Hochschule Rhein-Waal

Rhine-Waal University of Applied Sciences

Faculty of Communication and Environment

Prof. [Rolf Becker](https://moodle.hochschule-rhein-waal.de/user/view.php?id=524&course=1)

**Final Assignment**

Project Report

Winter Term 2019/2020

[M-IE\_2.02 Geoinformatics](https://moodle.hochschule-rhein-waal.de/course/view.php?id=10688)

By

Francisco Susana (26574)

Erick Rea Reasco (26525)

Submission Date:

25th February 2019

Abstract

Monitoring of mining areas with application of geodetic measurements is the basic task of

mining surveys. Classical geodetic measurements include linear probing of the terrain by

observation lines, and surface probing such as measurement with use of a regular grid

within the surveyed mining area [1]. Classical observation systems in open cast mines use

surface, photogrammetric and satellite GNSS geodetic measurements as a source of input

data for GIS 3D databases [2]. The paper presents selected applications of satellite remote

sensing in the monitoring of mining areas and their surroundings with regard to assessing

the influence of open cast mining on the environment [3, 4]. Issues of remote sensing

monitoring are presented in macro scale [5]. The processed remote sensed data have

allowed for large-surface analysis as a tool for identifying local anomalies and

identification of potential hazards [6].

Table of contents

[Introduction](#_heading=h.gjdgxs) **2**

[Material and methods](#_heading=h.30j0zll) **5**

[Results](#_heading=h.1fob9te) **8**

[Discussion](#_heading=h.3znysh7) **8**

[Conclusions and Outlook](#_heading=h.2et92p0) **8**

[References](#_heading=h.tyjcwt) **9**

[Annex](#_heading=h.3dy6vkm) **9**

[Statement of authorship](#_heading=h.4d34og8) **9**

List of figures

List of Tables

# Introduction

The aim of this report is to analyse the effects of the 2018 drought Germany suffered, on the vegetation of specific zones in the Xanten region. Different data sources and tools were used to investigate the phenomena. Mainly precipitation and temperature data collected by the German Weather Service (DWD) between the years 2016 to 2018.

In order to analyze the impact of the drought on the vegetation, satellite imagery of area was used to calculate the spectral indices on a summer date for the years 2016, 2017 and 2018. A comparison and correlation analysis between these variables analysis was made to investigate the following hypotheses:

1. The drought affects vegetation and the Normalized Difference Vegetation Index (NDVI) shows the effects on vegetation.

2. The drought is related to the cumulative precipitation CP (i.e. the precipitation sum) over the last year. There is a correlation between CP and NDVI.

3. The drought is related to the average temperature AT over the last year. There is a correlation between AT and NDVI as well as AT and CP.

4. Human activity, lake

# Fundamentals

**2.1 Remote sensing**

The National Ocean Service of the US defines remote sensing as the science of obtaining information about objects or areas from a distance, typically from aircraft or satellites. Remote sensors collect data by detecting the energy (sunlight) that is reflected from Earth.

Remote sensing has become a useful instrument in the study and monitoring of regions despite the challenges it presents. Field or in-situ data yields more accurate results. However, remote sensing technology is constantly improving and provides a good view for the focus area of this research. This research makes use of imagery collected by satellite sensors from the Copernicus program, the sensors collect high-resolution radiometric, spatial and temporal remote sensing data.

**2.1.1 Copernicus Earth Observation programme**

Copernicus is the European Union's Earth Observation Programme, looking at our planet and its environment for the ultimate benefit of all European citizens. It offers information services based on satellite Earth Observation and in situ (non-space) data.

The Sentinel-2 twin satellites, launched in 2015 and 2017, are part of the Copernicus programme and provide high resolution optical imagery containing the multi spectral indices needed for this research.

The Copernicus Open Access Hub website provides complete, free and open access to the satellite observations in order to support researchers working on different projects in fields such as agriculture, land monitoring, water monitoring, flood mapping, etc.

**2.1.2 Products available**

The Sentinel-2 user handbook refers to a downloadable compilation of one or more granules of fixed size, within a single orbit as a product. A granule is the minimum indivisible partition of a product and contains information about all possible spectral bands at a pixel level (Sentinel, 2018). The access hub provides 2 product types for most of the geographic granules.

|  |  |  |  |
| --- | --- | --- | --- |
| **Product** | **High-level Description** | **Production & Distribution** | **Data Volume** |
| Level-1C | Top-of-atmosphere reflectances in cartographic geometry | Systematic generation and on-line distribution | 600 MB (each 100x100 km2) |
| Level-2A | Bottom-of-atmosphere reflectance in cartographic geometry | Systematic generation and on-line distribution and generation on user side (using Sentinel-2 Toolbox) | 800 MB (each 100x100 km2) |

#### Table 1: Sentinel-2 product types

The Level-2A product is considered as the mission Analysis Ready Data (ARD) meaning that can be used directly in applications without the need for further processing. Products of Level-1C type need to be atmospherically corrected (process to remove the effects of the atmosphere on the reflectance values) using the Sen2Cor tool.

Sen2Cor is a processing software for Sentinel-2 Level 2A product generation and formatting; it performs the atmospheric-, terrain and cirrus correction of Top-Of- Atmosphere Level 1C input data. Sen2Cor creates Bottom-Of-Atmosphere, optionally terrain- and cirrus corrected reflectance images; additional, Aerosol Optical Thickness-, Water Vapor-, Scene Classification Maps and Quality Indicators for cloud and snow probabilities.

**2.2 Normalized Difference Vegetation Index (NDVI)**

The Normalized Difference Vegetation Index (NDVI) is an indicator of the greenness of the biomes. (Copernicus). The NDVI has become a widely used instrument to analyze remote sensing measurements. It is calculated using the formula:

NDVI = (NIR-RED)/(NIR+RED)

Being “NIR” the reflection of light in the near-infrared spectrum of wavelength, which is reflected by the vegetation, while “RED'' refers to the reflection of light in the red range of the spectrum, which is mostly absorbed by the vegetation.

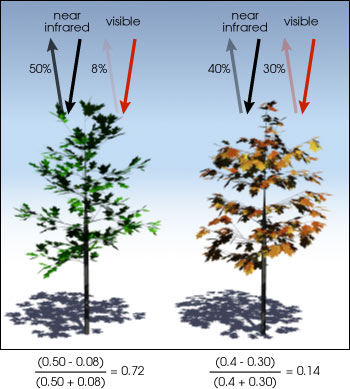


Fig. x (Nasa)

The NDVI is a tool used to represent the density of the vegetation in a specific area and its health. Various research papers make use of the NDVI to analyse agricultural production, forecast vegetation changes or study drought phenomena. The NDVI indicator was key for the correlation analysis made in the upcoming chapters.

The NDVI varies between -1 to +1. Even though there are not specific ranges to categorize land use according to NDVI values, it is widely agreed that some values likely represent certain environments. For instance, negative values likely represent water bodies, values close to +1 likely indicate the existence of dense green leaves. Finally, values close to 0 highly indicate that there are no green leaves or that the place is an urbanized area.

**2.3 Spatial autocorrelation**

The first law of geography states that “Everything is related to everything else. But near things are more related than distant things”. (Tobler, 1969)

measures how close objects are in comparison with other close objects. Moran’s I can be classified as positive, negative and no spatial auto-correlation.

* Positive spatial autocorrelation is when similar values cluster together in a map.
* Negative spatial autocorrelation is when dissimilar values cluster together in a map

Google book

**2.4 Geographic Information System (GIS)**

National Geographic defines a Geographic Information System (GIS) as a computer system for capturing, storing, checking, and displaying data related to positions on Earth’s surface.. GIS has also become an umbrella term referring to the processes, software and techniques used to study geographical data. The software used to analyze and process spatial information used in this research is QGIS which is the industry standard.

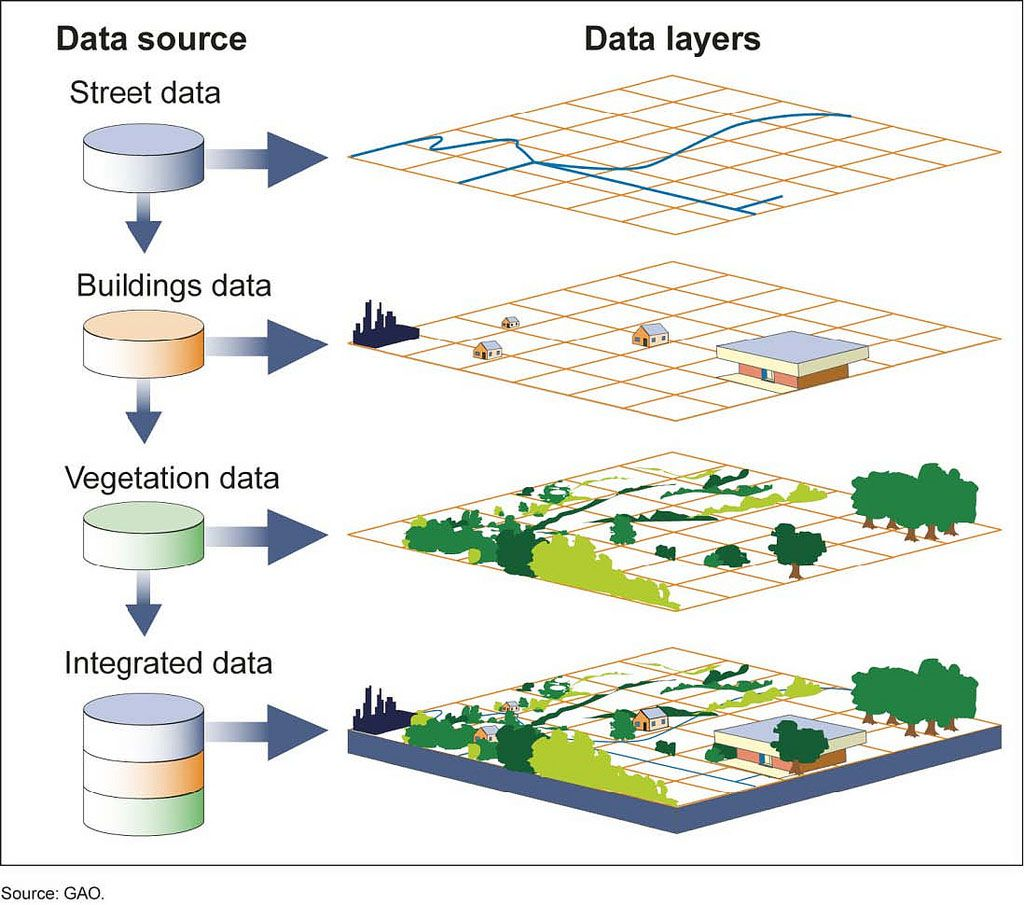


Fig. x GIS (US Government Accountability Office)

QGIS is a geographic information system (GIS) for multiple platforms. Like other geoinformation systems, QGIS allows the creation, processing, viewing and analysing of geospatial data and the composition of printable maps. The software is made available free-of-charge under the GNU General Public License. (QGIS-DE e.V., 2019)

In addition to the standard range of functions, QGIS can also be extended with plug-ins (e.g. OpenStreetMap). It can also process various file formats and work with several databases to include both vector data and raster data. For working with scripts, it is possible to use its integrated Python console. (QGIS-DE e.V., 2019)

# Material and methods

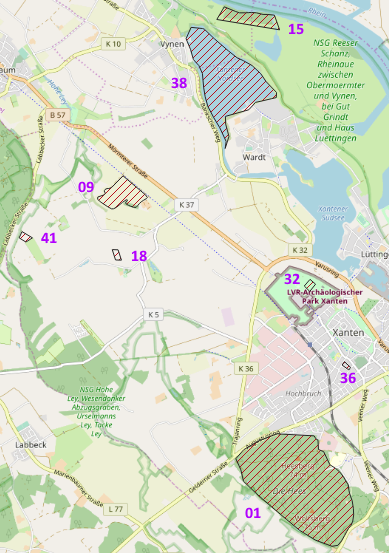
In order to analyse the 2018 drought that Germany suffered in comparison to the years 2016 and 2017, several regions of interest in the Xanten area have been identified. Specifically, 8 zones were selected as the subject of investigation for this research. In order to assess the drought in the 8 regions of interest (ROI), they were first classified according to their land use. Then the 3 variables used for this research were procured from different sources:

* The NDVI was calculated by downloading imagery from the Sentinel 2 mission satellites and then processing this data in QGIS
* Precipitation data was collected from an FTP server using Python code.
* Temperature data was collected from an FTP server using Python code.
* A correlation analysis was made

The following sections describe in detail the process used to reach obtain and analyze the data.

**3.1 Area description**

Xanten is a town in the state of North Rhine Westphalia, Germany. The 8 ROIs were first filtered from a shapefile containing several polygons. Every region of interest was exported individually as a new layer in order to have single shapefile layers. The regions were then classified depending on their land use into 5 categories: construction, farmland, forest, grassland and water.



The classification, per observation in Google maps and Open Street maps, is described in table x.

|  |  |
| --- | --- |
| **Region of interest id** | **Land use** |
| 1 | Forest |
| 9 | Farmland |
| 15 | Grassland |
| 18 | Farmland |
| 32 | Grassland |
| 36 | Construction |
| 38 | Water |
| 41 | Farmland |

**3.2 Sentinel-2 data**

Products for 6 dates in the summer periods from 2016 to 2018 were downloaded from the Copernicus Open Access Hub website: (a) 2016-05-08, (b) 2016-06-10, (c) 2017-05-26, (d) 2018-05-08, (e) 2018-05-11 and (f) 2018-06-30. The products were found by signing up for an account in the hub and using the advanced search interface to select the geographical target and the sensing dates A, C and F for the Sentinel-2 mission.

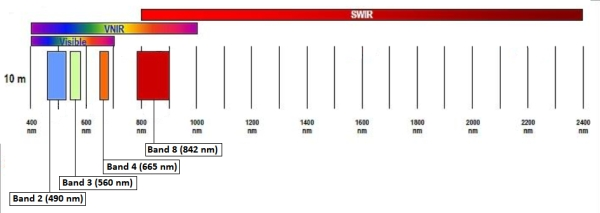
The granules obtained are composed of raster images and associated files containing multi-spectral data for areas of approximately 100 km2. The most suitable granules for this research were (a) 2016-05-08, (c) 2017-05-26 and (f) 2018-06-30 due to their almost non-existent cloud coverage which makes NDVI calculation more accurate for this scientific research. The area covered by each of the granules found for sensing dates A, C and F contain all the regions of interest within their borders.

Sen2cor uses a product 2A as input and outputs a product 1C with resolutions of 60, 20 and 10m. It has 2 versions available 2.5.5 and 2.8 (latest). Sen2Cor 2.5.5 is the previous release and it is needed if the user intends to process old Sentinel-2 L1C data with a PSD older than version 14.2 which was released in 2016-10-24. Therefore the tile (a) 2016-05-08 was corrected using Sen2Cor 2.5.5, the tile (c) 2017-05-26 and (f) 2018-06-30 did not need to be atmospherically corrected as the product downloaded already came in format 2A.

**3.3 NDVI calculation**

The files downloaded from the Copernicus Open Access Hub provide detailed information about the reflectance of the granules containing the areas of interest. This information comes in resolution sizes of 10, 20 and 60 m. The resolution used was 10 m since some of the regions of interest are small and benefit from more detailed data.

The 10 metre spatial resolution folder contained multiple rasters representing the different bands of wavelengths reflectance measurements. According to the user guide of the Sentinel-2 mission: the raster for Band 4 holds the readings regarding the red light and the one for Band 8 contains the readings for the near infrared light.



**Figure 1: SENTINEL-2 10 m spatial resolution bands: B2 (490 nm), B3 (560 nm), B4 (665 nm) and B8 (842 nm)**

The formula to calculate the NDVI in the context of the Sentinel-2 bands is:

**NDVI = (Band 8- Band 4)/(Band 8+ Band 4)**

The Band 4 and Band 8 rasters for each of the dates studied in 2016, 2017 and 2018 were clipped using the “Clip raster by mask layer” feature in QGIS. This was done in order to facilitate the calculation of the NDVI for each region of interest separately. It is important to keep track of the names of the layers as their amount increased rapidly. The “Raster calculator” feature was then used to calculate the NDVI with the pair of clipped rasters for each region of interest of 2016, 2017 and 2018.

In order to guarantee the accuracy of the results, the clipped rasters were also transformed to vectors by using the “Raster pixels to polygons” feature in the processing toolbox of QGIS. The algorithm converts a raster layer to a vector layer and creates a polygon feature for the pixels contained in the layer. The resulting vector layers featured an associated table with their band measurements that was copied to Excel and then used in a formula to calculate the NDVI. The resulting values are the following.

(INSERT TABLE)

**2.4 Climate data procurement**

The German Weather Service (DWD) provides the climate data needed for the regions of interest through the Open Data section in their website. The raw data is open and free to access via an FTP (File Transfer Protocol) server. Boilerplate code of Prof. Dr. Rolf Becker, available in his repository on the Spectors project website, was used to develop a Python script to download and pre-process this data.

The script connects to the DWD ftp server which contains daily temperature and precipitation measurements for various weather stations in Germany from different directories. The text files for temperature and precipitation stations were downloaded and converted to csv format. The csv file was filtered to contain only the stations in NRW.

The csv file was imported to QGIS to visualize the stations in the map and find the nearest ones to the ROI area by selecting the stations within a radius of 100 kilometers. The selected stations were filtered to find the ones containing data on the dates of interest. The next script downloaded the zip files containing the temperature (tmk column) and precipitation (rsk column) data of the stations filtered in the step before. The files were unzipped, transformed from text to csv, and clipped to contain only the information for the periods between:

* + - A: 2016-04-09 to 2016-05-08
    - C: 2017-04-27 to 2017-05-26
    - F: 2018-06-01 to 2018-06-30

The daily data was aggregated to get cumulative precipitation and average temperature for each month. Only 7 stations

* + After analyzing the data from the weather stations, from 10 stations selected only 7 had all the values that corresponded to our time of analysis.
* (5\_QGIS) In order to display the temperature and precipitation on the map the “IDW Interpolation” tool from QGIS was used to interpolate the climatic variables based on the area covered for 7 stations selected.
* (6\_EXCEL) Once the interpolated values were obtained, we could confirm that those values and the averaged values from step 4 values did not have much difference, reason why we decided to used them instead, same approach was made in a different study [**source**: paper#1]

we You will use Python to download and pre-process temperature and precipitation data from German Weather Service (DWD). Additionally, you will acquire the required datasets for your research, such as multispectral imagery, from reliable sources. Finally, you will use QGIS to load, process, and produce the

We will interpolate Geldern and Duisburg based on the homogeneity of the observations and the WMO recommendations regarding scales for representativeness. This is mentioned in the answer by [Janusz Pudykiewicz](https://www.researchgate.net/profile/Janusz_Pudykiewicz) citing the WMO in the doc

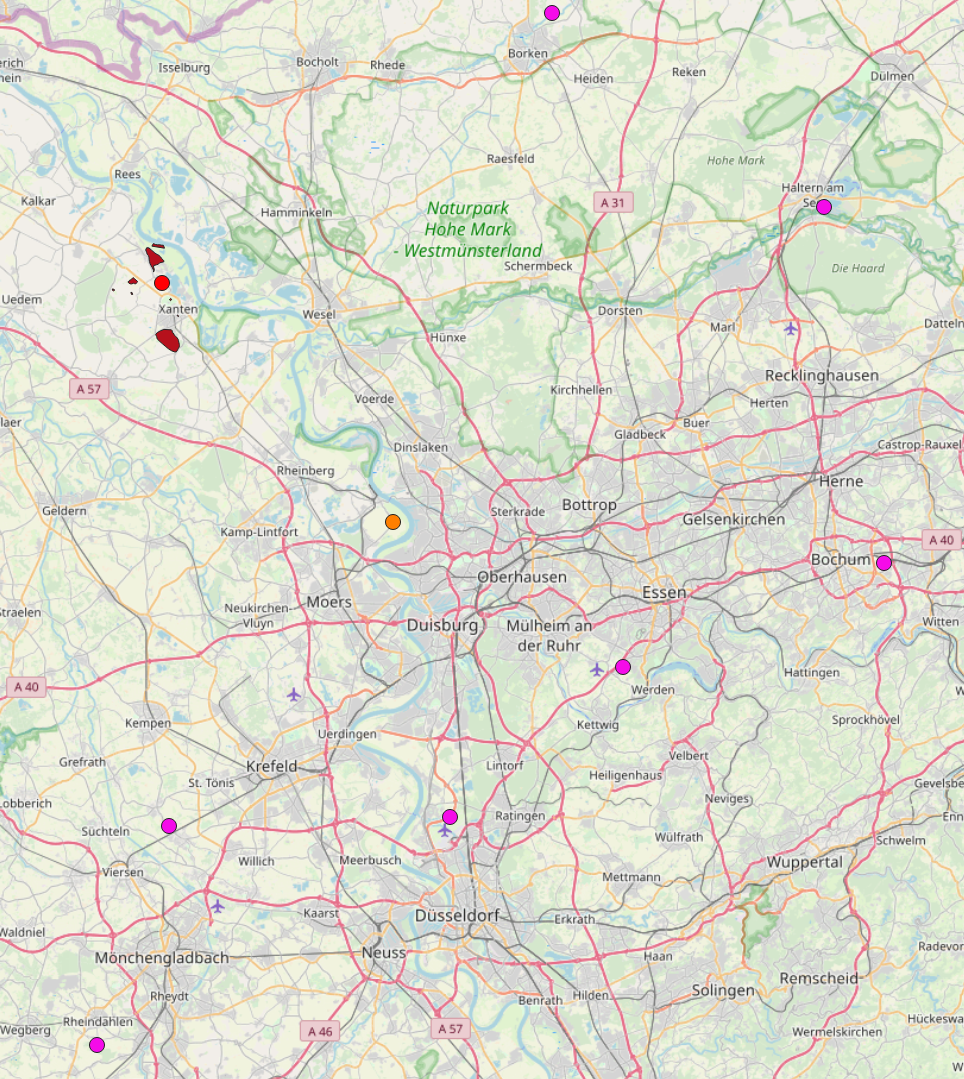
[https://www.wmo.int/pages/prog/www/IMOP/meetings/CB/Ed-Board-2/EdBd-2\_P-I\_Ch-1.doc](https://www.researchgate.net/deref/https%3A%2F%2Fwww.wmo.int%2Fpages%2Fprog%2Fwww%2FIMOP%2Fmeetings%2FCB%2FEd-Board-2%2FEdBd-2_P-I_Ch-1.doc)

Research in station selection based in station proximity

<https://www.researchgate.net/post/Data_and_weather_station_distance>

**bissextile yea**r

The downloaded rasters from the Copernicus Hub were clipped to fit the regions of interest. Each region was then converted into a vector to have granular data.



**Figure #1**. Nearest stations with valid dates for the analysis. (Use QGIS thematic mapping method to display the following text in the picture: Xanten station in red; Duisburg-Baerl station in orange; stations in pink represent block 2.)

Once the [stations\_nrw\_file].csv file was loaded in QGIS, all the 105 stations from NRW were displayed on the map. Two classifications of stations were made: the first one, a block of stations with a radius of 0-50 kilometers (block 1); and the second one, a block of stations with a radius of 50-100 kilometers (block 2). The selection of the stations within the mentioned distances was possible by using the “Select Features by Radius” on QGIS. Then, we classify them by exporting the selected stations on the CSV table to do further analysis. The number of stations was down to 25 (10 stations within the block 1, and 15 stations within the block 2), but, since not all the stations met the specified dates for the data analysis, the number of stations pre-selected decreased to 9 (1 station within the block 1, and 8 stations within the block 2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **station\_id** | **date\_from** | **date\_to** | **altitude** | **latitude** | **longitude** | **Block** |
| 1303 | 1935-01-01 | 2020-02-18 | 150 | 51.4041 | 6.9677 | 2 (100km) |
| 555 | 1940-01-01 | 2018-11-01 | 101 | 51.4789 | 7.2697 | 2 (100km) |
| 1078 | 1952-01-01 | 2020-02-18 | 37 | 51.296 | 6.7686 | 2 (100km) |
| 13901 | 1979-02-01 | 2020-02-18 | 41 | 51.7343 | 7.1998 | 2 (100km) |
| 617 | 2004-06-01 | 2020-02-18 | 47 | 51.873 | 6.8863 | 2 (100km) |
| 5064 | 2004-12-01 | 2020-02-18 | 37 | 51.2897 | 6.4437 | 2 (100km) |
| 7374 | 2006-03-01 | 2020-02-18 | 46 | 52.0814 | 6.941 | 2 (100km) |
| 13670 | 2007-05-31 | 2020-02-18 | 24 | 51.5088 | 6.7018 | 1 (50km) |
| 3321 | 2008-04-25 | 2020-02-18 | 77 | 51.1313 | 6.3609 | 2 (100km) |

**Table #1.** Pre-selection of stations around the ROI area.

However, the mentioned tool used to classify the different blocks throw an approximation from the area of interest to the stations demonstrated in table #1.

To calculate the exact distance, we had to use the QGIS vector function called “Distance Matrix”, where an input vector has to be given to measure its distance to a targeted vector, which means, measuring the precise distance between our area of analysis and the nearest stations shown on the previous table #1. As the input vector, we selected the Xanten station (ID 15190) because it is in the middle of our area of interest, the vector was created by duplicating the layer of the stations of the NRW and applying a filter with the ID of this station. The same process was repeated to create the targeted vectors or vectors that represent the stations that belong to the block 1 and 2.

|  |  |  |
| --- | --- | --- |
| **InputID** | **TargetID** | **Distance (m)** |
| 15190 | 13670 | 26,565.69 |
| 15190 | 617 | 37,826.11 |
| 15190 | 5064 | 43,450.37 |
| 15190 | 1303 | 48,061.69 |
| 15190 | 1078 | 48,624.81 |
| 15190 | 13901 | 53,215.06 |
| 15190 | 7374 | 56,632.07 |
| 15190 | 3321 | 61,285.13 |
| 15190 | 555 | 62,050.37 |

**Table #2**. The precise distance between the area of analysis (around the Xanten station) and the possible stations.

Table #2 illustrate the outcome of the “Distance Matrix” function, revealing the list of stations that are closer to the area of interest and that meet the dates for the analysis.

[Check the files from every station to check which one meet the data to analyze]

Station ID 13670 or Duisborg-Baerl ...

# Results

<https://www.researchgate.net/publication/233416443_Drought_Risk_Assessment_Using_Remote_Sensing_and_GIS_Techniques>

<https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php>

Learning outcomes / Competences and qualifications

profileStudents have expanded the ability to understand and apply basic and advanced concepts and techniques in geo-informatics and geographical information systems (GIS). With respect to their fields of application students have developed skills to recognize the assumptions, implications, values and limitations of different methods in spatial analysis. They have learned to deploy existing skills in modeling and programming for implementing spatial algorithms and extending existing GIS software.They are able to oversee the impacts of information modelling decisions on GIS feasibility and performance of GIS systems.Students are able to apply GIS methods in different fields such as technical, natural or economical systems in different contexts. They know the appropriate use of GIS as well as their limitations. Students are familiar with a range of modern GIS technologies as well as common tools. They are able to develop or apply their own ideas in this field in different contexts.

ContentBasic Concepts-

Information models, spatial data, spatial autocorrelation, error and uncertainty in spatial dataTechniques- Spatial data models and spatial data modelling, geographic databases, GeoWebAnalysis- Map production, geovisualisation, spatial data analysis (e.g. Network analysis. cluster detection, spatial interpolation), spatial modeling with GIS

Additional informationRecommended readings:

Acevedo, M. F.: Data Analysis and Statistics for Geography, Environmental Science, and Engineering, CRC Press, 2013de Smith, MJ, Goodchild M.F., and Longley, P.A.: Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools, 3rd Edition. 2009. Kemp, K.K.: Encyclopedia of Geographic Information Science, Sage Publications. ed. 2008. O’Sullivan, D. and Unwin, DJ.: Geographic Information Analysis, 2nd Edition. John Wiley & Sons. 2010.

4.1. NDVI analysis

* Consider reflecting materials

4.2. Temperature and precipitation analysis

4.2.1 Temperature and precipitation behaviour on the research area

[Remember to mention its relation according to drought definition]

4.2.2 Temperature and precipitation correlation

4.3. Relation between NDVI and climate variables

# Discussion

* Selection of measurement station (direct selection instead of interpolation)
* Yearly visualisations of meteorological data
* NDVI minimum and maximum from vectors differ from the rasters
* Visualisation between precipitation and temperature
* Analyze results

# Conclusions and Outlook

[Start writing here]

**TODO:** <https://trello.com/c/HbNnszOw>

# References

*(PDF) First experience with Remote Sensing methods and selected sensors in the monitoring of mining areas – a case study of the Belchatow open cast mine*. Available from:<https://www.researchgate.net/publication/322823258_First_experience_with_Remote_Sensing_methods_and_selected_sensors_in_the_monitoring_of_mining_areas_-_a_case_study_of_the_Belchatow_open_cast_mine> [accessed Feb 27 2020].

(Using interpolation) <https://www.youtube.com/watch?v=GoypLUmqPL4&feature=youtu.be>

(Correlation concept) <https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/>

<https://www.investopedia.com/terms/c/correlationcoefficient.asp>

(Correlation in excel) <https://www.excel-easy.com/examples/correlation.html>

(Used of pandas) <https://www.tutorialspoint.com/python_pandas/python_pandas_dataframe.htm>

# Annex

[Start writing here]

# Statement of authorship

*I* ***Francisco*** *hereby declare that my contribution to the work presented herein is my own work completed without the use of any aids other than those listed. Any material from other sources or works done by others has been given due acknowledgment and listed in the reference section. Sentences or parts of sentences quoted literally are marked as quotations; identification of other references regarding the statement and scope of the work is quoted. The work presented herein has not been published or submitted elsewhere for assessment in the same or a similar form. I will retain a copy of this assignment until after the Board of Examiners has published the results, which I will make available on request.”*

*I* ***Erick*** *hereby declare that my contribution to the work presented herein is my own work completed without the use of any aids other than those listed. Any material from other sources or works done by others has been given due acknowledgment and listed in the reference section. Sentences or parts of sentences quoted literally are marked as quotations; identification of other references regarding the statement and scope of the work is quoted. The work presented herein has not been published or submitted elsewhere for assessment in the same or a similar form. I will retain a copy of this assignment until after the Board of Examiners has published the results, which I will make available on request.”*