of the fire danger contributed by fuels available for burning".

Fire risk: "(1) the chance of fire starting as determined by the presence and activity of causative agents, (2) a causative agent (3) a number related to the potential of firebrands to which a given area will be exposed during the rated day" (FAO 1986).

There are variables changing value almost continuously during the day and variables having a variation noticeable only over a long period of time; week, month or even years. Respectively, these variables were classified as short-term variables and long-term variables. Evapotranspiration, relative humidity, wind and air temperature furnish an example of variables clearly unsteady during the day. Fuel type, fire history, amount of population, topography of the territory, soil type and proximity of roads are variables with a roughly stable behavior in a short period.

There are variables changing value almost continuously during the day and variables having a variation noticeable only over a long period of time; week, month or even years. Respectively, these variables were classified as short-term variables and long-term variables.

Evapotranspiration, relative humidity, wind and air temperature furnish an example of variables clearly unsteady during the day. Fuel type, fire history, amount of population, topography of the territory, soil type and proximity of roads are variables with a roughly stable behavior in a short period:

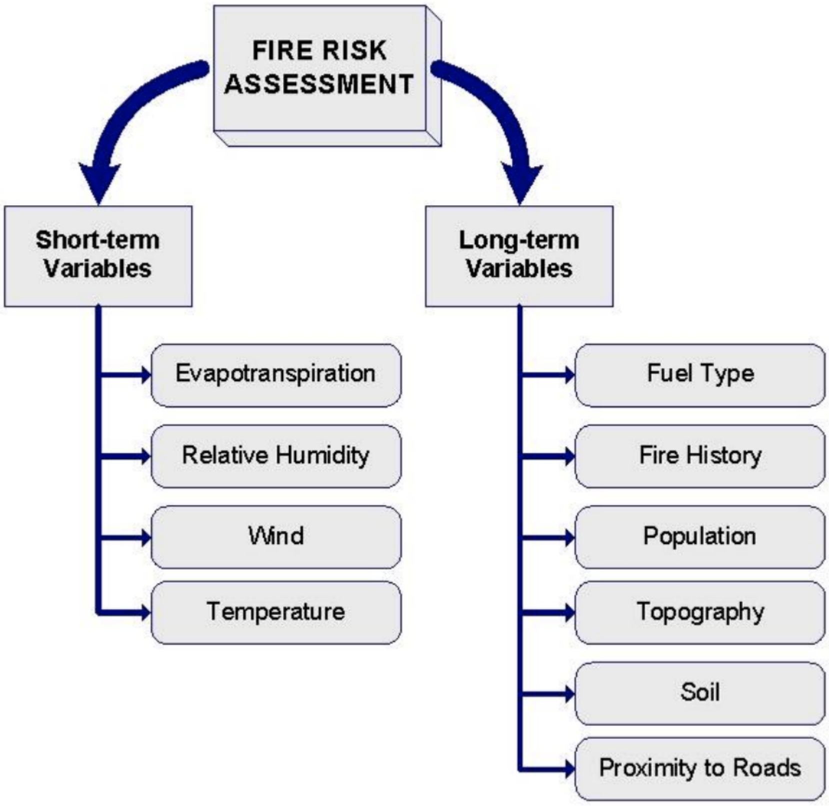


Figure 2.5: Fire assessments variables

In most cases variables on which fire risk depends exist. We can do variables classification into long and short term. Some of them changes continuously and some of them have a variation noticeable only over a long period of time; week, month or even years. We can extract such types of variables:

- Meteorological related - fire occurrences and propagation are strongly related to particular meteorological conditions: solar radiation, air temperature, relative humidity, precipitation, wind.
- Vegetation related - water retention in plants and in soil is basic to predict moisture content of vegetation; which plays an important role in fire ignition and propagation.
- Human behavior related - most of fires causes are directly linked to human behavior. The presence of settlements, agricultural burning, pyromaniacs, barbecues and cigarettes contribute to increase the risk of accidental fires.

## 2.7 Forest fire risk indices

An index of risk describes a composite indicator that identifies countries at risk of humanitarian crisis and disaster that would overwhelm national response capacity and it permits to better manage and compare information than using values directly. The values of variables identified as indicators of risk are managed by mathematical expressions. Thus, the result of these expressions is considered in order to extract an index which quantifies the risk throughout a numerical scale.

The indices of wildfire risk can be several. For the sake of simplicity it can be classified in such way:

- Long term.
- Short term.

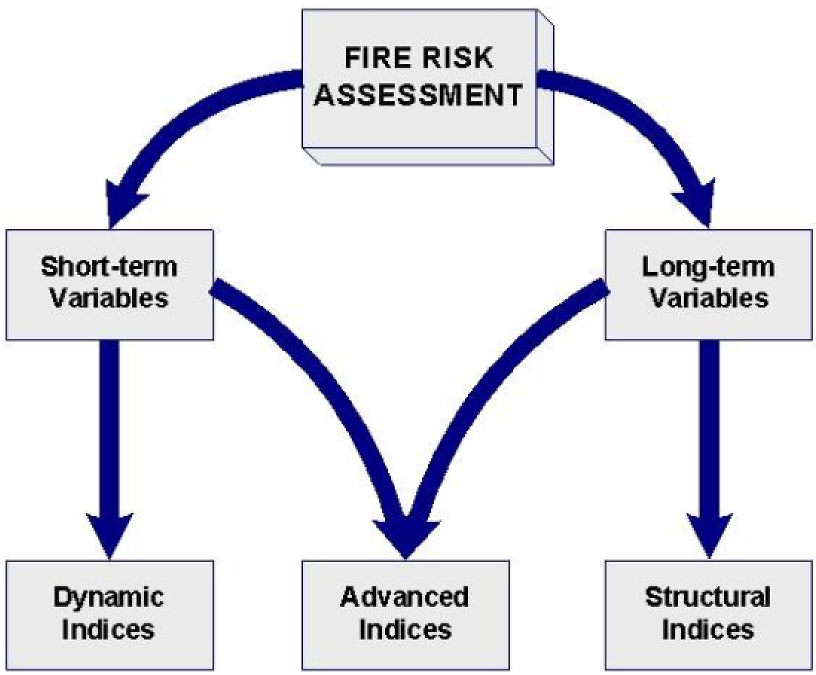


Figure 2.6: Classification of fire risk indices

## 2.8 Meteorological derived fire risk indices

Weather condition have a big impact for risk of wildfire. Depending on solar radiation or wind speed and direction fire risk can increase dramatically.

Meteorological factors such as dew point, soil temperature, air temperature, humidity, precipitation and wind speed have a major impact on the occurrence of forest fires as these climatic factors change with time and space rapidly (Liu et al., 2015).

The most common meteorological derived fire risk indices used in Europe are:

- The Canadian Fire Weather Index (FWI) (Van Wagner 1987).
- The Portuguese index (Goncalves and Lourenco 1990).
- The Spanish ICONA method - probability of ignition (ICONA 1993).
- The Sol Numerical Risk (Drouet and Sol 1993).

In most cases a correlative data analysis can be used to develop weather index that estimates the risk of forest fires in the interest area (Using correlative data analysis to develop weather index that estimates the risk of forest fires).

### 2.8.1 Canadian Fire Weather Index (FWI)

The Canadian Forest Fire Weather Index (FWI) was issued in 1970. It uses four meteorological parameters: noon relative humidity; noon temperature; precipitation during 24 h and the maximum speed of the average wind (Using correlative data analysis to develop weather index that estimates the risk of forest fires).

The FWI System is comprised of six components: three fuel moisture codes (Fine fuel moisture code, Duff Moisture code & Drought code) and three fire behavior indexes (Initial spread index, Buildup index & Fire weather index). The mathematical equations of FWI are given below:

$$B = 0.1Rf(d) \tag{2.1}$$

where $B$ is the B-scale of FWI readjusted by the factor 0.1, $R$ is the rainfall (mm) and $f(d)$ is the fuel availability ($ft^2$). The final S-scale FWI is given bellow:

$$ln(s) = 2.72 \cdot [0.434 \cdot lnB]^{0.647} \tag{2.2}$$

The Canadian model has been tested and adopted in New Zealand, Fiji, Alaska, Mexico, Chile, Argentina and Europe (Using correlative).

### 2.8.2 The Portuguese index

Derived from Nesterov model and based on the assessment of atmospheric conditions in the proximity of the fuel layer:

$$\begin{aligned} x_k &= y_{k-1} - S\nabla f(y_{k-1}) \\ y_k &= x_k + \frac{k-1}{k+2}(x_k - x_{x-1}) \end{aligned} \tag{2.3}$$

For any fixed step size $s \leq 1/L$, where $L$ is the Lipschitz constant of $\nabla f$, this scheme exhibits the convergence rate (A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights).

### 2.8.3 The Spanish ICONA method

ICONA (Nature Conservation National Institute) adopted one of the classification that was proposed by Rothermel in 1972. That classification distinguishes among 13 fuel models depending on the flame-spreading element. Those models are grouped in four categories:

- Pastures.
- Scrub.
- Leaf litter under tree.
- Cutting debris and forestry operations.

### 2.8.4 The Sol Numerical Risk

The Numerical risk index was developed by Sol in order to improve the prediction of fire occurrence and spread in southern France.

The Numerical risk takes air humidity, soil water reserve and wind speed into account. It requires therefore daily air temperature, dew point temperature, cloud cover, wind speed and potential evapotranspiration (Pereira, A.R., and W.O. Pruitt. 2004. Adaptation of the Thornthwaite scheme for estimating daily reference evapotranspiration. Agricultural Water Management 66: 251-257.) as input variables.

The Numerical risk RN 'is calculated as follows (https://wikifire.wsl.ch/tiki-indexe343.html?page=References):

$$RN = 25 - \frac{FHWRFWF}{15} + RSF \tag{2.4}$$

where FH is false relative humidity, WRF the soil water reserve factor, WF the wind factor, and RSF the rate of spread correction factor. FH is calculated as follows:

$$FH = 100\frac{e_s(T_{dew})}{e_s(T_{soil})} \tag{2.5}$$

where $e_s(T_{dew})$ is the saturation vapor pressure at the dew point temperature, and $e_s(T_{soil})$ the saturation vapor pressure at the soil (litter) temperature.

The $e_s(T_{soil})$ is derived as follows (temperature at the soil or litter temperature) (Camia & Bovio 2000):

$$T_{soil} = \begin{cases} 0.874 \cdot T - 0.189 \cdot U + 11.38, & Cc \leq 2 \\ \\ 1.36 \cdot T - 1.422 \cdot Cc - 0.22 \cdot T_{dew} + 13.42, & Cc \geq 3 \end{cases} \tag{2.6}$$

where $T$ is air temperature, $U$ wind speed, $Cc$ cloud cover, and $T_{dew}$ dew point temperature.

The soil water reserve factor $WRF$ is calculated as follows:

$$WRF = 3 + 2 \cdot \tanh(\frac{r - 50}{25}) \tag{2.7}$$

where $\tanh$ is the hyperbolic tangent and r the soil water reserve.

The wind factor WF is calculated as follows:

$$WF = 3 + 3 \cdot \tanh(\frac{45 - U}{50}) \tag{2.8}$$

The rate of spread correction factor RSF is determined as follows:

$$RSF = \begin{cases} -3, & ROS \leq 600 \\ \\ 0, & 600 < ROS < 1000 \\ \\ 2, & ROS \geq 1000 \end{cases} \tag{2.9}$$

where ROS, the rate of spread, is calculated as follows:

$$ROS = 180 \cdot e^{T1714} \cdot \tanh(\frac{100 - r}{150} \cdot \{1 + 2 \cdot (0.8483 + \tanh(\frac{U}{30} - 1.25))\}) \tag{2.10}$$

## 2.9 Vegetation derived fire risk indices

Vegetation indices are derived by remote sensing with the aim to attempt to evaluate the vegetation stress. They are formed from combinations of several spectral values indicating the amount or vigor of the observed vegetation. The simplest form of vegetation index is a ratio between measurements of reflections in separate portions of the spectrum (Comission).

Indices based on the vegetation stress estimate are called vegetation indices.

A list of the most common indices of vegetation includes:

- The Normalized Difference Vegetation Index (NDVI).
- The Soil-Adjusted Vegetation Index (SAVI).
- Normalized Difference Water Index.
- Relative Greenness Index.

### 2.9.1 The Normalized Difference Vegetation Index (NDVI)

The NDVI is the most commonly used index for forest fire risk assessment in which the difference in reflectance is divided by the sum. This compensates for changing illumination conditions, surface slope, aspect, and other extraneous factors and produces a number between -1 and +1. The typical range of actual values is about 0.1 for bare soils to 0.9 for dense vegetation. NDVI is thought to be more sensitive to low levels of vegetative cover (Rouse et al. 1974).

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \tag{2.11}$$

where $\rho_{NIR}$ - near infrared band numerical value and $\rho_{red}$ - red band numerical representation.

### 2.9.2 The Soil-Adjusted Vegetation Index (SAVI)

This index resembles the NDVI with some added terms to adjust for different brightness of background soil.

$$SAVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red} + L} \cdot (1 + L) \tag{2.12}$$

where $\rho_{NIR}$ - near infrared band numerical value and $\rho_{red}$ - red band numerical representation. In principle, the term L can vary from 0 to 1 depending on the amount of visible soil. The constant L is empirically determined to minimize the index sensitivity to soil background reflectance variation. However, 0.5 works as a reasonable approximation for L when the amount of soil in the scene is unknown and for intermediate vegetation cover ranges. The factor (1+L) set the range of SAVI values between -1 and +1, as the range of the NDVI (Huete 1988).

### 2.9.3 Normalized Difference Water Index

This index was proposed for remote sensing of vegetation liquid water from space as a complementary index for NDVI. The formula Is equivalent to NDVI with the visible channel replaced by a short wave infrared reflectance at 1.24 $\mu m$ (Gao 1996).

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR} + L} \cdot (1 + L) \tag{2.13}$$

where $\rho_{NIR}$ - near infrared band numerical value and $\rho_{SWIR}$ - short wave infrared band numerical value.

### 2.9.4 Relative Greenness Index

It is defined as the relative variation of NDVI, with respect to its maximum and minimum of a long period. In this way, the change due to the climatic conditions can be better discriminated, since the absolute value of the NDVI is sometimes more related to the landscape composition instead of seasonal dynamism (Goward et al. 1991).

$$RGI = \frac{NDVI_i - NDVI_{min}}{NDVI_{max} + NDVI_{min}} \cdot 1000 \tag{2.14}$$

### 2.9.5 Vegetation Moisture Stress

For calculation of short term or dynamic components as fire risk indices, the moisture stress could be used. The vegetation's moisture content is another decisive factor for the outbreak and behaviour of a fire. The Vegetation Indexes (VIs) are calculated based on the reflectance values in different wavelengths, trying to ensure that atmospheric and ground influences are minimal. In this paper, the NDII (Normalised Difference Infrared Index) was calculated first, and afterwards, the relative Vegetation Condition Index (VCI) can be calculated:

$$VCI_i = [(NDII_i - NDII_{min})/(NDII_{max} - NDII_{min})] \cdot 100 \tag{2.15}$$

where $VCI_i$ is the Vegetation Condition Index for each date; $NDII_i$ is the index for dryness on the date in question; $NDII_{min}$ is the index of minimum dryness and $NDII_{min}$ is the index of maximum dryness for given oeriod.

## 2.10 Processing of remote sensing data using distribution systems

At present, remote sensing and GIS methods used to process remote sensing data are a powerful tool for studying, predicting and non-invasive monitoring of environmental change. The term "remote sensing" is widely used to describe a method of collecting information about the Earth's surface without contact with it. The electromagnetic energy reflected or emitted during remote sensing is recorded by a sensor in the form of an image. These images are then processed and analyzed to obtain meaningful information about the objects and phenomena depicted. Remote sensing is a multi-stage process involving several components and the interactions between them. Remote sensing requires an energy source. Currently, two types of remote sensing sensors are used - passive and active. In terms of passive remote sensing, solar energy reaches the Earth, interacting with the atmosphere and objects on the Earth's surface. The reflected part of this energy is received by the sensor, encoded by electrical signals and transmitted to the ground station. To become valuable, this data needs to be prepared, corrected, and improved. Further processing involves intensive decryption and analysis, which allows the data to be turned into meaningful information. Increasingly, remote sensing data is processed using so-called "learning systems," where neural network methodologies create a system that can process large amounts of data automatically, without human intervention (P. Scheunders D. et al., 2017; Atharva Sharma et al., 2017).

With active remote sensing, the sensor sends a signal (in the case of a satellite, generated radio waves) that is reflected from the earth's surface, interacting with objects on it, and returns to the sensor's recorder.

Currently, passive remote sensing (using optical satellite imagery) is commonly used in conjunction with GIS to:

- Manage, monitor the consequences of natural disasters.
- Anticipate the possible consequences of future natural disasters.

Many tools and methods have been developed, in particular:

- Land management.
- Management and analysis of cultural and natural heritage areas.
- In geology.
- In hydrology.
- In navigation.
- In geodesy and cartography.

Many methodologies have been developed that are applicable to optical satellite imagery, but there are several problems with the application of passive remote sensing techniques:

- There is no possibility to perform tests at any time of the day.
- Weather conditions (cloudiness) limit the use of optical photographs.
- Studies based on the optical properties of light rays (reflection, absorption, deflection) do not.
- Always show accurate results.

Some of these problems can be solved by using active remote sensing - radiometric satellite imagery:

- Tests can be performed at any time of the day.
- Weather conditions do not restrict the use of satellite imagery.
- Radiometric photographs provide more accurate results.

The main disadvantage of using active remote sensing is that most methods and algorithms for processing radiometric images require larger computations, and the algorithms themselves are complex. In order to process remote sensing data quickly and efficiently, it is necessary to apply distributed calculations:

- Cluster of computers.
- Cloud solutions;

Many GIS software packages do not yet use the full capacity of computers to process remote sensing data. Modern processors are made up of cores that can process data in parallel, thus speeding up the processing of remote sensing data. The paper examines methods that allow the use of distributed computations, as well as a methodology to accelerate their processing independently of remote sensing data processing algorithms to provide real-time data that could be used to monitor and analyze environmental parameters (Michael McCool Arch D. et al. 2012).

By accelerating the processing of remote sensing data, active satellite imagery can be adapted to detect the following phenomena:

- Determination of water surface contamination by petroleum products using active remote sensing (Syntetic Aperture Radar).
- Determination of water surface contamination with plastic (plastic products) using active remote sensing.
- To monitor plastic contamination of the earth's surface using active remote sensing.
- To assess the damage caused by a forest fire and to determine the direction of fire expansion.
- Determination of nitrogen content in plants in order to reduce the use of fertilizers in crop production.
- Determination of air pollution using active remote sensing.

Remote sensing of satellite systems can provide ecologically sound long-term data sets suitable for analyzing changes in the ecosystem in a given area, in terms of structure, time and space, using appropriate risk assessment criteria. It is often complex and unclear how to select and effectively use remotely obtained data that could be used for environmental parameters and risk assessment (Jining Yana et al., 2017).

Data obtained by remote sensing are often difficult to adapt due to the complexity of their processing. Often, satellite survey data cannot be processed fully automatically and their results must be interpreted with caution. It is for these reasons that it is often necessary to adapt processing and outcome evaluation methodologies to the area in which the research is conducted (Atharva Sharma et al., 2017). Automatic interpretation of remote sensing images is a very complex problem and is fast becoming a necessity for many areas such as disaster management, forest mapping, urban planning, and more. Indeed, images with ever-increasing spatial resolution are becoming such that their adaptation becomes difficult without any computer aid (Hui He et al., 2016).

Sometimes it is not possible to process existing data by adapting computer programs because they do not use the full capacity of the computer. Despite the fact that computer hardware is naturally parallel, computer architects decided 40 years ago to accept serial programming abstraction for programmers (fig. 2.7).

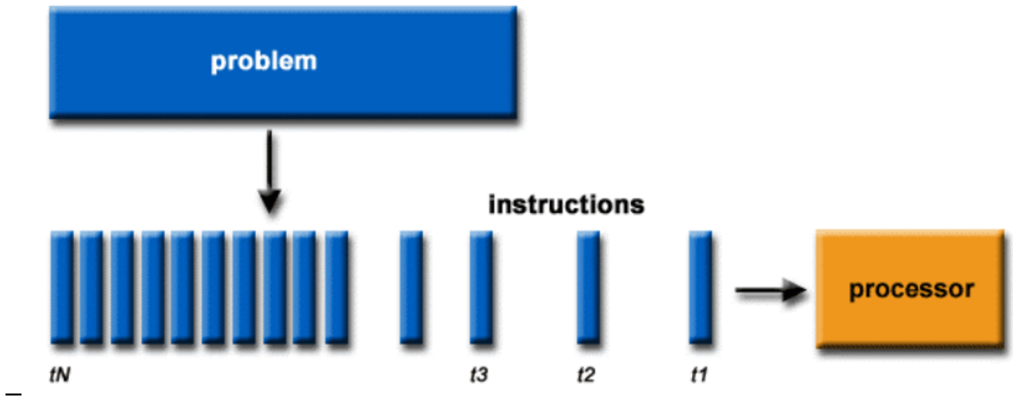


Figure 2.7: Serial execution of instructions on a simple processor

Decades of computer architecture have been designed to maintain the illusion of series execution. In modern processors, much effort is put into translating serial programs into parallel form so that they can run efficiently using parallel hardware inside the processor. Unfortunately, despite the increase in the number of transistors provided for in Moore's Law (which states that the number of transistors that can be integrated into a chip doubles every two years), the need for parallelism is now so great that the illusion of serial computing cannot be maintained (Michael McCool Arch D. et al., 2012).

Serial execution of computer operations is not optimized and GIS problem solving can be accelerated by applying algorithms that specify how operations should be performed in parallel.

Parallel computations and data processing are several computational resources at the same time when those resources are automatically allocated to solve the problem:

- The problem is broken down into several separate parts that can be solved at the same time.
- Each part is divided into separate series of instructions.
- Each instruction in the series is executed simultaneously using all computer resources in parallel.

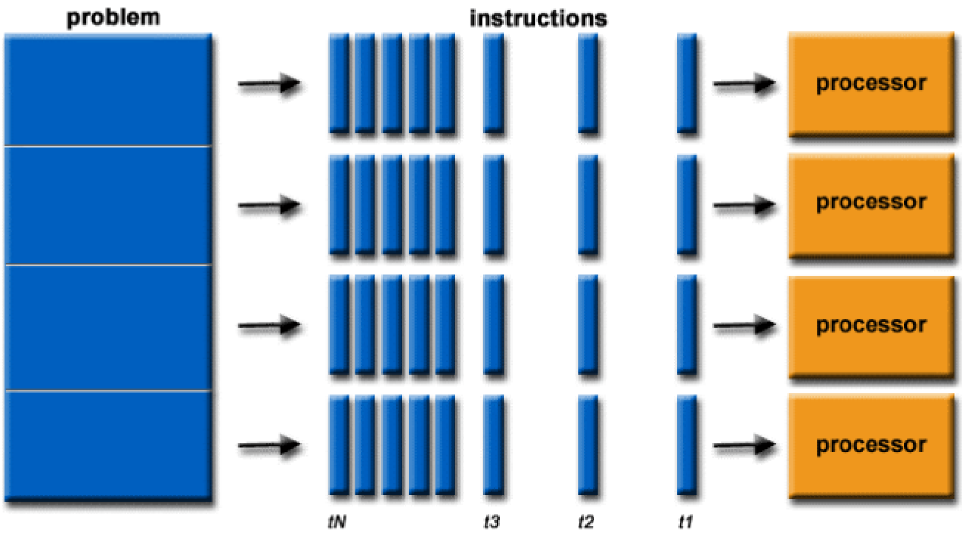


Figure 2.8: Parallel execution of instructions on a simple processor

Increasingly, remote sensing data is being processed using modern distributed computational techniques and automated computational scripts written in interpretable programming languages (Python, R language) that use optimized libraries written in lower-level programming languages (C, C ++, Fortran).

R-script programming language can be used for automated and distributed GIS data processing. R is a system of statistical analysis and graphing developed by Ross Ihaka and Robert Gentleman. R is software and language that is considered a dialect of the S language developed by AT&T Bell Laboratories. It is a de facto data language scripting languages that can orchestrate very complex analytical data streams (pipelines) using different data types. R has many packages that can be used to write high quality code using the packages that are written in them. Most of the code in the R programming language can be optimized using vectorized data structures with methods implemented in C / C++ low lever programming languages. In most cases, programming languages are used in conjuction with other programming languages

In most cases, parallel computations in the R programming language are realized using specific packages that simplify and solve parallelism. The most popular parallel R measures are snow, multicore, foreach, and rmpi. The first two of these are now part of the R core, which is named parallel (R-core. Package 'parallel', 2017). All four of the most popular packages mentioned earlier most commonly use the setter / getter (receiver) paradigm (Weston S. et al., 2017):

- Data set up - the manager breaks down the specified calculation into parts and sends them to the workers.
- Chunk computation - employees count each part of the operation and return the results to the manager.
- Data Acquisition and Merge - The manager receives those results and combines them to solve the specified problem.

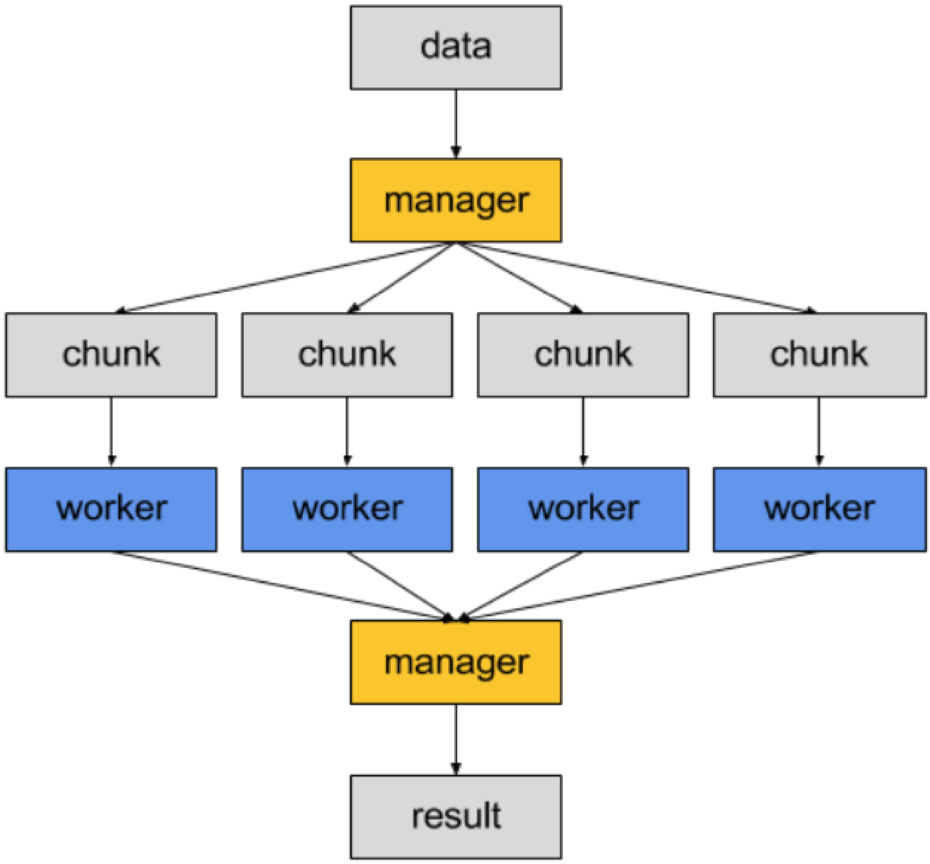


Figure 2.9: setter / getter paradigm for realizing parallel data processing in R programming language

One of the simplest examples of how many kernel computations could be used in the R programming language is the computation of the NDVI index. Calculating NDVI is fairly straightforward and it can be very simple to parallelize by depicting how parallel calculations can speed up the processing of large amounts of remote sensing (Dorman M., 2014).

When setting up parallel calculations, it is useful to know the number of processors or cores. Currently, almost all physical processors have two or more cores that run more or less independently (they can share a shared memory area and share RAM). However, in some processors, these kernels can perform multiple tasks on their own at the same time, and in some operating systems (e.g., Windows), there are logical processors that can exceed the number of kernels [2].

In many cases, the $detectCore()$ function can be used in an R programming language to try to determine the amount of CPU logical cores in a computer [3].

As mentioned earlier, in many cases, the identifier and receiver paradigm is used in the following steps:

- Determination of the number of logical cores.
- Creating and registering a cluster that will use individual computer cores.
- Creation of an employee (recipient of a task) responsible for the performance of a certain task.
- Establishment of a manager (task sender) responsible for sending data to employees.

With cloud services that becomes more popular and effective for distributed computing the processing of large amounts of data can be realized using Amazon cloud services. In the R programming language, a module is implemented that allows distributed data processing using the resources of the Amazon cloud service provider. The main advantage of using cloud services is that it is not necessary to have your own hardware to realize large amounts of data processing at a given time (Sloan T. et al., 2016).

## 2.11 Fire Risk Assessment Models Using Statistical Machine Learning

In the discrete probability distributions, Poisson and binomial distributions can be applied to analyze the fire occurrence data. In the proposed study, we focused on building a model for the possibility of fire, so we built a predictive model based on the binomial distribution. A random variable $Y$ is distributed to a binomial distribution with $n$ and $p$ when $Y$ is represented as follows [1]

$$P(Y = y) = \begin{pmatrix} n \\ \\ y \end{pmatrix} \cdot p^y(1-p)^{n-y}, y = 0, 1, ..., n \tag{2.16}$$

where n is the number of Bernoulli trials, and p is the probability of success. The expectation $E(Y)$ and variance $Var(Y)$ of $Y$ are $np$ and $np(1-p)$ respectively. Each y has a binary data value (1: occurred fire or 0: no occurred fire) representing whether a fire has occurred. So, we build a logistic regression model based on binary response variable $Y$ to forecast fire risk.

# Chapter 3

# Damage detection and predicting fire behavior and condition

Procedures for predicting the spread of fire include:

- Evaluation of installation fuel, fuel moisture, wind speed, surface slope.
- Estimation of fire spread speed and intensity.
- Methods for interpreting the speed and intensity of a fire to obtain the distance, perimeter, area, and conditions under which stains and crowns occur. An important feature is the determination of the probable time of fire growth on a map over a period of time.

The diagram below shows the flow in information system to implement the fire behavior prediction (Choi 2020):

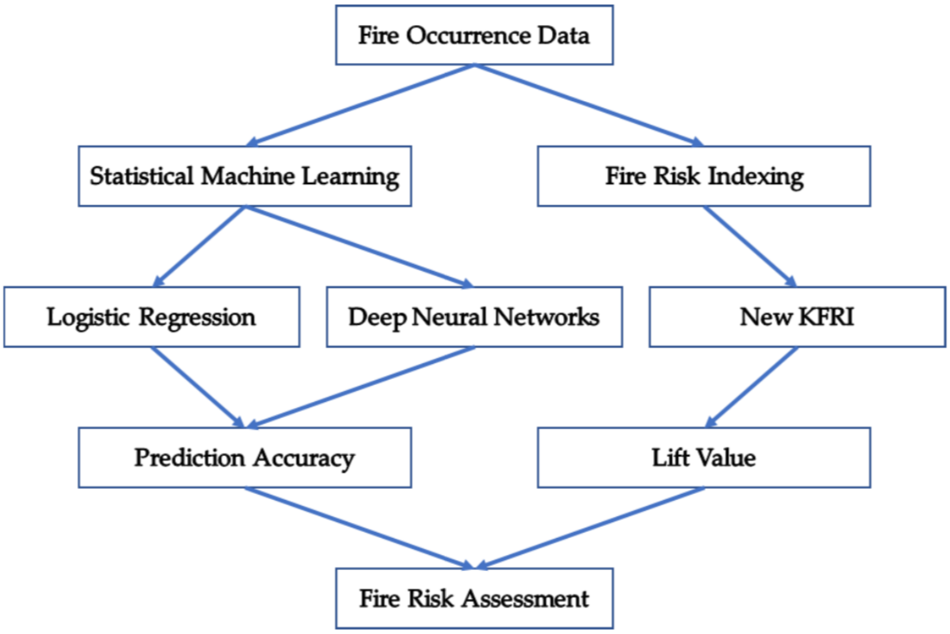


Figure 3.1: Procedure for fire risk assessment

## 3.1 Wildfire damage detection methodology using passive remote sensing data

To determine the damage from a forest fire, a relatively normalized installation index can be used, which allows to determine the intensity ranges of the damaged forest.

The application of the forest fire damage detection methodology was performed using the Sentinel-2 group satellite Sentinel-2A with passive remote sensing sensors, which was launched into orbit in 2015.

In order to determine whether the data provided is suitable for solving GIS tasks, a survey was conducted in 2017. Determination of fire damage in Portugal. Sentinel-2A data used for the study are from 04/06/2017 to 04/07/2017.

Forest fire damage was used in 2017. in central Portugal, in the Pedrógão Grand region, Sentinel-2A satellite data before and after the fire. The corresponding indices (before and after the fire) were calculated using raster algebra (pixel composition, subtraction, multiplication and division of different wavebands) and the results were compared with each other.

Before indexing, the initial data must be prepared by performing the appropriate image processing.

Sentinel-2A data has cloud information, so using SNAP software we can use this information for a cloud mask creation.

```
if (
  scl_cloud_medium_proba +
  scl_cloud_high_proba +
  scl_thin_cirrus
) < 255 then 0 else 1
```

where $scl\_cloud\_medium\_proba$, $scl\_cloud\_high\_proba$, and $scl\_thin\_cirrus$ cloud data, which distinguishes clouds from the original data.

This filter is used to perform raster algebra in SNAP software for removing clouds. After performing raster algebra and applying the above cloud filter, an additional band in the raster data set is obtained, which is required in the following data processing steps.

A normalized burn index is most commonly used to isolate vegetation damaged during a fire. This index is calculated according to the following formula (Parks et al., 2014):

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \tag{3.1}$$

where $NIR$ - infrared wavelenght band, $SWIR$ - shortwave infrared color band.

The higher the normalized implementation index, the vegetation in the study area is the least damaged and vice versa, the lower the index, the more damaged the vegetation (Coppoletta, M et al., 2016) (fig. 3.2).

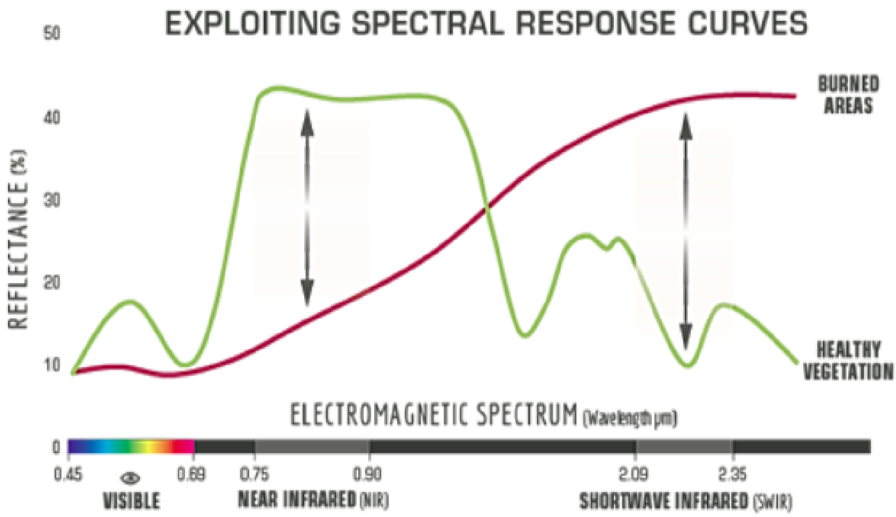


Figure 3.2: The NBR index was determined in the SNAP software

To refine the results, a normalized water index is determined that separates water from the forest vegetation. Calculation formula (Du et al., 2016):

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{3.2}$$

where $NIR$ - inrared color band and $Green$ - green color band

After performing raster algebra, the hydrological layer and raster data are separated.

To determine the final forest damage, a relative deployment index is calculated, which determines the ratio of the deployment index before and after the fire (Parks, S, et al. 2014):

$$dNBR = \frac{NBR_{prefire} - NBR_{postfire}}{NBR_{pre-fire} + 1.001} \tag{3.3}$$

where $NBR_{prefire}$ - normalized installation index before fire, $NBR_{postfire}$ - normalized installation index after fire.

The intervals of the $dNBR$ index allow to determine the extent of forest damage (tab. 3.1). These intervals are used for the final visualization of the processed data and presentation of the results.

Table 3.1: $dNBR$ index interval burn intervals (Tonbul, H et al., 2016)

| **dNBR** | **Burn severity** |
|---|---|
| < -0.25 | High regrowth after fire |
| frpm -0.25 to -0.1 | Low regrowth after fire |
| from -0.1 to 0.1 | Not burned |
| from 0.1 to 0.27 | Low |
| from 0.27 to 0.44 | Medium low burn |
| from 0.44 to 0.66 | Medium hight burn |
| > 0.66 | Hight burn |

The calculation of $dNBR$ is performed by applying normalized water index and cloud cover filters and applying formula 3.3. After calculating the $dNBR$, a raster data set was obtained, the pixel values of which determine the degree of forest burn (fig. 3.3).

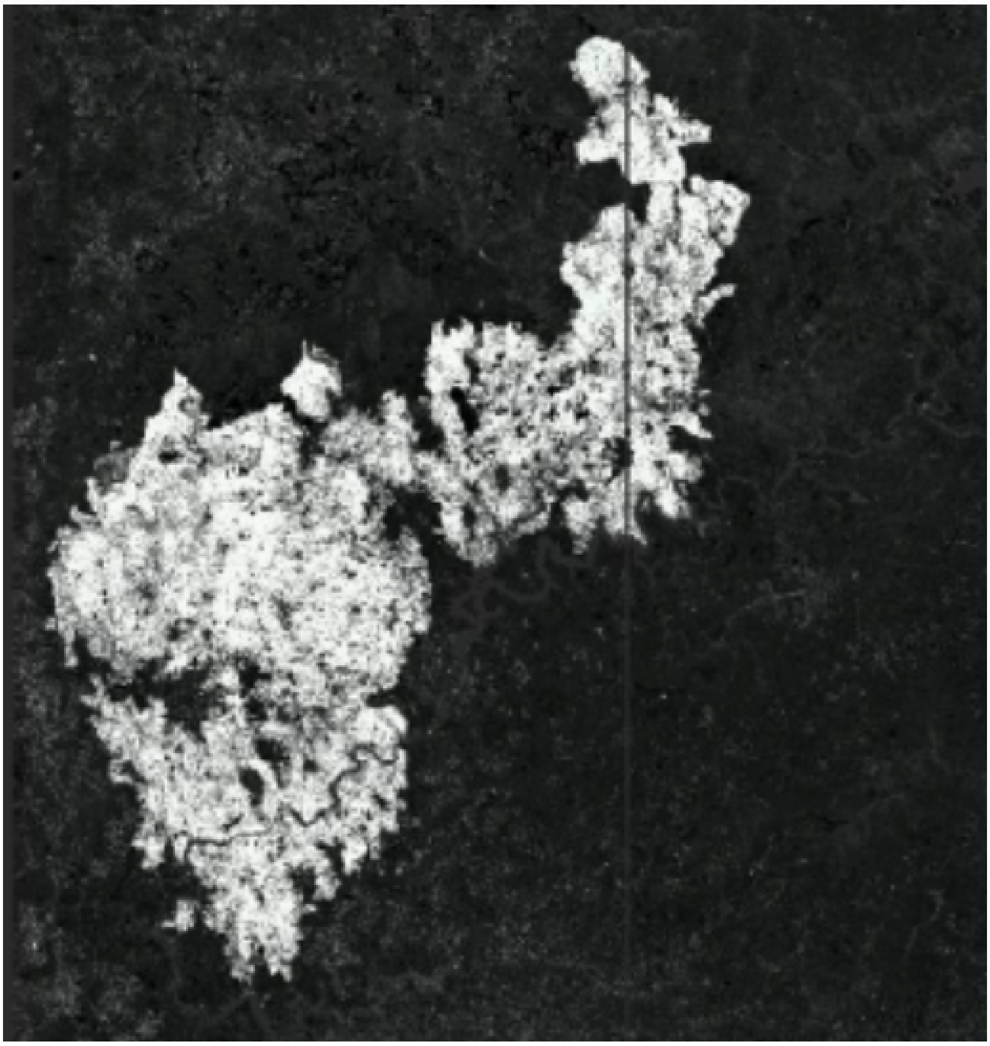

Figure 3.3: Calculated $dNBR$ index

# Appendix A

# Algorithms

In this Appendix we present pseudo codes and the settings used for the peer algorithms, which we implemented and used in experimental evaluation through the thesis.

# Appendix B

# Derivations and Computation Details

## B.1 Change Detection Using Hotelling T-test

...

# Bibliography

[1] Myoung-Young Choi Sunghae Jun Myoung-Young Choi, Sunghae Jun. Fire risk assessment models using statistical machine learning and optimized risk indexing.

[2] Matloff N. *Parallel Computing for Data Science: With Examples in R, C++ and CUDA.*

[3] Lim A. Tjhi W. *R High Performance Programming.*

# Publications by the Author

## Periodic

1. Tumelienė, Eglė; Nareiko, Viktor; Sužiedelytė-Visockienė, Jūratė. Photogrammetric measurements of heritage objects. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV5/W1. ISPRS Workshop "Geospatial Solutions for Structural Design, Construction and Maintenance in Training Civil Engineers and Architects", 4-6 December 2017, Kyiv, Ukraine: ISPRS. ISSN 2194-9042. eISSN 2194-9050. 2017, p. 71-76. DOI: 10.5194/isprs-annals-IV-5-W1-71-2017. [DOAJ; Scopus].

## Peer reviewed conference publications

2. Nareiko, Viktor; Sužiedelytė-Visockienė, Jūratė. Gaisro žalos nustatymo miškui metodologija pagal "Sentinel" nuotolinių tyrimų duomenis = Burned area determination of forest with Sentinel sattelite remote sensing data. // 21- osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis“ teminė konferencija Civilinė inžinerija ir geodezija, 2018 m. kovo 23 d. Vilnius : [straipsnių rinkinys] = Proceedings of the 21th conference for junior researchers "Science – future of Lithuania" Civil engineering and geodesy, 23 March, 2018, Vilnius, Lithuania. Vilnius : Vilniaus Gedimino technikos universitetas, 2018. eISBN 9786094761140. eISSN 2029-7149. p. 1-5. DOI: 10.3846/geo.2018.010. [M.kr.: 04T,10T] [Ind.: 1,000] [Ind. aut. lankais: 0,357]

3. Nareiko, Viktor; J. Sužiedelytė-Visockienė, Jūratė. Apply of open remote sensing data in environmental research = Atvirų nuotolinių tyrimų duomenų taikymas aplinkosaugoje. // 21-oji jaunųjų mokslininkų konferencija "Mokslas – Lietuvos ateitis" teminė konferencija Aplinkos apsaugos inžinerija, 2018 m. kovo 20 d. Vilnius.

# Curriculum Vitae

Viktor Nareiko graduated from ...

# Vocabulary - Žodynėlis

base learner - bazinis klasifikatorius
baseline - bazinis metodas
change point - pokyčio taškas
concept drift - koncepcijos pokytis
context aware - kontekstinis
data mining - duomenų gavyba
data source - duomenų šaltinis
gradual drift - palaipsnis pokytis
instance - vektorius
instance based learning - mokymas pagal vektorius
label - klasė
moving average - slenkantis vidurkis
peer methods - lyginamieji metodai
recurring concepts - pasikartojantis pokytis (pasikartojančios koncepcijos)
sequential learning - mokymas paeiliui
source - šaltinis
sudden drift - staigus pokytis
supervised learning - mokymas su mokytoju
training window - mokymo langas
unsupervised learning - mokymasis

# Summary in Lithuanian (Santrauka)

...

# Index