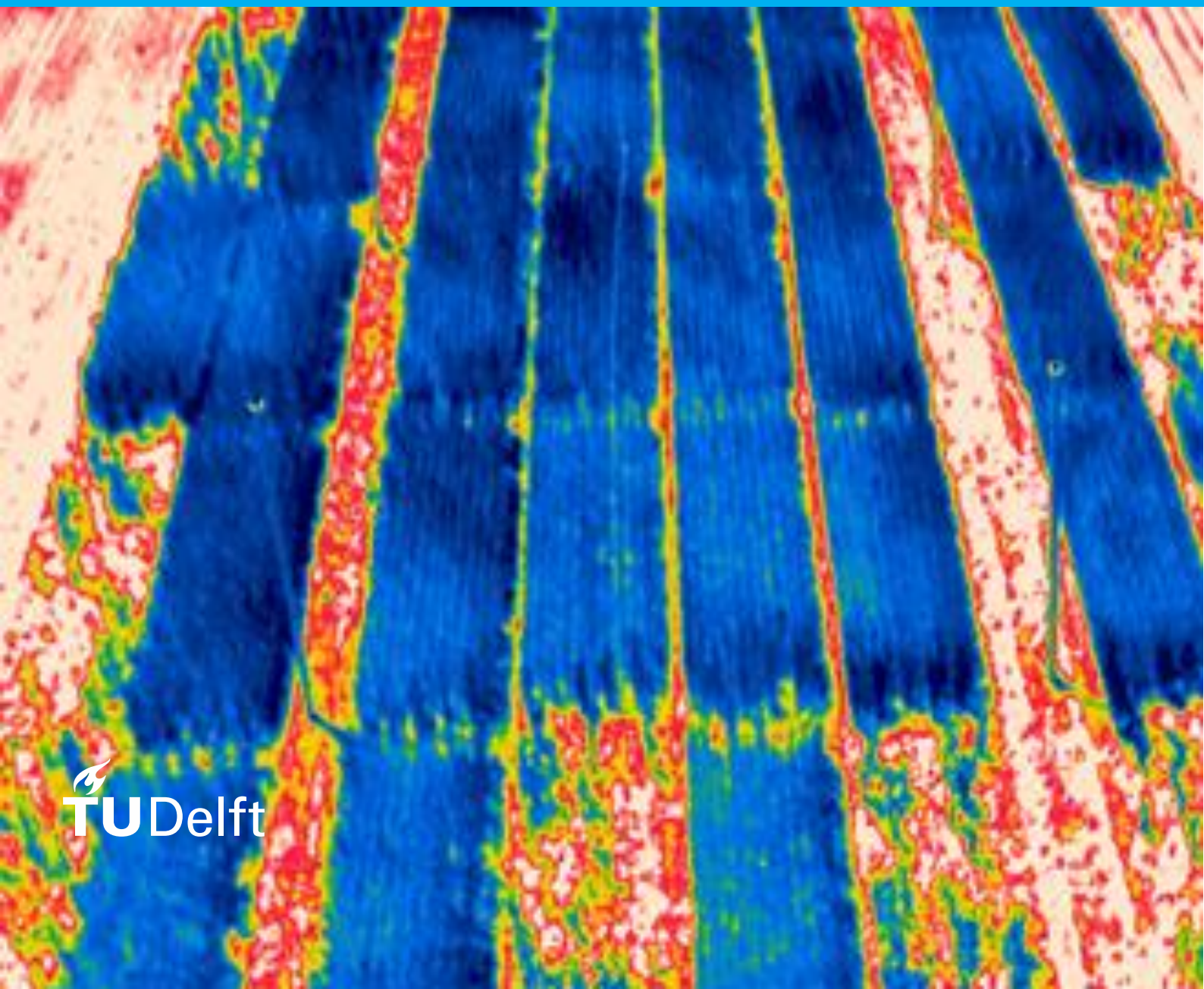
“Do the newly developed models of bare soil surface reflectance and thermal infrared image analysis improve the spatial estimation of soil water properties in the Noordoostpolder region, The Netherlands?”

Master thesis report

K. Groot

TU Delft
Geoscience and remote sensing



Do the newly developed models of bare soil surface reflectance and thermal infrared image analysis improve the spatial estimation of soil water properties in the Noordoostpolder region, The Netherlands?

Master Thesis report

By

Kevin Groot

in partial fulfilment of the requirements for the degree of

Master of Science
in Applied Physics

at the Delft University of Technology,
to be defended publicly on Friday July 5, 2019 at 10:00 AM.

Thesis committee:	Dr. S.L.M. Lhermitte,	Geoscience and Remote Sensing	TU Delft
	Dr. S.M. Alfieri,	Geoscience and Remote Sensing	TU Delft
	Prof. dr. ir. W.G.M. Bastiaanssen,	Water Management	TU Delft
	Ir. B. Rijk,	Company	Aurea Imaging

This thesis is confidential and cannot be made public until June 30, 2019.

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

Abstract

Methodology

Agriculture is an important sector to provide in our needs. Since 2000, the arable land per farmer increased with 66%. To help farmers managing their increasing amount of arable land, it is important to develop new techniques in example based on remote sensing data. One of the key parameters for agricultural management is the water availability in the area. The water availability can be clarified with help of two different soil water properties, soil water holding capacity and soil moisture content. The aim of this study is to estimate soil water properties at a scale of 30 meters based on bare soil surface reflectance and thermal infrared image analysis. In this study, two newly developed models are proposed to determine the spatial patterns of soil wetness. The first model is based on bare soil surface reflectance, named as SoilGrids30m, a spatial estimation model for clay content and organic matter content in the study area. These estimates are then used as input of pedotransfer functions to obtain spatial estimates of soil water holding capacity in the study area. The second model is based on the relation between the normalized difference vegetation index and the crop temperature, named as the soil wetness indicator. The soil wetness indicator is a representative of soil moisture content in the root zone of crops. Both models are evaluated against the soil moisture content estimates obtained from SEBAL. Soil water holding capacity is not directly related to soil moisture content. The hypothesis is therefore, soils with a high soil water holding capacity will tend to have higher soil moisture content estimates and vice versa during a long period of drought. The year 2018 has been used as reference year because of the extreme drought conditions in the months May until July.

Conclusions

The SoilGrids30m model improved the estimation of clay content and organic matter content in the study area compared to the existing coarser SoilGrids250m and SoilGrids1000m models. Especially, the organic matter content estimates significantly improves with help of the SoilGrids30m model. However, the influence of adding bare soil surface reflectance to the model was relatively low for both clay content (37%) and organic matter content (13%). The influence of bare soil surface reflectance only improved the clay content estimates compared to the SoilGrids1000m model. In all other cases, the target variable estimates did not improve by adding bare soil surface reflectance data. It can therefore be concluded that the improvement of the SoilGrids30m model is due to the use of a larger set of observation data from the study area and not because of the input of bare soil surface reflectance data.

The obtained soil water holding capacity estimates with help of pedotransfer functions did not show reliable results compared to the soil moisture estimates of SEBAL. The highly empirical pedotransfer functions are the main cause of the unreliable results but also seepage and irrigation are important factors that could have played a role. In future work it would therefore be recommended to avoid using pedotransfer functions calibrated in other regions than the study area.

The results of the soil wetness indicator showed a clear correlation with the soil moisture content estimates from SEBAL. The soil wetness indicator is therefore a simple and reliable tool to recognize spatial patterns of wetness. In example for precision agriculture, the soil wetness indicator could be used for irrigation management. The relative representation of the soil wetness indicator could determine the distribution of irrigation within a field.

Preface

This thesis serves to complete the study track MSc Geoscience and Remote Sensing in Civil Engineering.

First, I want to thank all committee members for their effort and time during the whole thesis process. In particular, my daily supervisor Silvia Alfieri who helped me throughout the whole thesis project with all my questions. Also, special thanks to Wim Bastiaanssen who introduced me to this thesis subject and directed me to the thesis as it is presented. I thank Stef Lhermitte for his critical view in the final stage of the thesis project.

I would also like to thank all colleagues at Aurea Imaging who took their time to help me anytime I asked. With special thanks to Bert Rijk for giving me the opportunity to be part of his company during the thesis project.

Last but not least, I would like to thank my parents, brothers and friends who supported and encouraged me during my thesis and all my study career.

K. Groot
Delft, June 2019

Contents

List of Figures

List of Tables

List of Abbreviations

ESA	European Space Agency
FAO	Food and Agriculture Organization of the United Nations
ISRIC	International Soil Reference and Information Centre
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LST	Land Surface Temperature
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
PCA	Principal Component Analysis
PTF	PedoTransfer Function
RCT	Relative Crop Temperature
SCORPAN	Soil classes or properties; Climate; Organisms, vegetation, fauna or human activity; Relief; Parent material; Age; N – spatial position
SEBAL	Surface Energy Balance Algorithm for Land
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Oceanic Salinity
TIR	Thermal-InfraRed
VNIR	Visible- and Near-InfraRed

1. Introduction

Agriculture is an important sector to provide in our needs. Since 2000, the amount of agriculture farmers in the Netherlands has dropped with more than 40% (CBS, 2018). While the total arable land has dropped with 6% (CBS, 2018). This results in an increase of arable land per farmer from 17.4 hectares to 28.9 hectares (+66%) (CBS, 2018). To help farmers managing their increasing amount of arable land, innovative developments have evolved quickly. The available data from satellites and the possibilities with drones could be useful tools to manage the increasing arable land per farmer.

At the moment, precision agriculture is facing a big problem because of the lack of good soil maps. Common soil maps, like Stiboka in the Netherlands, are too coarse for precision agriculture. Therefore, other techniques have been developed such as Veriscans and mole scans. These scans are labor-intensive and mainly based on empirical relations. Therefore, these techniques are not favorable for precision agriculture on larger scales. There have to be alternative ways, which should give a better representation of soil properties useful in precision agriculture.

One of the key parameters for agricultural management is the water availability in the area. The water availability can be clarified with help of two different soil water properties, soil water holding capacity and soil moisture content. Soil water holding capacity is defined as the maximum available water content in a soil for plant uptake (O'Geen, 2012). Both parameters are dependent on soil properties such as soil texture and organic matter content. It is therefore important to have “good” soil maps for a better understanding of the soil water properties in the area.

This study will use two newly developed models, a qualitative and a quantitative model, to determine the spatial patterns of soil wetness in the study area. The newly developed quantitative model will be named as SoilGrids30m. SoilGrids30m is inspired by the concepts of SoilGrids250m and SoilGrids1000m (Hengl et al., 2017; Hengl et al., 2014). The SoilGrids30m model uses bare soil surface reflectance to estimate physical soil properties. In this study, only clay content and organic matter content will be estimated. These physical soil properties will be translated into soil hydraulic properties with help of pedotransfer functions. The last step is to apply the water retention curve to obtain spatial estimates of soil water holding capacity.

The newly developed qualitative model will be named as the soil wetness indicator and is based on the concepts of the Trapezoid method (Yang et al., 2015). The soil wetness indicator uses thermal infrared satellite images, which are a reflection of root-zone soil properties. Crops react on a deficit of water content in the soil. Low water content will reduce the transpiration rate of a plant and causes the stomata to close. This process coincides with a temperature increase of the plant (Rutter et al., 1958; Xu et al., 2008; van den Bersselaar et al., 2005; Anderson et al., 2007). Numerous studies have shown that the temperature of the crops, obtained from thermal infrared data, compared with air temperature has a relationship with the soil moisture content (Bastiaanssen et al., 2006; Yang et al., 2015; Hatfield et al., 2008).

A third model, Surface Energy Balance Algorithm for Land, will be used to evaluate the two newly developed models. The Surface Energy Balance Algorithm for Land (SEBAL) uses the energy balance to estimate some aspects of the hydrological cycle such as soil moisture content (Bastiaanssen et al., 1995). Furthermore, SEBAL will be used to consistently pre-process all images used in this research.

The aim of this study is to estimate soil water properties at a scale of 30 meters based on bare soil surface reflectance and thermal infrared image analysis. With help of two newly developed models, the spatial patterns of soil wetness will be estimated. The output of SoilGrids30m model will be used to estimate the soil water holding capacity. Soil water holding capacity is not directly related to soil moisture content. The hypothesis is therefore, soils with a high soil water holding capacity will tend to have higher soil moisture content estimates and vice versa during a long period of drought. The year 2018 has been taken as reference year because of the extreme drought conditions in the months May until July. The soil wetness indicator is a qualitative measure of soil moisture conditions and is therefore directly related to soil moisture content. Both models will be evaluated with the soil moisture content estimates of SEBAL. SEBAL is a globally validated and applied model (Bastiaanssen et al., 2005) and will therefore be used as base for this study.

1.1. Problem statement

For this study, two parts of the electromagnetic spectrum will be used, visible and near infrared (VNIR) remote sensing and thermal infrared (TIR) remote sensing. One of the most common used techniques is VNIR remote sensing (Khanal et al., 2017), based on bare soil surface reflectance soil properties can be estimated. In contrary, TIR remote sensing is used in a lesser extent because of the limitation in high-resolution thermal images (Khanal et al., 2017). TIR remote sensing could be an interesting technique because the crop temperature is a direct reflection of the available water in the root zone. Both parts of the spectrum will be used to have an independent way of finding the spatial patterns of wetness in the study area.

Soil water holding capacity can be inferred from soil maps, which are usually not accurate enough to provide these data. The common used soil map in the Netherlands is Stiboka, which is available on a scale of 1:50.000. Farmers use additional techniques to get a more detailed map of their own fields such as Veriscans and mole scans. These methods are labor-intensive but gives, on a local-scale, information about soil moisture content and soil composition. The International Soil Reflectance and Information Centre (ISRIC) has introduced a new method to estimate several soil properties. ISRIC recently developed a global soil mapping model at a scale of 250 meters called SoilGrids250m. SoilGrids250m is based on 150.000 soil samples used for training and a stack of 158 remote sensing-based explanatory variables of soil (Hengl et al., 2017). The results give a good first estimate of soil properties at locations all over the world (Hengl et al., 2017). The freely accessible SoilGrids estimates, at a scale of 250 meters, are too coarse for precision agriculture. Therefore, a new model will be developed at a resolution of 30 meters based on Landsat 8 images. The soil water holding capacity can then be determined based on the obtained soil properties from the newly developed SoilGrids model.

Each crop field could be seen as a real time thermometer of the root zone soil properties. A deficit of water will cause an increase of the crop temperature (Rutter et al., 1958; Xu et al., 2008; van den Bersselaar et al., 2005; Anderson et al., 2007). VNIR remote sensing of the plant stress level is generally delayed because of the water storage in the plant itself. Due to this difference, TIR remote sensing is an interesting technique to directly measure the spatial variability of soil water properties.

1.2. Research objectives

The aim of this study is to estimate soil water properties at a scale of 30 meters based on bare soil surface reflectance and thermal infrared image analysis. This will be done with help of three different models. First, a newly developed soil mapping model will be introduced based on concepts of the SoilGrids250m and SoilGirds1000m models from ISRIC (Hengl et al., 2017; Hengl et al. 2014). Clay content and organic matter content will be estimated with help of remote sensing-based explanatory variables and soil samples. These soil properties are used as input for pedotransfer functions, which provides van Genuchten parameters from which the soil water holding capacity can be determined.

The second model is also a newly developed model named as the soil wetness indicator. Relative Crop Temperature (RCT) is set against the Normalized Difference Vegetation Index (NDVI). The RCT is crop temperature minus instantaneous air temperature. In this model, the theorem is that a high positive RCT suggest a deficit of water in the root zone of crops (Yang et al., 2015). The magnitude of high RCT changes with NDVI value. Higher NDVI values are areas with more vegetation cover and therefore areas with a higher demand of water. In these areas, the magnitude of high RCT is generally lower than in areas with lower NDVI values. Vice versa, low RCT suggest a proper amount of available water for plant uptake.

The third model is SEBAL, an existing model that will be used among other things to estimate soil moisture content. SEBAL estimates soil moisture content based on the energy balance. The soil moisture estimates of SEBAL will be used as reference to evaluate both newly developed models. All three models are a representation of the spatial variability of soil water properties in the study area. In this study, two different objectives can be pointed out. On the one hand the evaluation of a newly developed bare soil surface reflectance model along with the conversion to water holding capacity. On the other hand the qualitative determination of soil water properties based on thermal infrared image analysis.

The main research question for this study is formulated as follows:
"Do the newly developed models of bare soil surface reflectance and thermal infrared image analysis improve the spatial estimation of soil water properties in the Noordoostpolder region, The Netherlands?"

A number of sub-questions are formulated to help answering the research question:

- **Bare soil surface reflectance**
  - *What generic framework is recommended for the newly developed geostatistical SoilGrids30m model and under which conditions should the generic framework be applied?*
  - *How to obtain a reasonable set of explanatory variables to estimate physical soil properties?*
  - *To what extent can visible and near infrared bare soil surface reflectance provide information of physical soil properties (clay content and organic matter content)?*
  - *What is the performance of geostatistical model SoilGrids to estimate physical soil properties (clay content and organic matter content)?*
  - *Do the estimates of soil hydrological properties based on the SoilGrids model help to interpret the spatial patterns of observed water stress indicated by SEBAL?*

- ***Thermal infrared image analysis***
    - *Does the soil wetness indicator, based on LST and NDVI, help to interpret the spatial patterns of observed water stress indicated by SEBAL?*

1.3. Report structure

In chapter 2, background information is given about remote sensing of soil moisture. Furthermore, the SEBAL model is elaborately explained and previous work of the two newly developed models will be discussed. Chapter 3 provides information of the study area, characteristics of the used crop types and soil properties and a systematic explanation of the newly developed models. In chapter 4, the results of the newly developed models will be analyzed. The results will also be evaluated with the soil moisture estimates by SEBAL. The results given in chapter 4 will be discussed and recommendations will be given in chapter 5. The conclusions regarding to the overall project will be presented in chapter 6.

2. Background

2.1. Remote sensing of soil moisture

Soil moisture is a key parameter to monitor surface climatology, hydrology and ecology. It is used to monitor drought, predict floods, assist crop productivity, forecast weather and to link water, energy and carbon cycles (Brown et al., 2011). A variety of remote sensing techniques has been developed for soil moisture retrieval. These techniques are based on the characteristics of soil moisture content in different parts of the electromagnetic spectrum (Figure 2.1). Three different type of sensors will be discussed: radar, microwave and optical sensors. Radar is an active sensor; microwave and optical remote sensing are passive sensors.

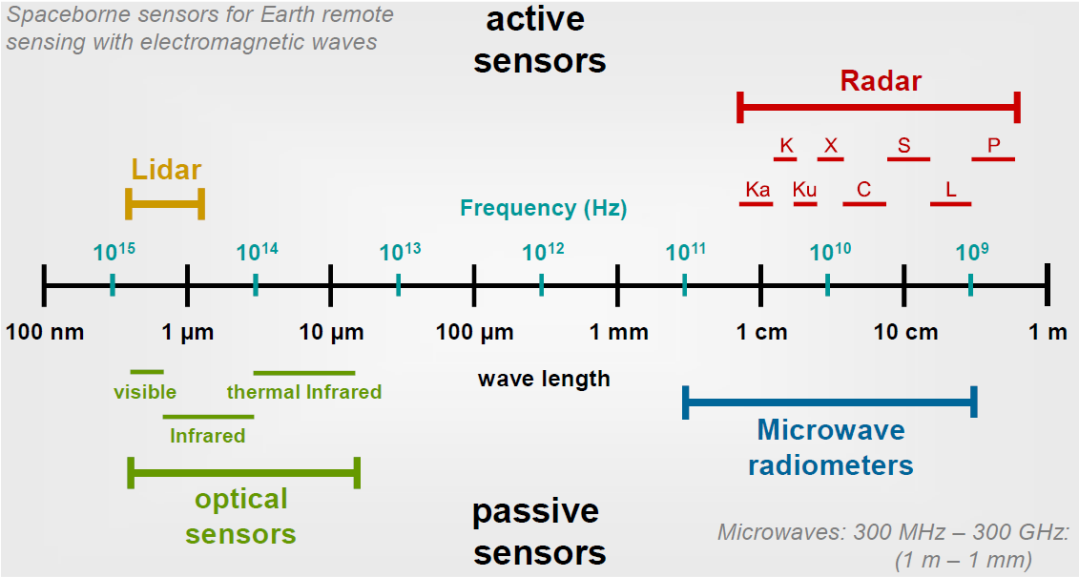


Figure 2.1 Remote sensing sensor types (Moreira, 2013)

2.1.1. Radar and microwave remote sensing

Radar and microwave remote sensing are widely used techniques mainly because of two reasons, it is independent on lighting conditions (day or night) and it can penetrate through clouds. Two main elements of radar and microwave remote sensing are the polarisation and frequency. Polarisation depends on the wavelength; the physical characteristics of the antenna; the reflecting material. The wavelength and the physical characteristics of the antenna are both known, therefore differences in physical characteristics of the reflected material can be assigned to the soil moisture conditions of the Earth's surface. Special designed satellite sensors to monitor soil moisture conditions with radar and microwave remote sensing are SMAP from NASA and SMOS from ESA.

Passive microwave sensors

A passive microwave satellite sensor measures radiated and reflected energy of the Earth. The measured energy can be, emitted by the atmosphere; reflected by the Earth's surface; emitted by the Earth's surface; transmitted from the subsurface. The radiated energy at microwave lengths is a function of the dielectric constant of the Earth's surface at that specific place. The dielectric constant depends on the soil-water configuration and thus on the soil moisture content (Wagner et al., 2007).

Active radar sensors

An active radar satellite sensor transmits pulses of microwave radiation. The intensity of the backscattered signal depends on the geometry and the dielectric properties of the Earth's surface at each specific location. The soil moisture content is determined by finding a relationship between the backscatter coefficient and the dielectric constant. Such a model has a high sensitivity to the geometry features of the Earth's surface, which makes it hard to find a correct relationship (Wagner et al., 2007).

2.1.2. Optical remote sensing

Radiated energy by the Sun is reflected by the Earth's surface, the reflected energy can be measured with optical sensors. An optical sensor is a passive sensor and thus sensitive to lighting and cloudy conditions. Optical sensors acquire images at different spectral bands at the same time (multispectral images). Optical remote sensing consists of visible and near infrared remote sensing and thermal infrared remote sensing.

Visible and near infrared remote sensing

Based on multispectral images it is possible to determine spectral characteristics of the Earth's surface. Each feature on the Earth's surface has its own spectral signature. The spectral signature depends on the chemical and physical properties of an object. Typical spectral signatures of different vegetation types are shown in Figure 2.2. The strong contrast between reflectance in the visible and near infrared part is for example an indicator for the greenness of an area or also known as the normalized difference vegetation index (NDVI). The greenness of an area could be an indicator of drought conditions. As shown in Figure 2.2, the spectral signature of dry-yellow grass is clearly different from the spectral signature of normal green grass. The NDVI rate for normal green grass would therefore be significant larger than the NDVI rate from dry-yellow grass. Among with the NDVI many different indices exists that are all based on spectral signatures.

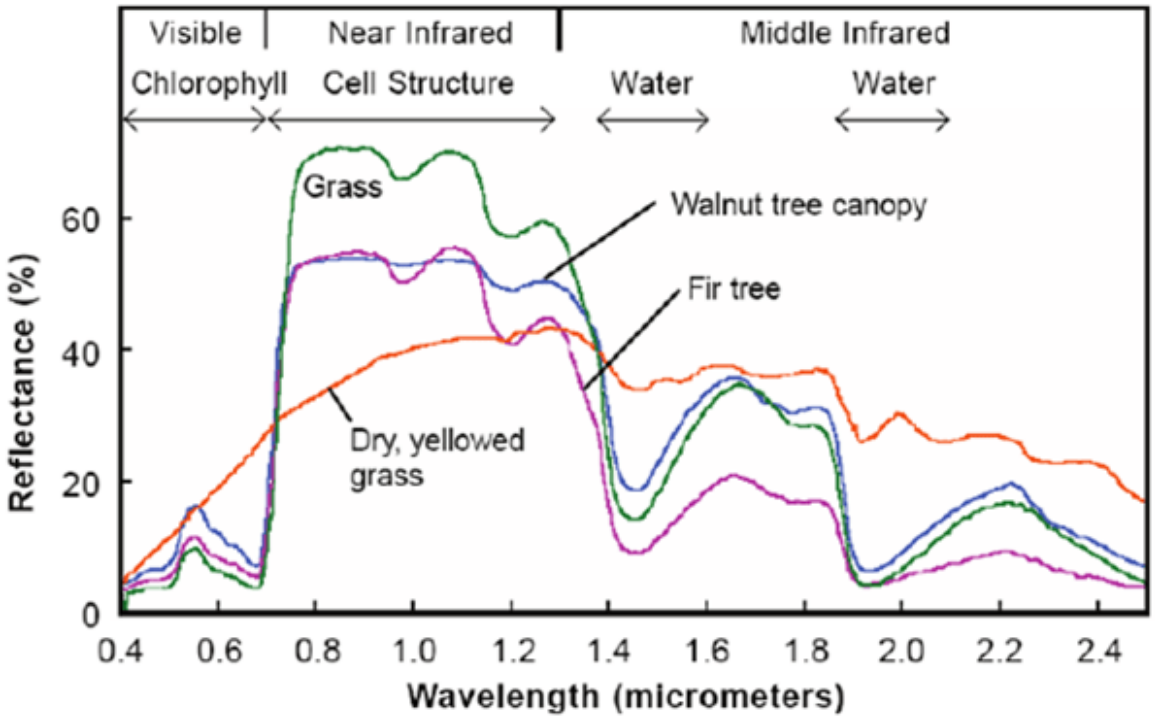


Figure 2.2 Spectral signatures for different types of vegetation ("Vegetation Spectral Signature Cheat Sheet", 2017)

Thermal infrared remote sensing

All objects with a temperature greater than zero Kelvin emits thermal radiation. This is measured in the thermal infrared part of the electromagnetic spectrum. Several algorithms are developed to convert thermal radiation into land surface temperatures. Land surface temperatures of vegetated areas depends on the amount of available water for plant uptake. A deficit of water could lead to stress in the plant that coincide in an increase of the temperature of the plant, this process has a clear relationship with the soil moisture content in the area as shown in numerous studies (Bastiaanssen et al., 2006; Yang et al., 2015; Hatfield et al., 2008).

2.2. The Surface Energy Balance Algorithm for Land model: SEBAL

The Surface Energy Balance Algorithm for Land model (SEBAL) uses the energy balance to estimate some aspects of the hydrological cycle (Bastiaanssen et al., 1995). The net energy driving the hydrological cycle is the incoming energy minus the energy heating the soil/air and the energy reflected back into space. SEBAL requires visible, near-infrared and thermal-infrared remote sensing data to estimate land surface characteristics. In addition, SEBAL requires meteorological data (air temperature, humidity, wind speed, solar radiation) and soil physical data (saturated soil moisture content, field capacity, wilting point). Important outputs from SEBAL are evapotranspiration, biomass growth, water deficit and soil moisture content. SEBAL has been applied and validated for different water management related purposes (Bastiaanssen et al., 2005) and can be applied at local scale (plot level) as well as at global scale.

2.2.1. Introduction to SEBAL

SEBAL is a sophisticated energy balance model, which calculates the energy exchanges between land and atmosphere. On each individual pixel, SEBAL computes a complete radiation and energy balance along with resistances for momentum, heat and water vapour transport. The resistances are a function of state conditions such as soil water potential, which is a measure for soil moisture content, wind speed and air temperature change. The SEBAL model is comprised of 25 computational steps to calculate hydrological aspects such as evapotranspiration, biomass growth, water deficit and soil moisture (Allen et al., 2002).

The principal steps of the SEBAL model to derive evapotranspiration values are shown in Figure 2.3. Evapotranspiration can be associated with the latent heat flux, which is part of the surface-energy balance. The surface-energy balance is shown in equation 2.1, where R_n (Wm^2) is the net radiation; G_0 (Wm^2) is the soil heat flux; H (Wm^2) is the sensible heat flux; λE (Wm^2) is the latent heat flux associated with evapotranspiration.

$$R_n = G_0 + H + \lambda E \quad [Wm^2] \qquad 2.1$$

SEBAL converts satellite radiances into land surface characteristics (surface temperature, vegetation index, surface albedo and leaf area index). The net radiation and soil heat flux are computed with simple conversions of these land surface characteristics and meteorological data. The sensible heat flux is calculated with a so-called “self-calibration” procedure. “Hot” and “cold” pixels are selected to set the boundary conditions for the energy balance. The “cold” pixels are full vegetated and well-irrigated crop surfaces. The land surface temperature and near-surface air temperature are assumed to be similar for “cold” pixels. The “hot” pixels are dry and bare soil surfaces where the evapotranspiration is assumed zero (Allen et al., 2002). SEBAL calculates the net radiation, soil heat flux and sensible heat flux to compute the latent heat flux. This procedure will be explained in more detail along with the basic formulas in

chapter 2.2.4. All formulas shown in paragraph 2.2 are obtained from the advanced training and users manual of SEBAL (Allen et al., 2002) unless stated otherwise.

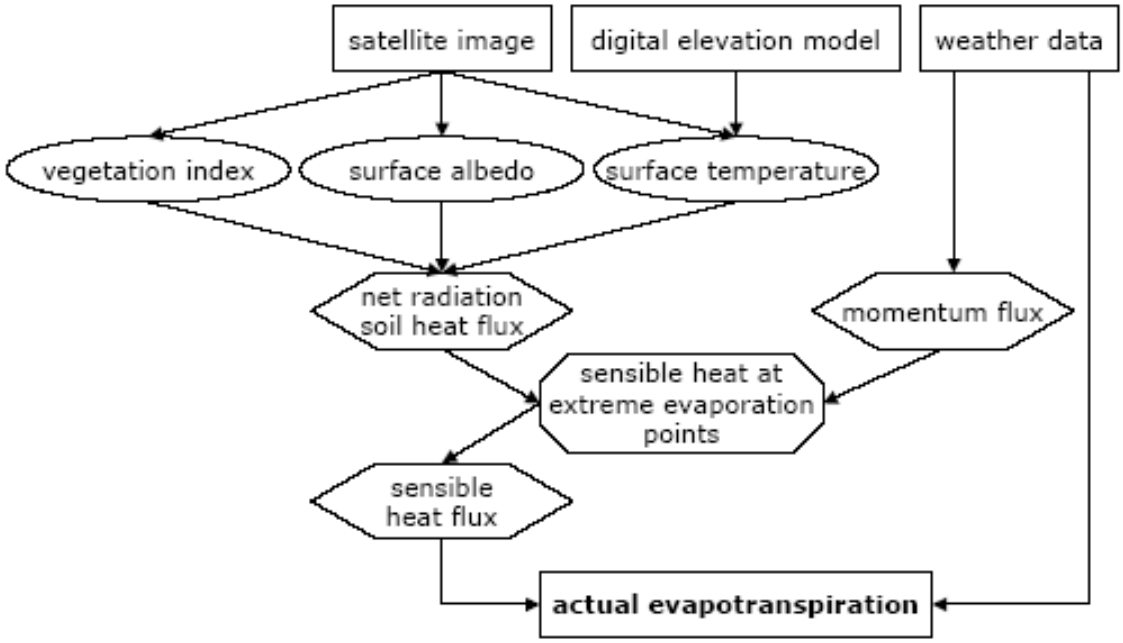


Figure 2.3 Schematic view of energy balance and evapotranspiration computations with SEBAL ("SEBAL a scientific description", n.d.)

2.2.2. Method of application of SEBAL in current research

The SEBAL model is used for three reasons: 1) to consistently pre-process all images used in this research; 2) to obtain estimates of the Normalize Differenced Vegetation Index (NDVI) and the Land Surface Temperature (LST) from the area of interest; 3) to obtain estimates of the soil moisture content from the area of interest. For evaluations between different models, it is important to pre-process all data in the same way; otherwise the differences found in the comparisons could also be due to differences in pre-processing the data. For this reason, all images used in this research are consistently pre-processed with SEBAL. The second application of SEBAL is to obtain NDVI and LST data that are used for the soil wetness indicator. The third application of SEBAL are the soil moisture estimates, which are used for calibrating the soil mapping model.

In this research, the latest version of pySEBAL (v3.4) is used. The pySEBAL code is continuously under development and changes from time to time. It is therefore important to use an up to date version of pySEBAL. The input of the pySEBAL code is an Excel file, which contains the folder path of the input files (digital elevation model and satellite images), folder path of the output files, meteorological parameters and soil parameters. The digital elevation model is used as the geospatial reference to clip all satellite images. The pySEBAL code generates georeferenced rasterized tiff-files for every computed step in the SEBAL process and stores the tiff-files in the specified output folder.

2.2.3. Image pre-processing SEBAL

This research uses level-1 Landsat 8 images for all of the analysis. Level-1 products of Landsat 8 images are radiometrically and terrain-corrected data products (USGS, 2016). These images are supplied in digital numbers which have to be converted to top of atmospheric spectral radiances first and then to top of atmospheric reflectance's. SEBAL uses a slightly different way to convert digital numbers into pixel reflectance data (Allen et al., 2002) than suggested in the USGS manual (USGS, 2016). The spectral radiance (L_λ) is computed as follows:

$$L_\lambda = \frac{L_{max} - L_{min}}{QCAL_{max} - QCAL_{min}} * (DN - QCAL_{min}) + L_{min} \quad [W/m^2/sr/\mu m] \qquad 2.2$$

Where, DN is the degree of greyness of each pixel; L_{max} and L_{min} are calibration constants of the sensor; $QCAL_{max}$ and $QCAL_{min}$ are the highest and lowest range of values for rescaled radiance in DN. All of these parameters are band specific and can be found in the product's metadata file.

The top of atmospheric reflectivity of a surface pixel is defined as the ratio of the reflected radiation flux to the incident radiation flux. It is computed using the following equation:

$$\rho_\lambda = \frac{\pi * L_\lambda}{ESUN_\lambda * cos\theta * d_r} \quad [-] \qquad 2.3$$

Where, L_λ is the spectral radiance for each band computed with equation 2.2; $ESUN_\lambda$ is the mean solar exo-atmospheric irradiance for each band ($W/m^2/\mu m$), see Table 2.1; $cos\theta$ is the cosine of the solar incidence angle from nadir; d_r is the inverse squared relative Earth-Sun distance.

Band	ESUN [W/m²/µm]
2	1973.28
3	1852.68
4	1565.17
5	963.69
6	245
7	82.106

Table 2.1 Mean solar exo-atmospheric spectral irradiances Landsat 8

2.2.4. NDVI, LST and soil moisture retrieval in SEBAL

Three outputs generated with SEBAL are key to this research: the NDVI, LST and soil moisture content. Within the multispectral remote sensing data, the NDVI is one of the most well-known and applied vegetation index. The NDVI is a simple and fast method to identify vegetated areas and their condition. The LST is used as thermometer of crops and is therefore a representative of the water availability in the root zone of crops. The soil moisture content defines the total water content in a soil and is therefore a useful indicator of drought conditions (Yang et al., 2015; Sreelash et al., 2017).

NDVI

The NDVI can be computed with the reflectivity values found in equation 2.4 and is the ratio of the differences in reflectivity for the near-infrared band (ρ_5) and the red band (ρ_4) to their sum.

$$NDVI = \frac{\rho_5 - \rho_4}{\rho_5 + \rho_4} \quad [-] \qquad 2.4$$

Values of NDVI ranges between -1 and +1, where negative values are related to waterbodies, snow, ice or clouds and positive values indicate vegetation cover. Dense vegetation, such as forests, have a NDVI of in between 0.6 and 0.9; shrubs or agriculture have a NDVI of in between 0.2 and 0.5; bare soils are indicated with a NDVI lower than 0.2 (Carlson et al., 1997).

LST

In this research the land surface temperature is estimated with help of a split window algorithm. In 1975, McMillin was the first who proposed to use a split window algorithm to accurately retrieve sea surface temperatures (McMillin, 1975). The algorithm takes advantage of the availability of two thermal infrared bands, which enables the atmospheric correction. The differing amounts of absorption occurring in the different thermal bands are used to estimate atmospheric effects (Sobrino et al., 1996). SEBAL computes the land surface temperature slightly different than proposed by Sobrino (Sobrino et al., 1996), see equation 2.5.

$$T_s = T_{B,10} + C_1\left(T_{B,10} - T_{B,11}\right) + C_2\left(T_{B,10} - T_{B,11}\right)^2 + C_0 + (C_3 + C_4 W)(1 - \varepsilon_{B,10}) \quad [K] \qquad 2.5$$

Where, $T_{B,10}$ and $T_{B,11}$ are the top of atmosphere brightness temperatures of band 10 and band 11 from Landsat 8 calculated according to equation 2.6; $\varepsilon_{B,10}$ is the land surface emissivity of band 10 from Landsat 8 calculated according to equation 2.7/2.8; W is the atmospheric water vapor pressure calculated according to equation 2.9; C_0-C_4 are split-window coefficient values according to Table 2.2.

Constant	Value
C_0	-0.268
C_1	1.378
C_2	0.182
C_3	54.3
C_4	-2.238

Table 2.2 Split-window coefficients (Jiménez-Muñoz et al., 2014)

The top of atmosphere brightness temperature is calculated in the same manner as proposed by USGS (2016). For both bands, 10 and 11, the top of atmosphere brightness temperature can be calculated as follows:

$$T_B = \frac{K_2}{LN\left(\frac{K_1}{L_\lambda} + 1\right)} \quad [K] \qquad 2.6$$

Where, K1 and K2 are thermal conversion constants that can be found in the product's metadata file; L_λ is the spectral radiance calculated according to equation 2.2.

The land surface emissivity is the ratio of thermal energy emitted from a surface to that of a black body with the same temperature. Assuming the Earth without vegetation cover is the black body, the Earth's emissivity will change according to the amount of vegetation cover per area. The amount of vegetation cover per area is indicated as the leaf area index (LAI). If the NDVI is greater than zero, two situations are distinguished to calculate the land surface emissivity (Allen et al., 2002). Equation 2.7 for $LAI < 3$ and equation 2.8 for $LAI \geq 3$.

$$\varepsilon = 0.95 + 0.01 * LAI \quad [-] \qquad 2.7$$

$$\varepsilon = 0.98 \quad [-] \qquad 2.8$$

The atmospheric water vapor pressure is computed with help of meteorological data and is based on the Penman-Monteith method (Allen et al., 1998). There are several ways to compute the atmospheric water vapor pressure according to the Food and Agriculture Organization of the United Nations (FAO). SEBAL does this procedure with help of the relative humidity, see equation 2.9.

$$W = \frac{RH_{inst}}{100} * e_{sat,inst} \quad [kPa] \qquad 2.9$$

Where. RH_{inst} is the relative humidity at satellite overpass time [%]; $e_{sat,inst}$ is the saturated water vapor pressure calculated according to equation 2.10.

$$e_{sat,inst} = 0.6108 * \exp\left(\frac{17.27 * T_{inst}}{T_{inst} + 237.3}\right) \quad [kPa] \qquad 2.10$$

Where, T_{inst} is the air temperature at satellite overpass [K].

Soil moisture content

SEBAL computes the soil moisture content as a function of the evaporative fraction (Allen et al., 2002). Bastiaanssen et al. (1997) was the first who found a relationship between soil moisture and evaporative fraction, which resulted in equation 2.11.

$$\theta = \theta_{sat} * \exp\left(\frac{\Lambda - a}{b}\right) \quad [cm^3/cm^3] \qquad 2.11$$

Where, θ_{sat} is the saturated soil moisture content fixed to 0.45 $[cm^3/cm^3]$ in this research; Λ is the evaporative fraction according to equation 2.12; a and b are curve-fitting parameters respectively 1.0 and 0.421 (Scott et al., 2003).

$$\Lambda = \frac{R_n - G_0 - H}{R_n * G_0} \quad [-] \qquad 2.12$$

Where, R_n is the net radiation at the surface $[W/m^2]$ according to equation 2.13; G_0 is the soil heat flux $[W/m^2]$ according to equation 2.14; H is the sensible heat flux $[W/m^2]$ according to equation 2.15.

The net radiation is the first computational step in the SEBAL procedure. The net radiation uses the surface radiation balance equation according to equation 2.13.

$$R_n = (1 - \alpha)R_{s,in} + R_{l,in} - R_{l,out} - (1 - \varepsilon)R_{l,in} \quad [W/m^2] \qquad 2.13$$

Where, α is the surface albedo; $R_{s,in}$ is the incoming shortwave radiation calculated based on meteorological data; $R_{l,in}$ and $R_{l,out}$ are the incoming and outgoing longwave radiation and are computed with the Stefan-Boltzmann equation; ε is the surface emissivity calculated with equation 2.7/2.8.

The soil heat flux is the rate of heat storage into the soil and vegetation due to conduction (Allen et al., 2002). It is a difficult term to determine because of the high dependency on soil type and land classification. Bastiaanssen (2000) has developed an empirical equation representing values near midday, see equation 2.14.

$$\frac{G}{R_n} = \frac{T_s}{\alpha} * (0.0038\alpha + 0.0074\alpha^2)(1 - 0.98 * NDVI^4) \quad [-] \qquad 2.14$$

Where, T_s is the surface temperature in ℃; α is the surface albedo; NDVI is the normalized differenced vegetation index calculated according equation 2.4. The soil heat flux (G) is then calculated by multiplying equation 2.14 with 2.13.

The sensible heat flux is the rate of heat loss to the air by convection and conduction due to temperature differences (Allen et al., 2002). The computation of the sensible heat flux is quite elaborate, as explained in the introduction of SEBAL it works with a self-calibration procedure to determine the boundary conditions of the energy balance. The sensible heat flux is computed as follows:

$$H = \frac{\rho * c_p * dT}{r_{ah}} \quad [W/m^2] \qquad 2.15$$

Where, ρ is the air density $[kg/m^3]$; c_p is the air specific heat and is fixed to 1004 $[\mathrm{J/kg/K}]$; dT is the temperature differences in Kelvin; r_{ah} is the aerodynamic resistance to heat transport $[s/m]$.

The self-calibration procedure takes place between the estimation of the temperature rate, the aerodynamic resistance and the sensible heat flux itself. The model starts with initial estimates of the temperature rate based on the selected “hot” and “cold” pixels and the aerodynamic resistance based on meteorological data. From there the iterative process starts, first the temperature rate is computed based on a linear relation between the temperature and temperature rate; secondly an estimate of the sensible heat flux is computed; lastly the aerodynamic resistance is corrected with help of the Monin-Obukhov theory. This iterative procedure takes place until the values for the temperature rate and the aerodynamic resistance stabilizes at the “hot” pixel. The final values for the temperature rate and the aerodynamic resistance are used to determine the sensible heat flux for each pixel. Further details of the procedure can be found in the SEBAL Advanced Training and Users Manual (Allen et al., 2002).

2.3. Previous work newly developed models

2.3.1. SoilGrids30m

The SoilGrids30m model is based on a widely used spatial estimation technique regression-kriging. In 1951, a South-African mining engineer, Krige, proposed the fundamentals of the ordinary-kriging technique. The ordinary-kriging technique initially has been developed to predict the most likely distribution of gold in a mine based on samples from a few boreholes (Krige, 1951). In 1963, a French mathematician, Matheron, formulated the theoretical fundamentals of the ordinary-kriging technique (Matheron, 1963). Burgess and Webster (1980) were the first who applied ordinary-kriging for soil study purposes and encouraged soil scientist to use the ordinary-kriging technique as a spatial estimation method. Since then, the ordinary-kriging technique has been widely used in various forms of soil sciences (McBratney et al., 2000). In 1969, Matheron introduced a new kriging technique called universal-kriging. Universal-kriging is a hybrid model that combines a simple or multi-linear regression model

with ordinary-kriging. It computes the trend along with residuals simultaneously and gives a combined kriging variance. In 1994, Odeha et al. suggested to extract the trend from the residuals, krige the de-trended residuals and sum them afterwards, this technique is known as regression-kriging (Odeha et al., 1994). The separation between the two models allows regression-kriging to use arbitrarily complex regression methods (Hengl et al., 2007).

2.3.2. Soil wetness indicator

Back in 1975, Idso et al. (1975) already found a distinct relation between the surface radiant temperature of bare soil and soil wetness. Vegetated areas are a lot more complex to relate to soil wetness. Not only because of the irregular geometry and spatial distribution of vegetation but also because leaf temperatures tends to remain close to air temperature (Idso et al., 1975). Nemani and Running (1989) were the first who demonstrated a physical relationship between NDVI and surface radiant temperature. They found a marked divergence between a dry and a moist day in the NDVI-LST plot. This difference could be assigned to a change in soil moisture content. In 1990, Price proposed a method to infer regional scale evapotranspiration by relating variations of satellite-derived surface temperature to a vegetation index (Price, 1990). Gillies and Carlson (1995) used this concept for the estimation of regional patterns of surface moisture availability and fractional vegetation in the presence of spatially vegetation cover. In 1997, this method for the first time was denoted as the triangle method (Gillies et al., 1997). The triangle method is based on the physical relationship between the land surface temperature and apparent vegetation cover, i.e. the NDVI. The boundaries of the pixel envelope are derived based on the scatter plot feature space between the inverse relationship of the land surface temperature and the NDVI (Gillies et al, 1997). Figure 2.4 shows an example of the triangle method used to validate the method by Gillies et al. (1997). The warm edge correlates with lower soil moisture content values and the cold edge correlates with higher soil moisture content values.

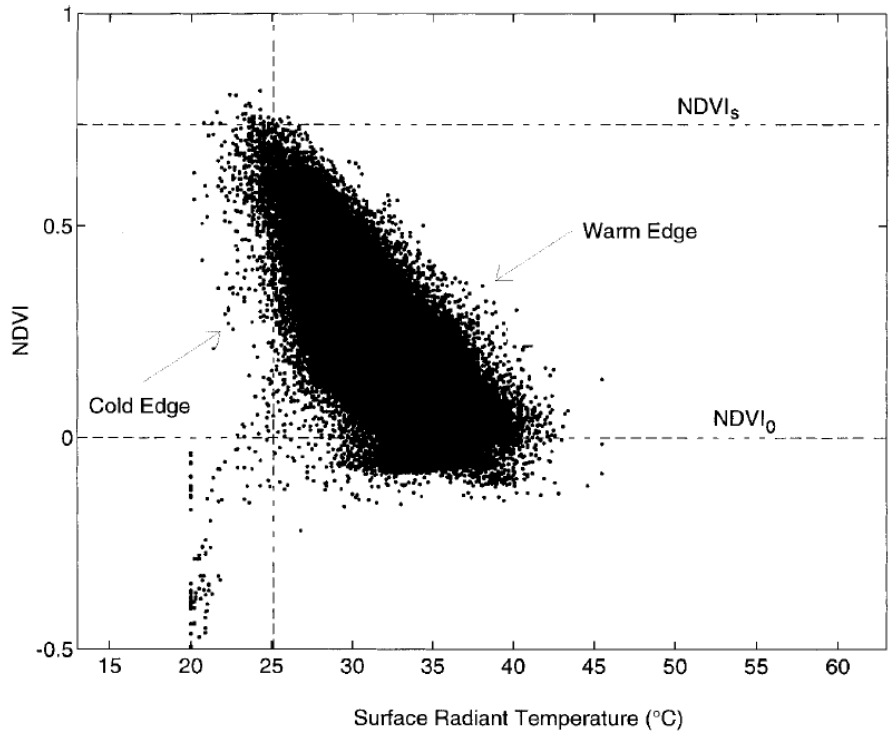


Figure 2.4 Example of scatter plot of NDVI versus surface radiant temperature (Gillies et al., 1997)

However, the triangle method also has its limitations. According to Moran et al. (1994), the distinction between well-watered vegetation and water-stressed vegetation is not represented in the triangular space. Both situation appear to have the same surface temperature in the triangular space. Moran et al. (1994) proposed a trapezoidal space instead of a triangular space. Where the temperature of full-vegetated areas represents a range between wet and dry areas, see Figure 2.5. In this study, the trapezoidal space will be used as the basis for a qualitative analysis of the soil moisture content in the area denoted as the soil wetness indicator.

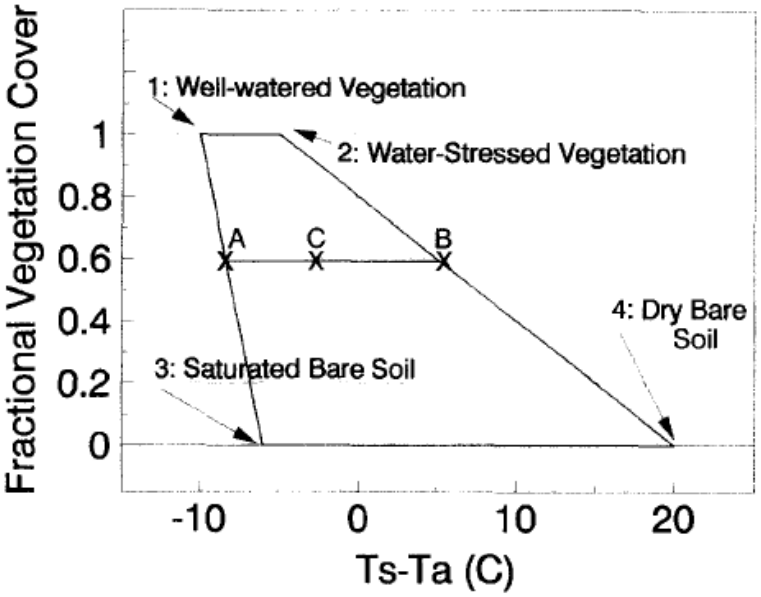


Figure 2.5 The hypothetical trapezoidal space (Moran et al., 1994)

3. Methods and materials

3.1. Study area

The study area is located in the central-northern part of the Netherlands called the “Noordoostpolder”, see Figure 3.1. It has been created in waters of the “Zuiderzee” and is artificially drained since 1942. The old seabed is generally full of minerals with good soil structures for agricultural purposes. Therefore, the area has a high density of agricultural fields/farmers. The area is chosen because of two reasons, the high density of agricultural fields and the variability in soil types.

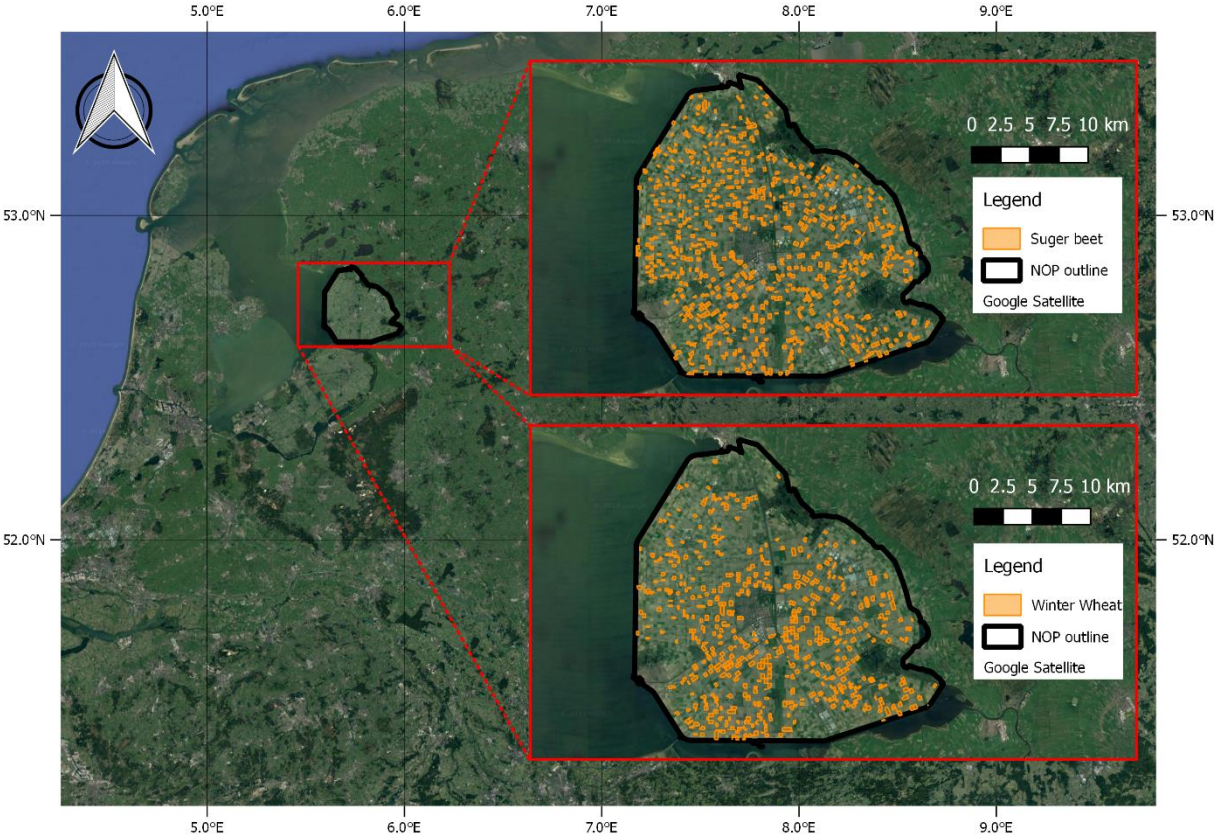


Figure 3.1 Geographic location study area and crop types in 2018

The study area consists of approximately 85% of fields for agricultural purposes. Approximately half of it is used for crops in open air such as sugar beet, onion, potato, maize, wheat, flowers, etc. The other half is used for cattle and greenhouses ("Landbouw; gewassen, dieren en grondgebruik naar gemeente", 2018). In this study, the focus will be on sugar beet and winter wheat of which the geographical distribution of the fields in the study area for 2018 are shown in Figure 3.1.

The variability in soil types in the study area is important in finding relationships between soil types and soil moisture content. The distributed soil map by the Dutch government contains a wide range of different soil types ("Dataset: Basisregistratie Ondergrond (BRO)", 2018). Based on De Vries et al. (2003), the wide range of soil types are rearranged into eight soil classes, shown in Figure 3.2. The area mainly consists of clayey soils along with calcareous extremely low clayey soils and calcareous sandy soils. For this study, it is important to have a wide range of clay content. According to Figure 3.2, there is enough variability in the area to meet this requirement.

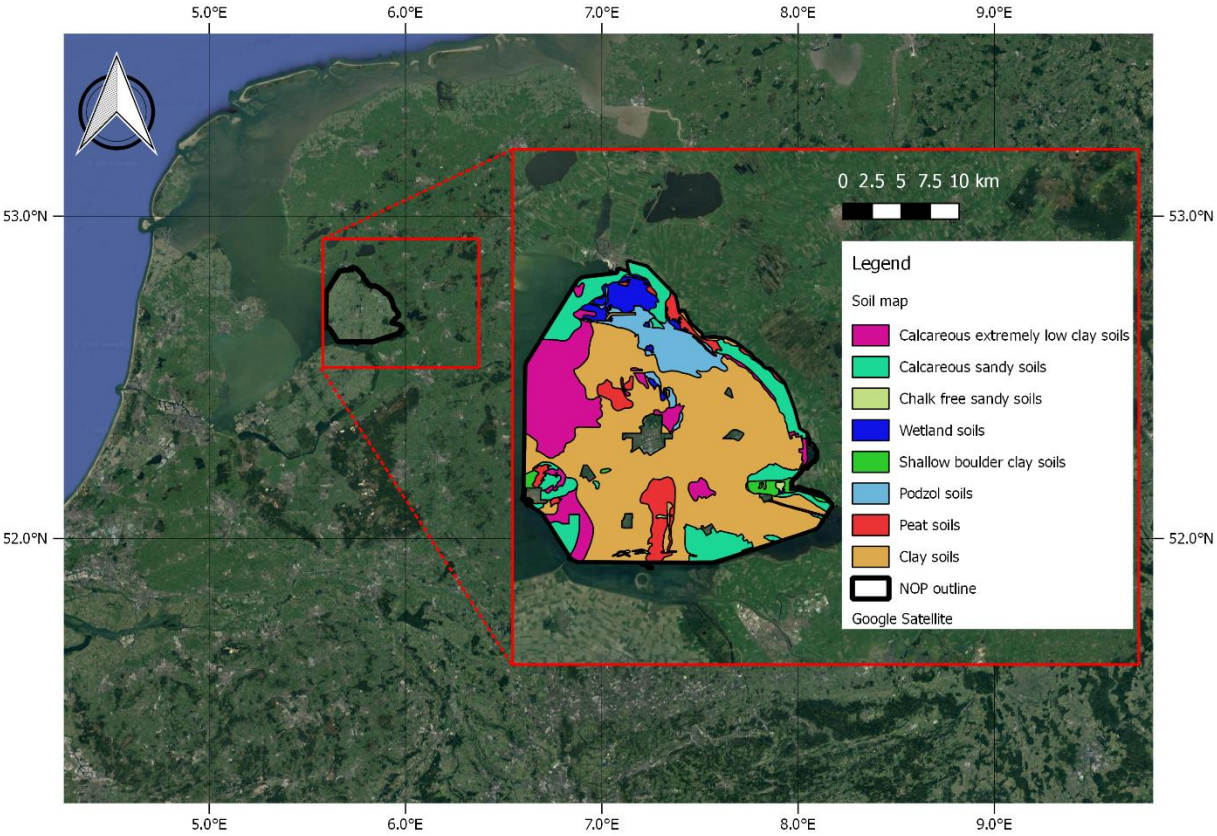


Figure 3.2 Geographic location of soil classes in study area

Climate

The climate of the Noordoostpolder region is a maritime climate with moderately warm summers and cool winters, and typically high humidity. In Figure 3.3, the monthly weather data of 2018 for the single KNMI station in the Noordoostpolder region is shown. The KNMI station is located in Marknesse, central-east in the Noordoostpolder region. The year 2018 has been used as reference year because of the extreme dry weather conditions. As shown in Figure 3.3, the months April to July all have higher evapotranspiration rates than precipitation rates. This results in a decrease of the groundwater storage with as consequence a decrease of the groundwater level. Crops react on a deficit of water content in the soil. Low water content will reduce the transpiration rate of the plant and causes the stomata to close. This process coincides with a temperature increase of the plant (Rutter et al., 1958; Xu et al., 2008; van den Bersselaar et al., 2005; Anderson et al., 2007). For this study, it is therefore important to measure the crop temperatures in dry conditions to see significant differences between different soil classes and crop types.

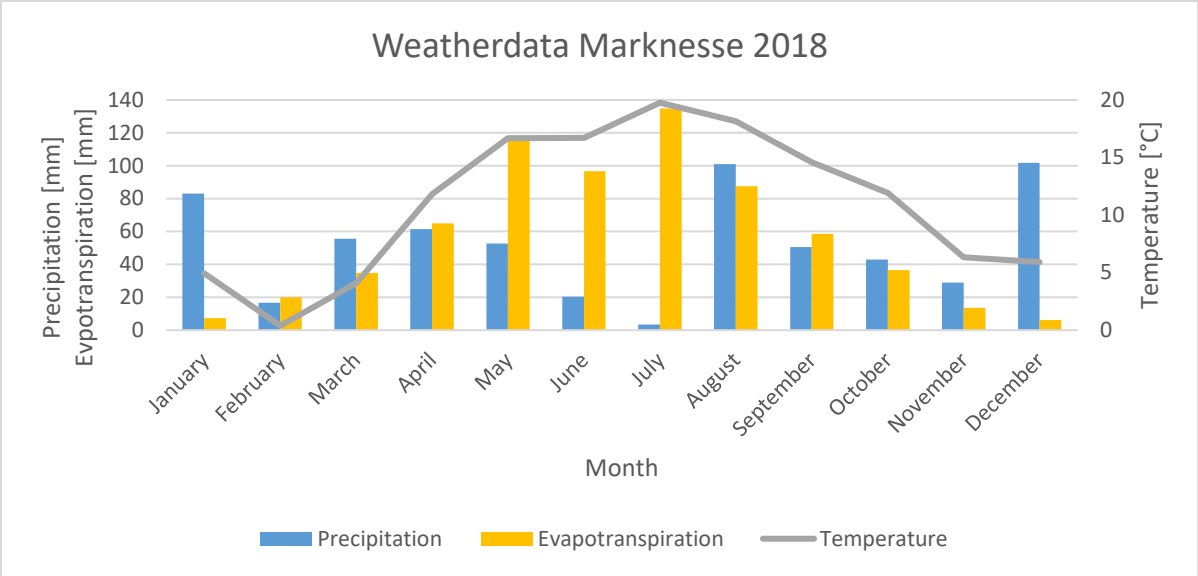


Figure 3.3 Monthly precipitation, evapotranspiration and mean temperature KNMI station Marknesse 2018

3.2. Crop characteristics

Sugar beet

In the Netherlands, sugar beet is the main resource for the production of sugar (FAO – Sugar beet, n.d.). Sugar beet originally has a life cycle of two years but for the production of sugar, the crops are harvested within one year. In the second development stage high concentrations of sugar is created in the leaves of the plant that is mainly used for the growth process (vegetative state, see Figure 3.4). Later on in the growth process, the sugar concentration is mainly stored in the roots of the plant. The sugar yield depends both on root size and sugar concentration, which mainly depends on climate, water supply and nitrogen level in the soil. In general, the sugar percentage in the roots lies between 15 and 20 percent of the root its weight (FAO – Sugar beet, n.d.). Sowing of the seeds occurs in spring (March-April) and normally needs a growing period of 140 to 160 up to 200 days. Harvesting commonly takes place in autumn (September-October).

In this study, it is favourable to measure the crop temperatures in dry periods along with a water deficit. In the vegetative and yield development stage, water deficits could lead to lower sugar yields. While, a surplus of water in the ripening stage could lead to a decrease of the sugar yield (FAO – Sugar beet, n.d.). As explained and shown in paragraph 2.1, in April to July the water supply balance was negative. During 2018, the vegetative and yield development stages coincide with relatively dry periods. Therefore, it should be taken into account that farmers possibly irrigated their crops during these development stages.

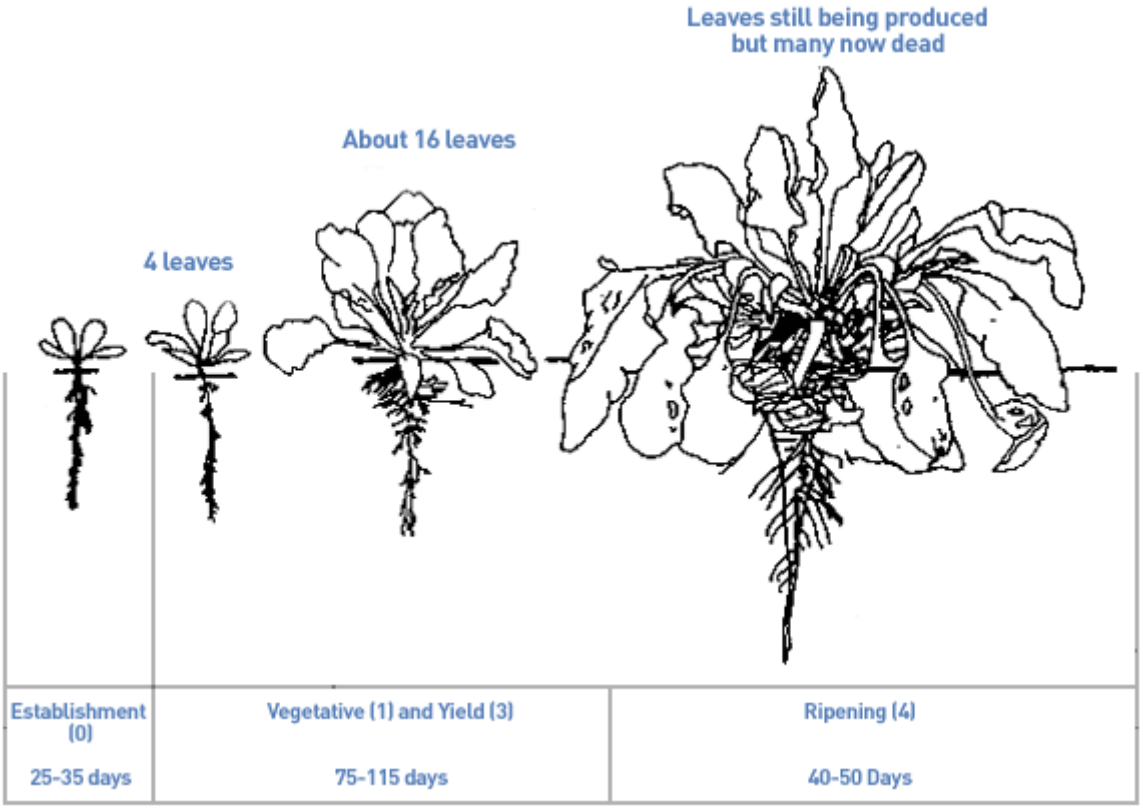


Figure 3.4 Duration of different growth periods sugar beet (FAO – Sugar beet, n.d.)

Winter wheat

In the Netherlands, winter wheat is by far the most cultivated type of wheat (FAO – Wheat, n.d.). Winter wheat seeds are sowed from October to half December. It requires a period of frost in the early development stages (Dormancy period, see Figure 3.5) in favor of the growth process. When the head development stage starts the strong resistance against frost is lost. The winter wheat yield depends on the number of heads per plant, the number of grains per head and the size of heads, which mainly depends on climate and water supply (FAO – Wheat, n.d.). The total growing period of winter wheat ranges from 180 to 250 days. Harvesting commonly takes place in the summer (end of July – August).

As mentioned with the sugar beet, it is favourable to measure the crop temperatures in dry periods along with a water deficit. In the establishment and tillering development stage it is important to have sufficient water supply in favour of the winter wheat yields. In the next development stages, dormancy and head development, slight water deficits may have little effect on production yields. The flowering period is most sensitive to water deficit, it will reduce the number of heads per plant, head size and number of grains per head. During the yield formation stage, water deficiency could reduce the grain weight and possibly causes shrivelling of grains (only combined with hot, dry and strong wind). The ripening stage is a drying-off period where a water deficit has a slight effect on the yield (FAO – Wheat, n.d.). During 2018, the flowering and yield formation development stages coincide with relatively dry periods. Therefore, it should be taken into account that farmers possibly irrigated their crops during

these development stages.

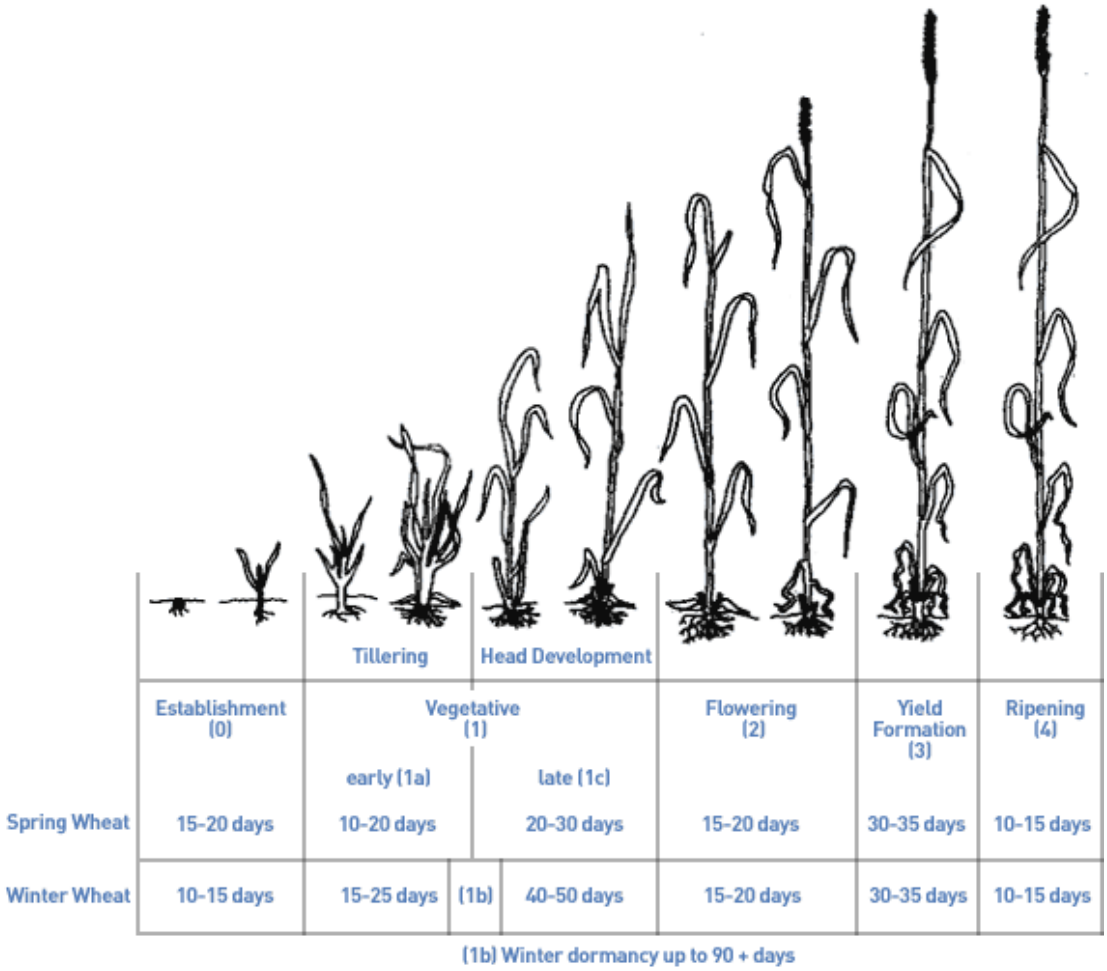


Figure 3.5 Duration of different growth periods wheat (FAO - Wheat, n.d.)

3.3. Soil characteristics

Soil texture

Soil texture determines the infiltration rate of a soil, which is important for i.e. irrigation management. Coarse textured soils (sands and loamy sands, see Figure 3.6) have a high infiltration rate and a low water retention rate. This can be explained by the pore size of the soils, which are generally large with limited ability to retain water (O'Geen, 2013). On the other hand, fine textured soils (clays, sandy clays and silty clays, see Figure 3.6) have a low infiltration rate and a high water retention rate. Remark, high water retention rate does not mean a higher availability of water for crop use. This has to do with the attraction force of a soil on the available water, which is at some point stronger than the plant water uptake force. Therefore, the best conditions for crops are generally loamy textured soils, which are in between coarse and fine textured soils (O'Geen, 2013).

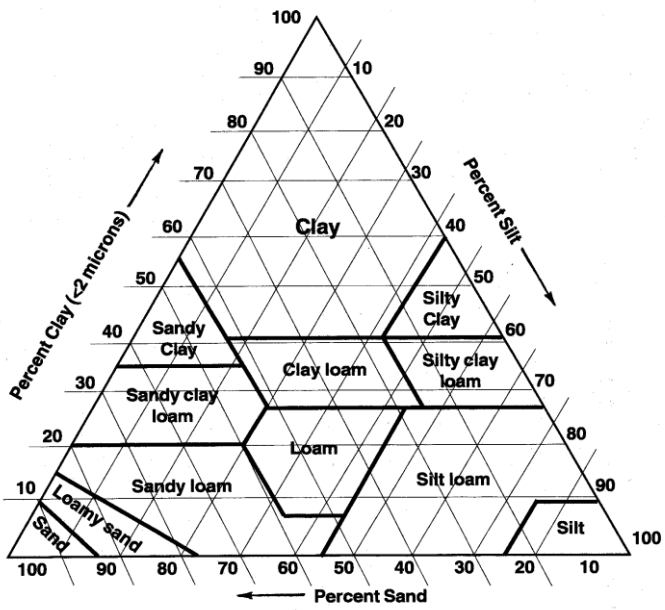


Figure 3.6 Soil texture triangle (Plant and Soil Sciences eLibrary, n.d.)

Soil organic matter

Organic matter is a source of nutrients, which stimulates the crop growth. It also acts like a sponge, therefore the amount of water a soil can hold will increase with increasing amount of organic matter. In contrary with for instance clayey soils, almost all water absorbed by organic matter will be available for crop use (Funderburg, 2016).

3.4. Overall methodology

The aim of this study is to estimate soil water properties at a scale of 30 meters based on bare soil surface reflectance and thermal infrared image analysis. With help of two newly developed models, the spatial patterns of soil wetness will be estimated. The output of SoilGrids30m model will be used to estimate the soil water holding capacity. Soil water holding capacity is not directly related to soil moisture content. The hypothesis is therefore, soils with a high soil water holding capacity will tend to have higher soil moisture content estimates and vice versa during a long period of drought. The soil wetness indicator is a qualitative measure of soil moisture conditions and is therefore directly related to soil moisture content. Both models will be evaluated with the soil moisture content estimates of SEBAL. Furthermore, a generic framework will be proposed for the SoilGrids30m model and the performance of the models will be tested. The methodology will be explained with help of Figure 3.7.

The first model, SoilGrids30m, is based on the SoilGrids250m and SoilGrids1000m model (Hengl et al., 2017; Hengl et al., 2014). SoilGrids30m is a model to translate bare soil surface reflectance into physical soil properties. A common way to obtain physical soil properties in the Netherlands is with help of the Stiboka soil map. The Stiboka soil map is based on a database of soil samples collected over the years therefore the spatial and temporal resolution are generally too coarse for precision agriculture. The disadvantages of the Stiboka soil map will be improved with the development of the SoilGrids30m model.

The obtained physical soil properties from the SoilGrids30m model are translated into van Genuchten parameters with help of pedotransfer functions. The van Genuchten parameters are the input of the water retention curve to estimate soil water holding capacity. The soil water

holding capacity is an important parameter for irrigation management and has a direct influence on the crop growth.

The SoilGrids30m model only gives estimates of fixed physical soil properties, which limits the model to be able to estimate soil moisture content. Soil moisture content is the available water present in a soil and depends on i.e. weather conditions and irrigation. Soil moisture content is therefore a dynamic process that cannot be determined with fixed physical soil properties. Instead, soil water holding capacity will be determined which is a fixed soil water property.

This study will be focused on dry conditions. Areas with a high water holding capacity should have a higher soil moisture content than areas with a low soil water holding capacity. An evaluation will be made to see if there is indeed a clear relationship visible between the soil water holding capacity and the total soil moisture content during dry periods in the study area.

The estimated soil water holding capacity will be determined with help of an automated soil mapping model, pedotransfer functions and the water retention curve, all of these steps have their inaccuracies and therefore the reliability of this approach depends on these inaccuracies. Furthermore, this approach is based on for instance the reflectance of the soil top layer, which does not have to have the same soil properties as in the root zone.

The second model is based on the concepts of the Trapezoid method (Yang et al., 2015), which will be named as the soil wetness indicator. The soil wetness indicator is a qualitative measure of moist conditions in the study area. For each pixel, the NDVI and the relative crop temperature (RCT) will be calculated with help of the SEBAL model. The RCT is the difference between land surface temperature and the instantaneous air temperature. The model is a relative representation of the at date situation/conditions. Each day the pixel envelope differs because it depends on the local weather conditions and the growth stage of the vegetation. With help of an RCT-NDVI plot, the wetness of each pixel can be determined. In this model, the wetness of a pixel is relative to the wetness of all pixels together. Based on the RCT-NDVI density plot, the boundaries of the pixel envelope will be determined. The boundaries are used to gradually divide the pixel envelope into classes from dry to wet. Each pixel can then be assigned to a class of wetness.

The third model used in this study is SEBAL, which will be used for different purposes. In the first place, SEBAL will be used to have a uniform way to pre-process all images used in this study. Furthermore, the total soil moisture content estimated by SEBAL will be used to evaluate the results of the SoilGrids30m model and the soil wetness indicator.

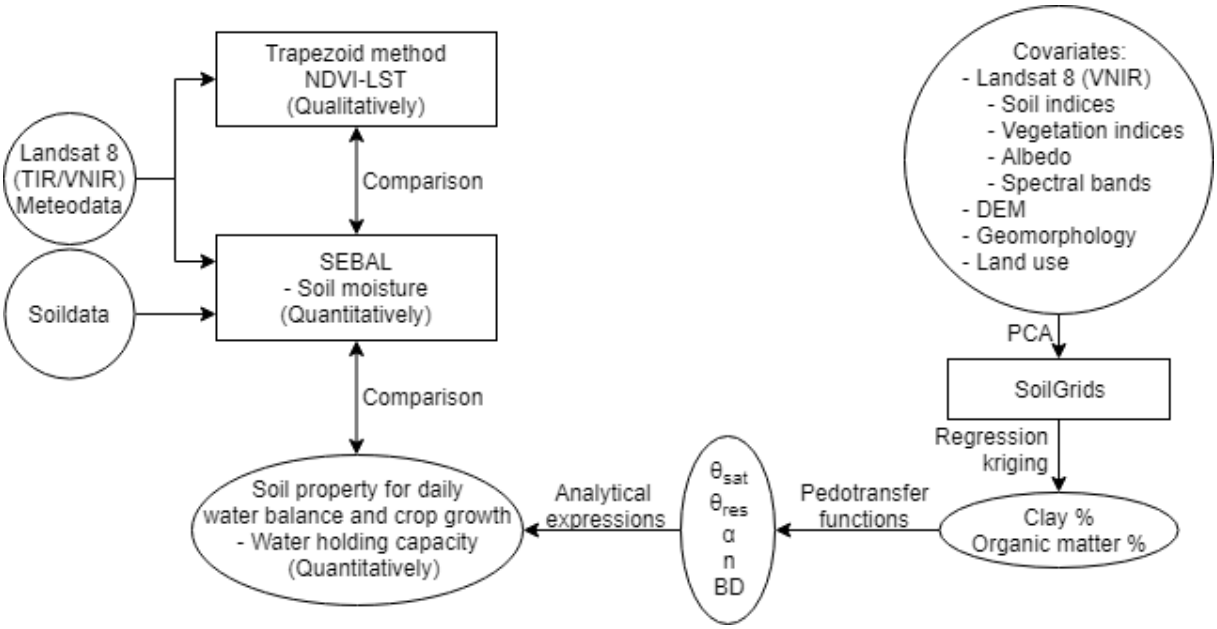


Figure 3.7 Flowchart overall methodology

3.5. Spatial estimation of surface soil properties using remote sensing data

3.5.1. Spatial estimation technique: regression-kriging

The fundamental of spatial estimation is predicting target values at unvisited locations. The spatial variation in predicted target values often correlates with environmental properties such as land use, elevation, slope, etc. These environmental properties can be summarized with the term SCORPAN: Soil classes or properties; Climate; Organisms, vegetation, fauna or human activity; Relief; Parent material; Age; n – spatial position (McBratney et al., 2003). Including SCORPAN properties to a spatial estimation technique often improves estimation accuracy (Hengl et al., 2019). In this study, this will be done with help of the regression-kriging technique, which uses explanatory variables along with soil samples. As mentioned in paragraph 2.3.1, regression-kriging consists of two separate parts, the ordinary-kriging model and the simple or multi-linear regression model.

Ordinary-kriging can be described as an interpolation technique where the estimation at a location is a linear combination of observations nearby. The weight that is given to each observation depends on the degree of (spatial) correlation (Hengl et al., 2007). Ordinary-kriging is denoted as a best linear unbiased predictor, “best” because it minimizes the variance of the errors; “linear” because its estimates are weighted linear combinations of the observations; “unbiased” because the mean residual is equal to zero. Furthermore, ordinary-kriging does not only provide an estimate but also the variance of an estimate. Ordinary-kriging is formulated according to equation 3.1 (Hengl et al., 2007).

$$\hat{z}(s_0) = \sum_{i=1}^{n} \lambda_i * z(s_i) \qquad 3.1$$

Where, $\hat{z}(s_0)$ is the predicted value of the target variable at an unvisited location (s_0); $z(s_i)$ are the observations; λ_i are the weights of each observation with respect to the unvisited location (s_0).

The SoilGrids30m model uses a multi-linear regression model to estimate the regression coefficients between the target variable and the explanatory variables at the sample location. At the unvisited locations, the estimated regression coefficients along with the known explanatory variables are used to predict the value of the target variables. The Multi-linear regression model is formulated according to equation 3.2 (Hengl et al., 2007).

$$\hat{z}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k * q_k(s_0); \quad q_0(s_0) \equiv 1 \qquad 3.2$$

Where, $q_k(s_0)$ are the values of the explanatory variables at the unvisited locations (s_0); $\hat{\beta}_k$ are the estimated regression coefficients; p is the number of explanatory variables.

As mentioned in paragraph 2.3.1, regression-kriging combines ordinary-kriging and multi-linear regression. The multi-linear regression model is used to fit the explanatory variations from the explanatory variables and ordinary-kriging is used to fit the residuals. The regression-kriging model is formulated according to equation 3.3 (Hengl et al., 2007).

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k * q_k(s_0) + \sum_{i=1}^{n} \lambda_i * e(s_i) \qquad 3.3$$

Where, $\hat{m}(s_0)$ is the fitted trend from the multi-linear regression model; $\hat{e}(s_0)$ is the interpolated residual from the ordinary-kriging model; $e(s_i)$ is the residual at location s_i.

3.5.2. Data input

Observations

A set of 518 soil samples from the study area have been extracted from "BISNederland" (n.d.). However, the amount of useful soil samples is significantly lower because of two constraints. Firstly, the SoilGrids30m model only focusses on bare soil surface pixels because these pixels can be directly related to physical soil properties of the top layer. Bare soil generally is denoted with an NDVI value within the range of 0.0 and 0.2 (Carlson et al., 1997), all values outside this range can be related to i.e. vegetation or waterbodies. Secondly, if at any location, at pixel level, information from one or more explanatory variable(s) is missing or it is an outlier the pixel will be left out of the analysis. Therefore, only profile observations that meet both constraints will be used for the analysis.

Explanatory variables

In this study, the applied data can be categorized based on SCORPAN properties. Table 3.1 shows an overview of all explanatory variables applied in this study. Remote sensing is a technique that allows to obtain soil characteristics from a study area in the order of hundreds of square kilometers. Therefore, the majority of the explanatory variables are defined by reflectance data obtained from Landsat 8 images. The amount of soil reflectance depends on the scattering and absorption properties of a soil (Weidong et al., 2002). Which in their turn depends i.e. on soil composition, physical structure, land cover, etc. A second set of data is related to the relief properties of the area. Although the study area can be considered as a flat area, the height of the area gradually decreases from East to West. Numerous studies have already shown a relationship between relief and soil properties (Pachepsky et al., 2001; Sobieraj et al., 2002; Ceddia et al., 2009). Therefore, the elevation, slope and aspect are implemented in terms of the relief parameters. Additional, categorical information about soil classes and soil parent material are added to complete the explanatory variables dataset.

Explanatory variable	Abbreviation	SCORPAN	Formulation	Short description	Source
Soil class		S	Categorical map	Type of soil	https://www.pdok.nl
Blue	B2	O		Clay spectral signature	Casa et al. (2013)
Green	B3	O		Clay spectral signature	Casa et al. (2013)
Red	B4	O		Clay spectral signature	Casa et al. (2013)
Near Infrared	B5	O		Organic matter absorption band	Summers et al. (2011)
Short-wave Infrared 1	B6	O		Clay and organic matter absorption band	Summers et al. (2011)
Short-wave Infrared 2	B7	O		Clay and organic matter absorption band	Summers et al. (2011)
Enhanced Vegetation Index	EVI	O	$2.5 * \frac{B5 - B4}{B5 + 6 * B4 - 7.5 * B2 + 1}$	Health and amount of vegetation	USGS (2017)
Modified Soil Adjusted Vegetation index	MSAVI	O	$B5 + 0.5 - 0.5 * \sqrt{(2 * B4 + 1)^2 - 8 * (B4 - B3)}$	Reduced soil background effect for health and amount of vegetation	USGS (2017)
Normalize Differenced Moisture Index	NDMI	O	$\frac{B5 - B6}{B5 + B6}$	Crop water stress level	USGS (2017)
Brightness Index	BI	O	$\sqrt{\frac{B4^2 + B3^2 + B2^2}{3}}$	Average reflectance magnitude	Forkuor et al. (2017)
Coloration Index	CI	O	$\frac{B4 - B3}{B4 + B3}$	Soil color	Forkuor et al. (2017)
Hue Index	HI	O	$\frac{2 * B4 - B3 - B2}{B3 - B2}$	Primary colors	Forkuor et al. (2017)
Redness Index	RI	O	$\frac{B4^2}{B2 * B3^3}$	Hematite content	Forkuor et al. (2017)
Saturation Index	SI	O	$\frac{B4 - B2}{B4 + B2}$	Spectral slope	Forkuor et al. (2017)
Albedo		O		Diffuse reflection of solar radiation	SEBAL
Emissivity		O		Surface effectiveness in emitting thermal radiation	SEBAL
Surface roughness		O		Deviation of surface reflectance in the direction of the normal vector	SEBAL
Land use	BRP	O	Categorical map	Type of land cover	https://www.pdok.nl
Nitrogen		O		Nitrogen level	SEBAL
Elevation		R		Height above sea level	SEBAL
Slope		R		Inclination of land surface from horizontal	SEBAL
Aspect		R		Direction the slope faces	SEBAL
Geomorphology		P	Categorical map	Parent material	https://www.pdok.nl

Table 3.1 Overview explanatory variables SoilGrids30m

Three SCORPAN properties, climate; age; n – spatial position, are not mentioned yet. The age and spatial position of each pixel has not been applied at all. The area has been evolved due to human activities; therefore, it does not contain i.e. high mountains, moderate hills or major rivers that could have an influence on the spatial position of the soil properties. Furthermore, the area is relatively small and therefore differences in spatial position are also relatively small. The information available about the age of the soils in the area is too coarse, with as result a constant value for the whole area. Without noticeable differences within an explanatory variable, the explanatory variable does not add any new information to the model and should be left out of the analysis. In terms of climate properties, an extra set of data has been created by taking the difference, for all obtained explanatory variables based on reflectance, between a moist day and a dry day image. Sandy soils i.e. have a higher water permeability than clayey soils. The difference between a moist and dry date, per explanatory variable, could give extra information about soil properties. Meteorological data obtained from the KNMI weather station in “Marknesse” has been used to select a moist date and a dry date. The dry date is on 5 January 2017 as shown in Figure 3.8. The moist date is on 10 March 2017 as shown in Figure 3.9.

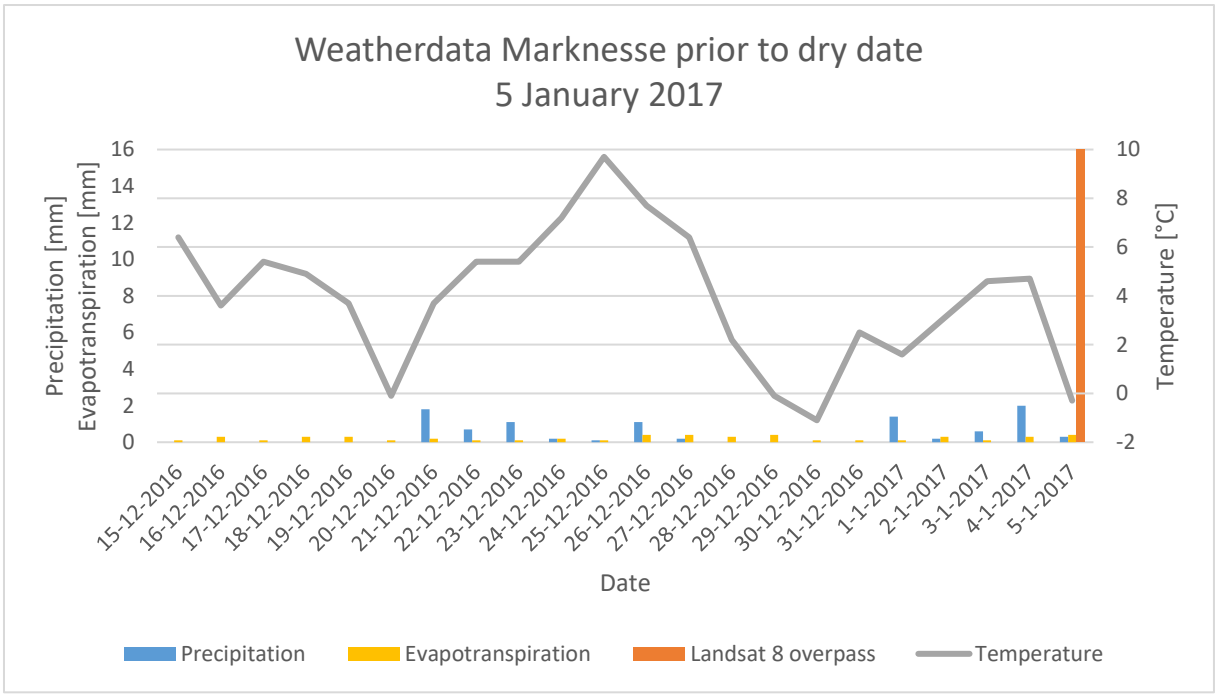


Figure 3.8 Weatherdata KNMI station Marknesse prior to dry date SoilGrids30m

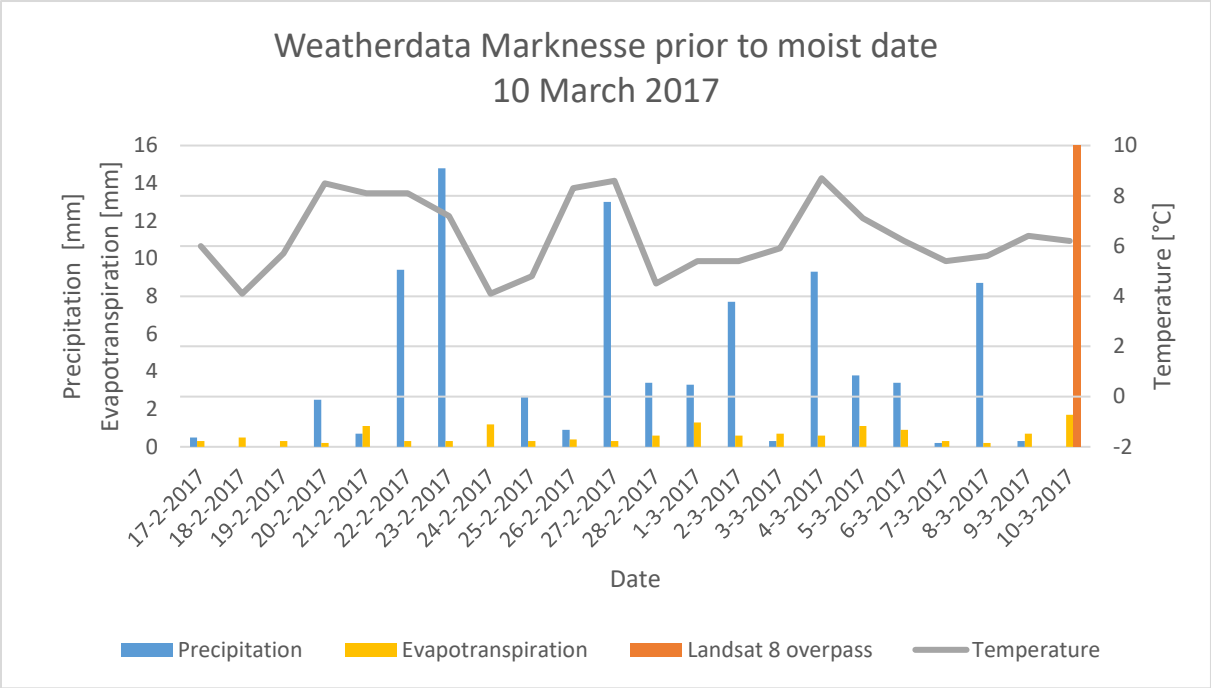


Figure 3.9 Weather data KNMI station Marknesse prior to moist date SoilGrids30m

3.5.3. Spatial estimation of surface soil properties

The goal of the SoilGrids30m model is to estimate surface soil properties. Due to the available observations, only target variables clay content and organic matter content will be estimated. Each target variable will be estimated separately with its own model created in the statistical programming language R. The model consists of six modules as shown in Figure 3.10, the R-scripts of all modules can be found in Appendix A until Appendix E.

Collecting and preparing data

The first module is collecting and preparing the data. Preparing the data is done in five different steps, the first step is a pre-prepping step. All data should, be rasterized (also categorical data); have the same resolution; have the same projection. The second step is to select only bare soil surface pixels, with a range between 0.0 NDVI and 0.2 NDVI, as explained in the observations part of paragraph 3.5.2. The third step is to detect all outliers based on the boxplot theory. In this model, all outliers and non-bare soil pixels are set to “NA”. The fourth step is to remove explanatory variables without any variation in the data. The fifth step is to standardize the data. All applied explanatory variables operate in different ranges of absolute values. To prevent the model to be dominated by explanatory variables with significant higher absolute values, they will all be standardized using a z-score according to equation 3.4 (Levi & Rasmussen, 2014).

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \qquad 3.4$$

Where, Z_{ij} is the z-score of pixel i in explanatory variable j; x_{ij} is the untransformed value of pixel i in explanatory variable j; μ_j is the mean of explanatory variable j; σ_j is the standard deviation of explanatory variable j.

Data correlation

The second module is to extract the explanatory data at the observation point locations and evaluate the explanatory data based on their correlation to the target variable. The first step is creating a spatial dataset that contain all observations along with all explanatory data at the

observation locations. The SoilGrids30m model uses 62 different explanatory variables of which not all of them has a clear relation with respect to the target variable. A correlation threshold will be applied to reduce the number of explanatory variables used for further analysis. The explanatory variables are arranged based on their correlation with the target variable. The correlation threshold is a percentile value that eliminates the n^{th} percentile explanatory variables with the lowest correlation to the target variable. After applying a correlation threshold, the first module should be executed again. As earlier mentioned, a soil sample location will be left out of the analysis if information at that location from one or more explanatory variable(s) is missing or it is an outlier. The correlation threshold will exclude explanatory variables from the analysis, which could also lead to an increase of useful soil sample locations for the analysis. A higher amount of soil sample locations will be beneficial for the results of the model.

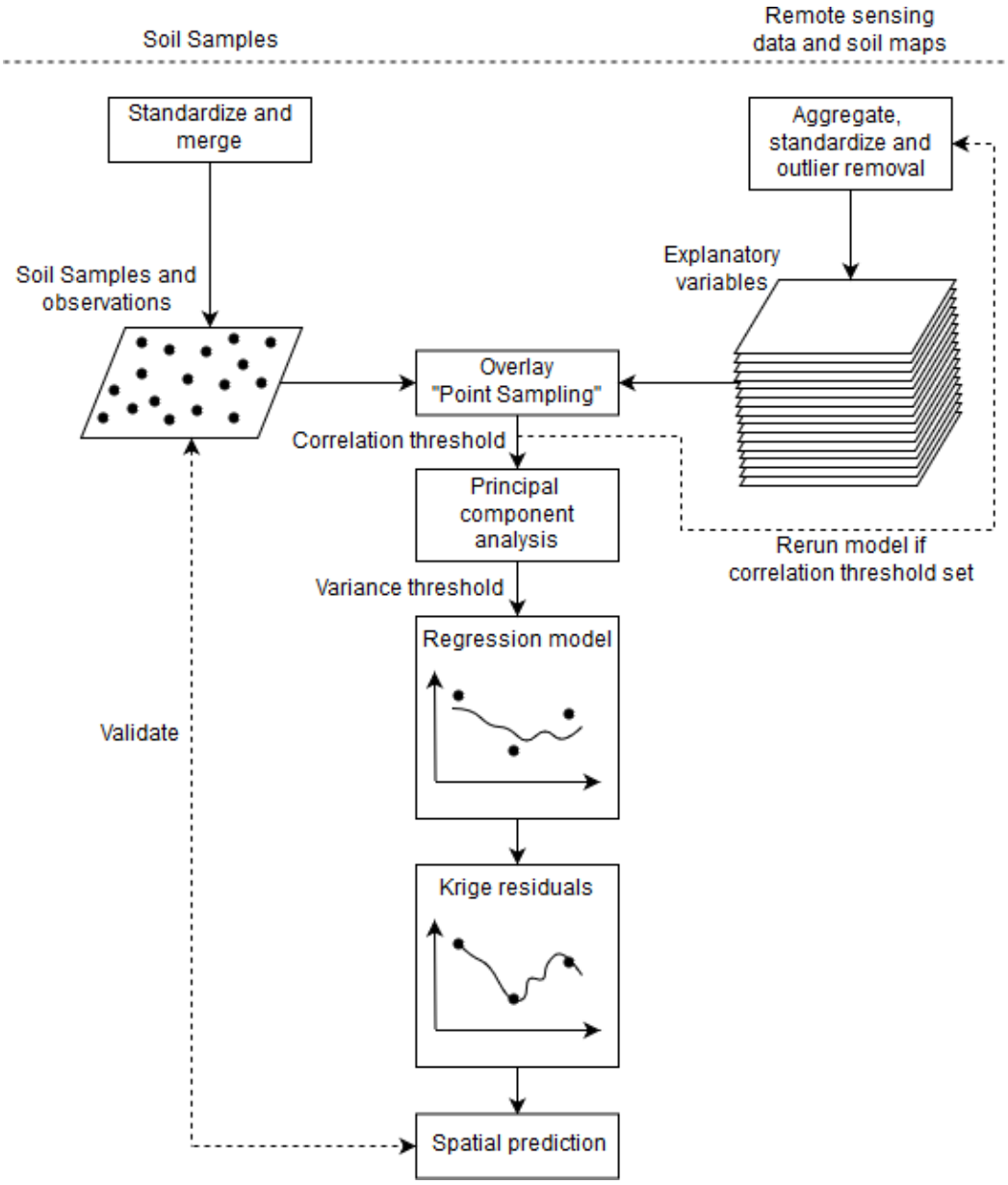


Figure 3.10 Flowchart SoilGrids30m model

Principal component analysis

The third module is to create and select principal components for the regression model. Principal component analysis (PCA) transforms a set of "correlated" observations into linearly uncorrelated variables, called principal components (Alice, 2016). There will be as many principal components created as explanatory variables used as input for the analysis. The SoilGrids30m model uses, without correlation threshold, 62 explanatory variables as input and therefore it will create 62 principal components. The first principal component accounts for the largest possible variance. Each subsequent component is orthogonal to all previous components and has the subsequent highest variance. All principal components together account for 100% of the variance in the model. There are multiple reasons to apply principal components instead of using the explanatory variables themselves. In the first place, by using PCA the number of input variables for the regression model could be significantly reduced (Alice, 2016). A variance threshold can be set to select a subset of principal components that account for the threshold amount of variance in the model. This reduction of input variables for the regression model could reduce the complexity and the computational effort of the model. A second significant benefit of PCA is that it is able to avoid multicollinearity (Alice, 2016). All principal components are uncorrelated linear combinations of the explanatory variables. Therefore, each principal component will add "new" information to the model without any overlap with other principal components. This is certainly not the case when the explanatory variables are used instead of the principal components. Furthermore, by using principal components the regression model will be less sensitive to overfitting (Alice, 2016). Potential benefits comes with potential risks and drawbacks. One of the drawbacks of PCA is the decoupling of the relation between explanatory variable and target variable (Alice, 2016). It would be more convincing to find relations between explanatory variables and target variables. By using principal components, it could make it more difficult to explain what is affecting what. Another drawback of PCA is that the principal components are obtained in an unsupervised way (Alice, 2016). Which means that the target variable has not been used to determine the direction of the principal components. It is therefore not certain that the configuration of the principal components is optimal for the estimation of the target variable.

Regression-kriging

The fourth and the fifth module together is the regression-kriging technique and has been elaborately explained in paragraph 3.5.1. To obtain the final spatial estimation of the target variable all soil sample observations, which meet the requirements explained in the first module, are used.

Validation

The sixth module is the validation of the obtained results based on a n-fold cross-validation. After the principal component analysis, the data will be split into a validation set and a trainings set. The trainings set will be used to predict the target variable for each pixel that meet all requirements as mentioned in the preparation module. The performance of the SoilGrids30m model will be evaluated with the validation set with help of three indicators. The first indicator is the coefficient of determination, which indicates the amount of variance explained by the model (see equation 3.5).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad 3.5$$

Where, R^2 is the coefficient of determination; SS_{res} is the sum of squares of the residuals; SS_{tot} is the total sum of squares. The second indicator is the mean absolute estimation error, which is the true estimation error of the model (see equation 3.6).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{z}(s_i) - z(s_i)| \quad 3.6$$

Where, MAE is the mean absolute estimation error; n is the number of observations; $\hat{z}(s_i)$ are the estimated values at validation point s_i; $z(s_i)$ are the estimated values at validation point s_i. The third indicator is the root mean square estimation error, which is a measure for the accuracy of the estimation (see equation 3.7).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{z}(s_i) - z(s_i)\right)^2} \quad 3.7$$

Where, $RMSE$ is the root mean square estimation error.

3.5.4. Spatial estimation of soil water holding capacity

The soil water holding capacity of a soil is the maximum amount of water a soil can hold for crop use. To optimize the crop production it is important to know what the maximum available amount of water for crop use in a soil is. The soil water holding capacity is defined as the amount of water at field capacity minus the amount of water at wilting point (see equation 3.8).

$$WHC = \theta_{FC} - \theta_{WP}, \quad [cm^3/cm^3] \quad 3.8$$

The exact definition of field capacity has not been established because field capacity should be reached at some equilibrium point, which it never does. This is because water in soil is a dynamic process and therefore continuously changing (Kirkham, 2014). In this study, field capacity will be seen as the amount of water a soil can hold against gravitational forces and when the drainage rate has been decreased (Kirkham, 2014). Wilting point on the other hand has a clear definition. Wilting point is related to a deficiency of water for crops in the area and is denoted as the point where crops for the first time undergo a permanent reduction in moisture content (Kirkham, 2014). Both field capacity and wilting point can be computed with help of the water retention curve (Van Genuchten, 1980). The water retention curve is an analytical expression between soil water content and soil water potential, also denoted as pressure head (see equation 3.9).

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha h|^n)^{1-1/n}}, \quad [cm^3/cm^3] \quad 3.9$$

Where, $\theta(h)$ is the water retention curve $[cm^3/cm^3]$; h is the suction pressure or pressure head $[cm]$; θ_r is the residual water content $[cm^3/cm^3]$; θ_s is the saturated water content $[cm^3/cm^3]$; α is related to the inverse of the air entry suction $\alpha > 0$ $[cm^{-1}]$; n is a measure of the pore-size distribution $n > 1$ $[-]$. Field capacity and wilting point are estimated to be the water content at a pressure head of 100cm and 16000cm respectively.

Input parameters of the water retention curve are also known as soil hydraulic properties. The most accurate way to determine soil hydraulic properties are with field measurements. However, these measurements are often inconvenient, costly, time consuming and labour intensive (Schaap et al., 2001; Wagner et al., 2001). For this study i.e., it is almost impossible to have a substantial spatial coverage of the hydraulic properties as large as the study area. To overcome these problems many indirect methods have been developed (Rawls & Brakensiek, 1989; Wösten et al., 1995; Stolte et al., 1996; Schaap et al., 2001). The development of an equation to indirectly estimate soil properties has been termed as

pedotransfer function (PTF) for the first time by Bouma (1989). In essence, PTFs translates data, which are generally easy to obtain (i.e. soil texture and organic matter), into what we need (i.e. soil hydraulic properties). Soil hydraulic databases are used to calibrate the PTFs it therefore strongly depends what database has been used, especially from which region. PTFs are therefore strongly empirical related models.

In this study, the available observations only contain clay content and organic matter content values. Therefore, the PTFs used to estimate soil hydraulic parameters are based on clay content and organic matter content but also on bulk density (introduced as unknown). Four unknown soil hydraulic parameters, $\theta_r, \theta_s, \alpha, n$, and the bulk density will be estimated with help of PTFs. The PTFs to compute bulk density, α and n are obtained from Wösten et al. (2001), see equations 3.10 until 3.12.

$$\alpha^* = -19.13 + 0.812 * OM + 23.4 * BD - 8.16 * BD^2 + 0.423 * OM^{-1} + 2.388 * LN(OM) - 1.338 * BD * OM \quad (3.10)$$

$$n^* = -0.235 + 0.972 * BD^{-1} - 0.7743 * LN(Clay) - 0.3154 * LN(OM) + 0.0678 * BD * OM \quad (3.11)$$

$$\frac{1}{BD} = 0.6117 + 0.003601 * Clay + 0.002172 * OM^2 + 0.01715 * LN(OM) \quad (3.12)$$

Where, OM is the organic matter content $[\%]$; BD is the bulk density $[g/cm^3]$; $Clay$ is the clay content $[\%]$. α^* and n^* are transformed parameters and should be converted according to equations 3.13 and 3.14.

$$\alpha = e^{\alpha^*} \quad (3.13)$$

$$n = e^{n^*} + 1 \quad (3.14)$$

The PTFs to compute θ_r and θ_s are obtained from Scheinost et al. (1997), see equations 3.15 and 3.16.

$$\theta_s = 0.85 * \left(1 - \frac{BD}{2.65}\right) + 0.13 * Clay \quad (3.15)$$

$$\theta_r = 0.0051 * Clay + 0.017 * C_{org} \quad (3.16)$$

Where, BD is the bulk density $[g/cm^3]$; $Clay$ is the clay content $[\%]$; C_{org} is the organic carbon content $[\%]$. To obtain organic carbon content from organic matter content, a division factor of two has been applied to the organic matter content, according to Pribyl (2010). The R-script for the pedotransfer functions can be found in Appendix F.

3.6. Soil wetness indicator

3.6.1. Data input

The fundamental idea of the soil wetness indicator is based on the relative crop temperature (RCT) and the NDVI. Both input parameters have been computed with help of SEBAL according to paragraph 2.2.4. The RCT is the land surface temperature minus the instantaneous air temperature. SEBAL computes the land surface temperature based on a split-window algorithm, where both band 10 and band 11 (TIR bands) of Landsat 8 are used. The NDVI has been determined based on the difference ratio between band 5 and band 4 (NIR and red band) of Landsat 8. The used data should meet a couple of requirements to be useful, the images should be cloudless; there should be a sufficient amount of vegetation in the area; they should be taken during a period of drought; there should be a sufficient amount of fields with the same crop type.

A cloudless image is necessary to be able to measure the reflectance of the Earth surface. To be able to measure the crop temperature there should be sufficient amount of vegetation cover in a pixel. Without vegetation, the temperatures measured are related to bare soil while it is important to measure crop temperatures. A plant with its roots can be seen as a thermometer where the crop temperature is a measure for the water availability in the root zone. Crops react on a deficit of water content in the soil. Low water content will reduce the transpiration rate of the plant and causes the stomata to close. This process coincides with a temperature increase of the plant (Rutter et al., 1958; Xu et al., 2008; van den Bersselaar et al., 2005; Anderson et al., 2007). Numerous studies have shown that the temperature of crops, obtained from thermal infrared data, compared with air temperature has a relationship with the soil moisture content (Bastiaanssen et al., 2006; Yang et al., 2015; Hatfield et al., 2008). The base of the analysis is therefore, crops with sufficient water will tend to be close to air temperature and crops with a deficit of water will show an increase in temperature with respect to air temperature. The water availability depends on soil type, which can only be related to crop temperature when external influences, such as precipitation and or irrigation, are eliminated. It is therefore important for the analysis to use images taken during a longer period of drought, as well for precipitation as for irrigation. Therefore, the year 2018 will be used as reference year because of the extreme dry conditions, see paragraph 3.1. As mentioned, the images should have sufficient vegetation that limits the date range to the growth season. Four dates have been found of cloudless satellite images during the growth season, see Figure 3.11 until Figure 3.14 for the meteorological summary per date. At three of the four dates, 21 days prior to the Landsat 8 overpass, the total evapotranspiration is greater than the total precipitation. At 7 May, the precipitation from 29 April to 1 May has not been evapotranspirated completely at overpass time of Landsat 8. Yet, this day could be useful because 4 days prior to satellite overpass time where without precipitation.

Furthermore, sufficient amount of fields with same crop type should be available at all images. Each crop type reacts different to drought conditions, which i.e. depends on the root depth. It is therefore important to evaluate the results per crop type to eliminate the influence of different crop type properties. Another reason has to do with irrigation management. Each crop type will need a different irrigation scheme to have the best growth curve. In this study, sugar beet and winter wheat are used for the analysis. Both crop types generally do not need irrigation throughout the season in the Netherlands. Furthermore, these crop types generally have a sufficient ground coverage to measure mainly the crop temperature instead of the surrounding soil. In addition, these crops are frequently present in the area.

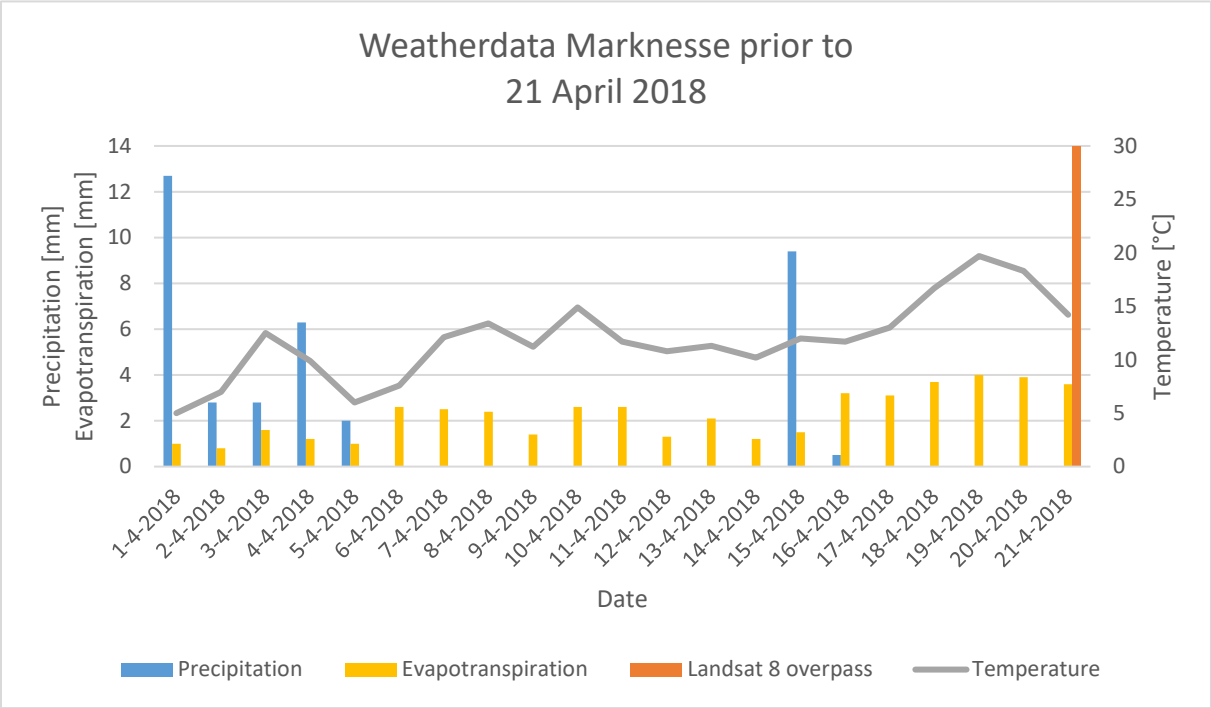


Figure 3.11 Weatherdata KNMI station Marknesse prior to 21 April 2018 SWI

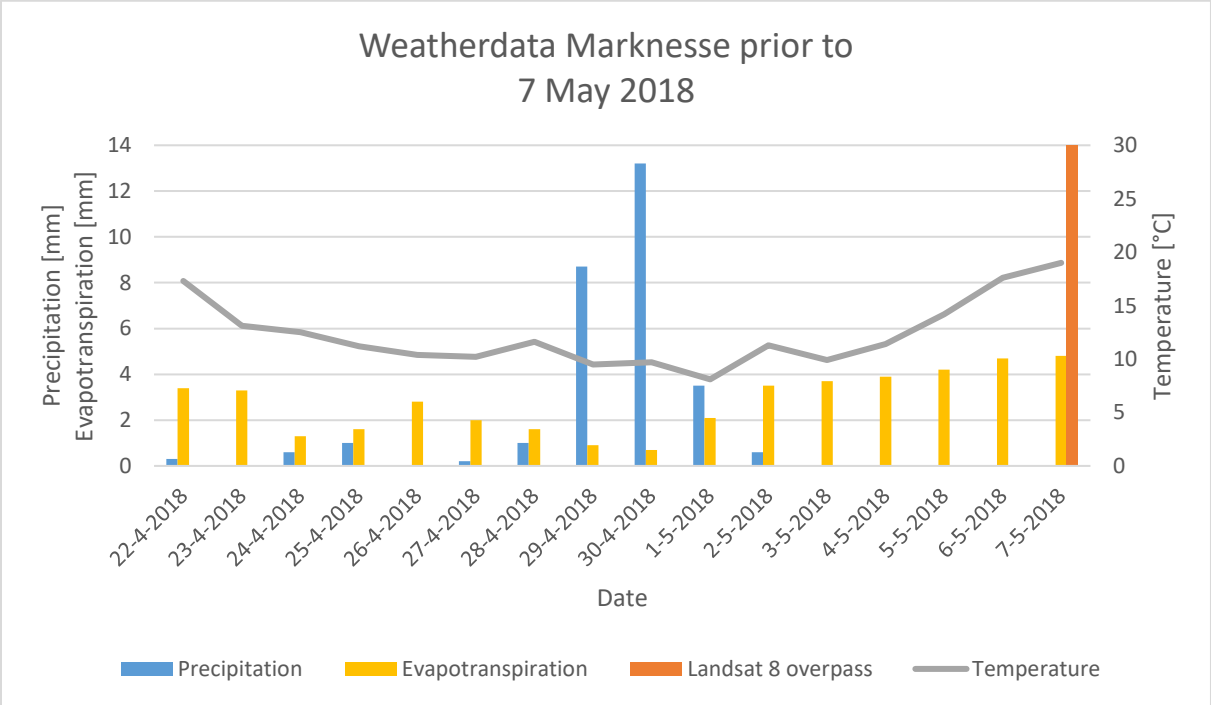


Figure 3.12 Weatherdata KNMI station Marknesse prior to 7 May 2018 SWI

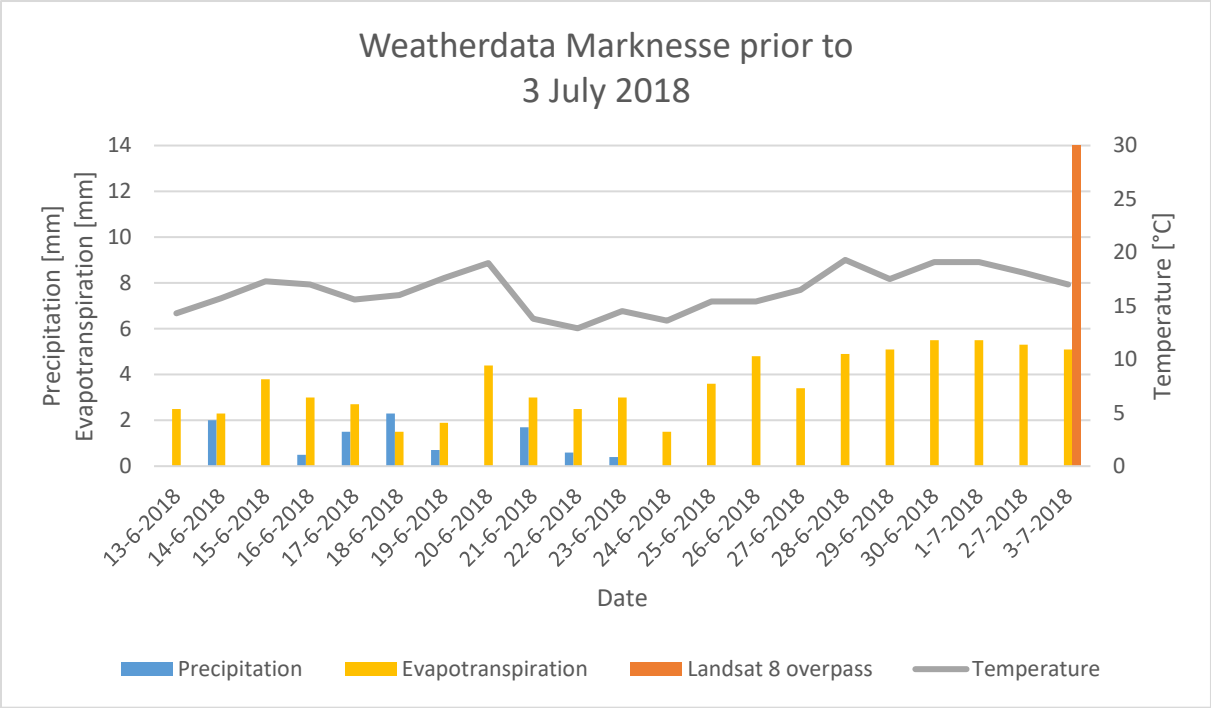


Figure 3.13 Weatherdata KNMI station Marknesse prior to 3 July 2018 SWI

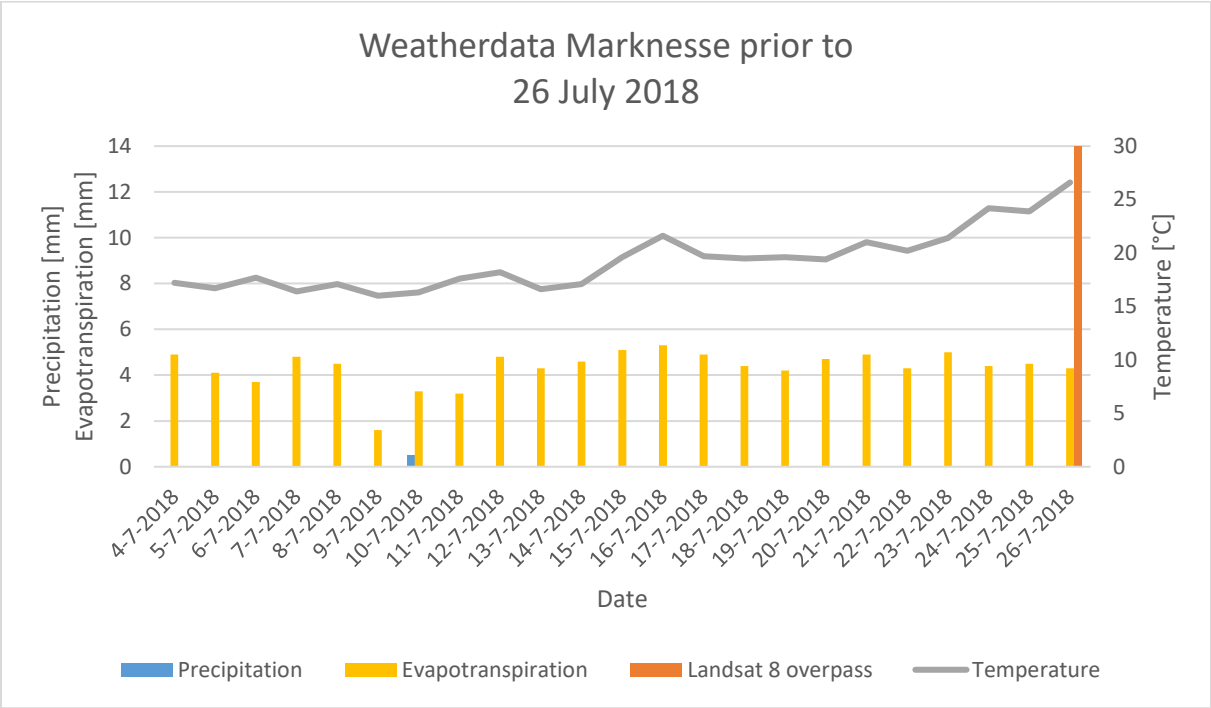


Figure 3.14 Weatherdata KNMI station Marknesse prior to 26 July 2018 SWI

3.6.2. Procedure to derive soil wetness indicator

The goal of the soil wetness indicator is to detect patterns of wetness that can be related to the total soil moisture content estimated by SEBAL. As explained in the previous paragraph only sugar beet and winter wheat will be used for the analysis. The soil wetness indicator can be divided into five modules according to Figure 3.15. The soil wetness indicator has been completely programmed in Python, the script of each module can be found in Appendix G until Appendix K. All modules have been developed in a variable way. This means that the modules can be used with multiple and different input parameters, i.e. dates; crop types; initial parameters, etc.

Create buffer zone per field per crop type

The first module of the soil wetness indicator extracts all pixels per crop type. The land use data is freely available in shapefile format offered by the Dutch government (https://www.pdok.nl). The analysis of the soil wetness indicator will be done with NDVI and LST values obtained from Landsat 8. Landsat 8 operates on a spatial resolution of 30 meters. A pixel will therefore cover an area of 30 square meters and could contain not only reflectance of vegetation but also parts of roads, houses and or ditches. These objects will have an influence in the reflectance magnitude of the pixel. To eliminate these artifacts and to obtain the reflectance of only crops of interest a buffer zone of 60 meters will be created around each field.

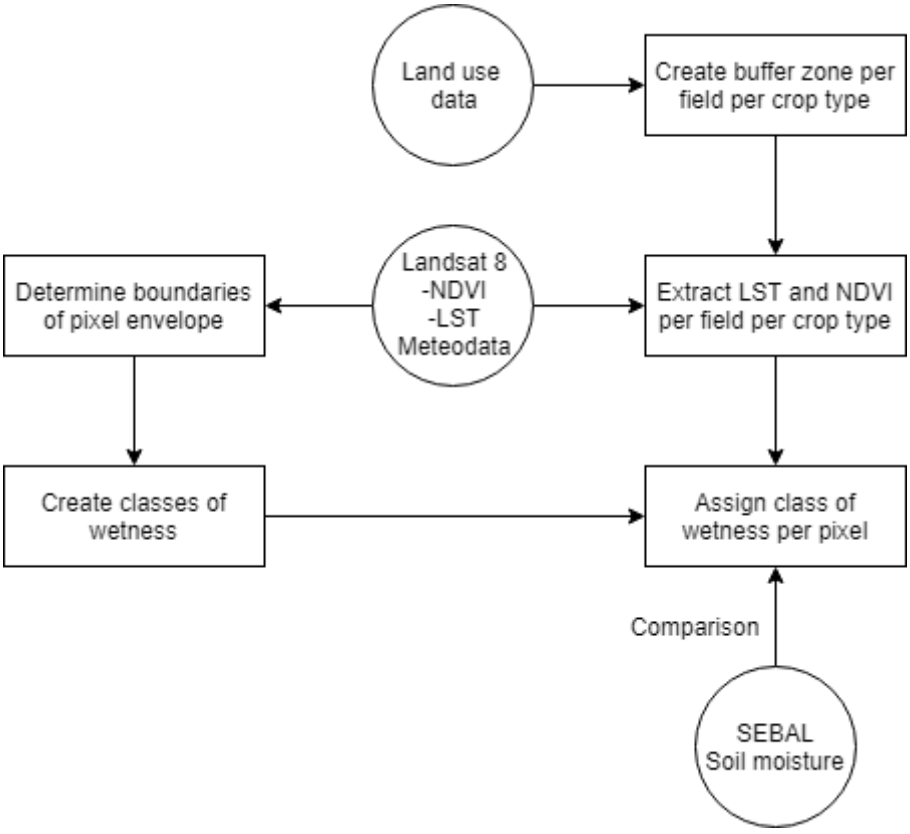


Figure 3.15 Flowchart soil wetness indicator

Extract LST and NDVI per field per crop type

The second module has been built with help of two developed functions, a function to extract for each field of interest the desired raster values; a function to merge all obtained raster values of each field into one overall raster layer. The first function extracts, based on land use information, per field the LST and NDVI raster values. Both properties are stored in a separate folder and each field has its own raster file. As explained in paragraph 3.6.1, for the analysis

the crop temperature relative to the air temperature will be used. Therefore, an extra step to obtain the RCT is applied by extracting the air temperature from the estimated land surface temperature. The second function merges all the individual field raster files per property into one raster file.

Determine boundaries of pixel envelope

The third module is to determine the upper and lower boundaries of the pixel envelope. It starts with preparing the data used for the analysis. The first step is to extract the air temperature from the estimated LST values of SEBAL. The second step is to remove all data that are not realistic and or meaningful for the analysis. Two constraints are used to select only realistic and meaningful pixels, only NDVI pixel values ranging between 0.0 and 0.9 (excludes waterbodies) are used for the analysis; RCT constraint is set based on the boxplot theorem. To determine the boundaries subsets with an increment of 0.1 NDVI are used. For each increment, the upper extreme RCT value has been determined based on the boxplot theorem. Finally, the upper RCT constraint has been set to the highest upper extreme RCT value of all increments. The lower RCT constraint has been set to the lowest lower extreme RCT value of all increments. After removing all outliers, the boundaries of the trapezoidal space are determined. The upper and lower n^{th} percentile has been determined for each increment. Based on the upper and lower n^{th} percentile an upper and lower boundary line will be estimated with help of linear regression. The final boundaries of the pixel envelope are determined in a somewhat subjective way. With help of two constraints the location of the boundaries of the pixel envelope can be manually adjusted. The n^{th} percentile is subjective in the way that it can be set to any arbitrary number. For the lower boundary the 2^{th} percentile has been applied. For the upper boundary, this is always the opposite of the spectrum respectively 98^{th} percentile. The second constraint is determined based on a visualization of the percentile values per increment. If the percentile value is visually not in line with the expected boundary line, it can be left out of the analysis for fitting the final boundary line. This method has been developed to have a semi-automated and quick way of detecting the boundaries of a pixel envelope based on the same principles.

Create classes of wetness

The fourth module is to create classes of wetness based on the boundary lines determined in the third module. Based on the number of classes, which is set to 20 for this study, two increments are determined. The first increment is determined at the minimum NDVI value of 0.0; the second increment is determined at the maximum NDVI value of 0.9. Each increment is determined as the difference between the upper and lower boundary line at mentioned NDVI location divided by the number of classes. Finally, the space between each set of lines will be converted to a polygon class and each polygon class can be assigned to a class of wetness. The third and fourth module are implemented in one Python script and are calculated in one go.

Assign class of wetness per pixel

The fifth module combines previous modules, with as result a qualitative representation of the wetness of a pixel. Each pixel per crop type, obtained from module 2, will be converted into a point class. The obtained polygon classes from module 4 are used to identify to which polygon class a point class belongs. Along with the wetness of each pixel, a density plot is made of the RCT-NDVI space. The density plot shows the overall pixel density of all pixels of the image on top of that the polygons obtained in module 4 are shown and the pixels per crop type obtained from module 2 are highlighted.

4. Results and analysis

4.1. Spatial estimation of surface soil properties using remote sensing data

In this paragraph, the results of three surface soil properties will be discussed: clay content; organic matter content; soil water holding capacity. In the first subparagraph, the model settings of the SoilGrids30m model will be discussed and the results of clay content and organic matter content will be analysed. In the second subparagraph, the results of the soil water holding capacity estimates will be analysed.

4.1.1. Spatial estimation of clay content and organic matter content

Model set up

A newly developed spatial estimation model, SoilGrids30m, is used to obtain spatial estimates of clay content and organic matter content in the study area. As explained in paragraph 3.5.3, SoilGrids30m consists of six modules. Each module is programmed in a variable way such that the model can be adjusted easily. Therefore, each step uses initial parameters to set up the module as desired.

In module 1A, all data used for analysis will be aggregated. The data used in this study is projected in WGS84/UTM zone 32N with a spatial resolution of 30 meters. Furthermore, an extra set of explanatory variables is obtained by taking the differences between a dry and a moist date. The dry image date used in this study is 5 January 2017 and the moist image date used in this study is 10 March 2017. The used initial parameters for module 1A are shown in Table 4.1.

Initial parameter	Parameter value	Unit
Module 1A	**Data collection**	
Projection	“EPSG32632”	WGS84/UTM zone 32N
Resolution input	“30”	[m]
Image date dry	“2017_01_05”	[-]
Image date wet	“2017_03_10”	[-]

Table 4.1 Initial parameters used for clay and organic matter content in SoilGrids30m module 1A

In module 1B, the aggregated data will be prepared for further analysis. Firstly, bare soil pixels will be selected based on a NDVI threshold. An upper limit of 0.2 NDVI is set to exclude vegetated areas and a lower limit of 0.0 NDVI is set to exclude waterbodies, clouds, snow and ice (see Table 4.2). Secondly, all outliers will be removed based on the boxplot theorem and lastly, all data will be standardized.

Initial parameter	Parameter value	Unit
Module 1B	**Data preparation**	
Bare soil NDVI threshold	0.2	[-]
Waterbodies NDVI threshold	0.0	[-]

Table 4.2 Initial parameters used for clay and organic matter content in SoilGrids30m module 1B

In module 2A, the model will be specifically set up for the desired target variable to obtain the correlation between the explanatory variables and the target variable. A correlation percentile threshold will be applied to reduce the number of explanatory variables used for further analysis. The explanatory variables are arranged based on their correlation with the target variable. The correlation percentile threshold is a percentile value that eliminates the n^{th} percentile explanatory variables with the lowest correlation to the target variable. In this study, the increment of the correlation percentile threshold is set to 0.05. In example, when a set of

60 explanatory variables is used for analysis, for each increment of 0.05 the three explanatory variables with the lowest correlation regarding to the target variable will be eliminated.

Initial parameter	Parameter value	Unit
Module 2A	**Data correlation**	
Correlation percentile threshold	[0.0-0.95]	[-] increments of 0.05

Table 4.3 Initial parameters used for clay and organic matter content in SoilGrids30m module 2A

Module 2B is comparable with module 1B but instead of all explanatory variables only the selected explanatory variables after setting the correlation percentile threshold are used. The initial parameters do not change compared to module 1B, see Table 4.4.

Initial parameter	Parameter value	Unit
Module 2B	**Data preparation**	
Bare soil NDVI threshold	0.2	[-]
Waterbodies NDVI threshold	0.0	[-]

Table 4.4 Initial parameters used for clay and organic matter content in SoilGrids30m module 2B

In module 3, a principal component analysis will be applied on the subset of explanatory variables from module 2B. The tuning parameter in module 3 is the minimum variance explained threshold. Each principal component explains partly the variance in the model. The minimum variance explained threshold sets a minimum value of the variance in the model that should be explained by the principal components. A low threshold will give at least one principal component for further analysis. A high threshold does not necessarily mean a selection of more principal components. This depends on how fine each principal component explains the variance of the model. If the variance of the model explained by the first principal component is already high then the number of principal components for further analysis could be still one. In this study, the threshold for the minimum variance explained has been fixed to 0.99 (see Table 4.5). In this way, 99% of the variance in the model will be used to estimate the target variables in the area. None of the remaining principal components are significant enough for further analysis, i.e. they represent the noise in the used Landsat 8 images.

Initial parameter	Parameter value	Unit
Module 3	**PCA**	
Minimum variance explained	0.99	[-]

Table 4.5 Initial parameters used for clay and organic matter content in SoilGrids30m module 3

In module 4-5, the regression-kriging technique will be applied. In this module, the semivariogram settings are set. These settings are used to fit a function that describes the degree of spatial dependence of the target variable residuals. For both, clay content and organic matter content, an exponential fit is used. The nugget, sill and range differ because of the difference in absolute value of both target variables (see Table 4.6).

Initial parameter	Parameter value	Unit
Module 4-5	**Regression kriging**	
Variogram nugget	5-0.1	[$\%^2$] Clay-OM
Variogram sill	13-0.15	[$\%^2$] Clay-OM
Variogram range	2000-4000	[m]
Variogram model	“Exp”	

Table 4.6 Initial parameters used for clay and organic matter content in SoilGrids30m module 4-5

In module 6 a cross-validation will be applied. The model will be split up in a training set and a validation set. In this study, 70% of the data is used for training and 30% is used for validation. The module is set to run 100 times and can therefore be denoted as a 100-fold cross-validation. Furthermore, the same semivariogram settings are used as in module 4-5 (see Table 4.7).

Initial parameter	Parameter value	Unit
Module 6	**Validation**	
Split coefficient	0.7	[-]
Number of calculation	100	[-]
Variogram nugget	5-0.1	[%^{2}] Clay-OM
Variogram sill	20-0.15	[%^{2}] Clay-OM
Variogram range	4000	[m]
Variogram model	"Exp"	Exponential

Table 4.7 Initial parameters used for clay and organic matter content in SoilGrids30m module 6

Correlation percentile threshold analysis

The correlation percentile threshold has been analyzed more extensively to find out what configuration gives the most reliable result. Increments of 5 percentile are used in the range from 0th percentile up to 95th percentile. Modules 2A until 4-5 are used for each increment. In module 4-5 all soil samples left (after outlier and non-bare soil removal in module 2B) are used to obtain the final estimation of the target variable in the area. To measure the goodness of fit of the results for each correlation percentile threshold, three performance indicators have been applied: coefficient of determination; root mean square error and mean absolute estimation error (see paragraph 3.5.3). The results of the analysis for the target variables are shown in Table 4.8. For the clay content estimation, the 15th percentile correlation threshold clearly shows the best fit of all analyzed percentiles. For the organic matter content, the 75th percentile correlation threshold shows the best fit of all analyzed percentiles. These results are also shown in Figure 4.1 and Figure 4.2 where the predicted values are plotted with the measured values of the soil samples.

Correlation percentile	Number of observations		R^2		RMSE [%]		MAE [%]	
Target variable	**Clay**	**OM**	**Clay**	**OM**	**Clay**	**OM**	**Clay**	**OM**
0th	143	143	0.90	0.61	2.7	0.32	2.3	0.24
5th	144	143	0.90	0.62	2.7	0.32	2.3	0.24
10th	144	153	0.93	0.74	2.3	0.27	2.0	0.20
15th	**146**	153	**0.97**	0.74	**2.1**	0.27	**1.9**	0.20
20th	146	155	0.89	0.72	2.7	0.28	2.3	0.22
25th	146	155	0.89	0.72	2.7	0.28	2.3	0.21
30th	146	155	0.88	0.72	2.9	0.28	2.5	0.20
35th	148	156	0.85	0.73	4.8	0.29	4.4	0.22
40th	148	156	0.91	0.76	14.0	0.28	14.0	0.20
45th	163	156	0.94	0.73	6.2	0.29	6.0	0.22
50th	163	156	0.94	0.73	6.2	0.29	6.0	0.21
55th	163	156	0.95	0.81	6.5	0.25	6.4	0.19
60th	163	157	0.93	0.82	9.1	0.24	9.0	0.19
65th	163	158	0.63	0.80	17.0	0.25	17.0	0.18
70th	163	158	0.64	0.82	17.0	0.24	17.0	0.18
75th	166	**160**	0.89	**0.83**	2.3	**0.24**	2.9	**0.17**
80th	169	161	0.66	0.80	17.0	0.26	16.0	0.19
85th	169	161	0.75	0.79	16.0	0.26	16.0	0.20
90th	169	166	0.66	0.76	17.0	0.28	17.0	0.21
95th	169	171	0.70	0.74	17.0	0.82	16.0	0.78

Table 4.8 Performance of target variable estimation for different correlation percentile thresholds SoilGrids30m

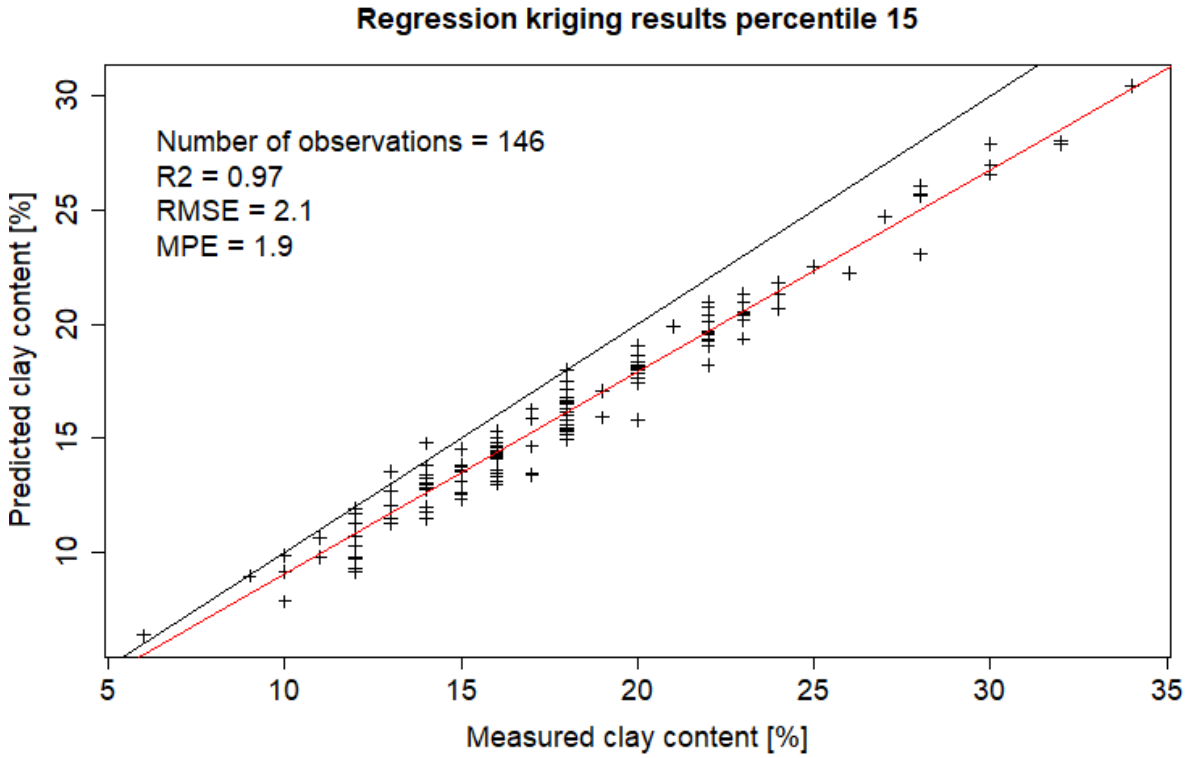


Figure 4.1 Predicted clay content regression kriging results against measured clay content soil samples for percentile 15, black line is a 1:1 ratio line and red line is the R^2 line

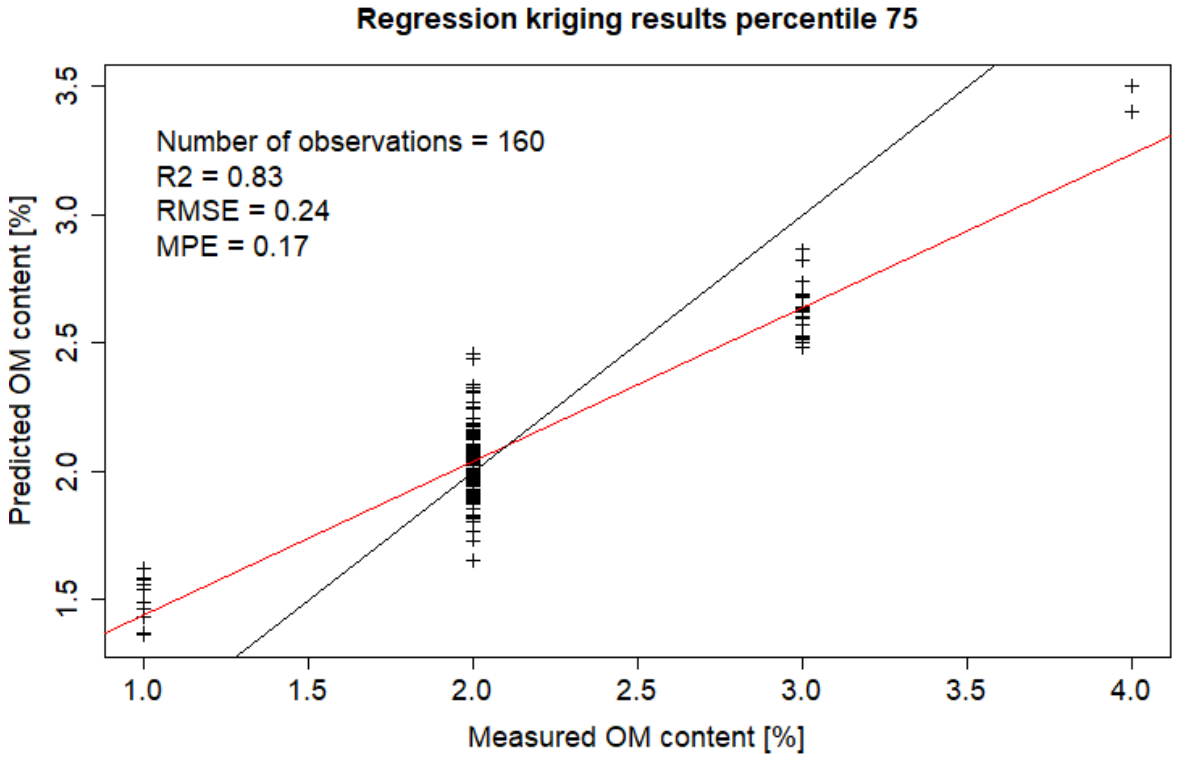


Figure 4.2 Predicted organic matter content regression kriging results against measured organic matter content soil samples for percentile 75, black line is a 1:1 ratio line and red line is the R^2 line

An important note to these results is the precision of the soil sample data, which are integers. Especially the organic matter content results clearly show the low quality of the soil sample data. In example, the root mean square error values are lower than the precision of the soil sample data. The available soil sample data are more indicative numbers than exact numbers. In addition, the variability of organic matter content in the selected soil sample set is too low. Therefore, the results of organic matter content are not significant enough to draw firm conclusions. It would be more convenient to use a simpler spatial estimation technique such as ordinary-kriging.

Results clay content

The results of clay content will be analyzed in three ways: based on the contribution of the significant explanatory variables; spatial visualization; 100-fold cross-validation results; comparison with SoilGrids250m and SoilGrids1000m. First, the contribution of significant explanatory variables. As mentioned above, the 15th percentile correlation threshold provides the best set of explanatory variables for the spatial estimation of clay content in the study area. In total 47 explanatory variables are used for the regression-kriging analysis. In module 3, the threshold of 99% variance what should be explained by the principal components has been applied. Nine principal components are needed to meet this threshold, see Figure 4.3. Principal component 1 explains the most variance in the model in which 14 explanatory variables have a significant contribution to this principal component, see Figure 4.4. The red dashed line corresponds to a uniform distributed contribution of all explanatory variables.

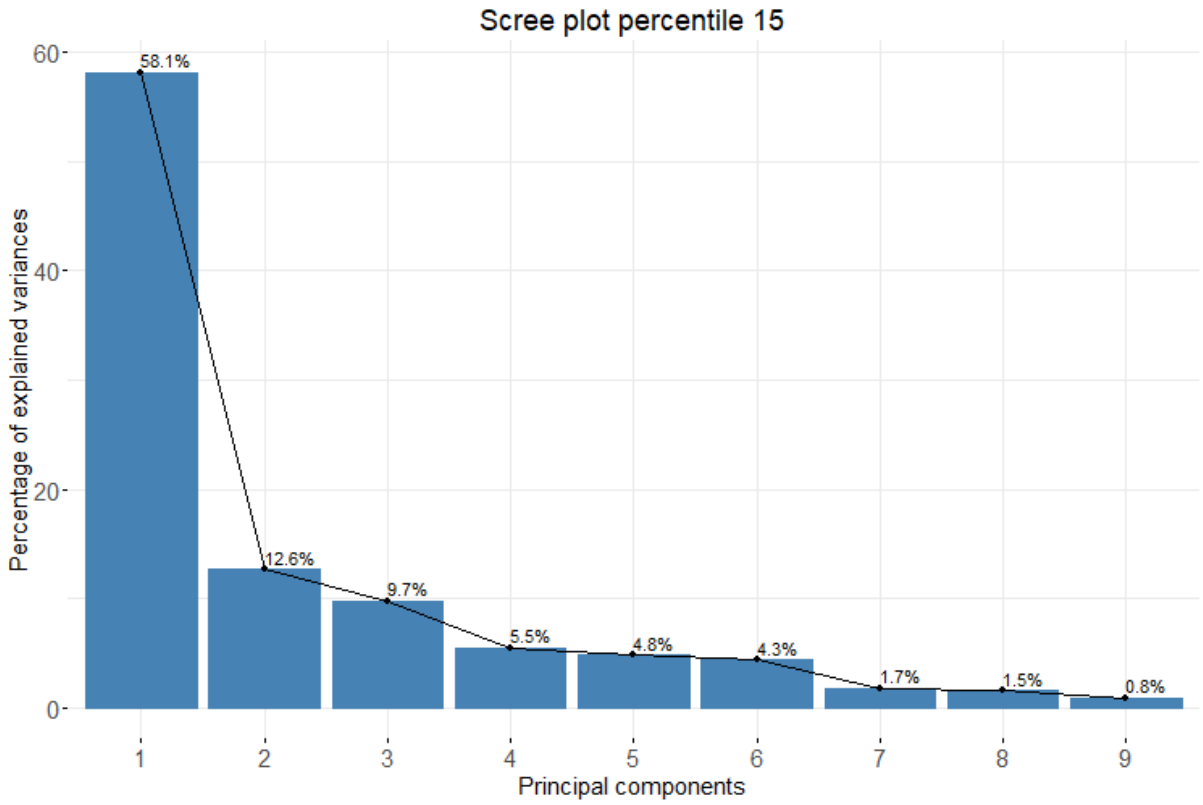


Figure 4.3 Percentage of explained variance for first nine principal components, together > 99% clay content

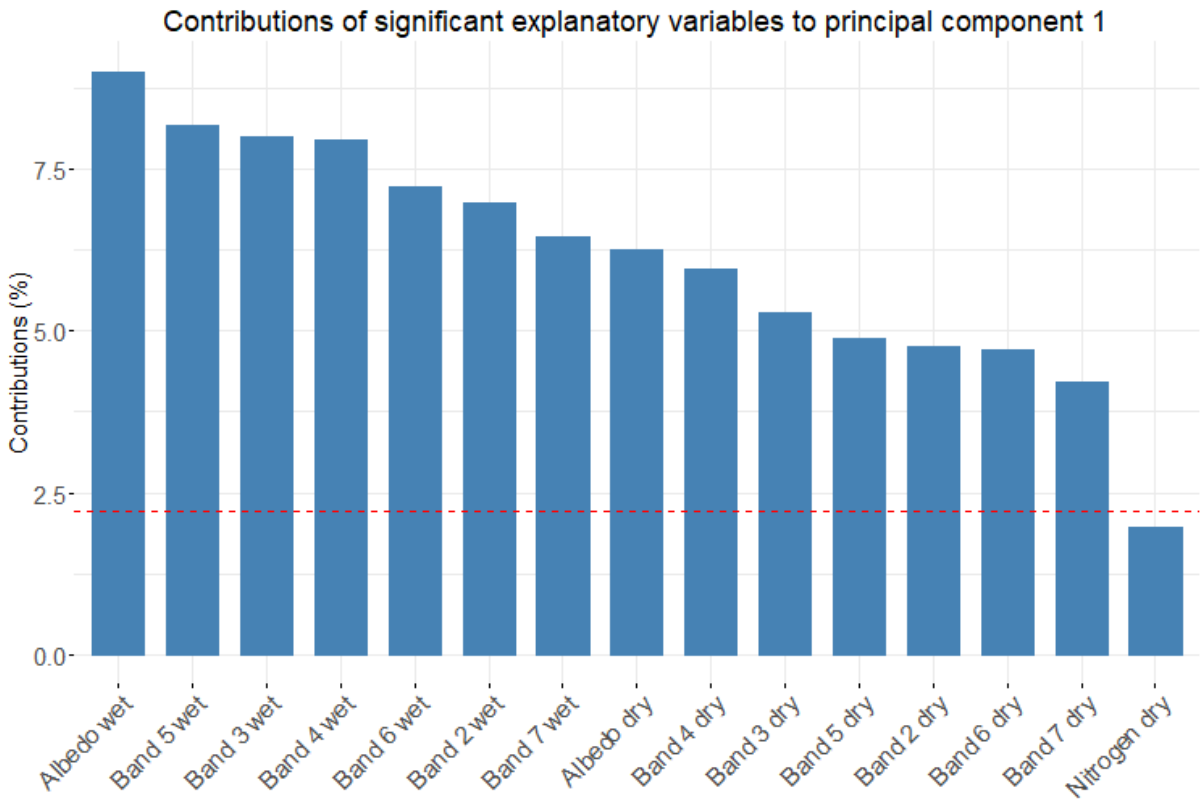


Figure 4.4 Contributions of significant explanatory variables to principal component 1 clay content percentile 15

The results of the contribution of significant explanatory variables for the other eight principal components can be found in Appendix L. Table 4.9 shows an overview of the three highest contributors per principal component and Figure 4.5 shows the total contribution of each significant explanatory variable according to all nine selected principal components.

	Total contribution	**Highest contributor**		**2nd highest contributor**		**3rd highest contributor**	
	[%]	**Name**	**[%]**	**Name**	**[%]**	**Name**	**[%]**
PC1	58.1	Albedo wet	9,0	Band 5 wet	8,2	Band 3 wet	8,0
PC2	12.6	Elevation	14,0	Nitrogen wet	13,1	Band 2 dry	12,9
PC3	9.7	Nitrogen dry	14,5	Elevation	12,3	Band 5 dry	9,7
PC4	5.5	Slope	72,4	Elevation	15,6	Band 7 dry	3,4
PC5	4.8	Band 7 dry	18,2	Slope	17,3	Band 6 dry	14,5
PC6	4.3	Elevation	36,8	Nitrogen wet	16,1	Band 7 dry	13,7
PC7	1.7	Band 7 wet	27,1	Band 6 wet	13,0	NDMI difference	10,5
PC8	1.5	Nitrogen dry	26,8	Nitrogen wet	18,4	Band 2 dry	13,9
PC9	0.8	Band 2 dry	36,2	Band 4 dry	16,4	SI dry	9,3

Table 4.9 Three highest contributors per principal component clay content percentile 15

According to Figure 4.5, the relief parameters elevation and slope are the most important explanatory variables for the spatial estimation of clay content in the study area. As mentioned in paragraph 3.5.2, numerous studies already have shown a relationship between relief and soil properties (Pachepsky et al., 2001; Sobieraj et al., 2002; Ceddia et al., 2009). In example, the elevation and the slope of a landscape controls the distribution of water and sediment throughout a region. However, the slope of the area is negligible (see Figure 4.7). The elevation in the study area shows a clear gradient from East to West of about 3 meters (see Figure 4.8) which could be related to the clay content in the study area.

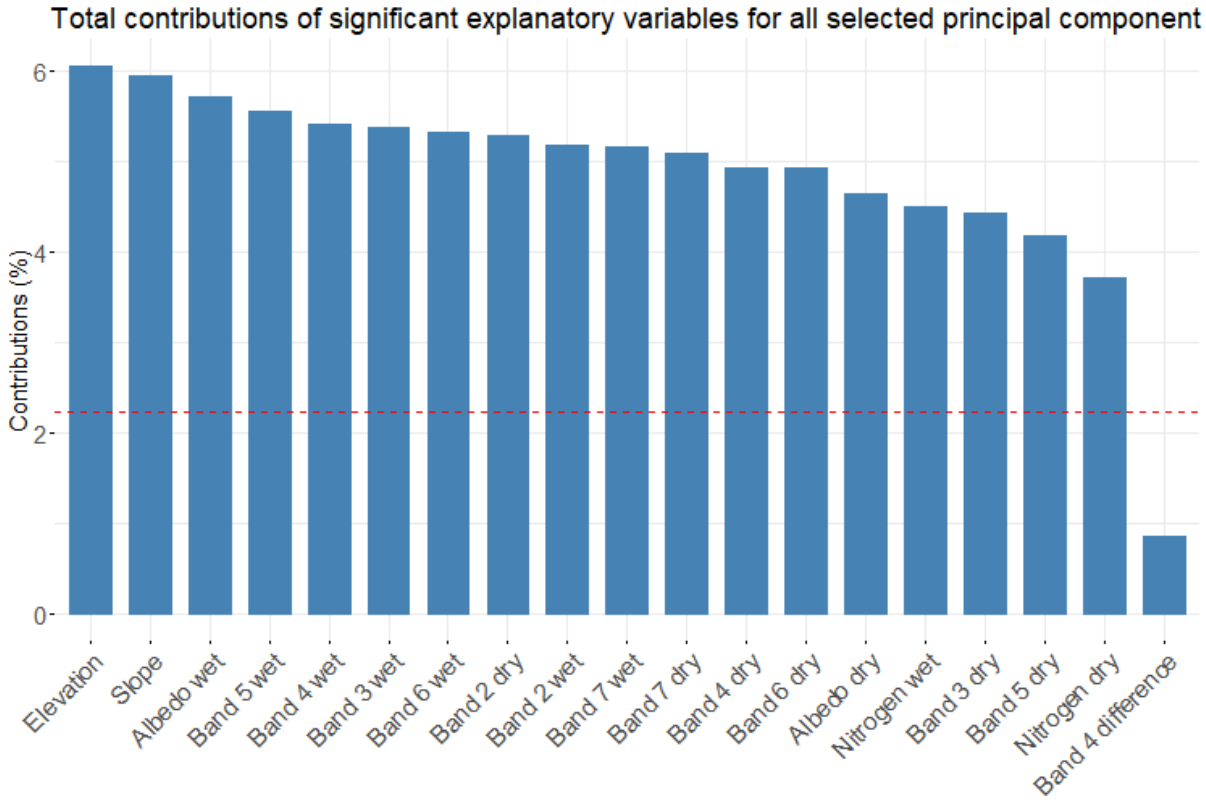


Figure 4.5 Total contribution of significant explanatory variables for principal components 1 to 9 clay content percentile 15

Figure 4.6 shows the correlation at the soil sample locations between the clay content and the selected explanatory variables. According to Figure 4.6, the correlation between clay content and the elevation/slope is relatively low. As mentioned in paragraph 3.5.3, one of the drawbacks of principal components is the decoupling of the relation between explanatory variables and target variable. The relative low correlation between elevation/slope and clay content and the high contribution of elevation/slope to the spatial estimation of clay content contradicts each other.

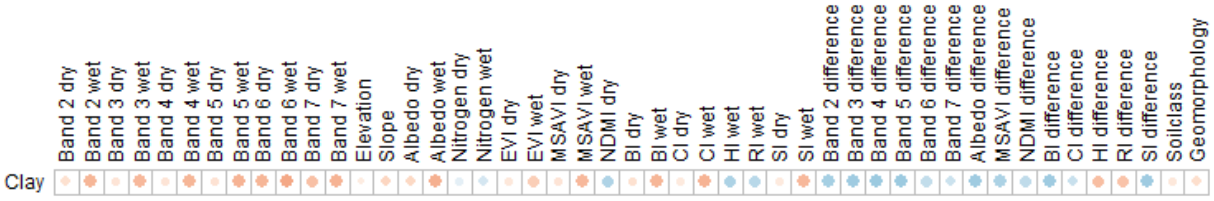


Figure 4.6 Correlation plot between clay content and the selected explanatory variables, dark red is a strong negative correlation and dark blue is a strong positive correlation

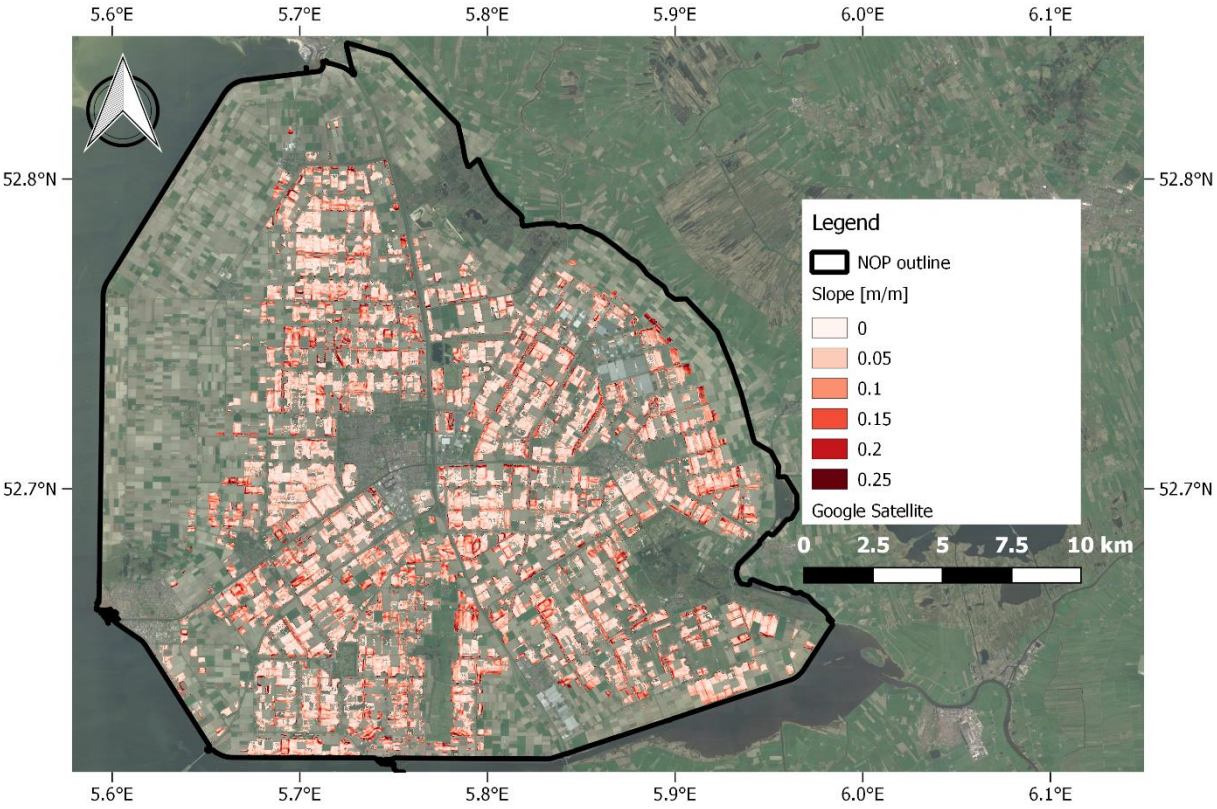


Figure 4.7 Slope map study area restricted to estimated clay content values

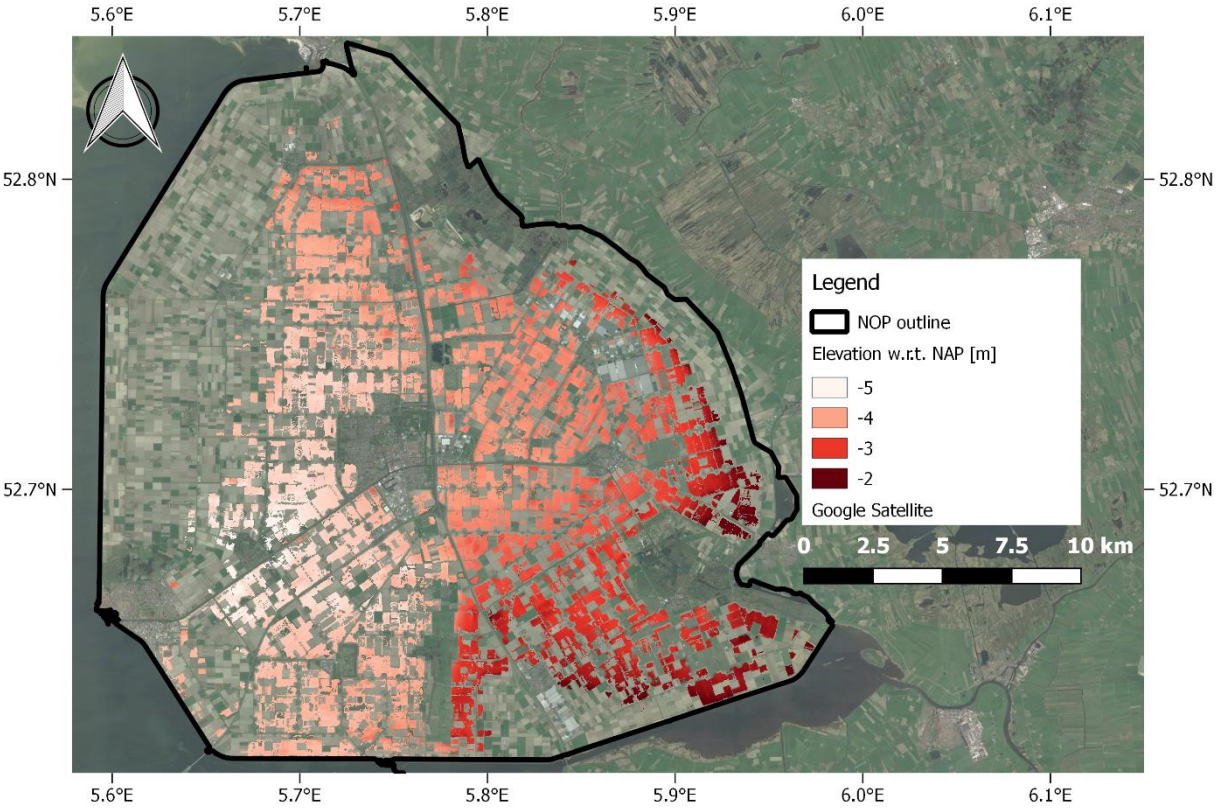


Figure 4.8 Elevation map study area restricted to estimated clay content values

Another interesting result is the absence of differenced explanatory variables (moist vs. dry date) as contributor to the spatial estimation of clay content in the study area. As stated in paragraph 3.5.2, the differenced explanatory variables could be a logical indicator to see differences in reflectance between clayey soils and sandy soils due to a difference in water permeability. The lack of influence of the differenced explanatory variables has probably to do with a too low contrast between the two dates used. At satellite overpass time, the top layer of the soil could already have been dried out as the satellite only measure reflectance of the top soil. Even though, the moist date clearly shows a higher contribution to the spatial estimation of clay content in the study area than the dry date.

The second analysis of the clay content estimates will be done based on visualization. The visualization has been done with module 4-5 where the regression-kriging model used all soil samples located at bare soils and without outliers. The predicted clay content in the study area is shown in Figure 4.9 and is based on 146 soil samples. The statistical summary of the estimated clay content throughout the study area is shown in Table 4.10.

Mean	Standard deviation	Minimum	Maximum
16.18	4.12	2.11	36.32

Table 4.10 Statistical summary of estimated clay content values in the study area in percentages

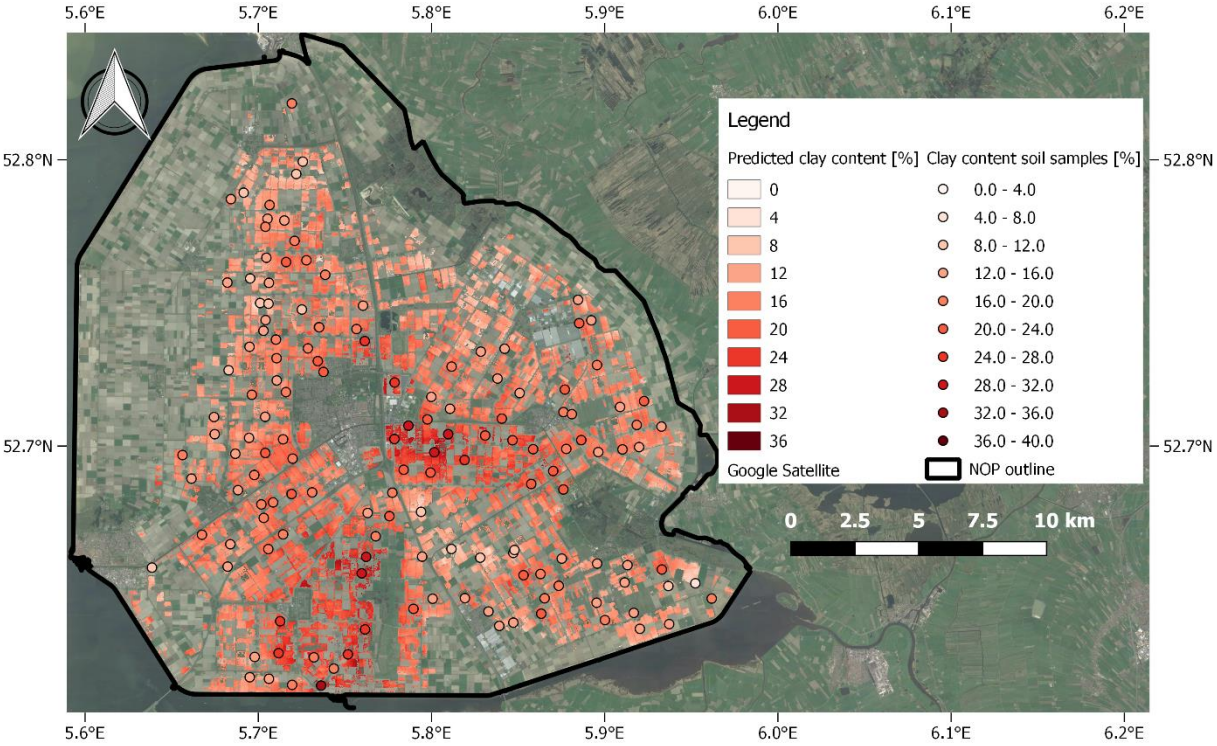


Figure 4.9 Predicted clay content in the study area along with the used soil samples

The regression-kriging model uses a semivariogram to describe the degree of spatial dependence of the target variable residuals. The results of the semivariogram can be used as indicator to determine if the target variable has a high, moderate or low spatial correlation in the study area. According to Cambardella et al. (1994), a low nugget to sill ratio indicates a high spatial correlation and vice versa. The semivariogram of the clay content residuals is fitted with an exponential model with a nugget of 3.1%^{2}, a sill of 14.1%^{2} and a range of 1353 meters. The nugget to sill ratio is quite low (22%) which indicates a high spatial correlation of clay content in the area. Therefore, a large part of the variance of the clay content in the area can be assigned to a spatial dependency.

Due to the presence of a nugget, kriging as a spatial interpolation technique is not an exact estimator anymore (Clarck, 2010). If the kriging model would be fixed to have a nugget of zero the predicted clay content map would show discontinuities at the soil sample locations. The discontinuities would visually give an unreliable feeling about the results. Clarck (2010) quoted: *"The less we trust our data the more confident we get in the results."* This quote clearly plays a role in the presence of a nugget effect or not. The results with a nugget effect visually give better results and therefore more confidence in the outcome of the model with as counterpart the loss of kriging as an exact spatial interpolation technique.

The third analysis is based on the 100-fold cross-validation. The validation results are shown in Figure 4.10 until Figure 4.12. For validation, 70% of the data (102 soil samples) has been used to train the model and 30% of the data (44 soil samples) has been used to validate the obtained results. The results shown are based on 100 runs of the validation module. In each run, the model selects randomly 44 validation soil samples. The coefficient of determination for 100 runs has a mean value of 0.37, which means that the explanatory variables on average only account for 37% of the variance in the model. In Table 4.11, the amount of variance explained by the explanatory variables is shown for all three SoilGrids models. The amount of variance explained by the explanatory variables of the SoilGrids30m model improves compared to the SoilGrids1000m model. However, compared to the SoilGrids250m the influence of the explanatory variables is significantly lower.

Target variable	SoilGrids30m	SoilGrids250m	SoilGrids1000m
Clay content	37.0%	72.6%	24.4%

Table 4.11 Amount of variance explained by the explanatory variables for clay content

The mean absolute estimation error for 100 runs has a mean value of 3.58%, which means that the estimated results on average are within a range of ± 3.58%. The root mean square error for 100 runs has a mean value of 4.56%, which means that the estimated results on average are within a range of ± 4.56%.

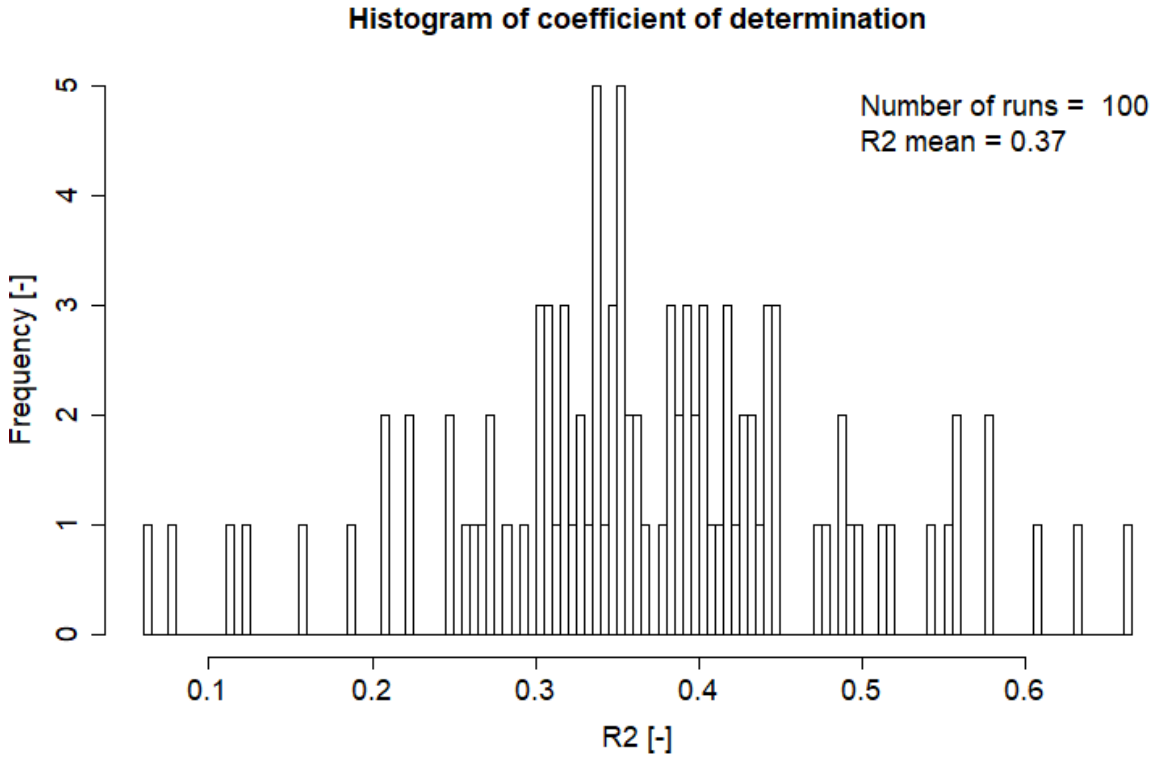


Figure 4.10 Coefficient of determination of validation set, 100 runs clay content

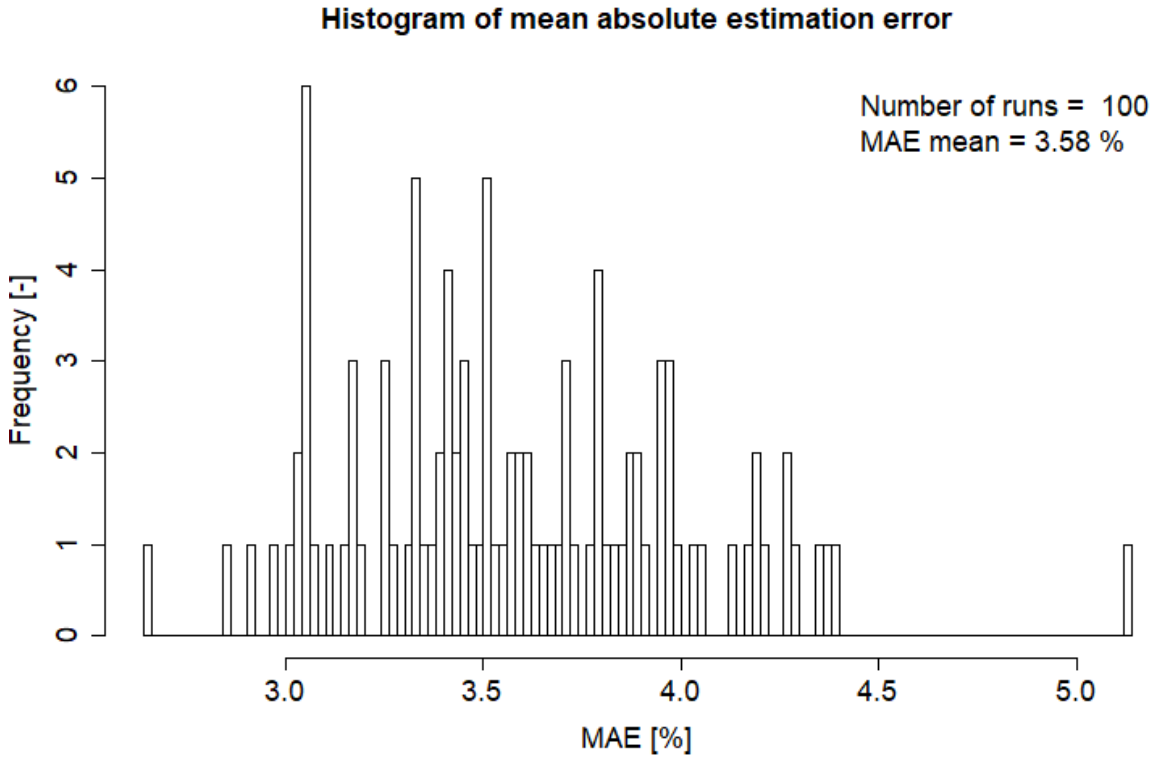


Figure 4.11 Mean absolute estimation error of validation set, 100 runs clay content

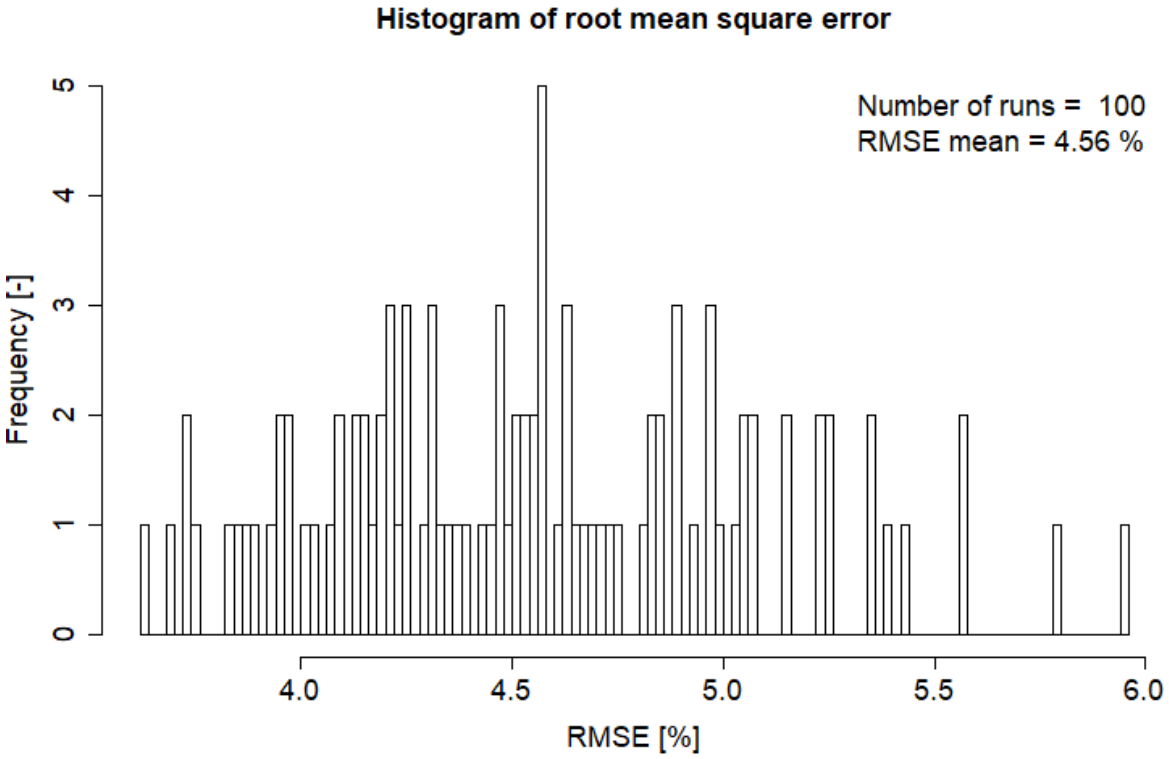


Figure 4.12 Root mean square error of validation set, 100 runs clay content

The fourth analysis is the comparison between SoilGrids30m, SoilGrids250m and SoilGrids1000m clay content estimates. In Figure 4.13 and Figure 4.14, the spatial prediction of clay content of respectively SoilGrids250m and SoilGrids1000m model are shown.

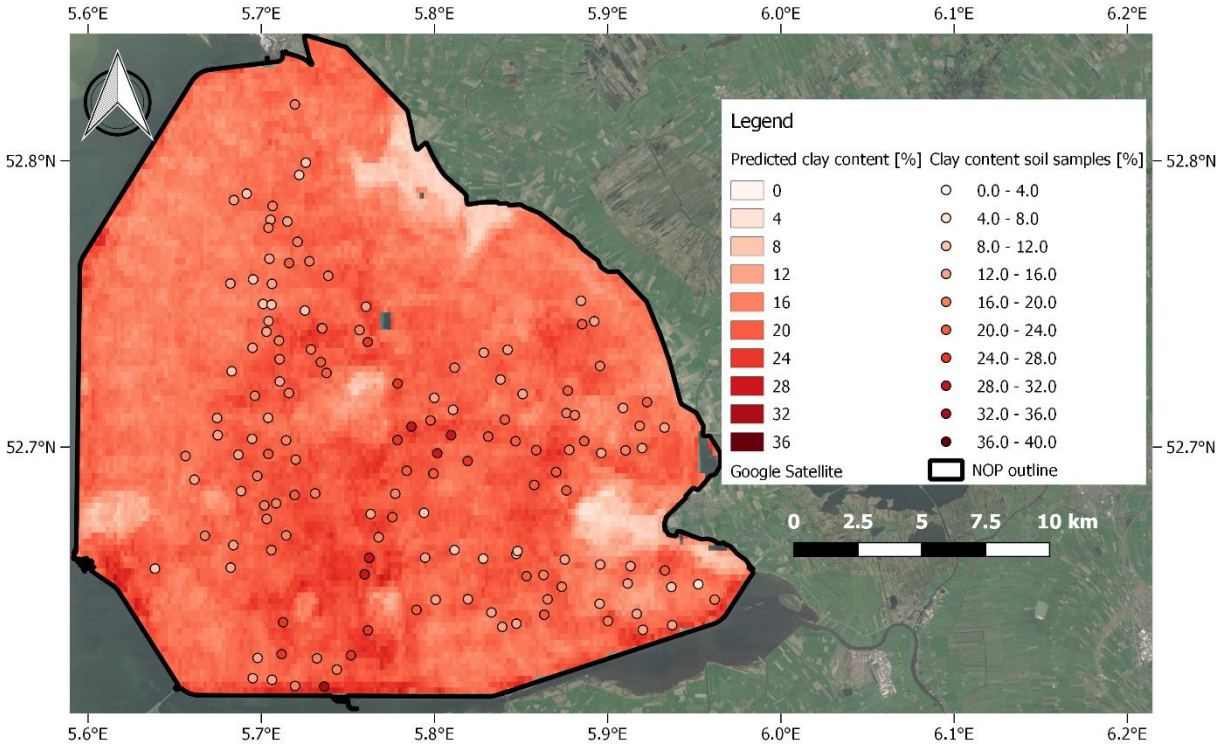


Figure 4.13 SoilGrids250m predicted clay content in the study area along with the used soil samples for SoilGrids30m

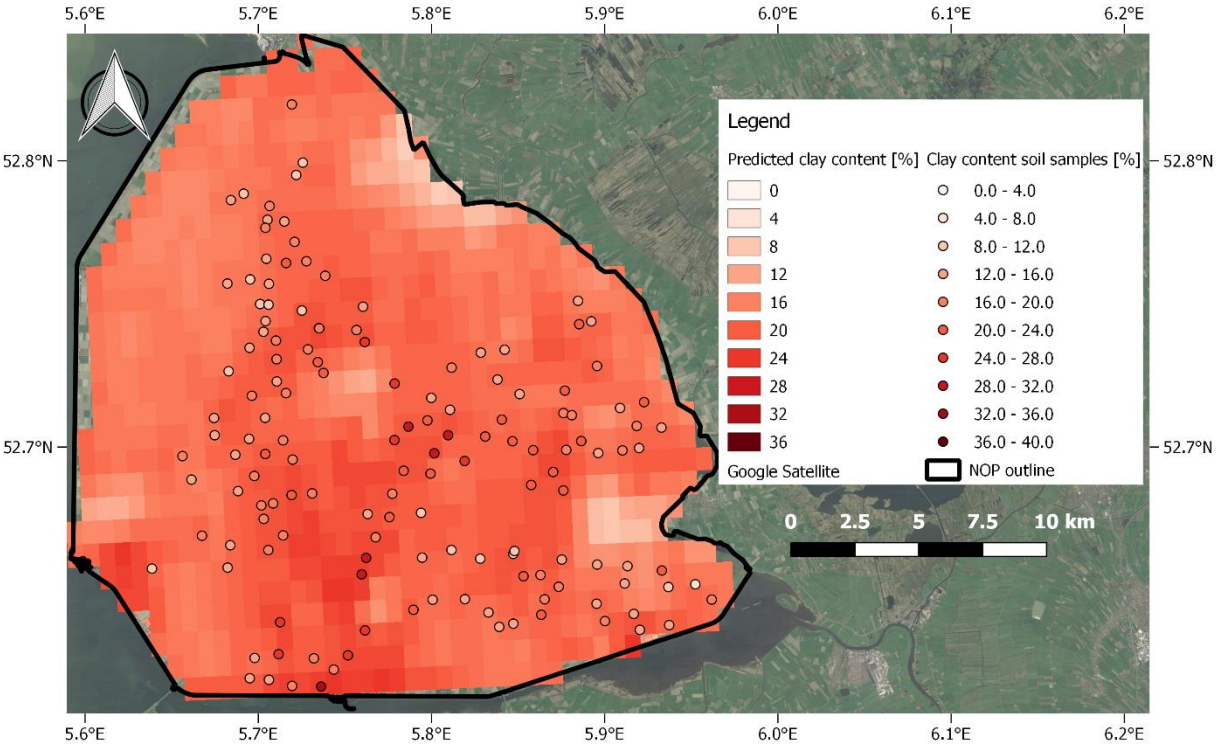


Figure 4.14 SoilGrids1000m predicted clay content in the study area along with the used soil samples for SoilGrids30m

In Table 4.12 and in Figure 4.15 and Figure 4.16, the performance of SoilGrids250m and SoilGrids1000m are shown. The performance of SoilGrids30m, in table 1.11, are the mean values of the 100-fold cross-validation. According to the performance values, the SoilGrids30m model slightly improves the spatial estimation of clay content in the study area. Interestingly, the coarser SoilGrids1000m model is a better representation of the clay content in the study area than the finer SoilGrids250m model.

Model	R^2	RMSE [%]	MAE [%]
SoilGrids30m	0.37	4.6	3.6
SoilGrids250m	0.14	4.8	3.8
SoilGrids1000m	0.18	4.7	3.8

Table 4.12 Performance of all three SoilGrids models clay content

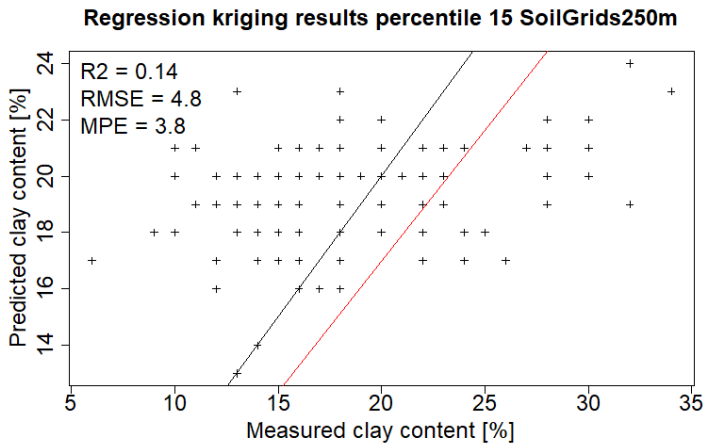


Figure 4.15 SoilGrids250m performance figure clay content

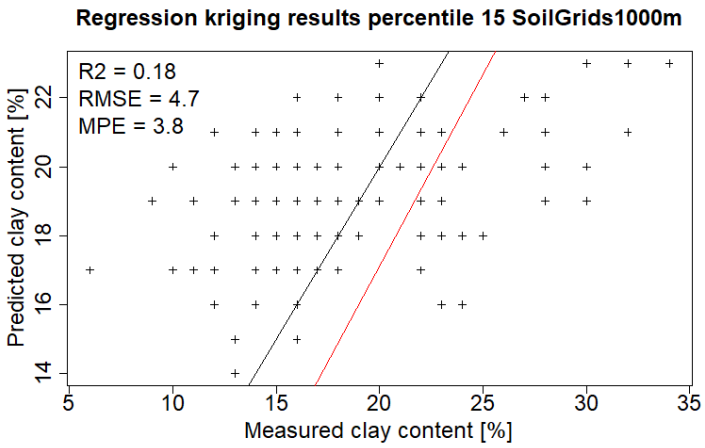


Figure 4.16 SoilGrids1000m performance figure clay content

Results organic matter content

The results of organic matter content will be analyzed in three ways: based on the contribution of the significant explanatory variables; spatial visualization; 100-fold cross-validation results; comparison with SoilGrids250m and SoilGrids1000m. First, the contribution of significant explanatory variables. As mentioned above, the 75th percentile provides the best set of explanatory variables for the spatial estimation of organic matter content in the study area. In total 14 explanatory variables are used for the regression-kriging analysis. In module 3, the threshold of 99% variance explained by the principal components has been applied. Six principal components are needed to meet this threshold, see Figure 4.17. Principal component 1 explains the most variance in the model in which four explanatory variables have a significant contribution to this principal component, see Figure 4.18. The red dashed line corresponds to a uniform distributed contribution of all explanatory variables.

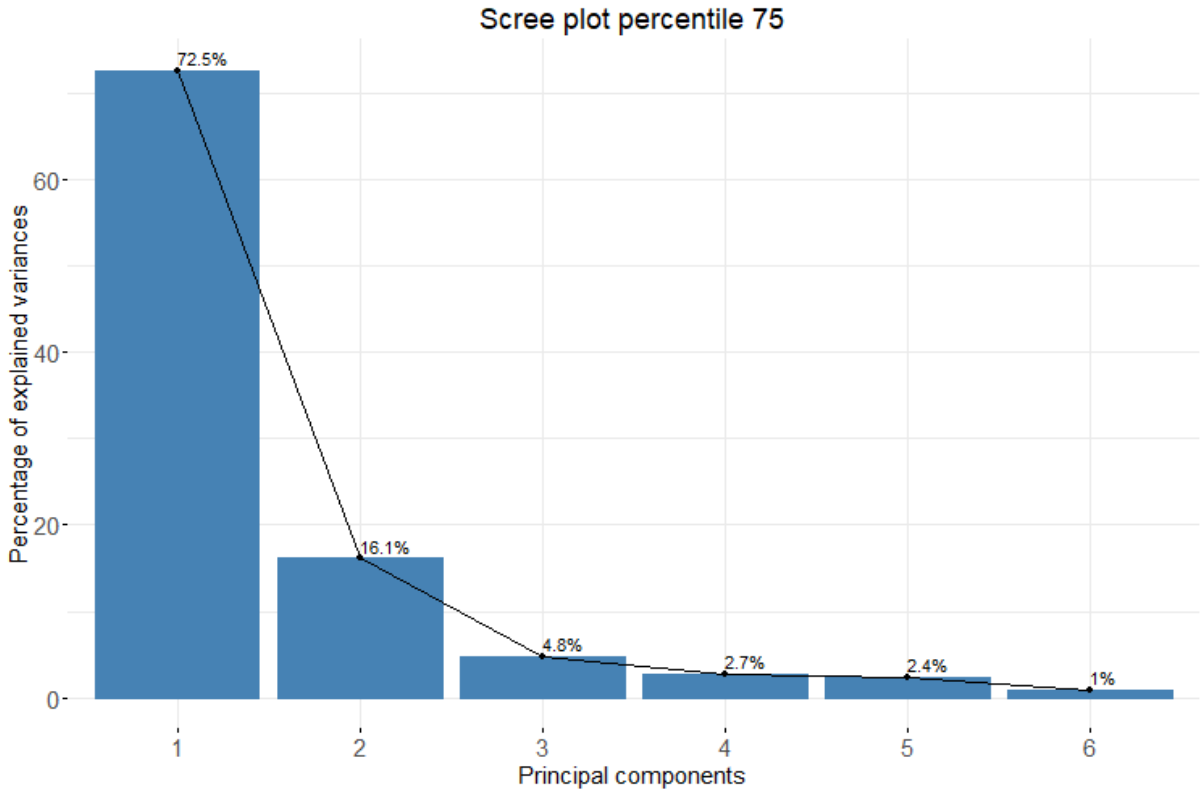


Figure 4.17 Percentage of explained variance for first six principal components, together > 99% organic matter content

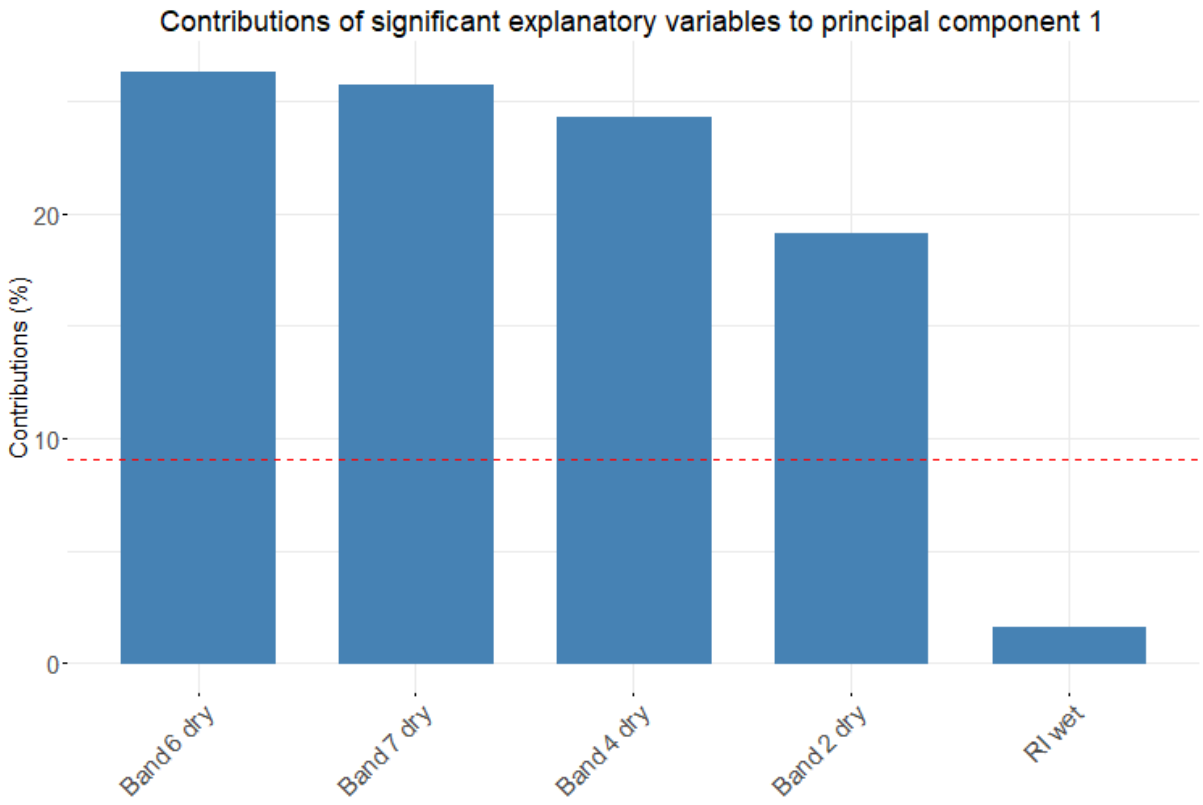


Figure 4.18 Contributions of significant explanatory variables to principal component 1 organic matter content percentile 75

The contribution of significant explanatory variables figures to the other five principal components can be found in Appendix L. Table 4.13 shows an overview of the three highest contributors per principal component and Figure 4.19 shows the total contribution of each explanatory variable according to all six selected principal components.

	Total contribution	**Highest contributor**		**2nd highest contributor**		**3rd highest contributor**	
	[%]	**Name**	**[%]**	**Name**	**[%]**	**Name**	**[%]**
PC1	72.5	Band 6 dry	26.3	Band 7 dry	25.7	Band 4 dry	24.3
PC2	16.1	Band 2 dry	34.4	Band 6 dry	22.5	Band 7 dry	18.3
PC3	4.8	Band 2 dry	36.3	Band 4 dry	27.5	NDMI difference	18.6
PC4	2.7	RI wet	28.6	NDMI difference	27.1	RI difference	26.0
PC5	2.4	Band 6 dry	24.7	RI wet	22.4	Band 7 dry	14.6
PC6	1.0	Band 7 dry	35.8	NDMI difference	35.2	Band 6 dry	12.4

Table 4.13 Three highest contributors per principal component for organic matter content percentile 75

According to Figure 4.19, the most important explanatory variables are both short-wave infrared bands, the blue band and the red band from Landsat 8. Many organic components can be assigned to absorption bands in the short-wave infrared region of the electromagnetic spectrum (Summers et al., 2011; Rossel & Behrens, 2010; Ertlen et al., 2010). This would be an explanation of the high contribution of band 6 and band 7 (Landsat 8 SWIR bands). In addition, the blue and red band play a significant role in the spatial estimation of organic matter content. According to He et al. (2009), darker soils can usually be related to higher organic matter content. This would mean a negative correlation between organic matter content and the visible bands of the electromagnetic spectrum. According to the data obtained in module 2A this is a valid statement. The high contribution of the blue and red band show that the color of the soil is an important component in estimating the organic matter content in the study area.

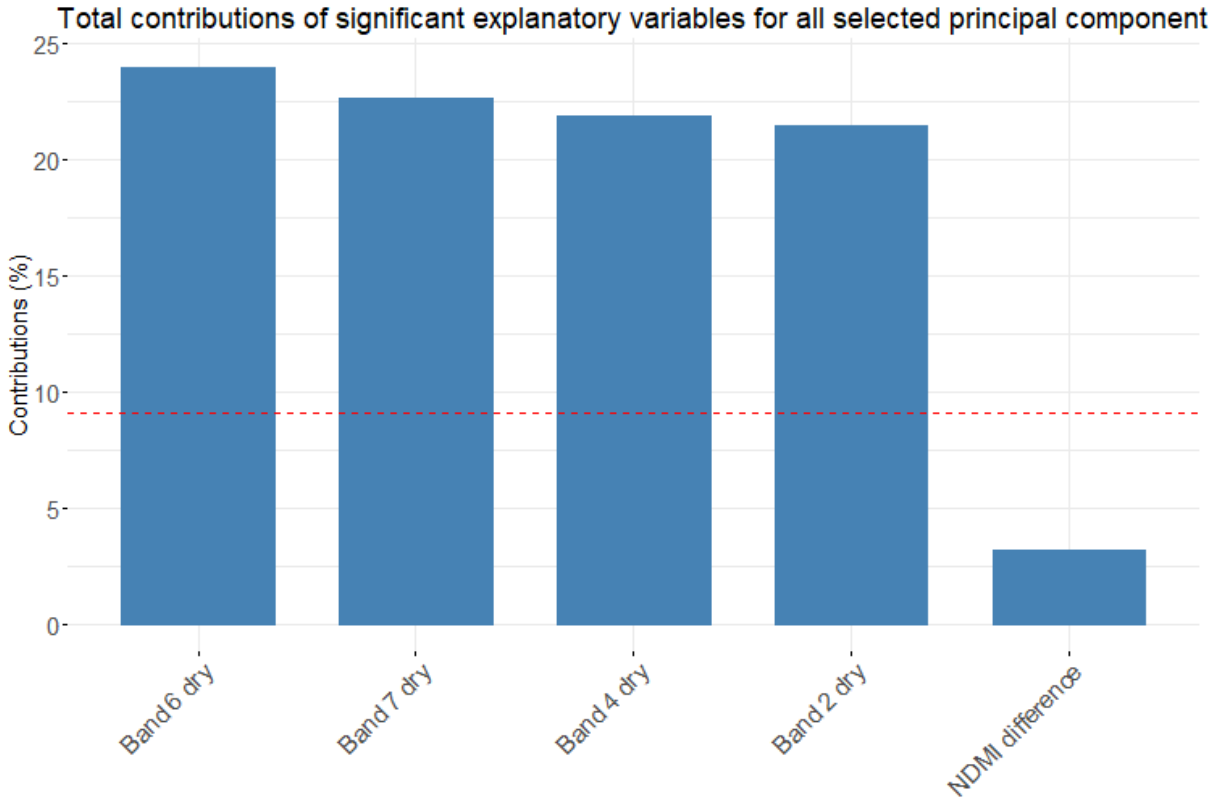


Figure 4.19 Total contribution of significant explanatory variables for principal components 1 to 6 organic matter content percentile 75

The second analysis of the organic matter content estimates will be done based on visualization. The visualization has been done with module 4-5 where the regression-kriging model used all soil samples located at bare soils and without outliers. The predicted organic matter content in the study area is shown in Figure 4.20 and is based on 160 soil samples. The statistical summary of the estimated organic matter content throughout the study area is shown in Table 4.14.

Mean	Standard deviation	Minimum	Maximum
2.101	0.274	1.185	3.520

Table 4.14 Statistical summary of estimated organic matter content values in the study area in percentages

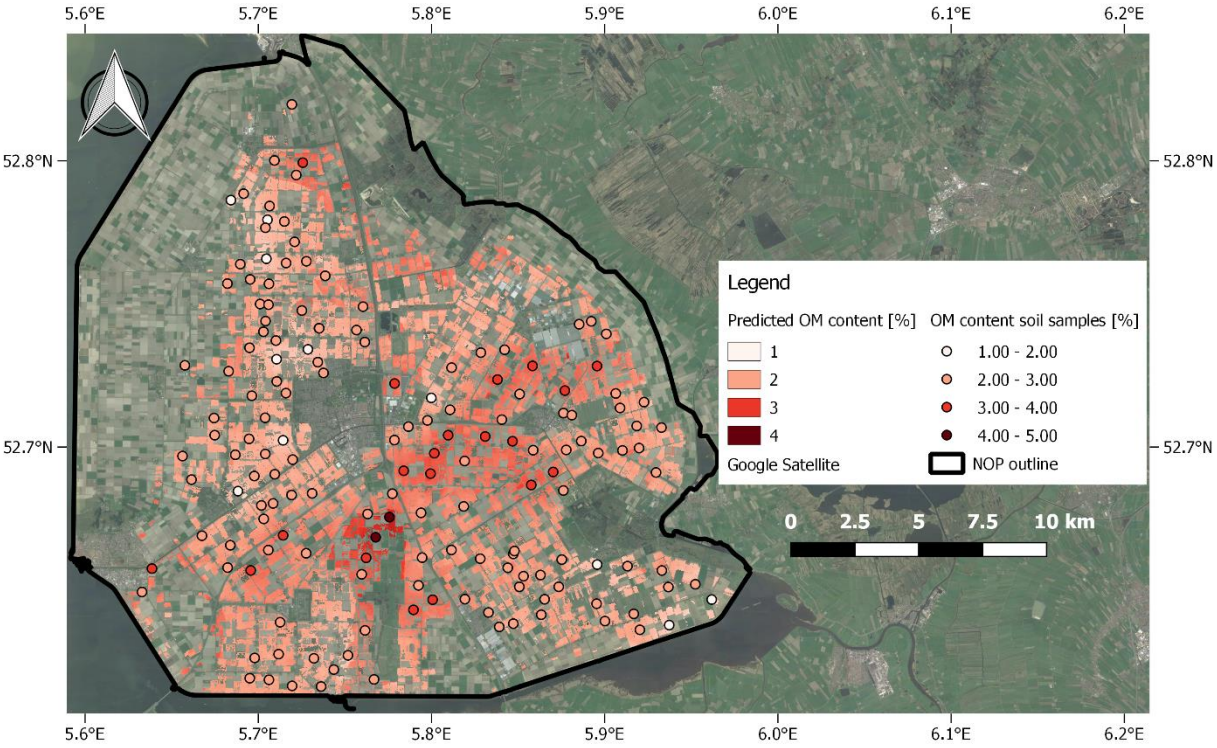


Figure 4.20 Predicted organic matter content in the study area along with the used soil samples

The regression-kriging model uses a semivariogram to describe the degree of spatial dependence of the target variable residuals. The results of the semivariogram can be used as indicator to determine if the target variable has a high, moderate or low spatial correlation in the study area. According to Cambardella et al. (1994), a low nugget to sill ratio indicates a high spatial correlation and vice versa. The semivariogram of the residuals was fitted with an exponential model with a nugget of 0.091%^{2}, a sill of 0.133%^{2} and a range of 1779 meters. The semivariogram shows a quite high nugget to sill ratio (68.4%) which indicates a moderate to low spatial correlation of organic matter content in the study area (Cambardella et al., 1994). Therefore, only a small part of the variance of the organic matter content in the area can be assigned to a spatial dependency. This can be well explained by the low variability in organic matter content in the study area. The soil samples used to train the model are in the range of 1%-4% of which the majority of soil samples have an organic matter content of 2%. Due to the low variability in the available soil samples, the model cannot find a clear spatial pattern. Also for the organic matter content, the nugget effect will introduce errors at the soil sample locations as explained with the clay content results.

The third analysis is based on the 100-fold cross-validation. The validation results are shown in Figure 4.21 until Figure 4.23. For validation, 70% of the data (112 soil samples) has been used to train the model and 30% of the data (48 soil samples) has been used to validate the obtained results. The results shown are based on 100 runs of the validation module. In each run, the model selects randomly 48 validation soil samples. The coefficient of determination for 100 runs has a mean value of 0.133, which means that the explanatory variables on average only account for 13.3% of the variance in the model. In Table 4.15, the amount of variance explained by the explanatory variables is shown for all three SoilGrids models. For the SoilGrids30m model, the percentage is related to organic matter content. For the other two models, the percentages are related to soil organic carbon. The amount of variance explained by the explanatory variables of the SoilGrids30m model did not improve compared to both models.

Target variable	SoilGrids30m	SoilGrids250m	SoilGrids1000m
Organic matter/ organic carbon content	13.3%	68.8 %	22.9%

Table 4.15 Amount of variance explained by the explanatory variables for organic matter/organic carbon content

The mean absolute estimation error for 100 runs has a mean value of 0.36%, which means that the estimated results on average are within a range of ± 0.36%. The root mean square error for 100 runs has a mean value of 0.50%, which means that the estimated results on average are within a range of ± 0.50%.

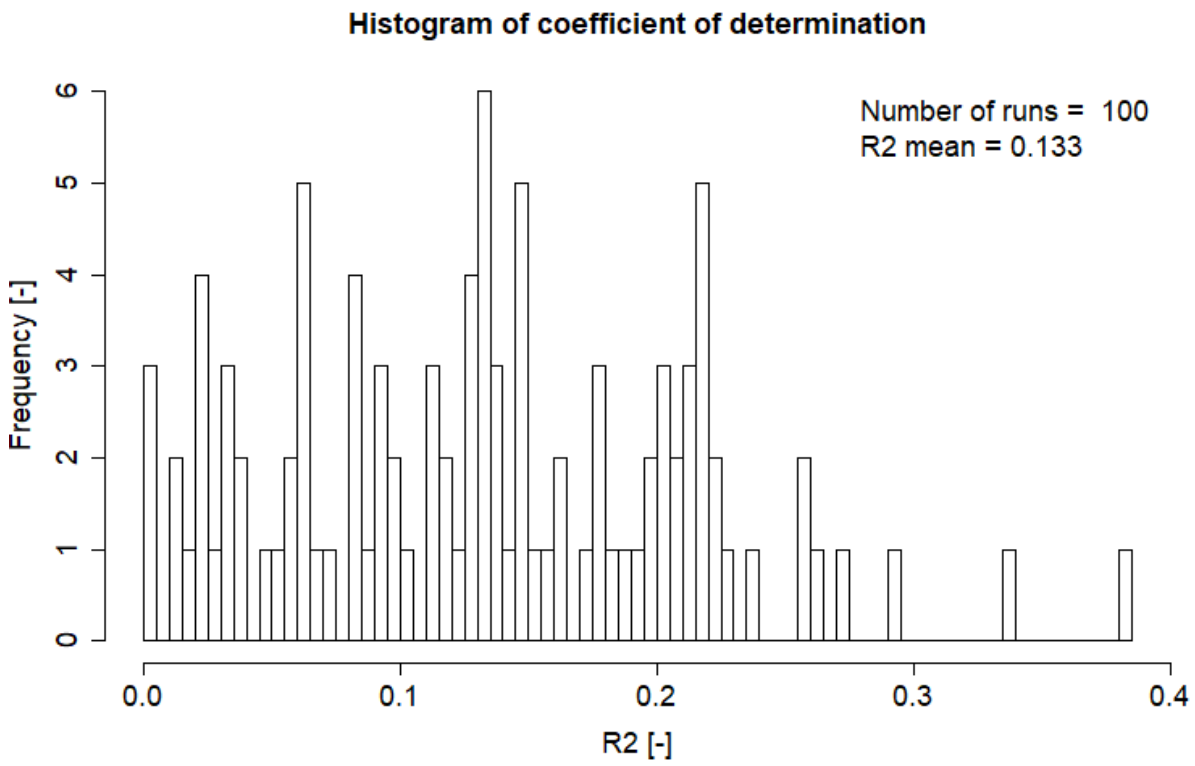


Figure 4.21 Coefficient of determination of validation set, 100 runs organic matter content

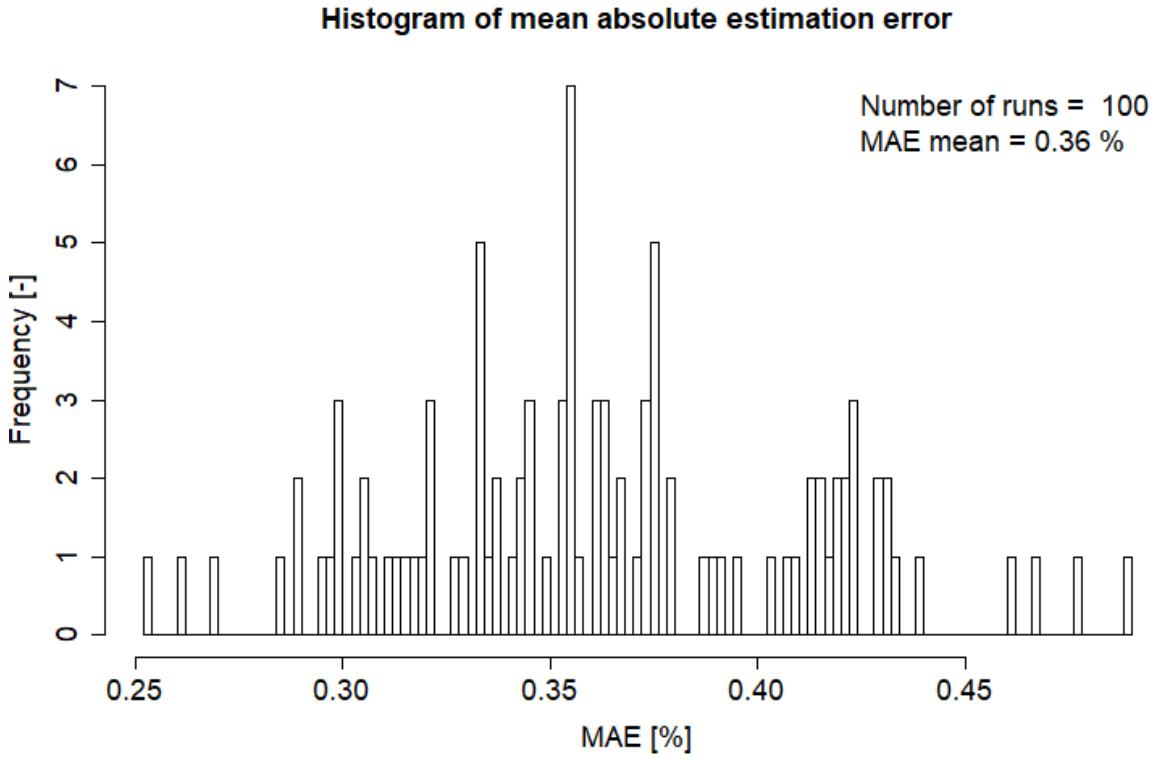


Figure 4.22 Mean absolute estimation error of validation set, 100 runs organic matter content

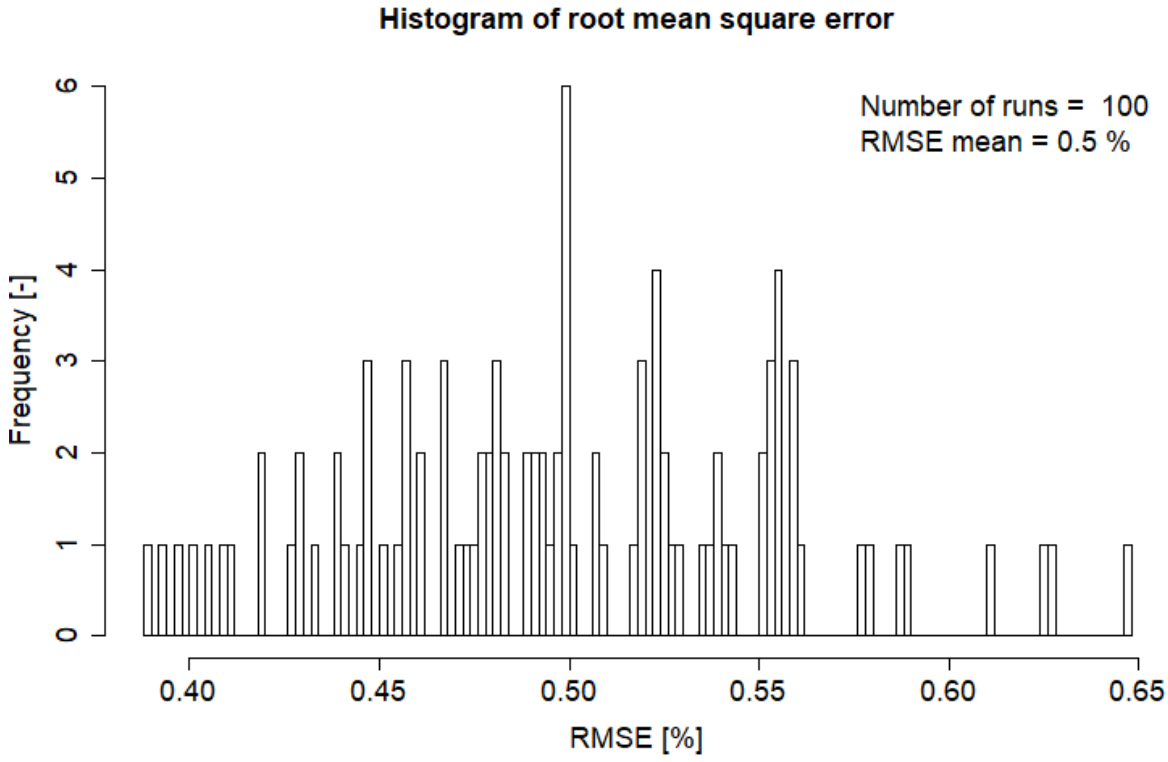


Figure 4.23 Root mean square error of validation set, 100 runs organic matter content

The fourth analysis is the comparison between SoilGrids30m, SoilGrids250m and SoilGrids1000m clay content estimates. The SoilGrids250m and SoilGrids1000m models do

not estimate organic matter content but organic carbon content. To obtain organic matter content from organic carbon content, a multiplication factor of two has been applied to the organic carbon content (Pribyl, 2010). In Figure 4.24 and Figure 4.25, the spatial prediction of organic matter content of respectively SoilGrids250m and SoilGrids1000m model are shown.

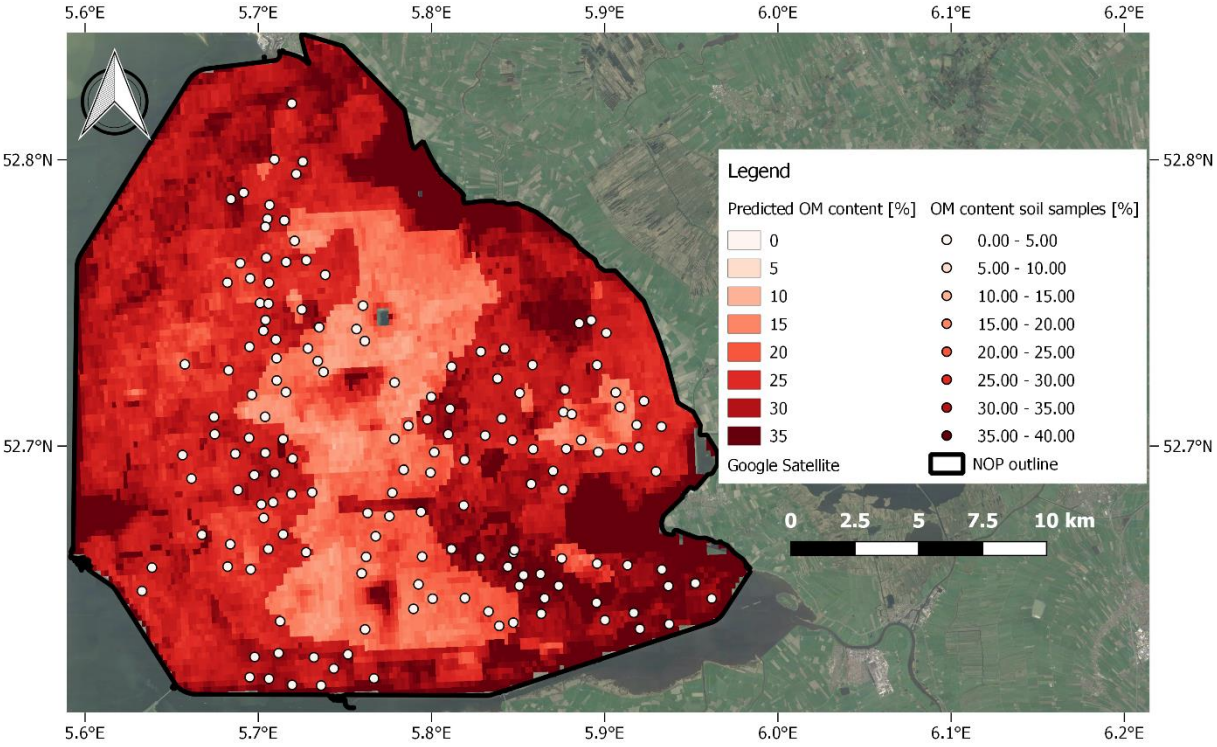


Figure 4.24 SoilGrids250m predicted organic matter content in the study area along with the used soil samples for SoilGrids30m

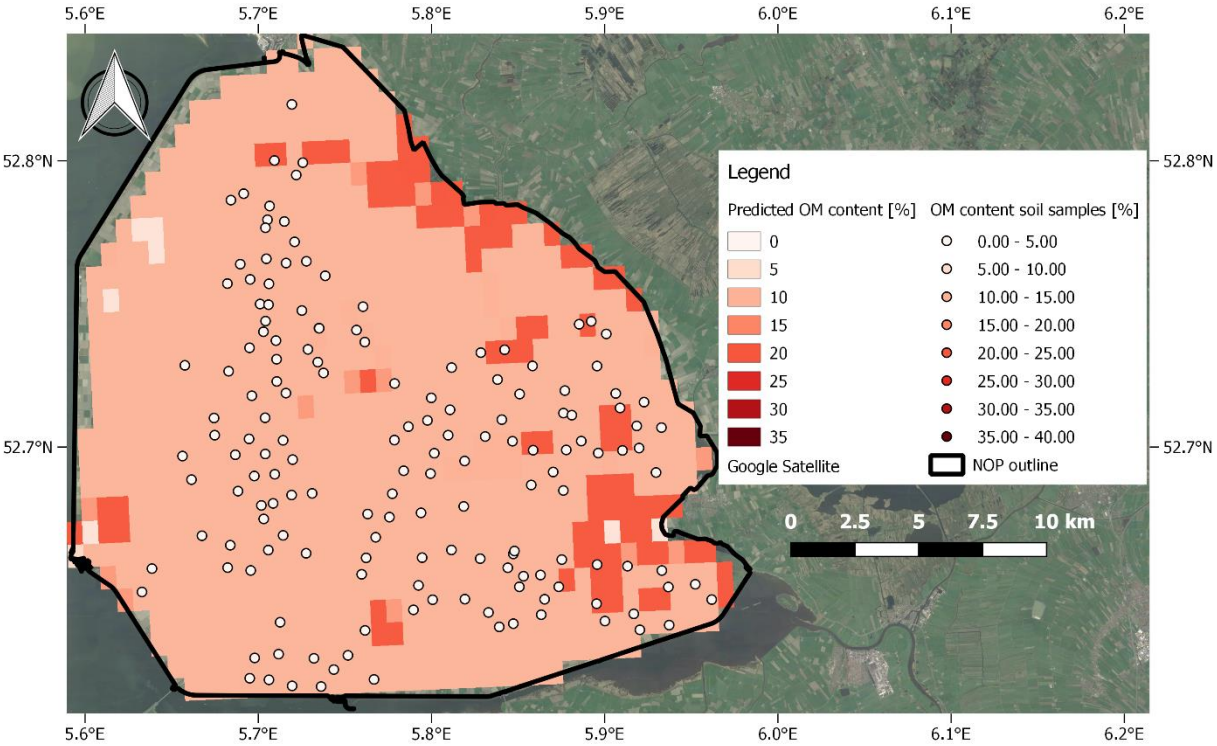


Figure 4.25 SoilGrids1000m predicted organic matter content in the study area along with the used soil samples for SoilGrids30m

In Table 4.16 and in Figure 4.26 and Figure 4.27, the performance of SoilGrids250m and SoilGrids1000m are shown. The performance of SoilGrids30m, in table 1.11, are the mean values of the 100-fold cross-validation. According to the performance values, the SoilGrids250m and SoilGrids1000m models give a bad spatial estimation of organic matter content in the study area. Interestingly, the coarser SoilGrids1000m model is a better representation of the organic matter content in the study area than the finer SoilGrids250m model.

Model	R^2	RMSE [%]	MAE [%]
SoilGrids30m	0.13	0.50	0.36
SoilGrids250m	0.01	24.31	23.48
SoilGrids1000m	0.01	8.42	8.09

Table 4.16 Performance of all three SoilGrids models

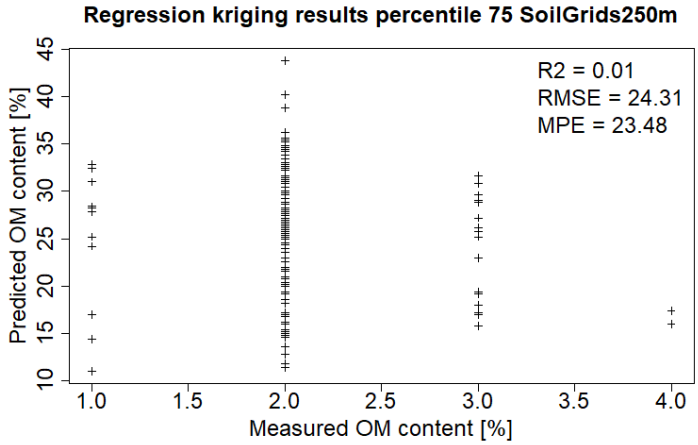


Figure 4.26 SoilGrids250m performance figure organic matter content

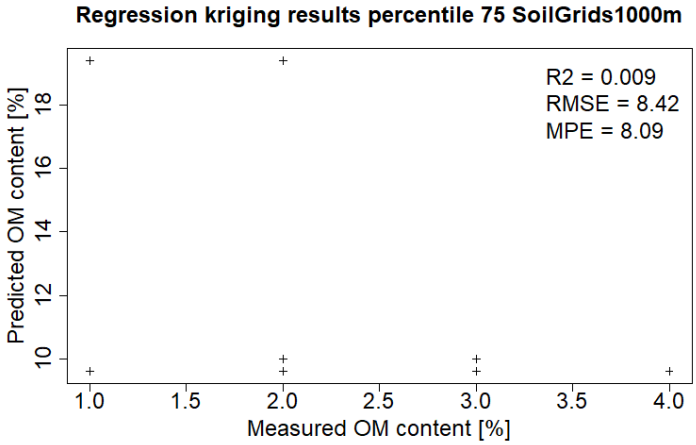


Figure 4.27 SoilGrids1000m performance figure organic matter content

4.1.2. Spatial estimation of soil water holding capacity

In this subparagraph, the spatial estimation of soil water holding capacity will be analyzed with help of some theoretical figures. The rate of soil water holding capacity is mainly determined by soil texture and organic matter content. In this study, the soil water holding capacity will be estimated only based on clay content and organic matter content obtained from the SoilGrids30m model. With help of pedotransfer functions the clay content and organic matter content can be converted to soil hydraulic properties from which the soil water holding capacity can be determined, see paragraph 3.5.4 for more details. In Figure 4.28, the estimated soil water holding capacity content is shown for the study area. In Table 4.17 the statistical summary of the soil water holding capacity is shown.

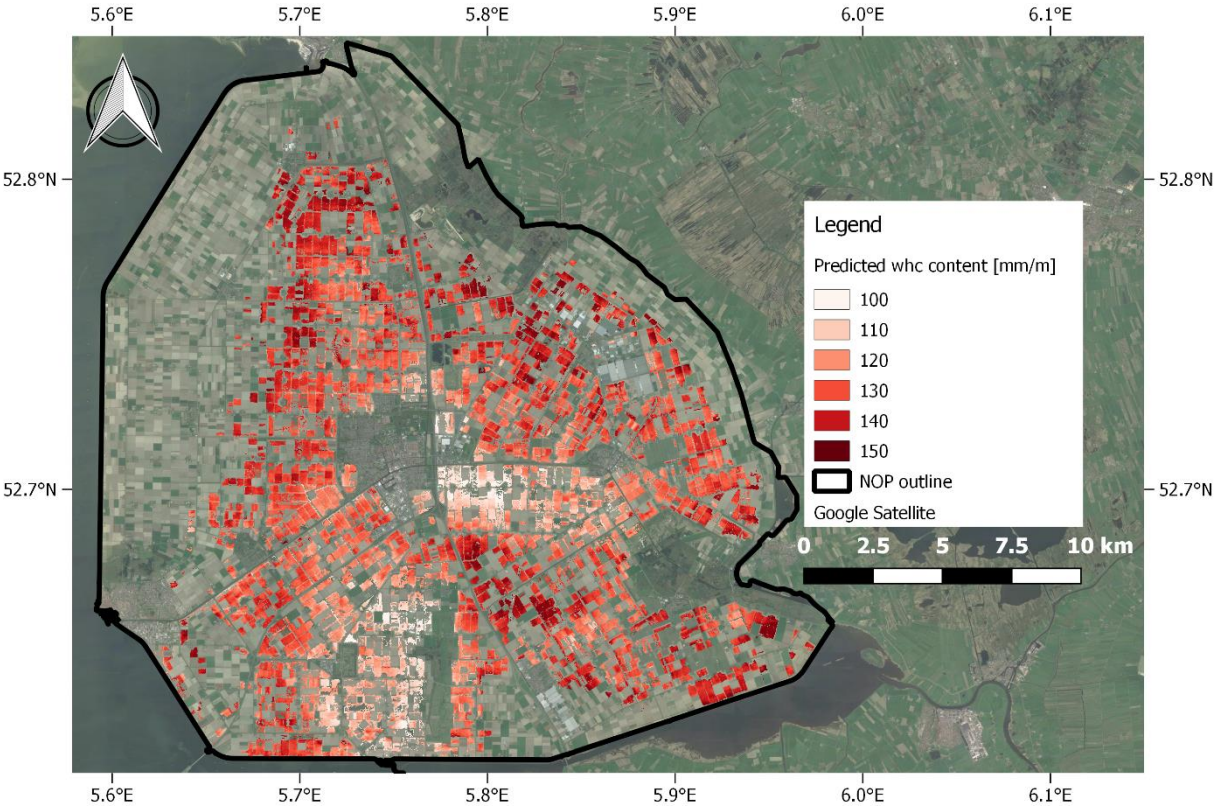


Figure 4.28 Predicted soil water holding capacity content in the study area

Mean	Standard deviation	Minimum	Maximum
128.6	9.7	98.4	149.6

Table 4.17 Statistical summary of estimated soil water holding capacity in the study area in mm/m

The spatial pattern of the soil water holding capacity estimates seems to be negatively correlated to the spatial pattern of clay content estimates (Figure 4.9, paragraph 4.1.1). This implies that high clay content estimates cause a low soil water holding capacity and the other way around. The high correlation between soil water holding capacity and clay content can also be explained by the low variation in organic matter content in the study area. Soil water holding capacity is the difference between soil water content at field capacity and at permanent wilting point. Figure 4.30 shows a generalized relationship between soil water holding capacity and soil textures. Based on Figure 4.29, a simplified soil texture map could be obtained. Low estimates of soil water holding capacity in the study area could be related to sandy soils and high estimates of soil water holding capacity in the study area could be related to loamy/clayey soils. In this study high clay content values are in the range of 25-36%, see Figure 4.9 in paragraph 4.1.1. According to the texture triangle in Figure 4.29, the range of high clay content

values could correspond to soil texture classes sandy clay loam, clay loam or silty clay loam. These texture classes can be used to find the theoretical values of the soil water holding capacity, which are respectively 140 mm/m, 160 mm/m and 180 mm/m (see Figure 4.30). High clay content values coincide with low soil water holding capacity estimates due to the negative spatial correlation between clay content and soil water holding capacity estimates. Low soil water holding capacity estimates are in the range of 100-110 mm/m, see Figure 4.28. The theory clearly does not correspond with the estimates, as the theory indicates significant larger soil water holding capacity values than estimated.

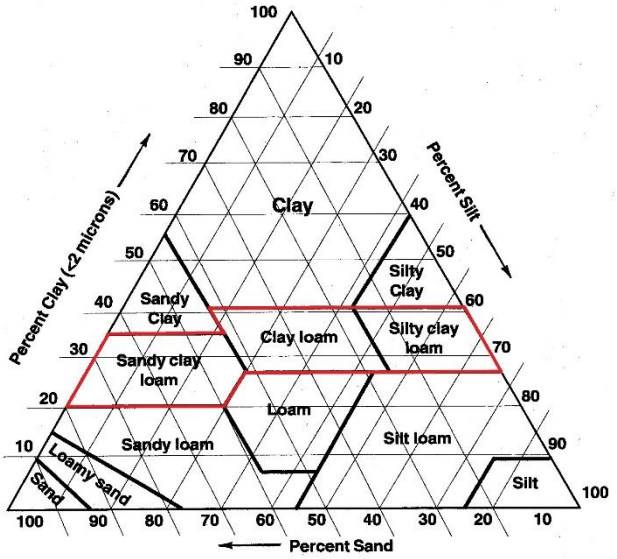


Figure 4.29 Soil texture triangle, in red box the soil textures for high clay content in study area (Plant and Soil Sciences eLibrary, n.d.)

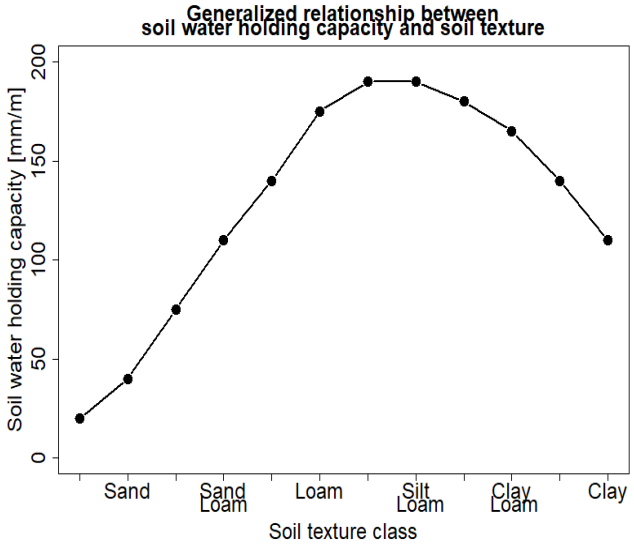


Figure 4.30 Generalized relationship between soil water holding capacity and soil texture (O'geen, 2012)

An extra check can be made to test if the soil water holding capacity estimates are reliable or not. To estimate the soil water holding capacity pedotransfer functions are used of which one is to obtain the bulk density based on clay content and organic matter content. Figure 4.32 shows the predicted bulk density of the study area obtained with help of a pedotransfer function. The bulk density is positively correlated with soil water holding capacity. Low values of bulk density therefore correspond to low values of soil water holding capacity. Low range of bulk density in the study area is between 1.3-1.4 g/cm^3 (1300-1400 kg/m^3). According to Table 4.18, this range of bulk density could correspond to silty loam, silt, clay loam, silty clay loam or clay. Again, these texture classes can be used to find the theoretical values of the soil water holding capacity, which are respectively 190 mm/m, 190 mm/m, 160 mm/m, 180 mm/m and 120 mm/m (see Figure 4.31). Low soil water holding capacity estimates are in the range of 100-110 mm/m, see Figure 4.28. Again, the theory does not correspond with the estimates, as the theory indicates larger soil water holding capacity values than estimated. According to the bulk density, clay content could be a possibility but related to the estimated clay content percentages the area does not contain clay soils.

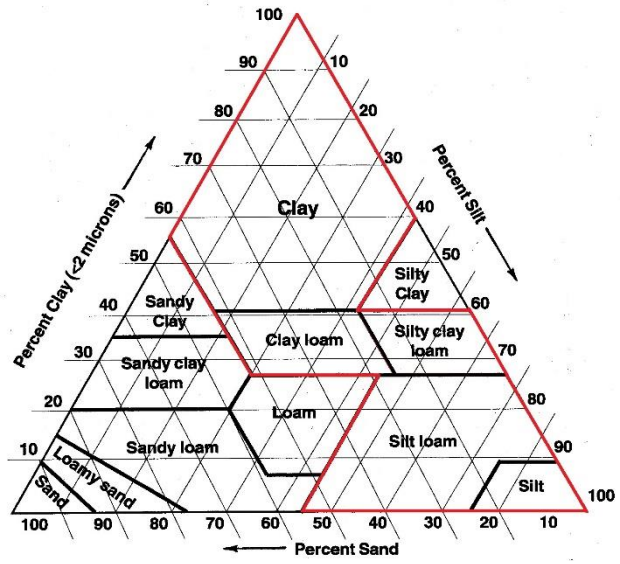


Figure 4.31 Soil texture triangle, in red box the soil textures for low bulk density in study area (Plant and Soil Sciences eLibrary, n.d.)

Soil Type	ρ (lb/ft^3) [kg/m^3]
Sand	(89) [1430]
Loamy Sand	(89) [1430]
Sandy Loam	(91) [1460]
Loam	(89) [1430]
Silty Loam	(86) [1380]
Silt	(86) [1380]
Sandy Clayey Loam	(94) [1500]
Clayey Loam	(87) [1390]
Silty Clayey Loam	(81) [1300]
Silty Clay	(79) [1260]
Sandy Clay	(92) [1470]
Clay	(83) [1330]

Table 4.18 Bulk density of different soil types (StructX, n.d.)

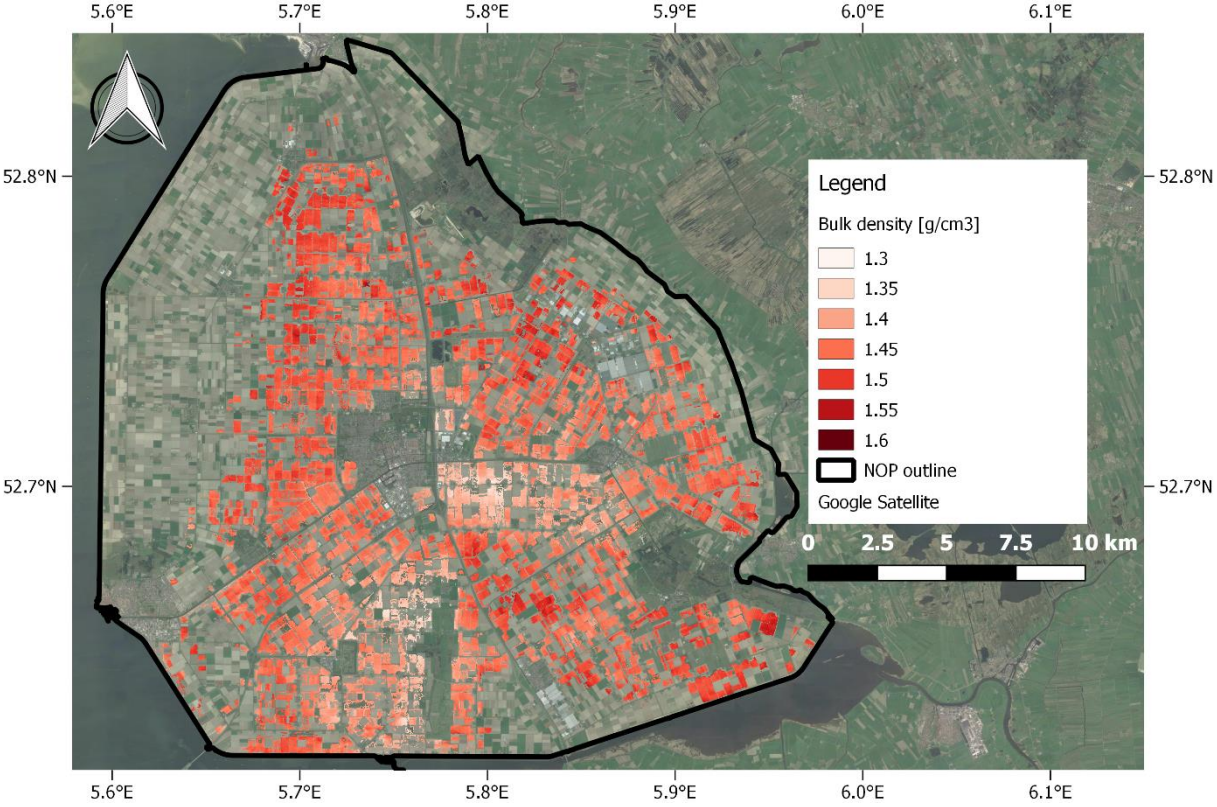


Figure 4.32 Predicted bulk density in the study area

4.2. Spatial evaluation of soil water holding capacity with soil moisture estimates

In this paragraph, the soil water holding capacity will be evaluated against the soil moisture estimates from SEBAL. For both sugar beet and winter wheat, the first step is to determine which dates are useful for analysis. The second step is to evaluate the results with help of a boxplot. The boxplot is used to find consistency between soil water holding capacity and soil moisture content. The third step is to evaluate the results with help of a visualization. The visualization is used to see if there are interesting spatial patterns visible between soil water holding capacity and soil moisture content.

The hypothesis in this study is that during long periods of drought the soil moisture content will reach its minimum; the patterns of minima could be related to the soil water holding capacity. As explained in paragraph 3.1, 2018 will be used as reference year because of the long period of drought in May until July. Five dates, that meet the requirements for useful images, are used for the analysis: 20 March, 21 April, 7 May, 3 July and 26 July. The precipitation rates prior to these dates are shown in Table 4.19. According to the precipitation rates, 3 July and 26 July clearly had to deal with a long period of drought.

Date	Week prior to date [mm]	2 weeks prior to date[mm]	4 weeks prior to date [mm]	6 weeks prior to date [mm]
20-03-2018	5.3	29.8	32.4	40.8
21-04-2018	9.9	9.9	57.7	73.5
07-05-2018	4.1	28.9	39.1	86.8
03-07-2018	0.0	2.9	13.7	59.1
26-07-2018	0.0	0.0	0.6	8.5

Table 4.19 Precipitation rates prior to selected image dates

Sugar beet

The first step is to determine which dates can be used to test the hypothesis. In Table 4.20, the statistical summary of NDVI values for sugar beet per date is shown. Sugar beet is sowed in March/April; the first three dates are therefore mainly based on bare soil surface reflectance, hence low mean NDVI values, and can be related to topsoil properties. The mean NDVI values in the month July are clearly during full crop coverage and therefore can be related to root zone soil properties. For the analysis, 20 March will be used for the top soil properties and 26 July will be used for the root zone soil properties.

Date	Mean	Standard deviation	Minimum	Maximum
20-03-2018	0.15	0.07	0.00	0.69
21-04-2018	0.14	0.03	0.08	0.62
07-05-2018	0.14	0.04	0.08	0.56
03-07-2018	0.80	0.04	0.45	0.86
26-07-2018	0.70	0.04	0.40	0.78

Table 4.20 Statistical summary NDVI values sugar beet

The second step is to analyze the results based on a boxplot. In Figure 4.33, a boxplot is shown of the relation between soil water holding capacity and soil moisture content on 20 March for sugar beet fields. The width of a box is a relative measure for number of pixels. According to Table 4.19, in the week prior to 20 March there was a cumulative amount of 5.3 mm precipitation. This precipitation rate actually fell 4 days prior to 20 March. It can therefore be assumed that the topsoil already dried out after the rainfall at satellite overpass time. The result clearly shows a relation between soil water holding capacity and soil moisture content

for the topsoil. An increasing soil moisture content coincides with an increasing soil water holding capacity.

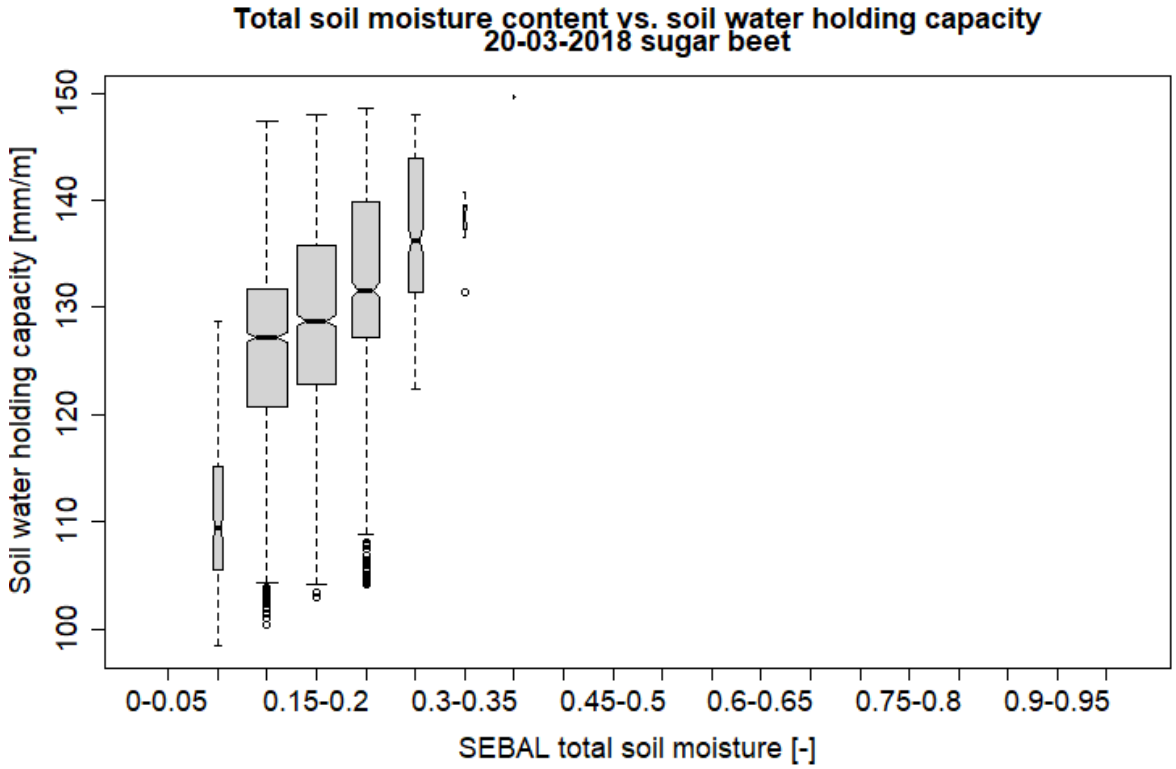


Figure 4.33 Boxplot soil moisture vs. soil water holding capacity, 20-03-2018 sugar beet

According to the hypothesis, the month July is the best month to find a relation between soil moisture and soil water holding capacity. In July, sugar beet will be in a transition phase from vegetative to ripening stage (see paragraph 3.2). The mean NDVI values show a full crop coverage of the sugar beet fields and therefore the soil moisture content could be related to root zone soil properties. In Figure 4.34, a boxplot is shown of the relation between soil water holding capacity and soil moisture content on 26 July for sugar beet fields. The width of a box is a relative measure for number of pixels. According to Table 4.19, 4 weeks prior to 26 July there was a cumulative precipitation amount of 0.6 mm and 6 weeks prior to 26 July a cumulative precipitation amount of 8.5 mm. These precipitation amounts clearly show a long period of drought in the area. The result however, shows a small correlation between soil water holding capacity and soil moisture content for the root zone. An increasing soil moisture content coincides with a small increase in soil water holding capacity. The results of all other dates for sugar beet can be found in Appendix M.

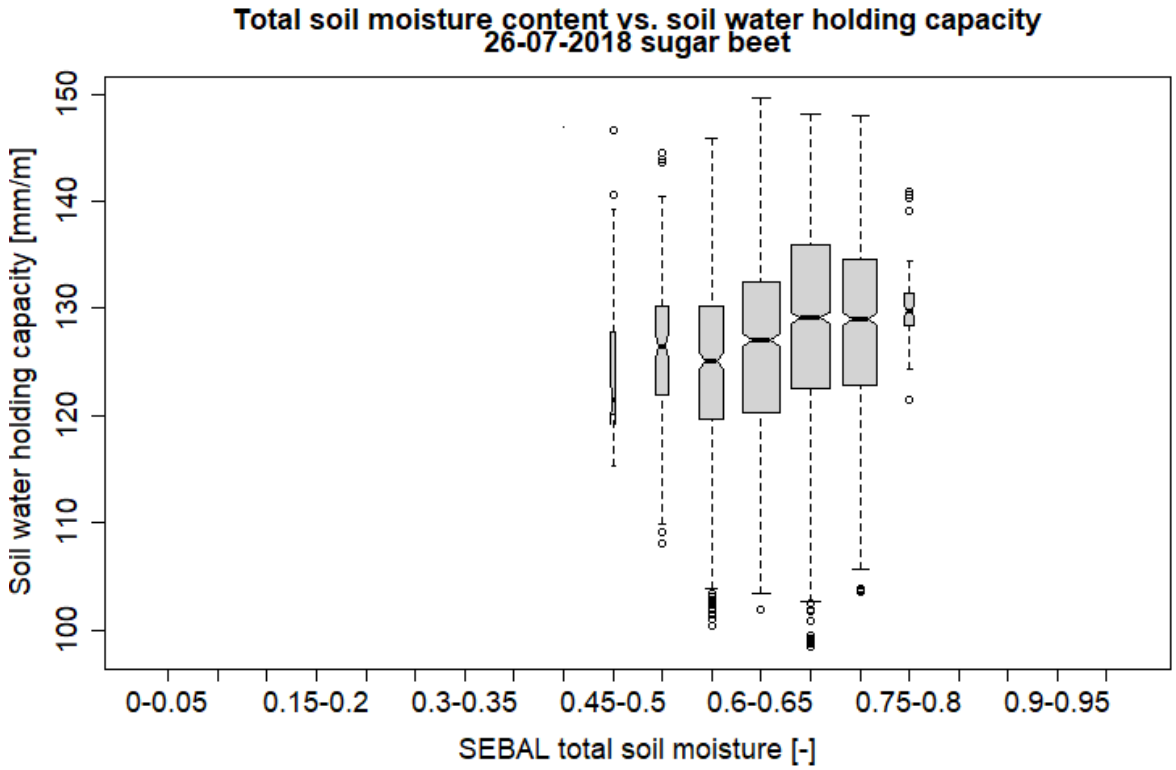


Figure 4.34 Boxplot soil moisture vs. soil water holding capacity, 26-07-2018 sugar beet

The third step is to analyse the results based on a visualization. In Figure 4.35 and Figure 4.36, respectively the soil water holding capacity and soil moisture content of 20 March are shown. Interestingly, the soil water holding capacity in the northern and to a lesser extent southwestern part of the study area show relative high values compared to the soil moisture content. In all other regions of the study area, the relation between soil water holding capacity and soil moisture content is clearly visible.

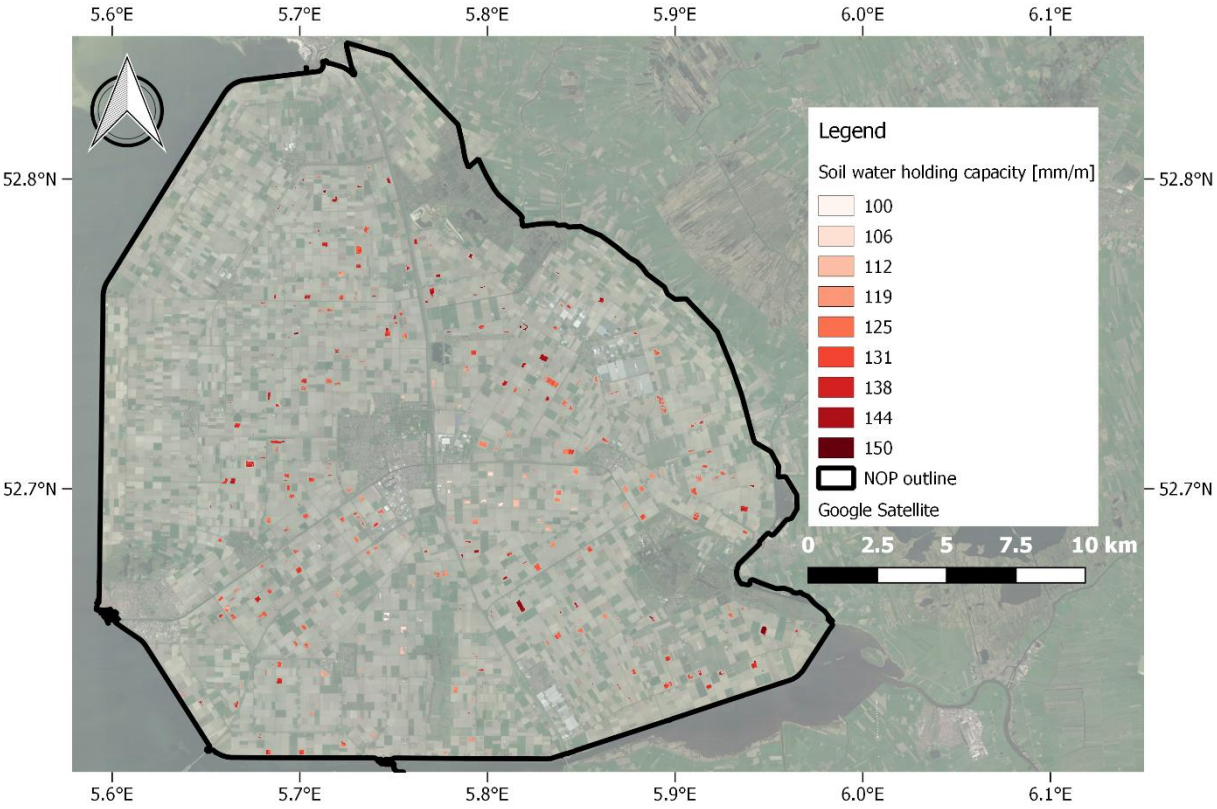


Figure 4.35 Soil water holding capacity map for sugar beet fields

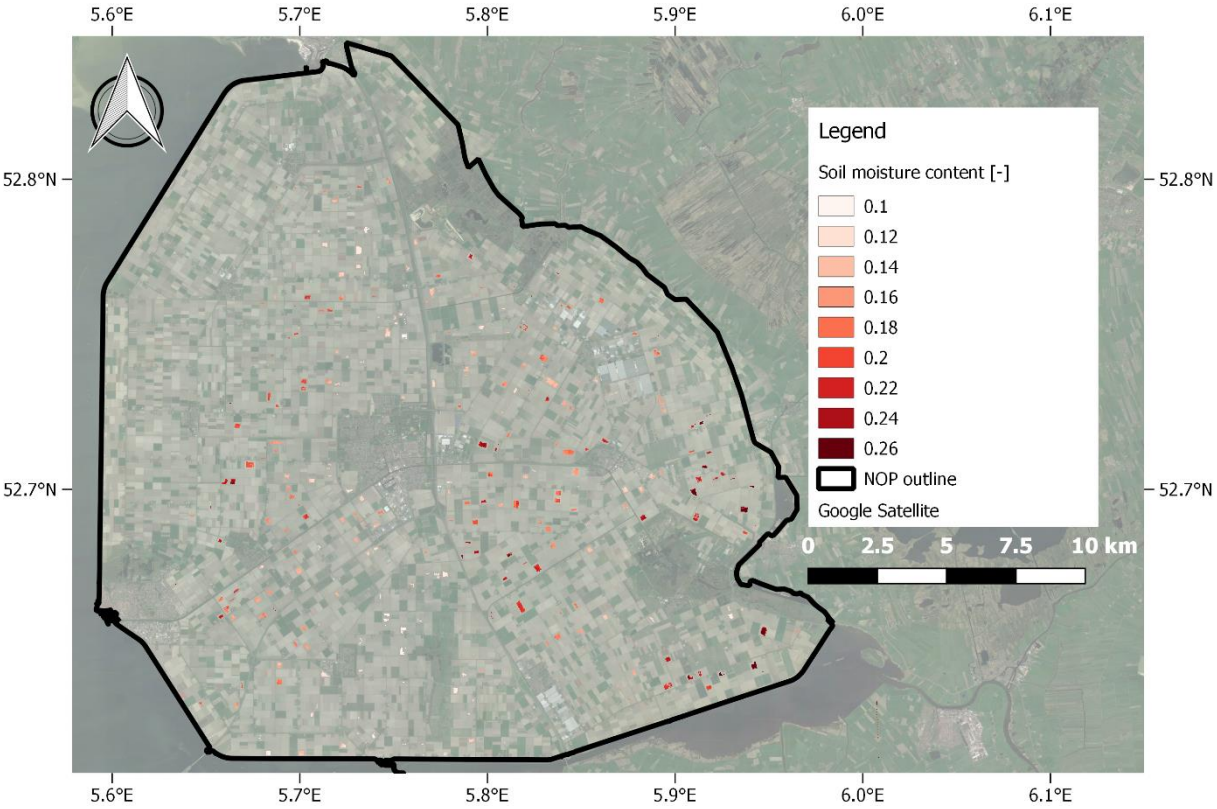


Figure 4.36 Soil moisture content sugar beet fields 20-03-2018

In Figure 4.37 and Figure 4.38, respectively the soil water holding capacity and soil moisture content of 26 July are shown. The spatial visualization does not show any clear differences in spatial patterns between soil moisture content and soil water holding capacity.

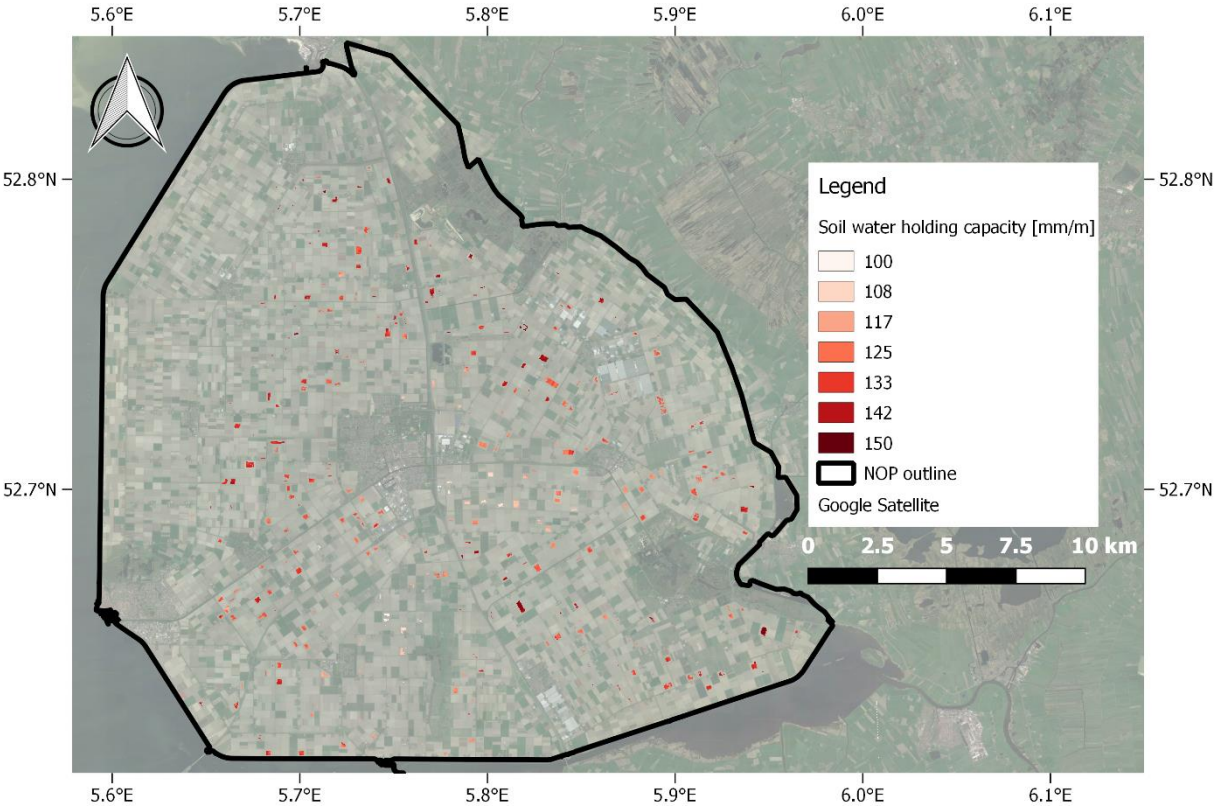


Figure 4.37 Soil water holding capacity map for sugar beet fields

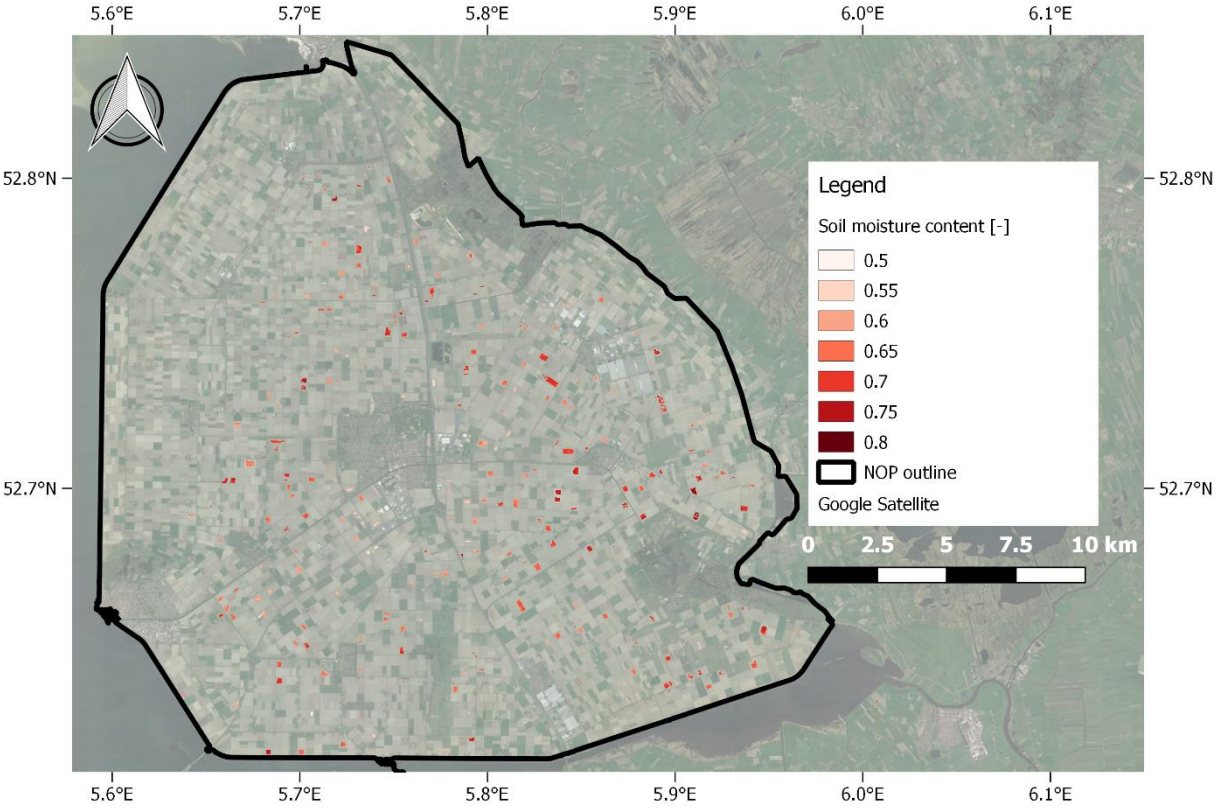


Figure 4.38 Soil moisture content sugar beet fields 26-07-2018

Winter wheat

The first step is to determine which dates can be used to test the hypothesis. In Table 4.21, the statistical summary of NDVI values for winter wheat per date is shown. Winter wheat has been sowed in October/December. Therefore, a sufficient crop coverage will be early in the season with as result a high mean NDVI starting from April. The harvesting period of winter wheat is at the end of July/August. On 26 July most of the winter wheat has probably already been harvested hence the low mean NDVI value. For winter wheat, there is not a clear bare soil image. On none of the dates the mean NDVI is lower than 0.2 which is the threshold set for bare soil. Therefore, the results are not directly related to the topsoil properties. As already mentioned, winter wheat probably already has been harvested on 26 July. The best day to test the hypothesis is therefore 3 July

Date	Mean	Standard deviation	Minimum	Maximum
20-03-2018	0.25	0.11	0.07	0.73
21-04-2018	0.64	0.16	0.13	0.84
07-05-2018	0.77	0.11	0.11	0.87
03-07-2018	0.58	0.09	0.21	0.80
26-07-2018	0.31	0.03	0.18	0.72

Table 4.21 Statistical summary NDVI values winter wheat

The second step is to analyze the results based on a boxplot. In Figure 4.39, a boxplot is shown of the relation between soil water holding capacity and soil moisture content on 3 July for winter wheat fields. The width of a box is a relative measure for number of pixels. According to Table 4.19, a cumulative precipitation rate of 2.9 mm has fell 2 weeks prior to 3 July, which indicate a long period of drought. The spread of the soil moisture content estimates show a wide range of soil moisture conditions that can be related to a deficit of available water content in the area. However, there does not seem to be a positive correlation between soil water holding capacity and soil moisture content as expected. The hypothesis therefore does not seem to hold for winter wheat fields during long periods of drought. The results of all other dates for winter wheat can be found in Appendix M.

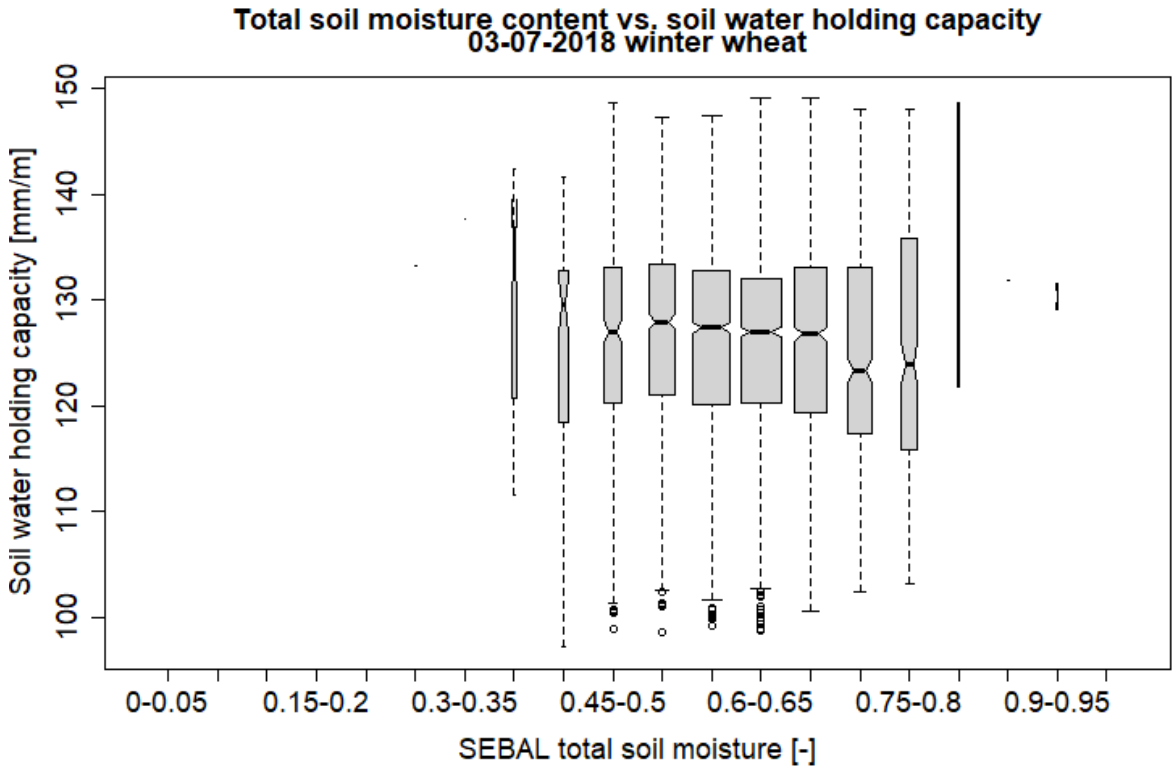


Figure 4.39 Boxplot soil moisture vs. soil water holding capacity, 03-07-2018 winter wheat

The third step is to analyse the results based on a visualization. In Figure 4.40 and Figure 4.41, respectively the soil water holding capacity and soil moisture content of 3 July are shown. Interestingly, the soil water holding capacity south and east from the middle of the study area show relative low values compared to the soil moisture content estimates. In all other regions of the study area, there are no clear differences between soil water holding capacity and soil moisture content.

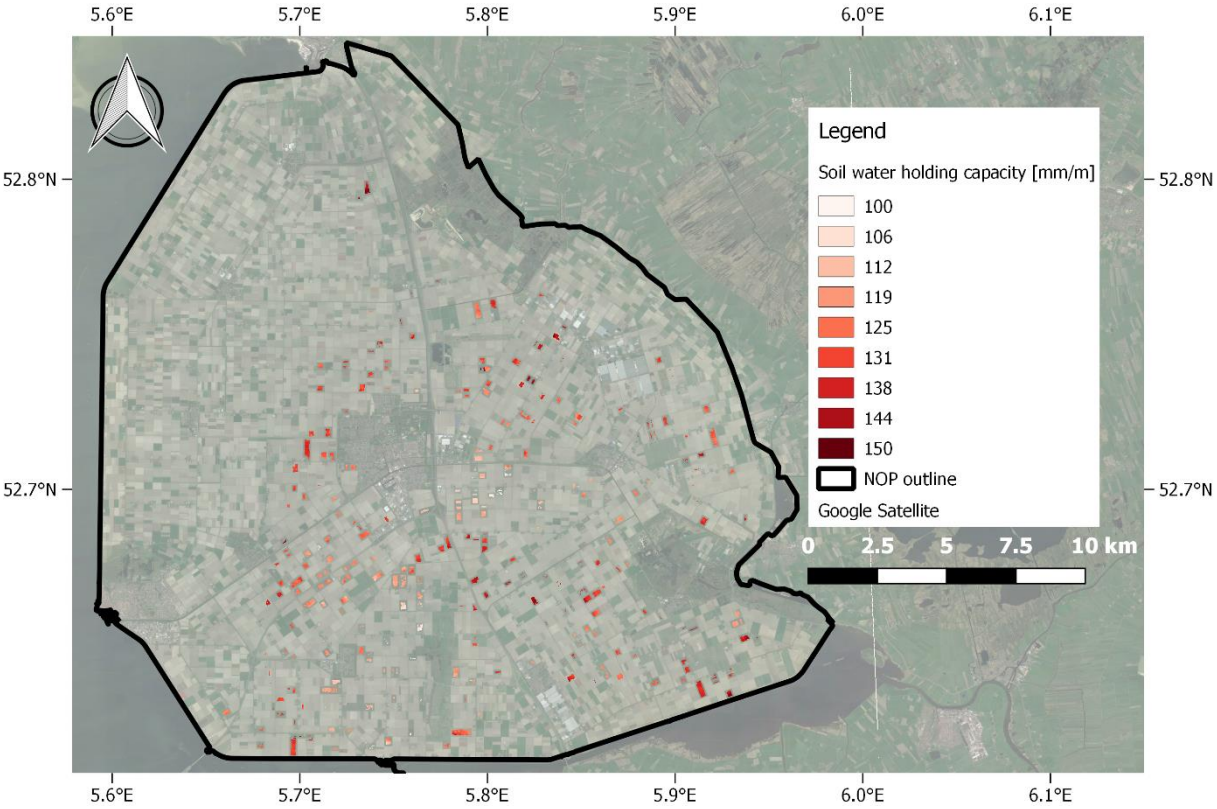


Figure 4.40 Soil water holding capacity map for winter wheat fields

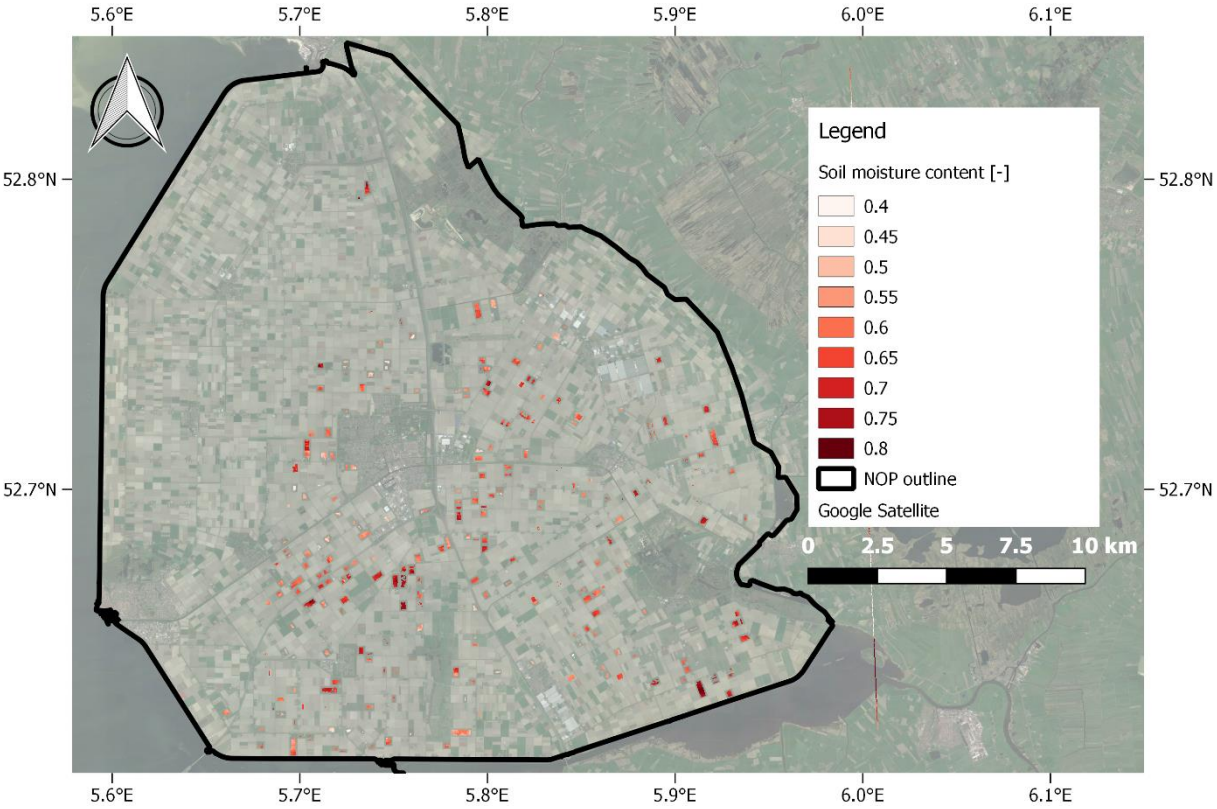


Figure 4.41 Soil moisture content winter wheat fields 03-07-2018

4.3. Soil wetness indicator

For the analysis of the soil wetness indicator, four different dates have been analyzed as mentioned in paragraph 3.6.1. The trapezoidal space for each date has been determined semi-automatically by visually select the 2nd and 98th percentile values that are in line with the density plot. The soil wetness indicator is based on the temperature profile of crops. As explained, crops can be seen as local thermometers of the root-zone water availability. It is therefore important for the analysis to measure only vegetated areas and not bare soil surfaces. In this paragraph, the dates for analysis will be selected for each crop type based on the crop coverage (NDVI).

Sugar beet

The vegetation coverage of sugar beet for the four available dates is shown in Table 4.22. On 3 July and 26 July, the vegetation coverage of sugar beet is sufficient for further analysis. It is interesting to see the decrease in NDVI from the 3 July to 26 July. According to paragraph 3.2, in July sugar beet will be in the end of the yield stage or beginning of the ripening stage. Therefore, there could be two explanations of the decrease in NDVI in the month July. One, at the end of July the ripening stage has started and leaves fail to recover with as result leaves turning into yellow/brownish colors (FAO, 2019). Two, there is a severe deficiency of water in the area due to the drought period in the months before. According to the FAO (2019), a deficiency of water causes the leaves to become dark green that coincides with an increase in the NDVI. Only severe deficiency of water would result in a decrease in NDVI due to the dying process of the leaves.

Date	Mean	Standard deviation	Minimum	Maximum
21-04-2018	0.14	0.03	0.08	0.62
07-05-2018	0.14	0.04	0.08	0.56
03-07-2018	0.80	0.04	0.45	0.86
26-07-2018	0.70	0.04	0.40	0.78

Table 4.22 Statistical summary NDVI values sugar beet

To determine root-zone soil properties, it is important to measure only crop temperatures. Therefore, it is important to have a dense crop coverage on the fields of interest. According to the NDVI, it could be concluded that the crop coverage is high for the month July and therefore the measured LST can be directly related to the root-zone water availability. In Table 4.23, the statistical summary of the instantaneous air temperature and RCT is shown for two dates in July picked based on the NDVI. The RCT values are relative crop temperatures with respect to instantaneous air temperature. The hypothesis of the soil wetness indicator is that the temperature of a crop will increase relatively to the air temperature if there is a deficiency of water in the root-zone of the crops. According to Table 4.23, the mean air temperature on 3 July is 8 Kelvin lower compared to 26 July but the mean crop temperature is 4 Kelvin higher. This negative correlation would suggest that indeed on 26 July the sugar beet has started the ripening phase. On 3 July the mean relative crop temperature is significantly higher than the air temperature, which could indicate a deficiency of water in the area. Both dates will be evaluated against the SEBAL soil moisture estimates. The resulting trapezoidal space for 3 July 2018 and 26 July 2018 are shown in Figure 4.42 and Figure 4.43.

Date	Instantaneous air temperature	Mean (RCT)	Standard deviation (RCT)	Minimum (RCT)	Maximum (RCT)
03-07-2018	295.30	7.45	1.02	4.15	13.98
26-07-2018	303.43	3.45	1.27	0.56	9.98

Table 4.23 Statistical summary sugar beet temperatures in Kelvin, RCT=relative crop temperature

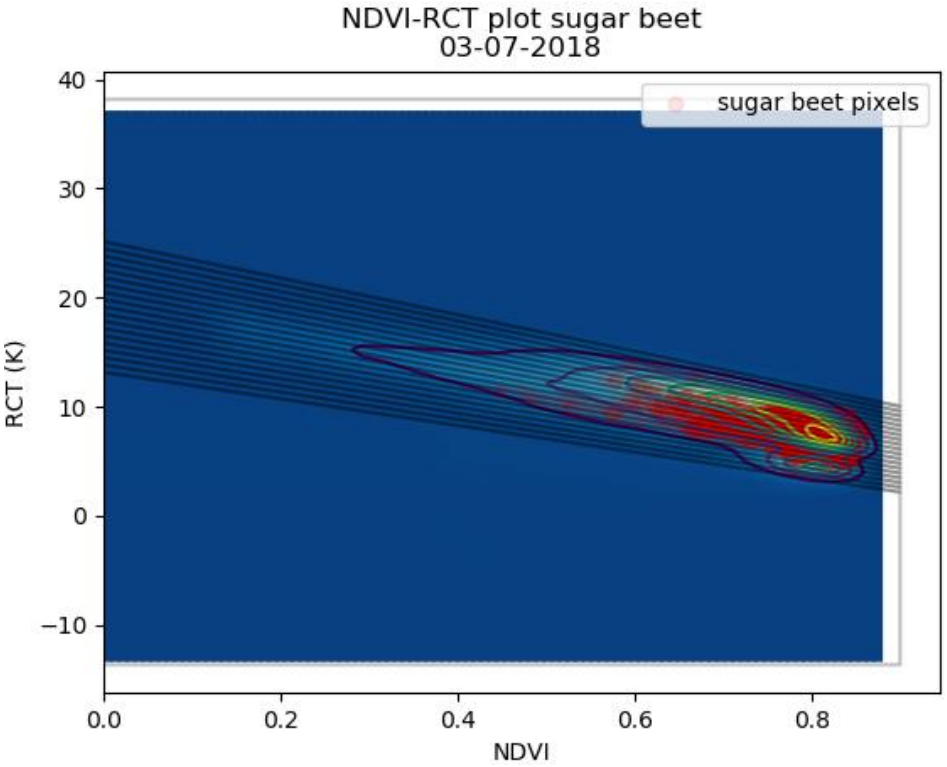


Figure 4.42 NDVI-RCT density plot of all pixels in study area along with sugar beet pixels on 03-07-2018

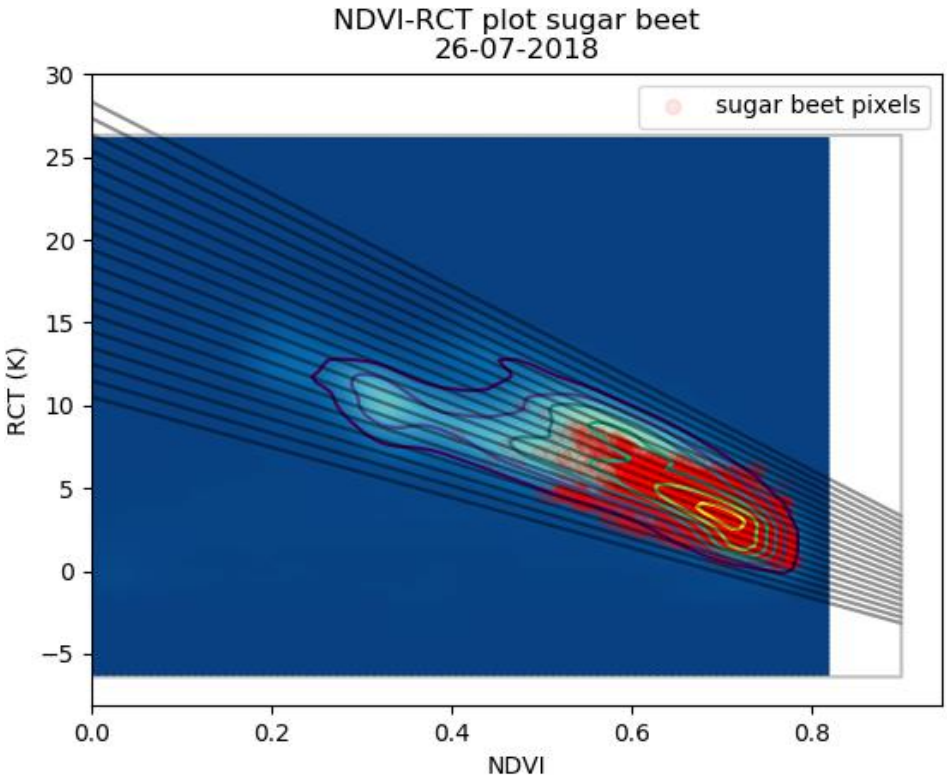


Figure 4.43 NDVI-RCT density plot of all pixels in study area along with sugar beet pixels on 26-07-2018

Winter wheat

The vegetation coverage of winter wheat for the four available dates is shown in Table 4.24. On 21 April, 7 May and 3 July, the vegetation coverage of winter wheat is sufficient for further

analysis. The NDVI increases as the winter wheat is developing in time. Harvesting of winter wheat generally takes place at the end of July or in August. The decrease of the NDVI values suggests that the ripening phase took place in July.

Date	Mean	Stand. dev.	Minimum	Maximum
21-04-2018	0.64	0.16	0.13	0.84
07-05-2018	0.77	0.11	0.11	0.87
03-07-2018	0.58	0.09	0.21	0.80
26-07-2018	0.31	0.03	0.18	0.72

Table 4.24 Statistical summary NDVI values winter wheat

As mentioned before, it is important to have a dense crop coverage on the measured fields of interest. According to the NDVI it could be concluded that the field crop coverage is high starting end April onwards until begin July. Therefore, the measured LST can be directly related to the root zone water availability. In Table 4.25, the statistical summary of the instantaneous air temperature and RCT is shown for three dates picked based on the NDVI. According to Table 4.24 and Table 4.25, the crop temperature seems to correlate negatively with the NDVI. Decreasing NDVI values coincide with increasing crop temperatures and vice versa. Based on this it could be concluded that indeed the ripening phase has started in the month July. Therefore, 21 April, 6 May and 3 July will be evaluated against the SEBAL soil moisture estimates to find if there is consistency between the two models. The resulting trapezoidal space for the selected three dates are shown in Figure 4.44 until Figure 4.46.

Date	Instantaneous air temperature	Mean (RCT)	Stand. dev. (RCT)	Minimum (RCT)	Maximum (RCT)
21-04-2018	291.70	6.63	2.32	2.98	14.77
07-05-2018	295.36	4.02	1.89	1.09	12.99
03-07-2018	295.30	9.20	1.53	3.35	16.34

Table 4.25 Statistical summary winter wheat temperatures in Kelvin, RCT=relative crop temperature

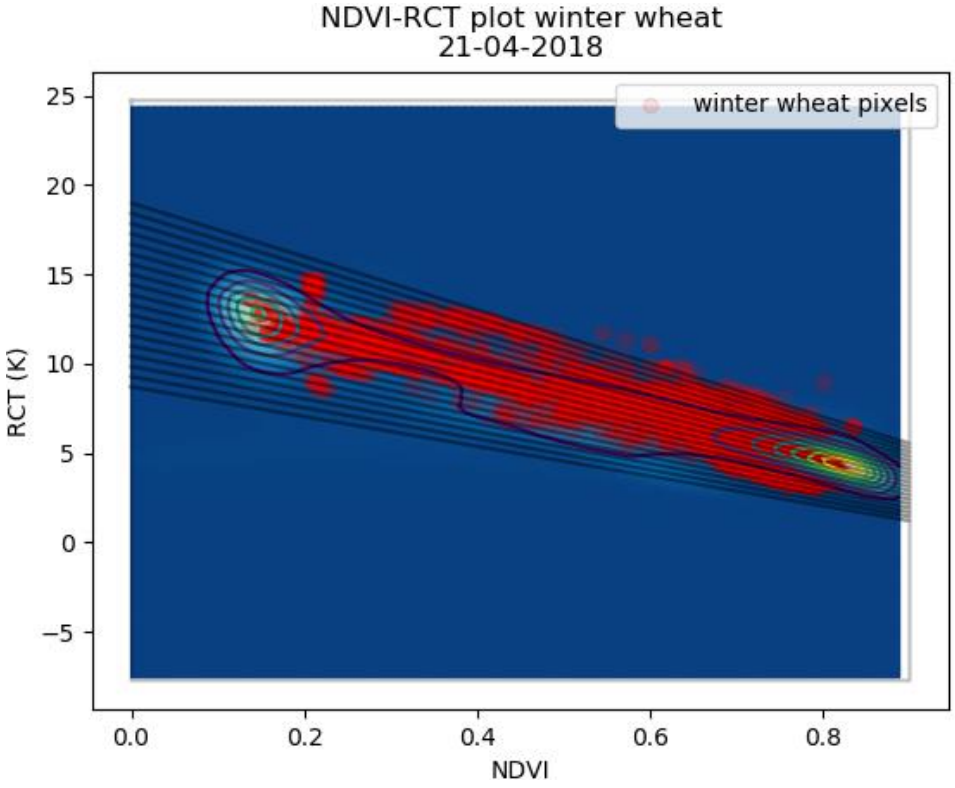


Figure 4.44 NDVI-RCT density plot of all pixels in study area along with winter wheat pixels on 21-04-2018

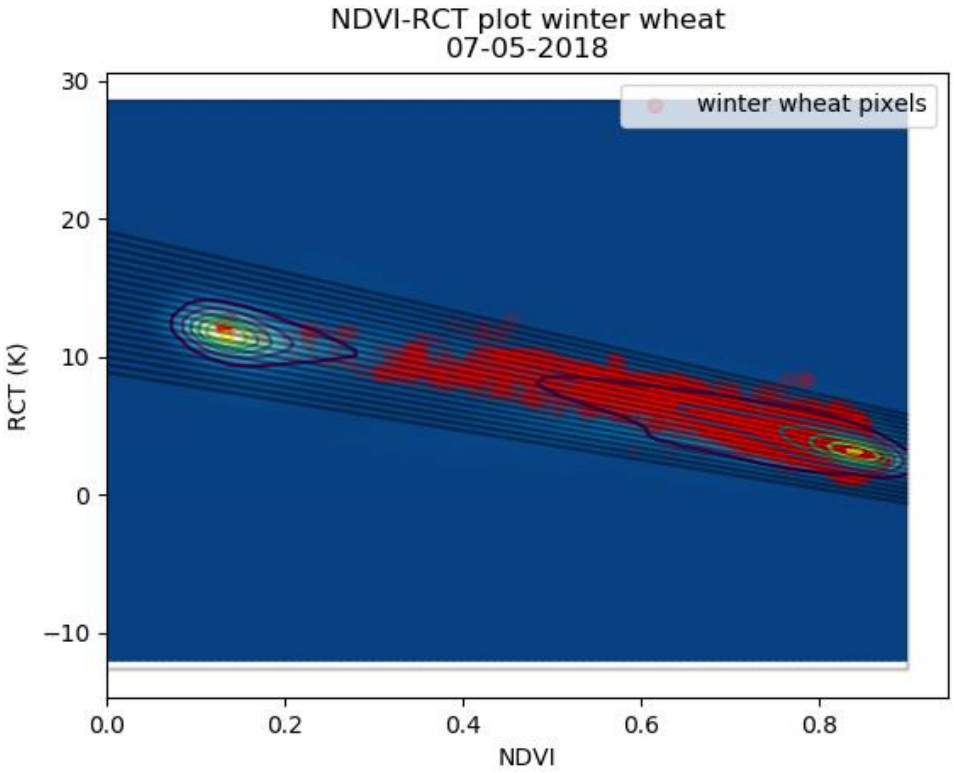


Figure 4.45 NDVI-RCT density plot of all pixels in study area along with winter wheat pixels on 07-05-2018

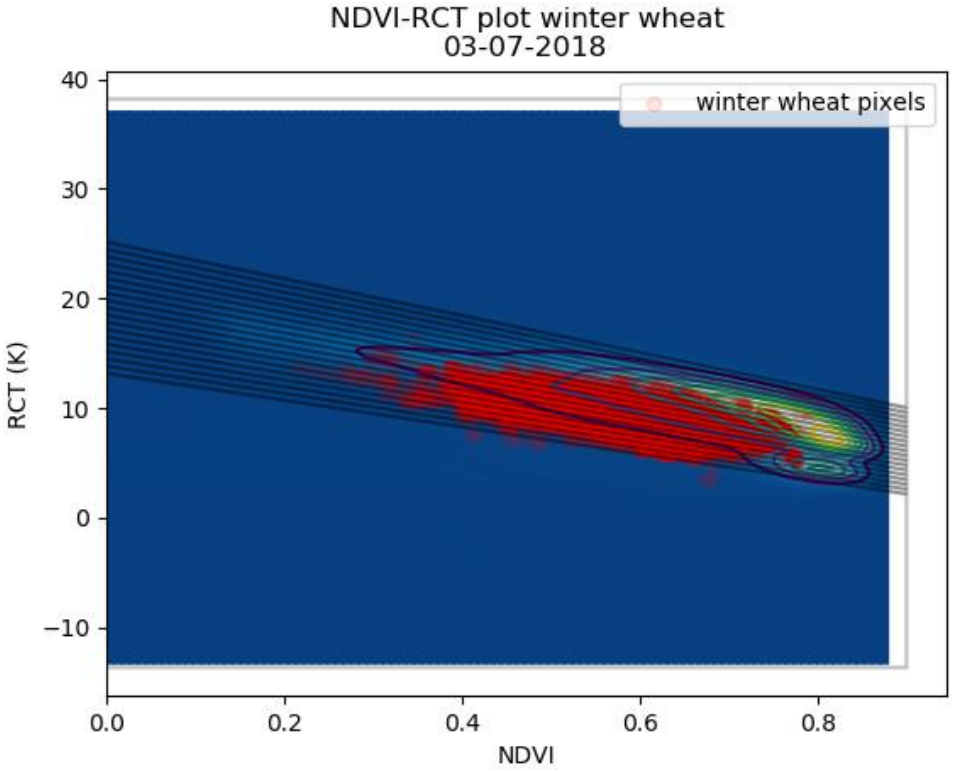


Figure 4.46 NDVI-RCT density plot of all pixels in study area along with winter wheat pixels on 03-07-2018

4.4. Spatial evaluation of soil wetness indicator with soil moisture estimates

In this paragraph, the soil wetness indicator will be evaluated with the soil moisture estimates from SEBAL. The hypothesis of the soil wetness indicator is that the crop temperature will increase relatively to the air temperature if there is a deficiency of water in the root-zone. Based on the RCT-NDVI density plot shown in paragraph 4.3, classes of wetness are determined. Low classes of wetness indicate sufficient amount of water availability in the area and vice versa. To find consistency between both methods the spatial patterns of wetness will be compared with help of a density plot and a spatial visualization for each crop type. First, the evaluation with a boxplot will be analyzed and second the spatial visualization will be analyzed.

Sugar beet

The first step is to evaluate the soil wetness indicator with soil moisture content with help of a density plot. In paragraph 4.3, two dates have been selected for the evaluation of sugar beet fields: 3 July and 26 July 2018. In Table 4.26 and Table 4.27, the statistical summaries of soil moisture content and the soil wetness indicator are shown for 3 July and 26 July.

	Mean	Standard deviation	Minimum	Maximum
Soil moisture (SEBAL) [-]	0.313	0.025	0.151	0.396
Classes soil wetness indicator	10.334	1.772	3.000	18.000

Table 4.26 Statistical summary SWI-SM plot sugar beet fields 03-07-2018

	Mean	Standard deviation	Minimum	Maximum
Soil moisture (SEBAL) [-]	0.291	0.025	0.160	0.353
Classes soil wetness indicator	8.609	1.706	3.000	16.000

Table 4.27 Statistical summary SWI-SM plot sugar beet fields 26-07-2018

In Figure 4.47 and Figure 4.48, the density plots are shown between the soil wetness indicator and soil moisture content for 3 July and 26 July. For both dates, the results clearly show a negative correlation. According to both figures, low soil classes are related to high soil moisture content and vice versa that is in agreement with the hypothesis. According to the precipitation rates in Table 4.19 paragraph 4.2, the drought period prior to 26 July is much longer than for 3 July. The statistical summary of soil moisture content also shows a decrease of available water in the study area for sugar beet fields. However, the statistical summary of classes of wetness indicate that there is more water available in the study area on 26 July compared to 3 July. It can be concluded that the soil wetness indicator is a good indicator to find areas of wetness in sugar beet fields. However, the classes are determined relative to the at date pixel envelope and therefore the changes in soil moisture content due to a long period of drought are not comparable with the changes in classes of wetness.

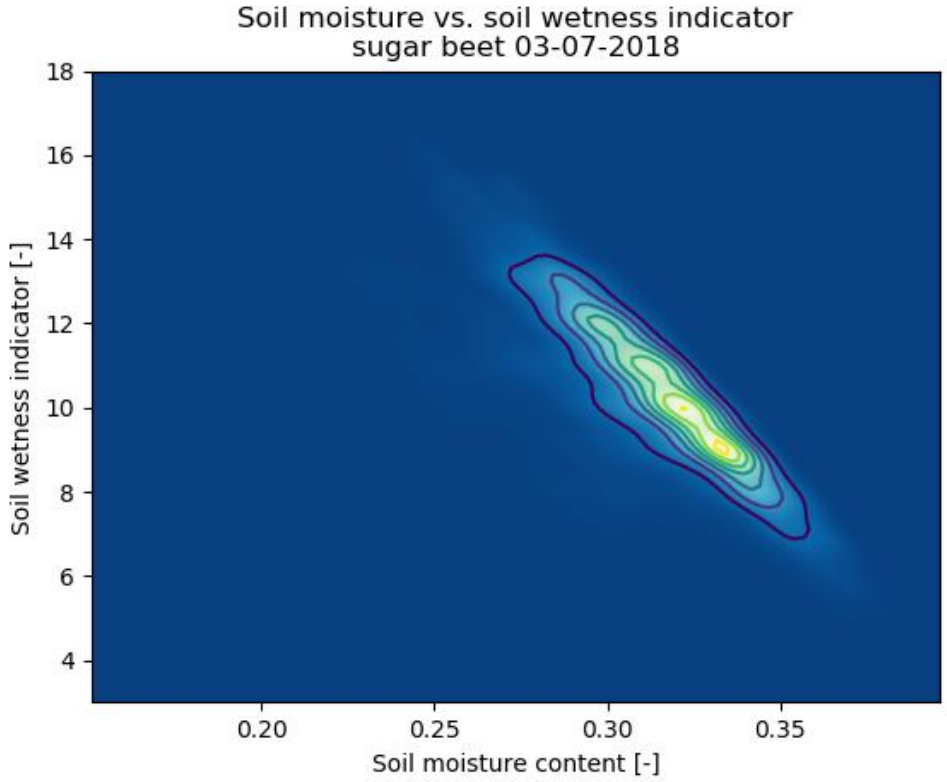


Figure 4.47 Density plot of soil moisture content vs. soil wetness indicator for sugar beet fields on 03-07-2018

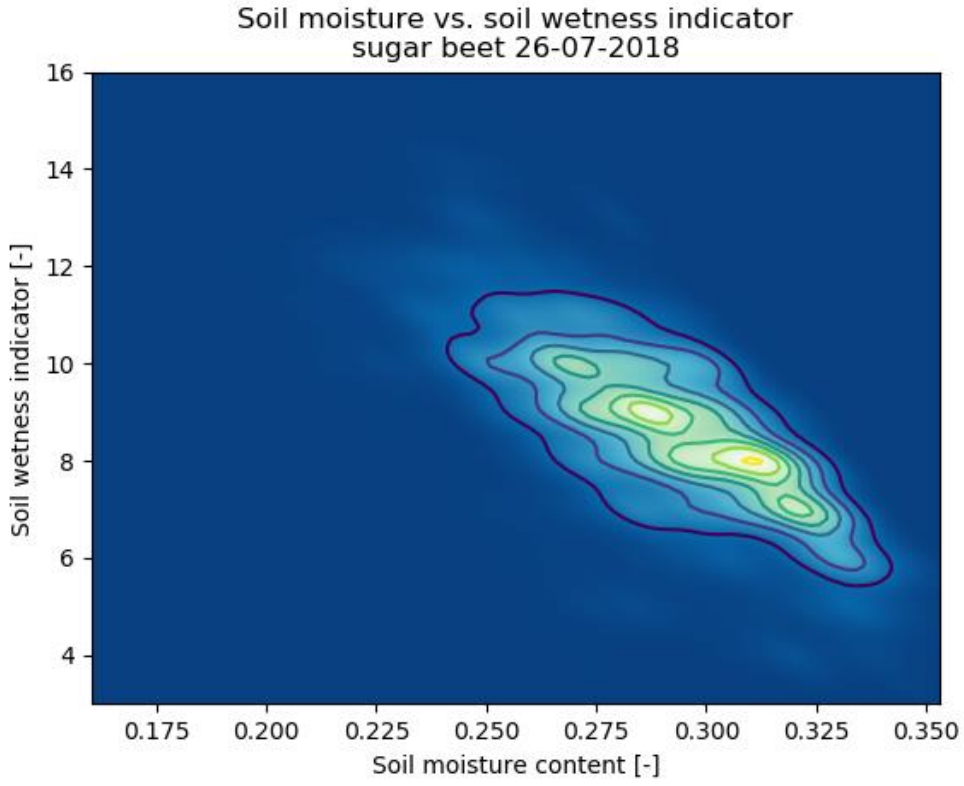


Figure 4.48 Density plot of soil moisture content vs. soil wetness indicator for sugar beet fields on 26-07-2018

The second step is to evaluate the results with help of a spatial visualization, see Figure 4.49 until Figure 4.52. For a better visualization, the color palette of the soil wetness indicator has been reversed because of the negative correlation between soil wetness indicator and soil moisture content. On both dates, the spatial patterns of wetness correspond to each other. It

could therefore be concluded that the soil wetness indicator for sugar beet fields give good results.

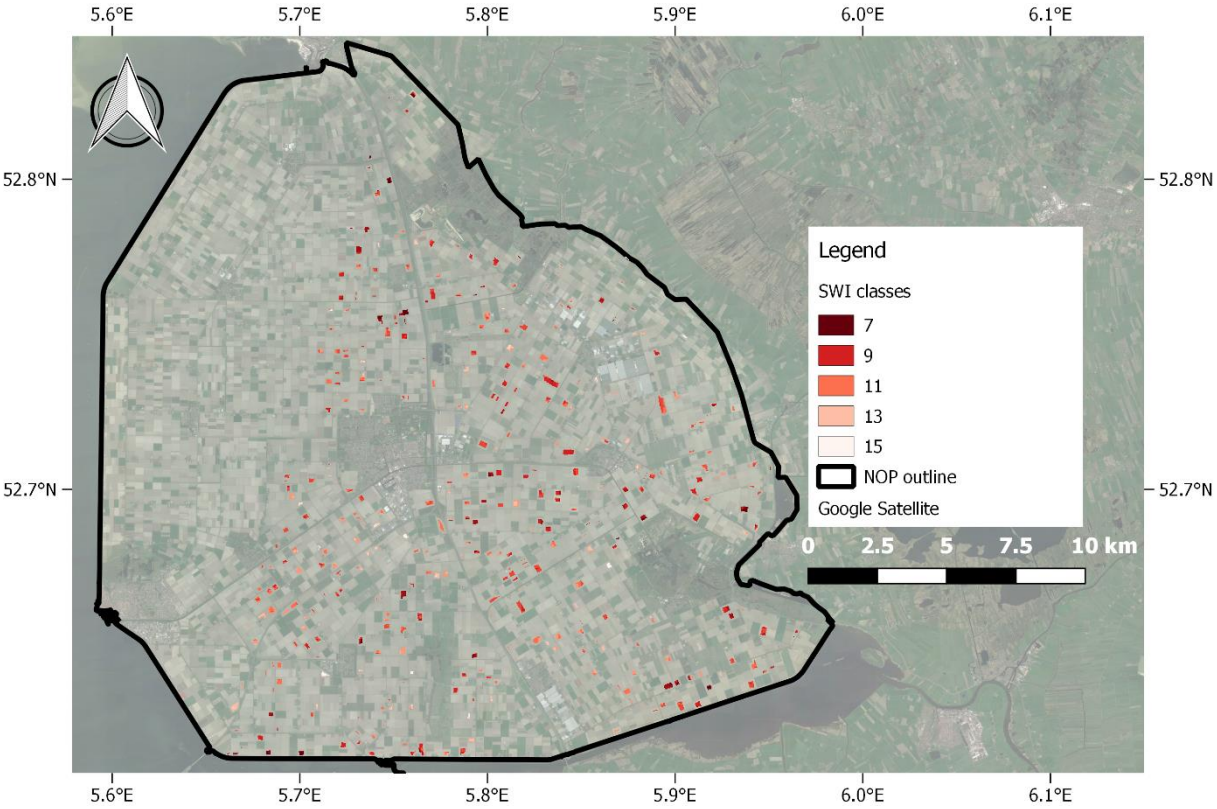


Figure 4.49 Soil wetness indicator map for sugar beet on 03-07-2018

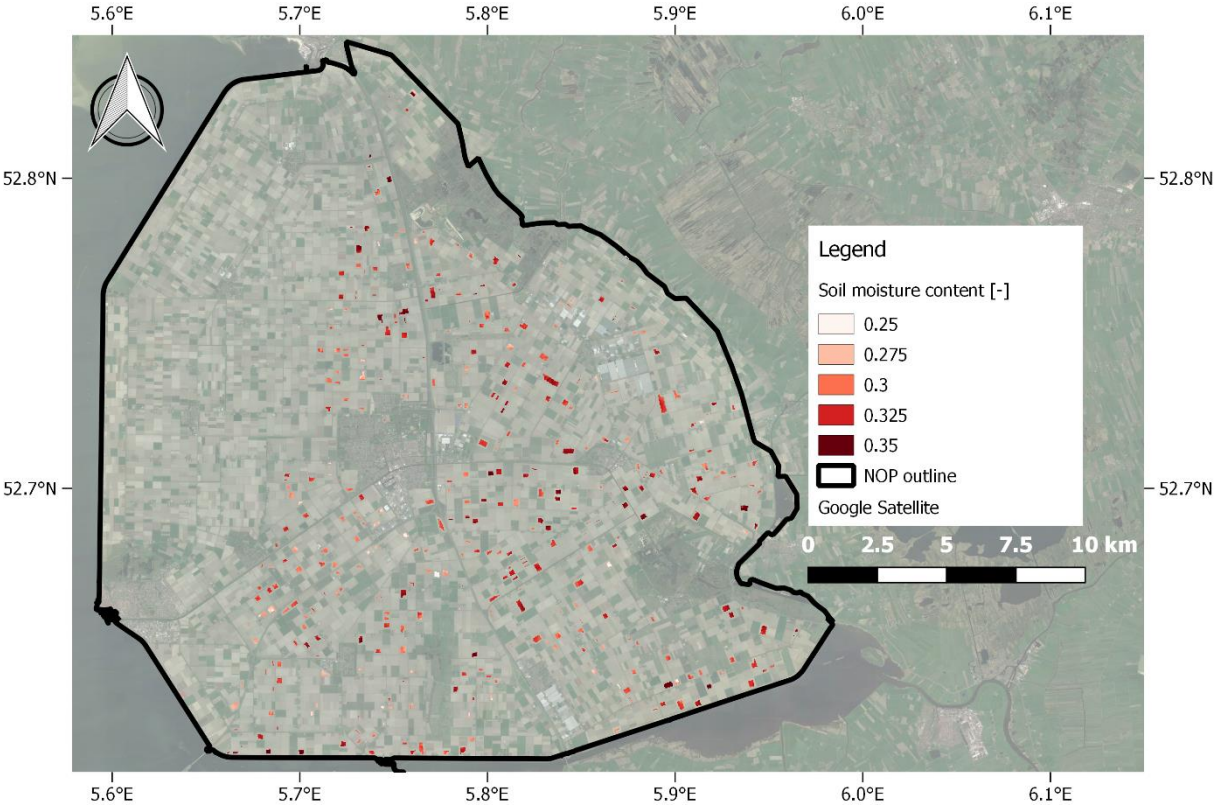


Figure 4.50 Soil moisture content map for sugar beet on 03-07-2018

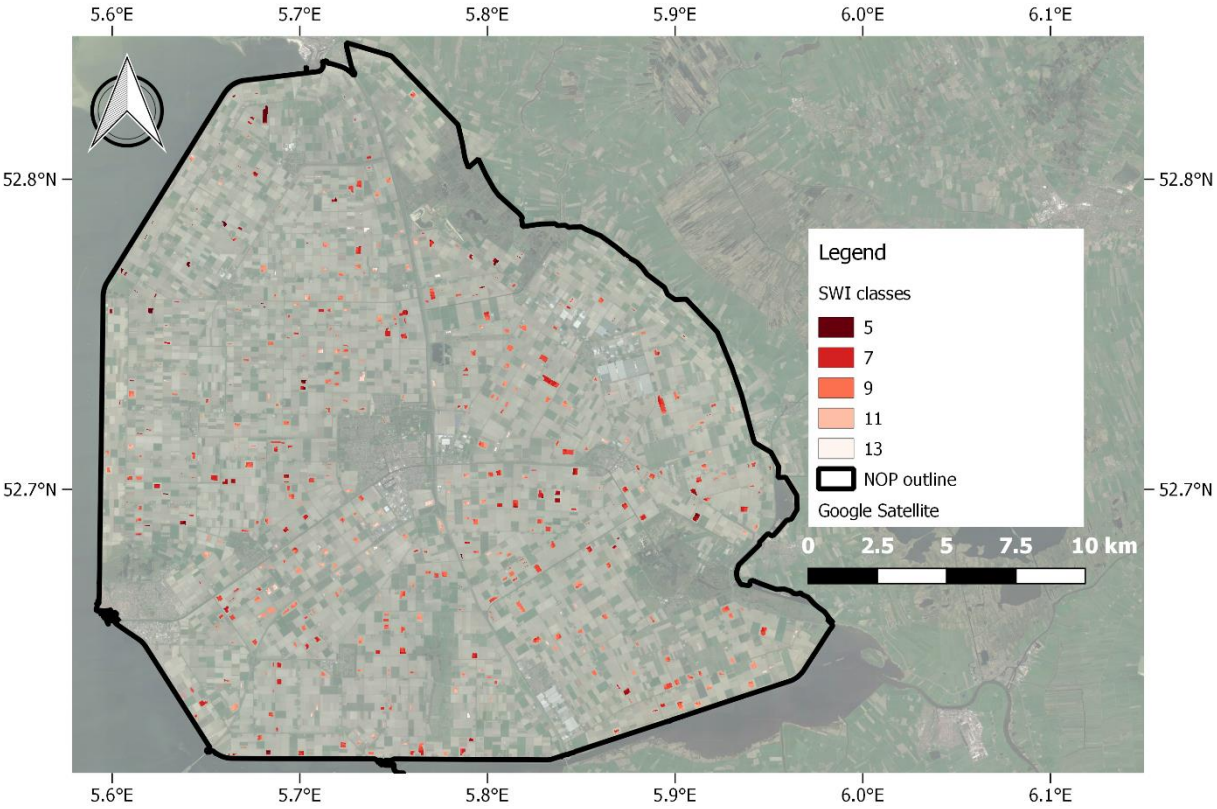


Figure 4.51 Soil wetness indicator map for sugar beet on 26-07-2018

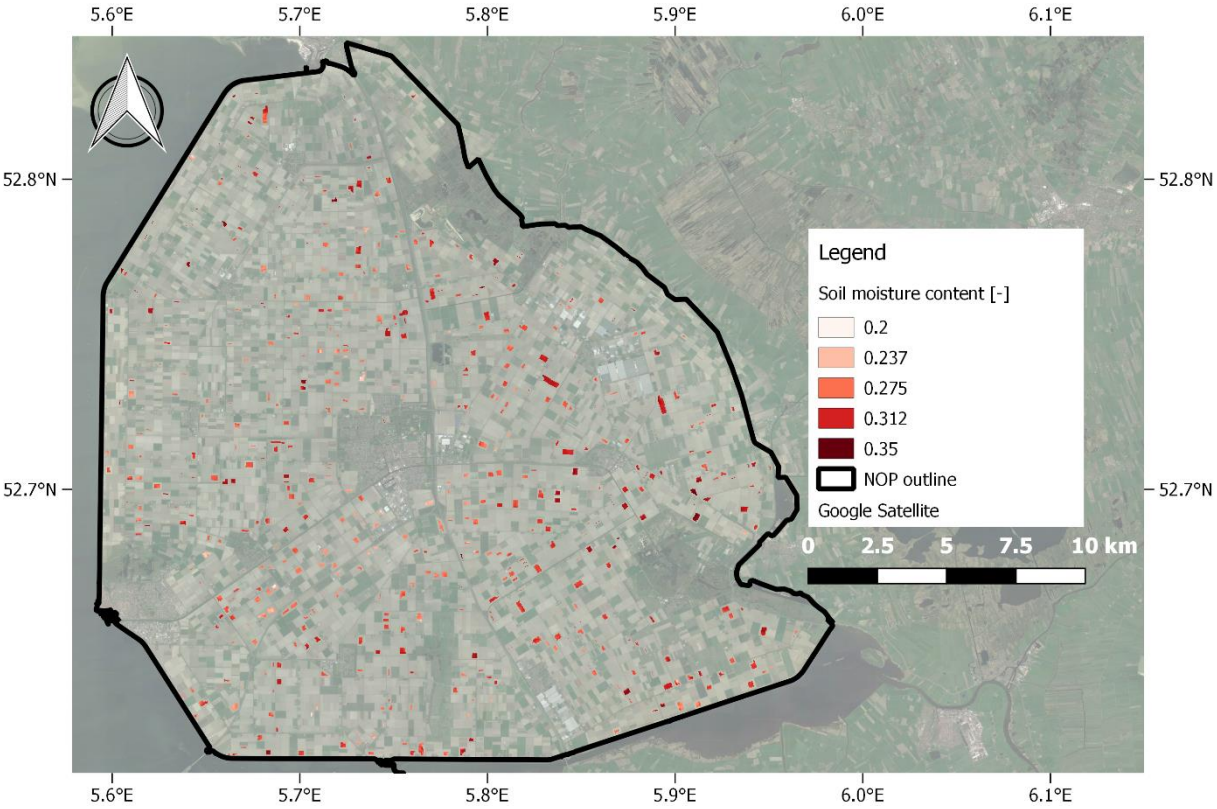


Figure 4.52 Soil moisture content map for sugar beet on 26-07-2018

Winter wheat

The first step is to evaluate the soil wetness indicator with soil moisture content with help of a density plot. In paragraph 4.3, three dates have been selected for the evaluation of winter wheat fields: 21 April, 7 May and 3 July 2018. In Table 4.28 until Table 4.30, the statistical summaries of soil moisture content and the soil wetness indicator are shown for 21 April, 7 May and 3 July.

	Mean	Standard deviation	Minimum	Maximum
Soil moisture (SEBAL) [-]	0.224	0.061	0.055	0.329
Classes soil wetness indicator	11.239	2.875	4.000	20.000

Table 4.28 Statistical summary SWI-SM plot winter wheat fields 21-04-2018

	Mean	Standard deviation	Minimum	Maximum
Soil moisture (SEBAL) [-]	0.228	0.043	0.042	0.300
Classes soil wetness indicator	10.037	2.359	3.000	20.000

Table 4.29 Statistical summary SWI-SM plot winter wheat fields 07-05-2018

	Mean	Standard deviation	Minimum	Maximum
Soil moisture (SEBAL) [-]	0.278	0.034	0.124	0.415
Classes soil wetness indicator	7.609	3.112	1.000	17.000

Table 4.30 Statistical summary SWI-SM plot winter wheat fields 03-07-2018

In Figure 4.53 until Figure 4.55, the density plots are shown between the soil wetness indicator and soil moisture content for 21 April, 7 May and 3 July. For all three dates, the results clearly show a negative correlation. According to all three figures, low soil classes are related to high soil moisture content and vice versa, that is in agreement with the hypothesis. According to the precipitation rates in Table 4.19 paragraph 4.2, the drought period prior to 3 July is much longer than the drought periods prior to 7 May and 21 April. However, the statistical summary of soil moisture content shows an increase of available water in the study area for winter wheat fields.

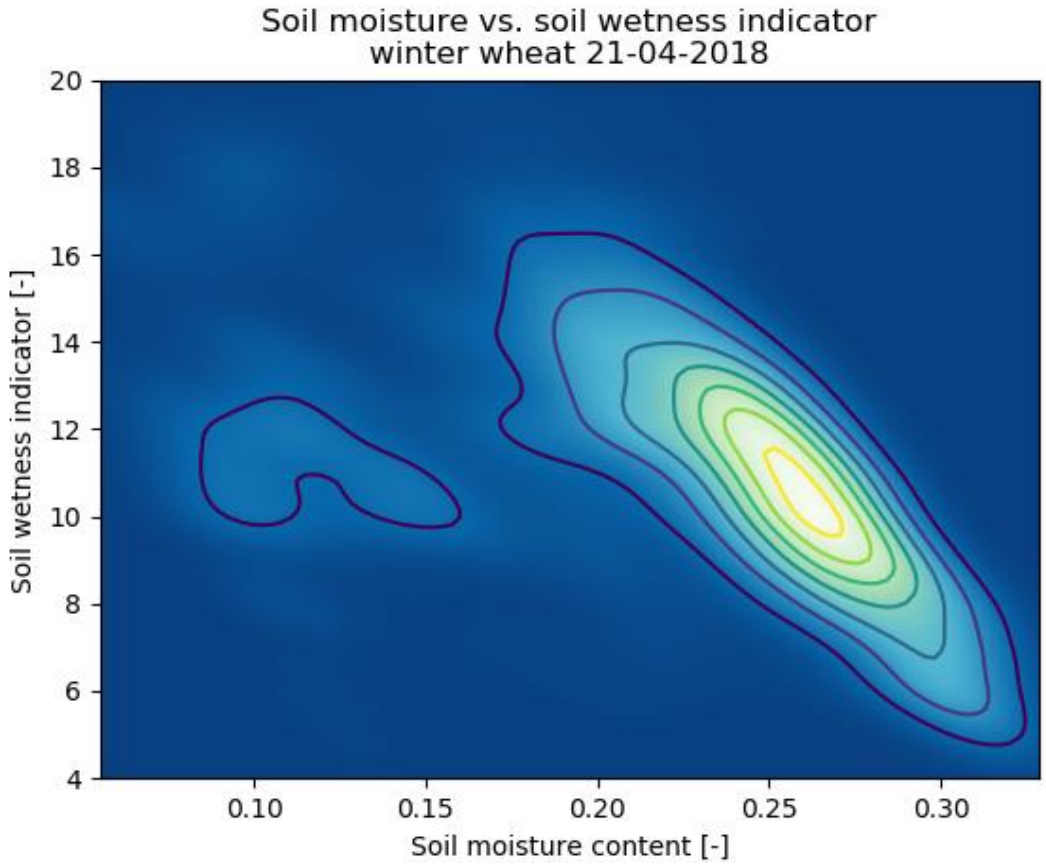


Figure 4.53 Density plot of soil moisture content vs. soil wetness indicator for winter wheat fields on 21-04-2018

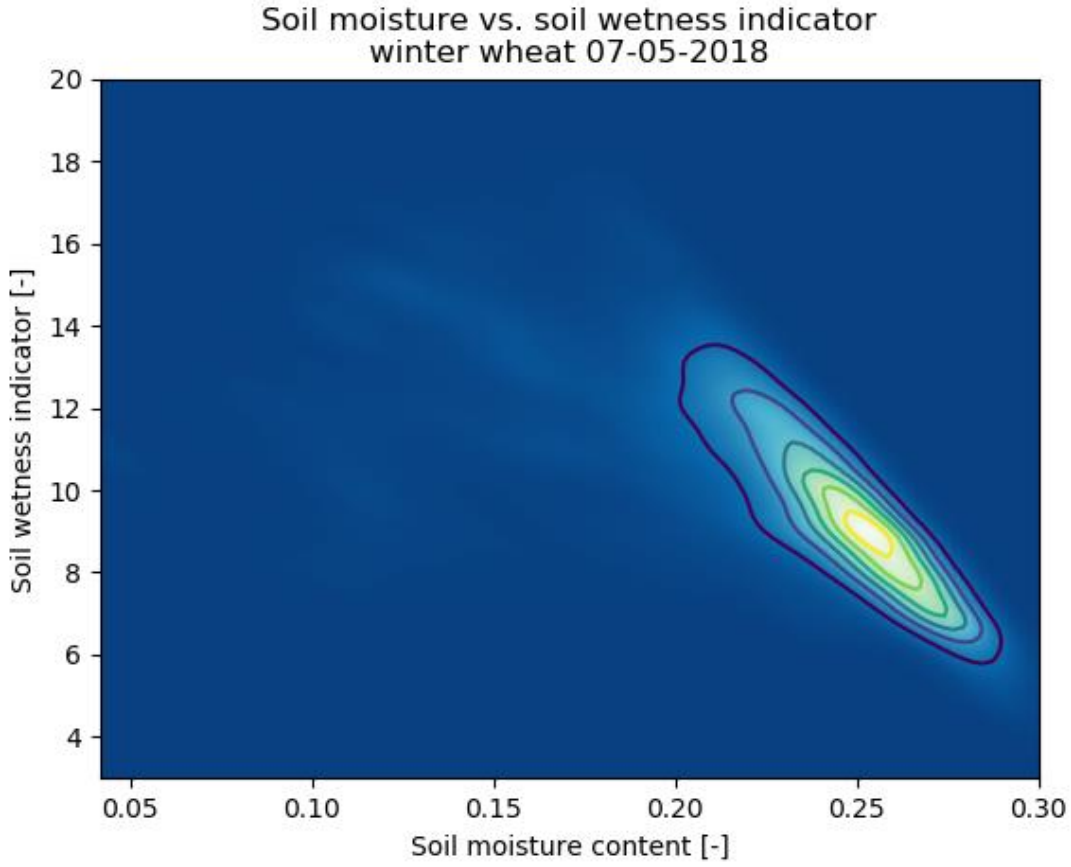


Figure 4.54 Density plot of soil moisture content vs. soil wetness indicator for winter wheat fields on 07-05-2018

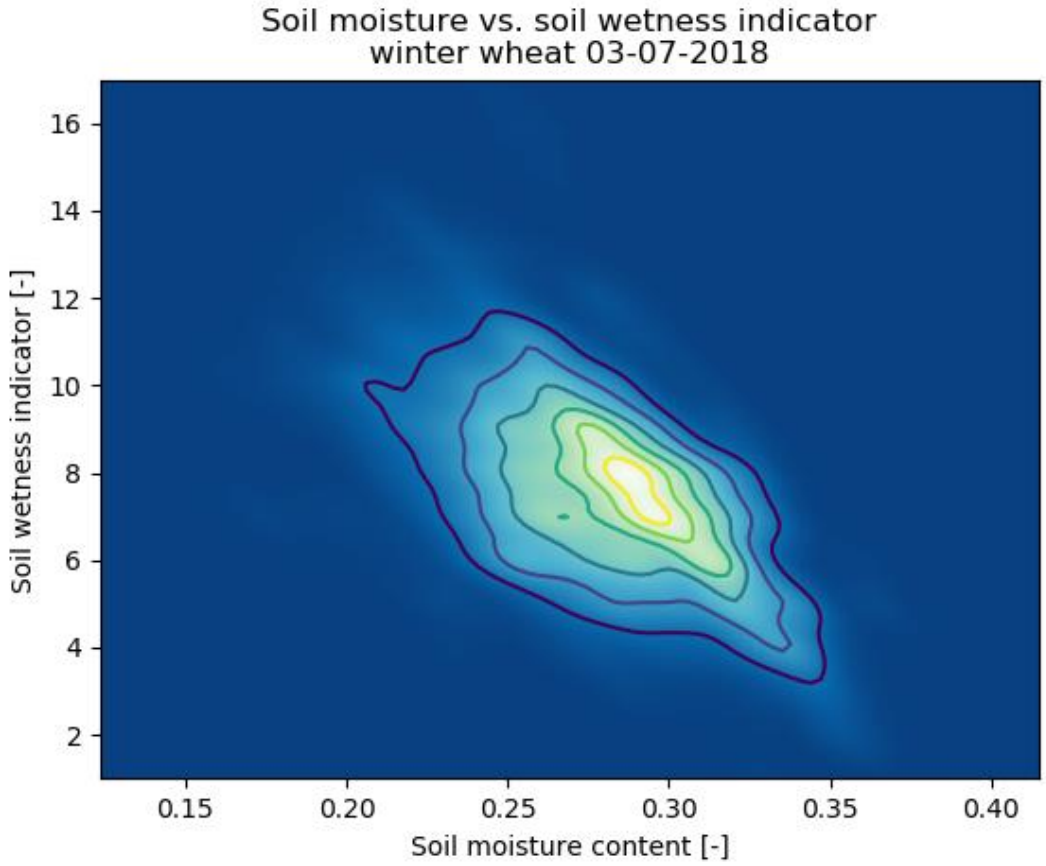


Figure 4.55 Density plot of soil moisture content vs. soil wetness indicator for winter wheat fields on 03-07-2018

The second step is to evaluate the results with help of a spatial visualization, see Figure 4.56 until Figure 4.61. For a better visualization, the color palette of the soil wetness indicator has been reversed because of the negative correlation between soil wetness indicator and soil moisture content. On all three dates, the spatial patterns of wetness correspond to each other. It could therefore be concluded that the soil wetness indicator for winter wheat fields give good results.

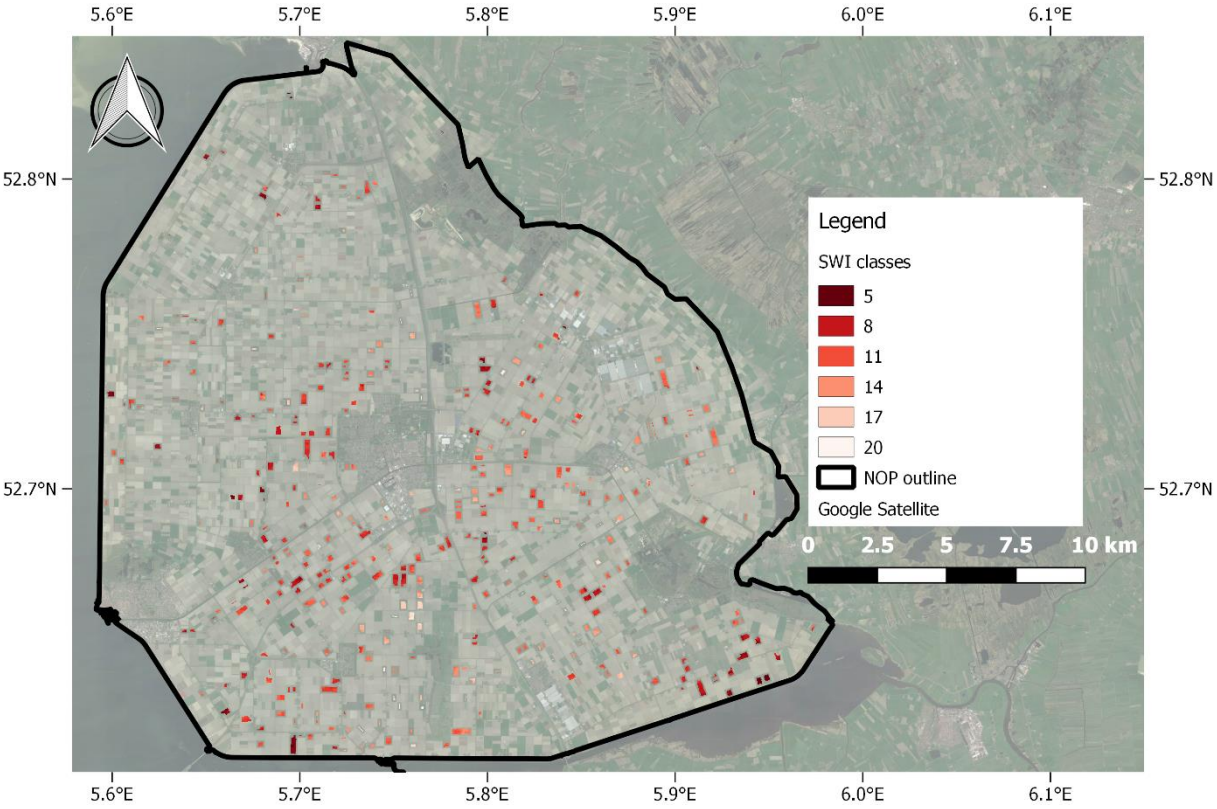


Figure 4.56 Soil wetness indicator map for winter wheat on 21-04-2018

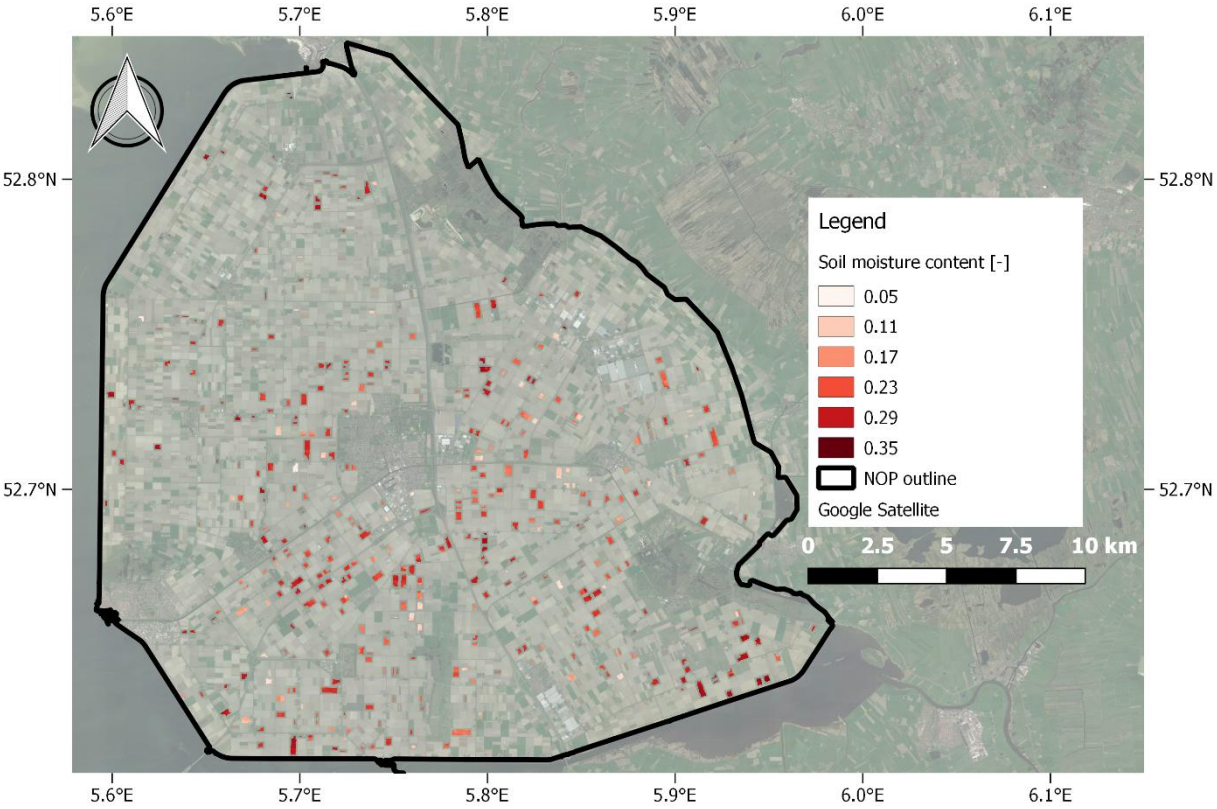


Figure 4.57 Soil moisture content map for winter wheat on 21-04-2018

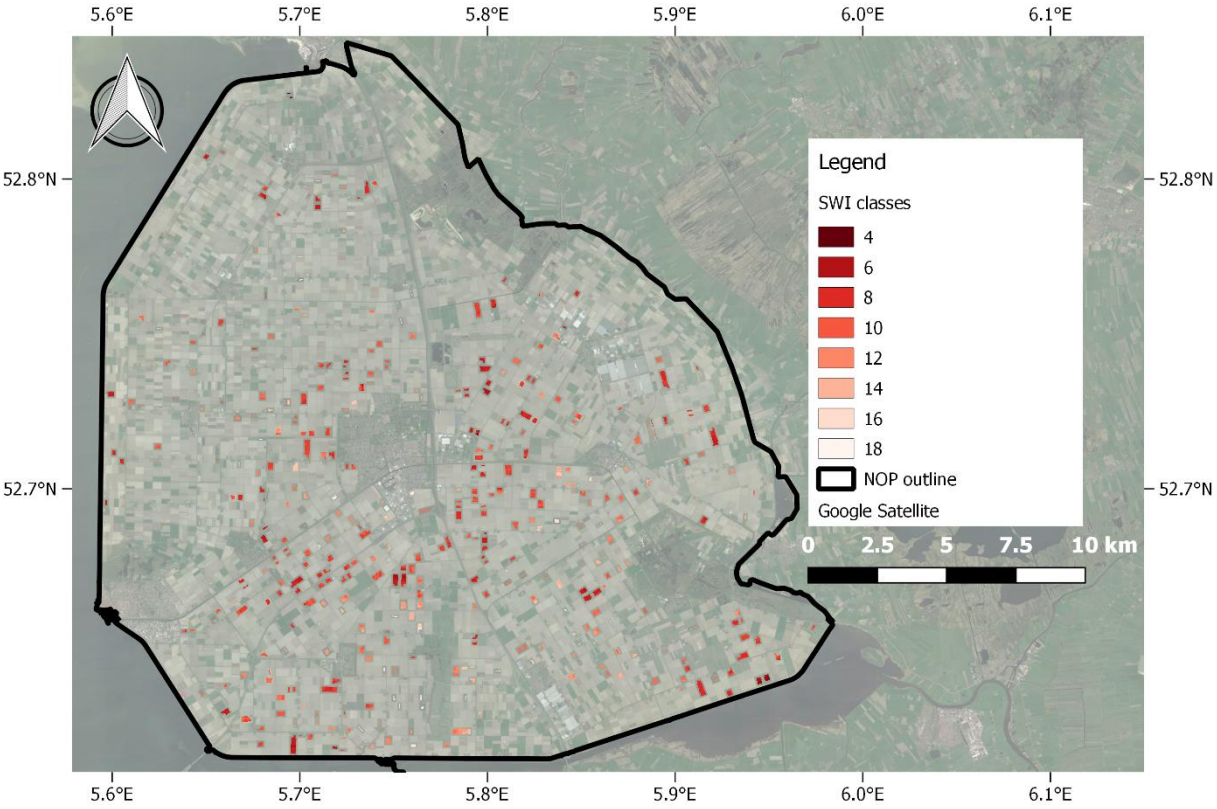


Figure 4.58 Soil wetness indicator map for winter wheat on 07-05-2018

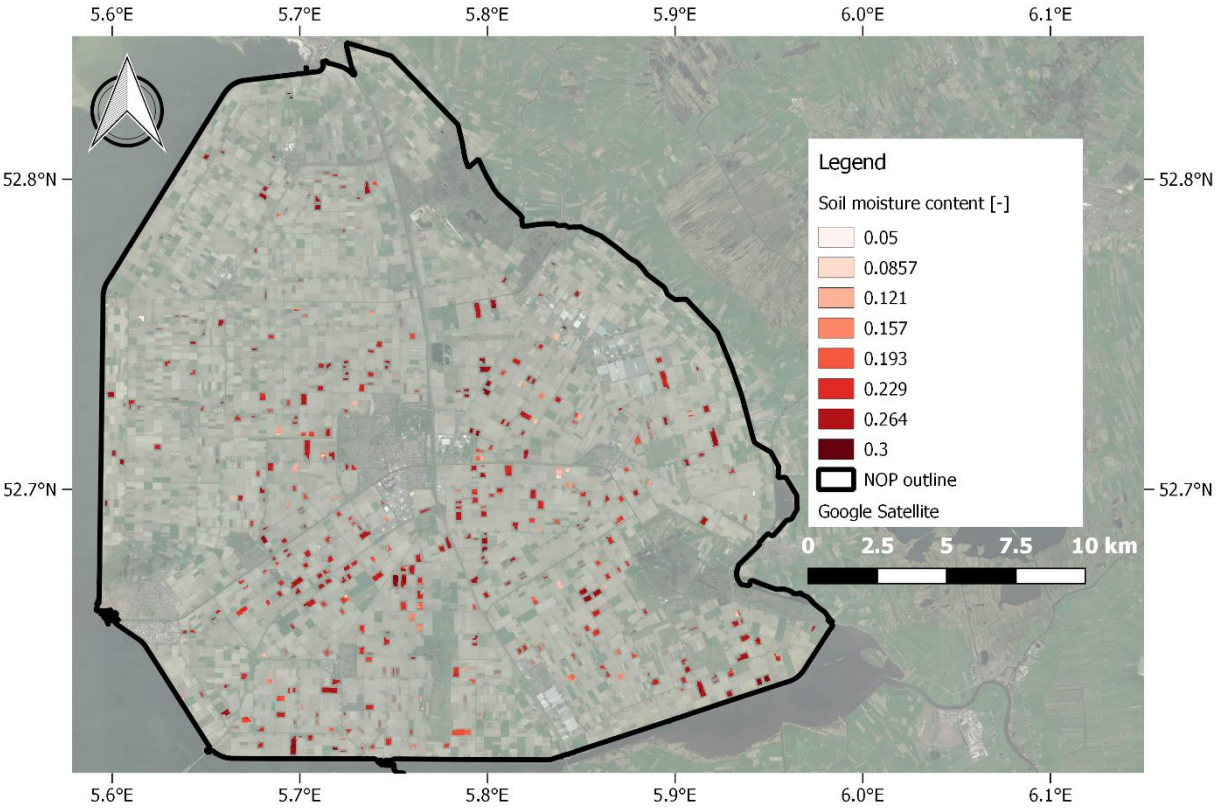


Figure 4.59 Soil moisture content map for winter wheat on 07-05-2018

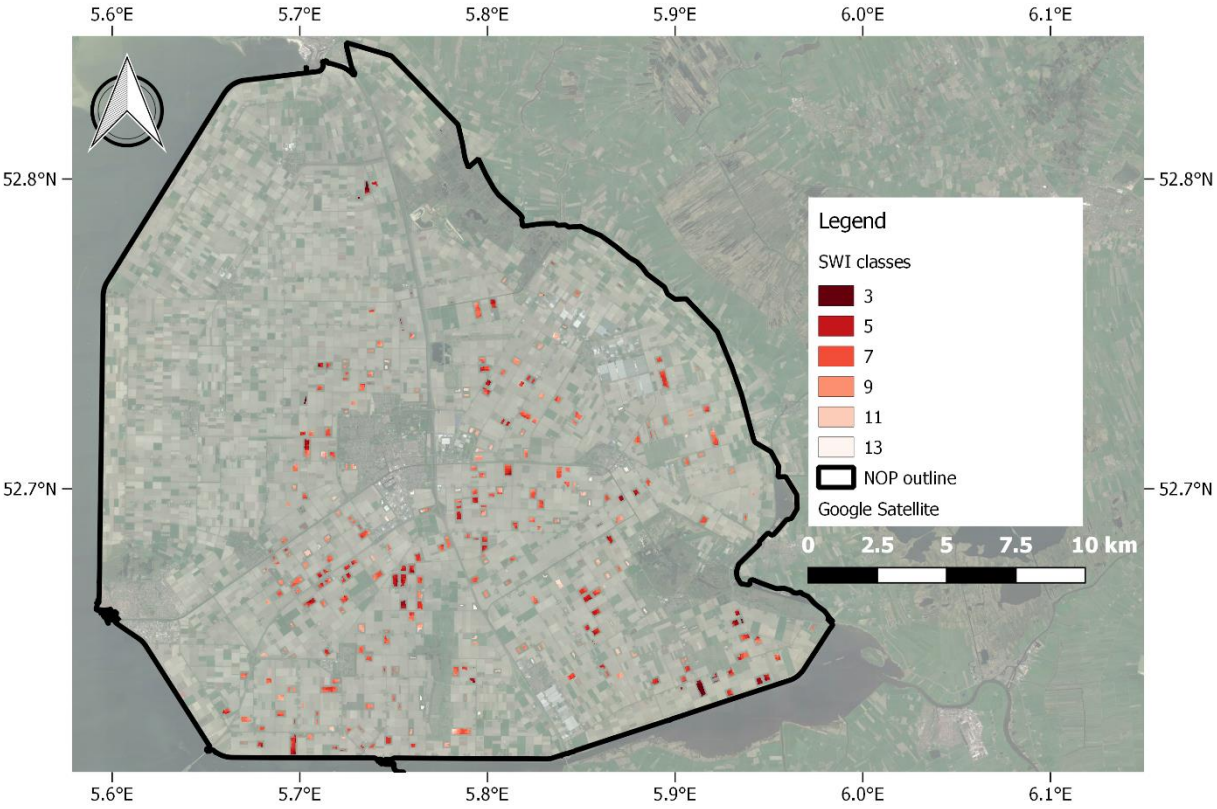


Figure 4.60 Soil wetness indicator map for winter wheat on 03-07-2018

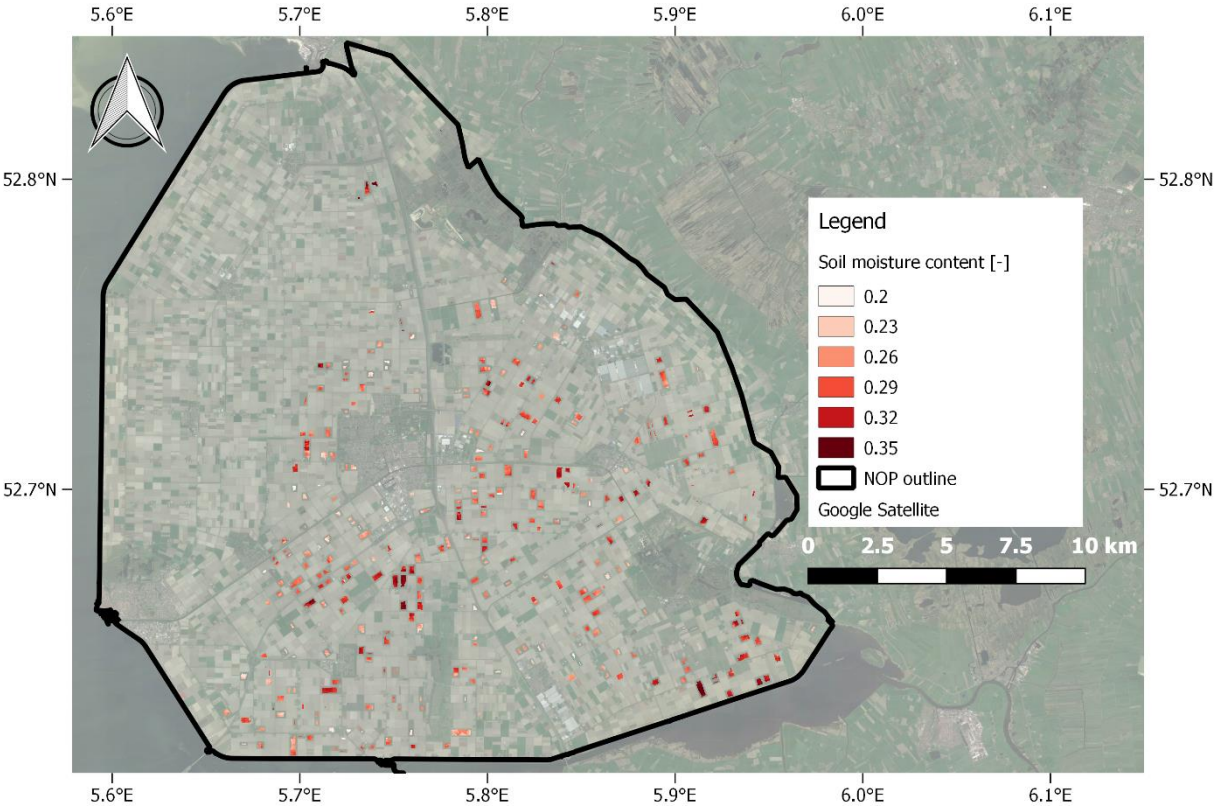


Figure 4.61 Soil moisture content map for winter wheat on 03-07-2018

5. Discussions and recommendations

5.1. Spatial estimation of surface soil properties using remote sensing data

Clay and organic matter content

The results of the SoilGrids30m model outperformed the existing SoilGrids250m and SoilGrids1000m models. Especially the result for organic matter content are significantly better than the two coarser existing SoilGrids models. However, the improvement for the clay content model is small. Therefore, it is arguable whether it is worth it to apply a newly developed SoilGrids30m model instead of the existing models for clay content estimates in other regions. It would be interesting to test the model in different regions to see if the results consistently improve the clay content and organic matter content estimates compared to the two coarser SoilGrids models.

The SoilGrids30m model is based on bare soil surface reflectance remote sensing data. According to Carlson et al. (1997), bare soils are indicated with an NDVI value in between 0.0 and 0.2. According to Gandhi et al. (2015), bare soils are indicated with an NDVI value in between 0.0 and 0.1. In this study, the suggestion of Carlson et al. (1997) has been used but it is arguable if a lower NDVI would be better for the estimation of clay and organic matter content. With an NDVI value of 0.2, there could already be small parts of vegetation in a square pixel of 30 meters. To avoid the influence of vegetation completely it would be better to use a lower bare soil NDVI threshold of in example 0.1.

The spatial estimation of clay and organic matter content is done with help of explanatory variables and soil samples. The used explanatory variables are chosen based on the SCORPAN properties but not all properties are applied in this study. For the clay content estimates, the elevation and slope played an important role based on the contribution to the selected principal components. However, the correlation between the elevation/slope and clay content is low. Allice (2016) stated that due to the use of principal component analysis, the relation between an explanatory variable and the target variable is decoupled. The decoupling is clearly shown with the high contribution of elevation and slope in the selected principal components and the low correlation between elevation/slope and clay content. For organic matter content estimates, band 6 and band 7 played an important role based on the contribution to the selected principal components. According to Summers et al., 2011; Rossel & Behrens, 2010; Ertlen et al., 2010, many organic components can be assigned to absorption bands in the short-wave infrared region of the electromagnetic spectrum (band 6 and band 7 of Landsat 8) that could be an explanation of the high contribution of these bands. As mentioned above, the decoupling between the explanatory variables and the target variable due to the use of principal component analysis makes it difficult to find out what is affecting what. The found relations between the target variables and the explanatory variables from the principal component analysis are therefore somehow coincidental relations. Bottom-line, for every outcome there could be a possible explanation found.

In this study, a selection of explanatory variables has been made based on the SCORPAN properties. There are dozens of explanatory variables that can be used as input for the SoilGrids30m model that makes the selection of explanatory variables used for this study somehow subjective. The model could be extended with several explanatory variables some examples are shown below.

- Time series of climate conditions, vegetation indices and soil indices;
- More detailed lithology and geomorphology maps;
- Hyperspectral images to use the absorption bands related to clay and or organic matter;

- Ground water table depth.

Instead of time series, the difference between a wet and a dry date has been used. The results of the differenced explanatory variables did not had as much influence as initially was expected. It is difficult to measure the reflectance of a wet soil because there should be a cloudless moment when the soil is still wet due to precipitation in the whole study area. To test the hypothesis of differenced explanatory variables it would be better to use reflectance data obtained from drones or airplanes that do not have to deal with cloudless circumstances and do not depend on specific overpass time.

The analysis also depends on the available soil samples in the study area. Due to outlier removal and the restriction to use only bare soil pixels, the amount of soil samples available for analysis decreases significantly to approximately 150 soil samples (30% of total). Coincidently, the variability in organic matter content of the soil samples left is very low. According to McBratney et al. (2003), a perfect set of soil samples should, one have enough variability, two have a homogeneous spread throughout the study area and three be dense enough. Therefore, it could be argued if the set of soil samples available in this study was sufficient to do the analysis with. Due to the low variability in organic matter content, it would be recommended to use a simpler estimation technique such as ordinary kriging, which can use all soil samples available in the area.

The model itself uses two thresholds to adjust the model as desired, the correlation threshold and the minimum variance explained threshold. The correlation threshold is to select a set of explanatory variables to use for further analysis based on their correlation to the target variable. For this study, a range from 0 to 95 percentile with an increment of 5 percentile has been analyzed. With an increment of 5 percentile, two or three explanatory variables will be left out of the analysis each time. The model could be analyzed more extensively by setting the percentile value in such a way that an increment removes only one explanatory variable each time. Furthermore, the minimum variance explained threshold is to select a set of principal components that accounts for the threshold amount of variance in the model. In this study, this threshold has been fixed to 99 percent. According to Kim et al. (2005), the obtained set of principal components can be decomposed into three parts: the first principal component represent an effect that influences all explanatory variables, followed by a set of principle components representing synchronized fluctuations affecting groups of explanatory variables, all remaining principal components represents randomness in for example the Landsat 8 images. The question is how many principal components should be selected to have a minimum loss of information and a maximum reduction of data. Rea et al. (2016) suggested two new methods based on a heat map and change in eigenvector angle that could be applied to find the optimal number of principal components. The SoilGrids30m model could be further automated by applying a method proposed by Rea et al. (2016) that make the model less subjective.

Soil water holding capacity

To estimate soil water holding capacity, pedotransfer functions are used. Pedotransfer functions translate “easily” to obtain data into data that is not so easily to obtain. In this study, clay content and organic matter content are used as input for the pedotransfer functions to transform them into soil hydraulic parameters. Databases of known soil hydraulic parameters are used to calibrate the pedotransfer functions; the outcome therefore strongly depends on what database has been used. Each area will have its own characteristics in soil hydraulic parameters depending on what type of soils there are. Pedotransfer functions are therefore strong empirical functions. The soil samples available in the study area only contain clay content and organic matter content. However, the pedotransfer functions related to only clay content and organic matter content are scarce. For this study, only one set of pedotransfer

functions has been found to obtain soil hydraulic parameters. Three of five pedotransfer functions are calibrated based on a database of Dutch soils; the other two are calibrated based on a database of local German soils. The use of pedotransfer functions calibrated on databases from two different regions could already lead to wrong assumptions. Furthermore, the pedotransfer functions related to Dutch soils are specifically calibrated for clayey soils. According to the soil samples in the study area, the range of clay content is 2-36%. There are clearly areas that do not fall in the category of clayey soils. The use of pedotransfer functions calibrated with different databases and the restriction to use them for clayey soils makes the pedotransfer functions even more unreliable. The reliability of the pedotransfer functions could be improved in two different ways. One, a new set of pedotransfer functions can be developed and calibrated on a database related to the study area. Two, use pedotransfer functions based on more input parameters such as sand content or silt content. The latter point is only possible if there is more information available at the soil sample locations, which is not the case for the soil samples used in this study. A new survey should be executed to obtain the additional input parameters.

The drought period during May until July should result in a significant water deficiency throughout the study area, assuming no irrigation was applied to the fields of interest. According to the soil moisture estimates from SEBAL, the crops had a sufficient amount of water available during the drought period. There are five possible explanations: the soil water holding capacity is extremely high in the study area, the crop rooting depth increases significantly in time, local precipitation, farmers did irrigate their fields or seepage played a significant role in the study area. According to FAO – sugar beet (n.d.), the evapotranspiration rate of sugar beet is 5 to 6 mm/day in normal conditions. The drought period took place from May until July, with a total precipitation rate of 71.7 mm in 87 days. Based on these numbers there should be at least 360 mm of water in the soil for crop growth, during the drought period (assuming the available water in the soil was in equilibrium state before May). In addition, the FAO – winter wheat (n.d.) and FAO – sugar beet (n.d.), also stated that for both crop types normally all water would be extracted from the first 1.2 meters of soil (rooting depth). Assuming a fixed root depth of 1.2 meters during the drought period, at least a soil water holding capacity of 300 mm/m is necessary to have a sufficient water supply for crop growth in the drought period. According to Rousseva et al. (2017), a soil water holding capacity of +300 mm/m is extremely high. Therefore, it is not to be expected that the soil water holding capacity in the area has a magnitude of in the range of +300 mm/m. It would also mean that the estimated soil water holding capacity in the study area would be off with more than a factor of two.

Another possibility is the rooting depth that has been fixed to 1.2 meters in the previous calculation. Crops uses their roots to provide themselves with sufficient water. The root depth of a crop will increase as the growth season progresses. The increasing root depth provides the crops of water from deeper soil layers. However, the rooting depth mainly depends on the bulk density of a soil (Lipiec et al., 2003) in which soil compaction plays an important role (Lipiec et al., 2003; Håkansson et al., 2000). Roots will find their way into the ground with help of existing cracks, pores and wormholes (Lipiec et al., 2003). Due to compaction these rooting paths could be affected which makes it harder for roots to reach greater depths. Therefore, soils with a high bulk density could reduce the rooting depth of crops. This statement would suggest that there should be a negative correlation between the soil water holding capacity and the bulk density. The results of this study show a positive correlation that could indicate that there are no clear indications of compaction throughout the study area. Furthermore, deep roots often follow the same existing cracks, pores and wormholes and therefore the area to extract water from will decrease (Brown et al., 1987). The estimated mean soil water holding capacity in the study area is 130 mm/m. To reach the requirement of having 360 mm of water available for crop use, the root depth should be at least 2.75 meters. According to the FAO –

winter wheat (n.d.) and FAO – sugar beet (n.d.), these root depths are not common. In addition, the estimated soil water holding capacity shows significant lower magnitudes. This would indicate that the rooting depth also did not play a significant role in the available water content during the drought period.

Three external factors are left that could explain the water availability according to the soil moisture estimates of SEBAL: local precipitation, irrigation and or seepage. None of the mentioned external factors has been accounted for in this study. In the study area, there is only one KNMI weather station located that has been used for precipitation analysis. Therefore, local precipitation cannot be ruled out. However, the months May until July were extremely dry and it can be assumed that in these months local precipitation has been minimal. The extreme drought conditions during these months brings up the next possible explanation. Farmers always try to create perfect crop development conditions. During long periods of drought, farmers could decide to irrigate their crops. Irrigation could artificially keep the water supply at the right level for crop growth. To minimalize the probability of irrigation throughout the season, sugar beet and winter wheat has been used for the analysis. These crop types have a low demand of water and generally do not need irrigation in the Dutch climate. The last explanation is the natural process seepage, which is present in the area. The study area is an artificial drained area and is mainly located below sea level. Groundwater in subsoil layers from higher located places will, under pressure, flow to lower located areas where it could reach the surface layer. This natural process could take place on larger scales but also on smaller scales. According to a paper by Bastiaanssen (2005), the Noordoostpolder area is having a seepage rate of approximately 340 millimetre per year. This is a significant positive seepage rate in the study area, which means that it should be taken into consideration. In this study, the assumption has been made that seepage does not play a role because of the extreme dry conditions in the months May until July. However, seepage and irrigation both could play an important role in the supply of water for crop growth in the study area and kept the soil moisture content in the area at a sufficient level for crop growth. Furthermore, the estimation of soil water holding capacity could be too low. With for example a soil water holding capacity of 200 mm/m the rooting depth should be minimal 1.8 meters, which is reasonable for both crop types (FAO – winter wheat, n.d.; FAO – sugar beet, n.d.).

As stated by van der Kwast (2009), hydrological parameters cannot directly be extracted from remote sensing data and therefore ground observations are a necessity to validate the results. The estimated soil water holding capacity in the study area could not be validated because there are no ground observations available. Therefore, it cannot be determined what the reliability of the soil water holding capacity estimates is. This can only be done when there are measurements of soil hydraulic parameters available in the study area. It would therefore be recommended to validate the model in an area with known soil hydraulic parameters.

5.2. Soil wetness indicator

For both crop types, sugar beet and winter wheat, the soil wetness indicator is a good spatial indicator of relative wetness. It is a relative indicator because the classes of wetness only represents the wetness compared to the at date crop/weather conditions. Therefore, it is not possible to obtain soil moisture estimates from the soil wetness indicator. However, the soil wetness indicator gives a quick first indication of the wetness conditions in the area without any difficult computations.

The goal of the soil wetness indicator is to find patterns of wetness based on the relation between NDVI and relative crop temperatures. The relative crop temperature is a measure of the root-zone soil properties that give an insight in the water availability at a greater depth than the topsoil. To have an unbiased representation of the water availability in the root zone it is important to measure only crop temperatures. However, the spatial resolution of 30 meters causes an aggregation of temperatures consisting of vegetation and bare soil. This problem has partly been solved by introducing a buffer space of 60 meters around each crop field. It is though not possible to avoid the temperature influences of uncovered soils like spray paths. It is one of the reasons why for this study sugar beet and winter wheat has been used because of their great soil coverage. It would be interesting to see what the results of the model would be with centimetre scale images from drones in example. At centimetre scale, almost all soil pixels can be eliminated and the relative crop temperature will be a better representative of the root zone soil water availability.

If there is a sufficient water supply available, crops will not show signs of stress and the relative crop temperature will stay low. As consequence, there will be no significant differences noticeable in the study area. Three external factors could provide the crops of sufficient water supply that are not taken into account in this study: local precipitation, irrigation and or seepage. All three factors are already explained in the previous paragraph.

The model itself is subjective while the user can use tuning parameters to obtain the “best” results. The visually inspection of boundary percentile values that are not in line of expectation is dependent on the interpretation of the user of the model. With this subjectivity, it is possible to obtain different results by different users. The model is therefore only a simple indicator of the soil wetness patterns in the study area. To improve the model and to take away the subjectivity of the model, the boundaries of the pixel envelope should be determined automatically.

Other additional recommendations can be made about the crop type and the available images. In this study only sugar beet and winter wheat has been used for analysis because of the mentioned reasons. It could be interesting to see if there are significant differences between other crops such as potato, onion, maize, etc. Furthermore, the moment in time of the images are fixed because of the dependency of satellite overpass time and cloud cover. To have a better representation of the soil wetness indicator it could be interesting to have more images spread throughout the season in example with drone images.

6. Conclusions

In this study, two newly developed methods are proposed to determine the spatial patterns of soil wetness in the study area. On the one hand, the SoilGrids30m model to estimate clay content and organic matter content that are used as input of pedotransfer functions to estimate soil water holding capacity. On the other hand, the soil wetness indicator to determine spatial patterns of wetness in the study area.

The SoilGrids30m model improved the estimation of clay content and organic matter content in the study area compared to the existing coarser SoilGrids250m and SoilGrids1000m models. Especially, the organic matter content estimates significantly improves with help of the SoilGrids30m model. However, the influence of adding bare soil surface reflectance to the model was relatively low for both clay content (37%) and organic matter content (13%). The influence of bare soil surface reflectance only improved the clay content estimates compared to the SoilGrids1000m model. In all other cases, the target variable estimates did not improve by adding bare soil surface reflectance data compared to the SoilGrids250m and SoilGrids1000m models. It can therefore be concluded that the improvement of the SoilGrids30m model is due to the use of a larger set of observation data from the study area and not because of the input of explanatory variables.

After a long period of drought, the soil moisture estimates of SEBAL showed a relative high and wide range of soil moisture content in the study area per crop type. Three factors could be appointed as the cause of the relative high and wide range of soil moisture content: seepage, irrigation and or a high and wide range of soil water holding capacity in the study area. According to the Stiboka soil map (https://www.pdok.nl/), a wide range of soils are available in the study area, from sandy soils to clayey soils. The wide range of soils corresponds to a wide range of soil moisture content from SEBAL. The soil water holding capacity estimates in the study area are in the range of 100 to 150 mm/m, which is a relatively low and small range. These estimates therefore do not correspond to the soil moisture content estimates from SEBAL and the Stiboka soil map. An important step in obtaining soil water holding capacity is the use of pedotransfer functions. Only one set of pedotransfer functions could be found that is applicable for this study. This set was calibrated based on different soil hydrological databases and some of them are specifically calibrated for clayey soils only. It can therefore be concluded that the pedotransfer functions used in this study are not sufficient for reliable results. The high empirical level of the pedotransder functions makes them highly unreliable. Therefore, the use of pedotransfer functions in general would not be recommended. Only if the pedotransfer functions are calibrated based on soil properties from the study area itself it would be useful for further analysis.

With the soil wetness indicator an easy and simple model has been introduced based on two parameters: normalized differenced vegetation index and the land surface temperature. The results of the soil wetness indicator showed a clear correlation with the soil moisture content estimates from SEBAL. The soil wetness indicator is therefore a simple and reliable tool to recognize spatial patterns of wetness. In example for precision agriculture, the soil wetness indicator could be used for irrigation management. The relative representation of the soil wetness indicator could determine the distribution of irrigation needs within a field.

In theory, both developed models are promising but in practice there are always unforeseen factors that play an important role. Especially in the estimation of soil water holding capacity there is a lot of space for improvement. In example, the contribution of explanatory variables to the spatial estimation of clay content and organic matter content is too low. The relation between environmental properties and clay content and organic matter content is proven but

clearly difficult to obtain with the SoilGrids30m model. Furthermore, the use of pedotransfer functions are unreliable and therefore new approaches or pedotranfer functions specifically obtained from the study area are highly recommended to derive soil hydrological properties. Overall, it can be concluded that there is still a long way to go for precision agriculture and the development of data driven estimations of natural processes.

Bibliography

A scientific description of SEBAL. (n.d.). Retrieved January 28, 2019, from http://www.waterwatch.nl/tools0/sebal/sebal-a-scientific-description.html

Alice, M. (2016, July 21). Performing Principal Components Regression (PCR) in R. Retrieved March 30, 2019, from http://www.milanor.net/blog/performing-principal-components-regression-pcr-in-r/

Allen, R. G., Tasumi, M., Trezza, R., Waters, R., & Bastiaanssen, W. (2002). Surface Energy Balance Algorithm for Land (SEBAL)–Advanced training and Users Manual. *Kimberly: Idaho Implementation*.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, *300(9),* D05109.

Anderson, M. C., Norman, J. M., Mecikalski, J. R., Otkin, J. A., & Kustas, W. P. (2007). A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research: Atmospheres*, *112*(D10).

Bastiaanssen, W. G. M., Pelgrum, H., Soppe, R. W. O., Allen, R. G., Thoreson, B. P., & de C. Teixeira, A. H. (2006, August). Thermal-infrared technology for local and regional scale irrigation analyses in horticultural systems. In *V International Symposium on Irrigation of Horticultural Crops 792* (pp. 33-46).

Bastiaanssen, W. G. M., van den Bersselaar, D., Jaarsma, M., & Zwart, S. (2005). *Opsporen hydrologische knelpunten in Noordoostpolder met remote sensing*. *H2O*, 22, 38–41.

Bastiaanssen, W. G. M., Noordman, E. J. M., Pelgrum, H., Davids, G., Thoreson, B. P., & Allen, R. G. (2005). SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of irrigation and drainage engineering*, *131*(1), 85-93.

Bastiaanssen, W. G. (2000). SEBAL-based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. *Journal of hydrology*, *229*(1-2), 87-100.

Bastiaanssen, W. G., Menenti, M., Feddes, R. A., & Holtslag, A. A. M. (1998a). A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of hydrology*, *212*, 198-212.

Bastiaanssen, W. G., Pelgrum, H., Wang, J., Ma, Y., Moreno, J. F., Roerink, G. J., & Van der Wal, T. (1998b). A remote sensing surface energy balance algorithm for land (SEBAL).: Part 2: Validation. *Journal of hydrology*, *212*, 213-229.

Bastiaanssen, W. G., Pelgrum, H., Droogers, P., De Bruin, H. A. R., & Menenti, M. (1997). Area-average estimates of evaporation, wetness indicators and top soil moisture during two golden days in EFEDA. *Agricultural and Forest Meteorology*, *87*(2-3), 119-137.

Bastiaanssen, W. G. M. (1995). Regionalization of surface flux densities and moisture indicators in composite terrain. *A remote sensing approach under clear skies in Mediterranean climates*.

BISNederland [Dataset]. (n.d.). Retrieved November 12, 2018, from http://maps.bodemdata.nl/bodemdatanl/index.jsp

Bouma, J. (1989). Using soil survey data for quantitative land evaluation. In *Advances in soil science* (pp. 177-213). Springer, New York, NY.

Brown, M., Moran, S., Escobar, V., & Entekhabi, D. (2011). Soil Moisture Active Passive (SMAP) Mission Applications Plan. *NASA Jet Propulsion Lab. y, Pasadena, CA, USA*

Brown, K. F., Messem, A. B., Dunham, R. J., & Biscoe, P. V. (1987). Effect of drought on growth and water use of sugar beet. *The Journal of Agricultural Science*, *109*(3), 421-435.

Burgess, T. M., & Webster, R. (1980). Optimal interpolation and isarithmic mapping of soil properties. *Journal of soil science*, *31*(2), 315-331.

Cambardella, C. A., Moorman, T. B., Parkin, T. B., Karlen, D. L., Novak, J. M., Turco, R. F., & Konopka, A. E. (1994). Field-scale variability of soil properties in central Iowa soils. *Soil science society of America journal, 58*(5), 1501-1511.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), 241-252.

Casa, R., Castaldi, F., Pascucci, S., Palombo, A., & Pignatti, S. (2013). A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma, 197*, 17-26.

Ceddia, M. B., Vieira, S. R., Villela, A. L. O., Mota, L. D. S., Anjos, L. H. C. D., & Carvalho, D. F. D. (2009). Topography and spatial variability of soil physical properties. *Scientia Agricola*, *66*(3), 338-352.

Christensen, R. (2001). Plane answers to complex questions: the theory of linear models. *Springer Science & Business Media*, 124-130

Clark, I. (2010). Statistics or geostatistics Sampling error or nugget effect. *Journal of the Southern African Institute of Mining and Metallurgy*, *110*(6), 307-312.

Dataset: Basisregistratie Ondergrond (BRO). (2018). Retrieved March 19, 2019, from https://www.pdok.nl/introductie/-/article/basisregistratie-ondergrond-bro-

De Vries, F., De Groot, W. J. M., Hoogland, T., & Denneboom, J. (2003). De Bodemkaart van Nederland digitaal; toelichting bij inhoud, actualiteit en methodiek en korte beschrijving van additionele informatie (No. 811). *Alterra.*

Ertlen, D., Schwartz, D., Trautmann, M., Webster, R., & Brunet, D. (2010). Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy. *European Journal of Soil Science*, *61*(2), 207-216.

FAO – Sugar beet (Food and Agriculture Organization of the United Nations). (n.d.). Retrieved March 20, 2019, from http://www.fao.org/land-water/databases-and-software/crop-information/sugarbeet/en/

FAO – Wheat (Food and Agriculture Organization of the United Nations). (n.d.). Retrieved March 20, 2019, from http://www.fao.org/land-water/databases-and-software/crop-information/wheat/en/

Forkuor, G., Hounkpatin, O. K., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, *12*(1), e0170478.

Funderburg, E. (2016). Organic matter serves important role in soil health. *Ag News and Views*, *34* (2), 1-2.

Gandhi, G. M., Parthiban, S., Thummalu, N., & Christy, A. (2015). NDVI: vegetation change detection using remote sensing and GIS–a case study of Vellore District. *Procedia Computer Science*, *57*, 1199-1210.

Gillies, R. R., Kustas, W. P., & Humes, K. S. (1997). A verification of the'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index (NDVI) and surface radiant temperature. *International journal of remote sensing*, *18*(15), 3145-3166.

Gillies, R. R., & Carlson, T. N. (1995). Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *Journal of Applied Meteorology*, *34*(4), 745-756.

Håkansson, I., & Lipiec, J. (2000). A review of the usefulness of relative bulk density values in studies of soil structure and compaction. *Soil and Tillage Research*, *53*(2), 71-85.

Hatfield, J. L., Gitelson, A. A., Schepers, J. S., & Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, *100*(Supplement_3), S-117.

He, T., Wang, J., Lin, Z., & Cheng, Y. (2009). Spectral features of soil organic matter. *Geo-spatial Information Science*, *12*(1), 33-40.

Hengl, T., MacMillan, R.A., (2019). Predictive Soil Mapping with R. *OpenGeoHub foundation, Wageningen, the Netherlands,* 370 pages, www.soilmapper.org, ISBN: 978-0-359-30635-0

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, *12*(2), e0169748.

Hengl, T., Heuvelink, G. B., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & geosciences*, *33*(10), 1301-1315.

Idso, S. B., Schmugge, T. J., Jackson, R. D., & Reginato, R. J. (1975). The utility of surface temperature measurements for the remote sensing of surface soil water status. *Journal of Geophysical Research*, *80*(21), 3044-3049.

Jiménez-Muñoz, J. C., Sobrino, J. A., Skoković, D., Mattar, C., & Cristóbal, J. (2014). Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Geoscience and Remote Sensing Letters*, *11*(10), 1840-1843.

Khanal, S., Fulton, J., & Shearer, S. (2017). An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture, 139*, 22-32.

Kim, D. H., & Jeong, H. (2005). Systematic analysis of group identification in stock markets. *Physical Review E*, *72*(4), 046133.

Kirkham, M. B. (2014). Principles of soil and plant water relations. *Academic Press.*

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(6), 119-139.

Landbouw; gewassen, dieren en grondgebruik naar gemeente. (2018, November 20). Retrieved March 19, 2019, from https://opendata.cbs.nl/statline/

Levi, M. R., & Rasmussen, C. (2014). Covariate selection with iterative principal component analysis for predicting physical soil properties. *Geoderma*, *219*, 46-57.

Lipiec, J., Medvedev, V. V., Birkas, M., Dumitru, E., Lyndina, T. E., Rousseva, S., & Fulajtar, E. (2003). Effect of soil compaction on root growth and crop yield in Central and Eastern Europe. *International agrophysics*, *17*(2), 61-70.

Matheron, G. (1969). Le krigeage universel (Vol. 1). Paris: *École nationale supérieure des mines de Paris.*

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246-1266.

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1-2), 3-52.

McBratney, A. B., Odeh, I. O., Bishop, T. F., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, *97*(3-4), 293-327.

McMillin, L. M. (1975). Estimation of sea surface temperatures from two infrared window measurements with different absorption. *Journal of Geophysical Research*, *80*(36), 5113-5117.

Moran, M. S., Clarke, T. R., Inoue, Y., & Vidal, A. (1994). Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote sensing of environment*, *49*(3), 246-263.

Moreira, A. (2013). Synthetic aperture radar (SAR): principles and applications. *4th Advanced Training Course in Land Remote Sensing*. *ESA*

Nemani, R. R., & Running, S. W. (1989). Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied meteorology*, *28*(4), 276-284.

Odeha, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1994). Spatial estimation of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, *63*(3-4), 197-214.

O'Geen, A. T. (2013) Soil Water Dynamics. *Nature Education Knowledge 4(5):9*

O'Geen, A. T. (2012) Soil Water Dynamics. *Nature Education Knowledge 3(6):12*

Pachepsky, Y. A., Timlin, D. J., & Rawls, W. J. (2001). Soil water retention as related to topographic variables. *Soil Science Society of America Journal*, *65*(6), 1787-1795.

Plant and Soil Sciences eLibrary. (n.d.). Soil texture triangle. Table. Retrieved from http://passel.unl.edu/pages/informationmodule.php?idinformationmodule=1130447039

Pribyl, D. W. (2010). A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, *156*(3-4), 75-83.

Price, J. C. (1990). Using spatial context in satellite data to infer regional scale evapotranspiration. *IEEE transactions on Geoscience and Remote Sensing*, *28*(5), 940-948.

Rawls, W. J., & Brakensiek, D. L. (1989). Estimation of soil water retention and hydraulic properties. In *Unsaturated flow in hydrologic modeling* (pp. 275-300). Springer, Dordrecht.

Rea, A., & Rea, W. (2016). How Many Components should be Retained from a Multivariate Time Series PCA. *arXiv preprint arXiv:1610.03588*.

Rossel, R. V., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1-2), 46-54.

Rousseva, S., Kercheva, M., Shishkov, T., Lair, G. J., Nikolaidis, N. P., Moraetis, D., ... & Banwart, S. A. (2017). Soil water characteristics of European SoilTrEC critical zone observatories. In *Advances in Agronomy* (Vol. 142, pp. 29-72). Academic Press.

Rutter, A. J., & Sands, K. (1958). The relation of leaf water deficit to soil moisture tension in pinus Sylvestris L. *New Phytologist*, *57*(1), 50-65.

Schaap, M. G., Leij, F. J., & Van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of hydrology*, *251*(3-4), 163-176.

Scheinost, A. C., Sinowski, W., & Auerswald, K. (1997). Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. *Geoderma*, *78*(3-4), 129-143.

Scott, C. A., Bastiaanssen, W. G., & Ahmad, M. U. D. (2003). Mapping root zone soil moisture using remotely sensed optical imagery. *Journal of Irrigation and Drainage Engineering*, *129*(5), 326-335.

Sobieraj, J. A., Elsenbeer, H., Coelho, R. M., & Newton, B. (2002). Spatial variability of soil hydraulic conductivity along a tropical rainforest catena. *Geoderma*, *108*(1-2), 79-90.

Sobrino, J. A., Li, Z. L., Stoll, M. P., & Becker, F. (1996). Multi-channel and multi-angle algorithms for estimating sea and land surface temperature with ATSR data. *International Journal of Remote Sensing*, *17*(11), 2089-2114.

Song, W., Mu, X., Ruan, G., Gao, Z., Li, L., & Yan, G. (2017). Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method. *International journal of applied earth observation and geoinformation*, *58*, 168-176.

Sreelash, K., Buis, S., Sekhar, M., Ruiz, L., Tomer, S. K., & Guerif, M. (2017). Estimation of available water capacity components of two-layered soils using crop model inversion: Effect of crop type and water regime. *Journal of hydrology*, *546*, 166-178.

Stolte, J., Wesseling, J. G., & Wösten, J. H. M. (1996). Pedotransfer functions for hydraulic and thermal properties of soil and the tool HERCULES. Wageningen: *DLO Winand Staring Centre.*

StructX. (n.d.). Densities of Different Soil Types. Table. Retrieved from https://structx.com/Soil_Properties_002.html

Summers, D., Lewis, M., Ostendorf, B., & Chittleborough, D. (2011). Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators, 11*(1), 123-131.

USGS. (2017). Product guide: Landsat surface reflectance-derived spectral indices (Version 3.6). Retrieved from https://landsat.usgs.gov/documents/si_product_guide.pdf

USGS. (2016). LANDSAT 8 (L8) Data users handbook (2nd ed.). Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-8-data-users-handbook

Van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., Kabat, P., Van Walsum, P. E. V., ... & Van Diepen, C. A. (1997). Theory of SWAP version 2.0; Simulation of water flow, solute transport and plant growth in *the soil-water-atmosphere-plant environment* (No. 71). DLO Winand Staring Centre.

van den Bersselaar, D., Jaarsma, M., Zwart, S. J., & Bastiaanssen, W. G. M. (2005). Opsporen hydrologische knelpunten in Noordoostpolder met remote sensing. *H2O*, *22*, 38-41.

Van der Kwast, J. (2009). Quantification of top soil moisture patterns: Evaluation of field methods, process-based modelling, remote sensing and an integrated approach. Utrecht University, Royal Dutch Geographical Society.

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil science society of America journal*, *44*(5), 892-898.

Vegetation Spectral Signature Cheat Sheet [Illustration]. (2017, May 16). Retrieved March 21, 2019, from https://grindgis.com/remote-sensing/vegetation-spectral-signature-cheat-sheet

Wagner, B., Tarnawski, V. R., Hennings, V., Müller, U., Wessolek, G., & Plagge, R. (2001). Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set. *Geoderma*, *102*(3-4), 275-297.

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J. C., Bizzarri, B., Wigneron, J. P., & Kerr, Y. (2007). Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrology Research, 38*(1), 1-20.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., & Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote sensing of environment*, *81*(2-3), 238-246.

Wösten, J., Veeman, G., De Groot, W., & Stolte, J. (2001). Waterretentie-en doorlatendheidskarakteristieken van boven-en ondergronden in Nederland: de Staringreeks. *Alterra-rapport No. 153,* Vernieuwde uitgave.

Wösten, J. H. M., Finke, P. A., & Jansen, M. J. W. (1995). Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma*, *66*(3-4), 227-237.

Xu, Z., & Zhou, G. (2008). Responses of leaf stomatal density to water status and its relationship with photosynthesis in a grass. *Journal of experimental botany*, *59*(12), 3317-3325.

Yang, Y., Guan, H., Long, D., Liu, B., Qin, G., Qin, J., & Batelaan, O. (2015). Estimation of surface soil moisture from thermal infrared remote sensing using an improved trapezoid method. *Remote Sensing*, *7*(7), 8250-8270.

Appendix A. SoilGrids30m R-script module 1

Module 1A: data collection

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(rgdal)
library(raster)

## Set working directory
setwd("~/Thesis/Data")

## Input parameters
Projection = "EPSG32632"
Resolution_input = "30"
ImageDate_dry = "2017_01_05"
ImageDate_wet = "2017_03_10"

## Read-in covariates remove outliers and fill outliers with nearest
neighbour function
ImageDate_dry_date <- as.Date(ImageDate_dry, "%Y_%m_%d")
DOY_dry = strftime(ImageDate_dry_date, format = "%j")
DOY_dry = ifelse(substr(DOY_dry, 1,1) == 0, substr(DOY_dry, 2,3), DOY_dry)
DOY_dry = ifelse(substr(DOY_dry, 1,1) == 0, substr(DOY_dry, 2,2), DOY_dry)
Year_dry = strftime(ImageDate_dry_date, format = "%Y")
ImageDate_wet_date <- as.Date(ImageDate_wet, "%Y_%m_%d")
DOY_wet = strftime(ImageDate_wet_date, format = "%j")
DOY_wet = ifelse(substr(DOY_wet, 1,1) == 0, substr(DOY_wet, 2,3), DOY_wet)
DOY_wet = ifelse(substr(DOY_wet, 1,1) == 0, substr(DOY_wet, 2,2), DOY_wet)
Year_wet = strftime(ImageDate_wet_date, format = "%Y")

## Load data
NDVI_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_ndvi_",Resolution_input,"m_",Year_dry,"_",DOY_dry,".tif"))
NDVI_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_ndvi_",Resolution_input,"m_",Year_wet,"_",DOY_wet,".tif"))

B2_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B2_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
Exp_var_stack <- B2_dry["band1"]
names(Exp_var_stack)[1] <- "SPC_SEBAL_B2_dry"

B2_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B2_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B2_wet <- B2_wet$band1

B3_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B3_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
```

```
Exp_var_stack$SPC_SEBAL_B3_dry <- B3_dry$band1

B3_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B3_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B3_wet <- B3_wet$band1

B4_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B4_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
Exp_var_stack$SPC_SEBAL_B4_dry <- B4_dry$band1

B4_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B4_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B4_wet <- B4_wet$band1

B5_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B5_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
Exp_var_stack$SPC_SEBAL_B5_dry <- B5_dry$band1

B5_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B5_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B5_wet <- B5_wet$band1

B6_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B6_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
Exp_var_stack$SPC_SEBAL_B6_dry <- B6_dry$band1

B6_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B6_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B6_wet <- B6_wet$band1

B7_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_spectral_reflectance_B7_",Resolution_input,"m_",Year_dry,"_",D
OY_dry,".tif"))
Exp_var_stack$SPC_SEBAL_B7_dry <- B7_dry$band1

B7_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_spectral_reflectance_B7_",Resolution_input,"m_",Year_wet,"_",D
OY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_B7_wet <- B7_wet$band1

Elevation <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/proy_DEM_",Resolution_input,"m.tif"))
Exp_var_stack$SPC_SEBAL_Elevation <- Elevation$band1
```

```
Aspect <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/aspect_",Resolution_input,"m.tif"))
Exp_var_stack$SPC_SEBAL_Aspect <- Aspect$band1

Slope <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/slope_",Resolution_input,"m.tif"))
Exp_var_stack$SPC_SEBAL_Slope <- Slope$band1

Surface_roughness_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_LS8_surface_roughness_",Resolution_input,"m_",Year_dry,"_",DOY
_dry,".tif"))
Exp_var_stack$SPC_SEBAL_Surface_roughness_dry <-
Surface_roughness_dry$band1

Surface_roughness_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_LS8_surface_roughness_",Resolution_input,"m_",Year_wet,"_",DOY
_wet,".tif"))
Exp_var_stack$SPC_SEBAL_Surface_roughness_wet <-
Surface_roughness_wet$band1

Emissivity_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_tir_emissivity_",Resolution_input,"m_",Year_dry,"_",DOY_dry,".
tif"))
Exp_var_stack$SPC_SEBAL_Emissivity_dry <- Emissivity_dry$band1

Emissivity_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_tir_emissivity_",Resolution_input,"m_",Year_wet,"_",DOY_wet,".
tif"))
Exp_var_stack$SPC_SEBAL_Emissivity_wet <- Emissivity_wet$band1

Albedo_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_surface_albedo_",Resolution_input,"m_",Year_dry,"_",DOY_dry,".
tif"))
Exp_var_stack$SPC_SEBAL_Albedo_dry <- Albedo_dry$band1

Albedo_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_surface_albedo_",Resolution_input,"m_",Year_wet,"_",DOY_wet,".
tif"))
Exp_var_stack$SPC_SEBAL_Albedo_wet <- Albedo_wet$band1

Nitrogen_dry <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_dry,"/Co
variates/LS8_nitrogen_",Resolution_input,"m_",Year_dry,"_",DOY_dry,".tif"))
Exp_var_stack$SPC_SEBAL_Nitrogen_dry <- Nitrogen_dry$band1

Nitrogen_wet <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/",ImageDate_wet,"/Co
variates/LS8_nitrogen_",Resolution_input,"m_",Year_wet,"_",DOY_wet,".tif"))
Exp_var_stack$SPC_SEBAL_Nitrogen_wet <- Nitrogen_wet$band1

Exp_var_stack$SPC_EVI_dry <- 2.5*(B5_dry$band1-
B4_dry$band1)/(B5_dry$band1+6*B4_dry$band1-7.5*B2_dry$band1+1)
```

```
Exp_var_stack$SPC_EVI_wet <- 2.5*(B5_wet$band1-
B4_wet$band1)/(B5_wet$band1+6*B4_wet$band1-7.5*B2_wet$band1+1)

Exp_var_stack$SPC_MSAVI_dry <- B5_dry$band1 + 0.5 -
0.5*sqrt((2*B4_dry$band1+1)^2 - 8*(B4_dry$band1-B3_dry$band1))

Exp_var_stack$SPC_MSAVI_wet <- B5_wet$band1 + 0.5 -
0.5*sqrt((2*B4_wet$band1+1)^2 - 8*(B4_wet$band1-B3_wet$band1))

Exp_var_stack$SPC_NDMI_dry <- (B5_dry$band1-B6_dry$band1) /
(B5_dry$band1+B6_dry$band1)

Exp_var_stack$SPC_NDMI_wet <- (B5_wet$band1-B6_wet$band1) /
(B5_wet$band1+B6_wet$band1)

Exp_var_stack$SPC_BI_dry <-
sqrt((B4_dry$band1^2+B3_dry$band1^2+B2_dry$band1^2) / 3)

Exp_var_stack$SPC_BI_wet <-
sqrt((B4_wet$band1^2+B3_wet$band1^2+B2_wet$band1^2) / 3)

Exp_var_stack$SPC_CI_dry <- (B4_dry$band1-B3_dry$band1) /
(B4_dry$band1+B3_dry$band1)

Exp_var_stack$SPC_CI_wet <- (B4_wet$band1-B3_wet$band1) /
(B4_wet$band1+B3_wet$band1)

Exp_var_stack$SPC_HI_dry <- (2*B4_dry$band1-B3_dry$band1-B2_dry$band1) /
(B3_dry$band1-B2_dry$band1)

Exp_var_stack$SPC_HI_wet <- (2*B4_wet$band1-B3_wet$band1-B2_wet$band1) /
(B3_wet$band1-B2_wet$band1)

Exp_var_stack$SPC_RI_dry <- B4_dry$band1^2 / (B2_dry$band1*B3_dry$band1^3)

Exp_var_stack$SPC_RI_wet <- B4_wet$band1^2 / (B2_wet$band1*B3_wet$band1^3)

Exp_var_stack$SPC_SI_dry <- (B4_dry$band1-B2_dry$band1) /
(B4_dry$band1+B2_dry$band1)

Exp_var_stack$SPC_SI_wet <- (B4_wet$band1-B2_wet$band1) /
(B4_wet$band1+B2_wet$band1)

Exp_var_stack$SPC_diff_B2 <- B2_dry$band1 - B2_wet$band1

Exp_var_stack$SPC_diff_B3 <- B3_dry$band1 - B3_wet$band1

Exp_var_stack$SPC_diff_B4 <- B4_dry$band1 - B4_wet$band1

Exp_var_stack$SPC_diff_B5 <- B5_dry$band1 - B5_wet$band1

Exp_var_stack$SPC_diff_B6 <- B6_dry$band1 - B6_wet$band1

Exp_var_stack$SPC_diff_B7 <- B7_dry$band1 - B7_wet$band1

Exp_var_stack$SPC_diff_Surface_roughness <- Surface_roughness_dry$band1 -
Surface_roughness_wet$band1

Exp_var_stack$SPC_diff_Emissivity <- Emissivity_dry$band1 -
Emissivity_wet$band1
```

```
Exp_var_stack$SPC_diff_Albedo <- Albedo_dry$band1 - Albedo_wet$band1

Exp_var_stack$SPC_diff_Nitrogen <- Nitrogen_dry$band1 - Nitrogen_wet$band1

Exp_var_stack$SPC_diff_EVI <- Exp_var_stack$SPC_EVI_dry -
Exp_var_stack$SPC_EVI_wet

Exp_var_stack$SPC_diff_MSAVI <- Exp_var_stack$SPC_MSAVI_dry -
Exp_var_stack$SPC_MSAVI_wet

Exp_var_stack$SPC_diff_NDMI <- Exp_var_stack$SPC_NDMI_dry -
Exp_var_stack$SPC_NDMI_wet

Exp_var_stack$SPC_diff_BI <- Exp_var_stack$SPC_BI_dry -
Exp_var_stack$SPC_BI_wet

Exp_var_stack$SPC_diff_CI <- Exp_var_stack$SPC_CI_dry -
Exp_var_stack$SPC_CI_wet

Exp_var_stack$SPC_diff_HI <- Exp_var_stack$SPC_HI_dry -
Exp_var_stack$SPC_HI_wet

Exp_var_stack$SPC_diff_RI <- Exp_var_stack$SPC_RI_dry -
Exp_var_stack$SPC_RI_wet

Exp_var_stack$SPC_diff_SI <- Exp_var_stack$SPC_SI_dry -
Exp_var_stack$SPC_SI_wet

#Soilmaps
#Import mask layer
mask <-
readGDAL(paste0("C:/Users/TUDelftSID/Documents/Thesis/Data/BRP/Mask/Mask_NO
P_",Resolution_input,"m_",Projection,".tif"))
mask_raster <- raster(mask)

#Create soilmap raster
Soilclass <-
readOGR(paste0("Bodemkaart/Bodemkaart_1_50000_is_25m_resolutie/SoilMap_NOP_
",Projection,".shp"))
Soilclass_r <- rasterize(Soilclass, mask_raster, field="Soil_ID")
Exp_var_stack$SMU_Soilclass <- Soilclass_r@data@values

Geomorphology <-
readOGR(paste0("Bodemkaart/Geomorfologie_1_50000/Geomorfologisch_NOP_",Proj
ection,".shp"))
Geomorphology_r <- rasterize(Geomorphology, mask_raster, field="GENESE_ID")
Exp_var_stack$SMU_Geomorphology <- Geomorphology_r@data@values

BRP_2018 <-
readOGR(paste0("BRP/brpgewaspercelen_2018/",Projection,"/BRP_Gewaspercelen_
2018_NOP_",Projection,".shp"))
BRP_2018_r <- rasterize(BRP_2018, mask_raster, field="CROP_ID")
Exp_var_stack$SMU_BRP_2018 <- BRP_2018_r@data@values

BRP_2017 <-
readOGR(paste0("BRP/brpgewaspercelen_2017/",Projection,"/BRP_Gewaspercelen_
2017_NOP_",Projection,".shp"))
BRP_2017_r <- rasterize(BRP_2017, mask_raster, field="CROP_ID")
Exp_var_stack$SMU_BRP_2017 <- BRP_2017_r@data@values
```

```
BRP_2016 <-
readOGR(paste0("BRP/brpgewaspercelen_2016/",Projection,"/BRP_Gewaspercelen_
2016_NOP_",Projection,".shp"))
BRP_2016_r <- rasterize(BRP_2016, mask_raster, field="CROP_ID")
Exp_var_stack$SMU_BRP_2016 <- BRP_2016_r@data@values

#Save multiple objects
save(Exp_var_stack, NDVI_dry, NDVI_wet, Resolution_input, Projection, mask,
     file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Module_1/1A_Data_collection
.RData")
```

Module 1B: data preparation

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(rlist)

## Set working directory
setwd("~/Thesis/Data")

load("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Module_1/1A_Data_colle
ction.RData")

## Count different type of predictors
SPC_SEBAL_number <- grep("SEBAL", names(Exp_var_stack), fixed = TRUE)
SMU_number <- grep("SMU", names(Exp_var_stack), fixed = TRUE)

## Remove vegetation, NA and 0 valued pixels use only SEBAL variables and
SMU
NDVI_thrh_high = 0.2 #Bare soil threshold
NDVI_thrh_low = 0.0 #Waterbodies threshold
Exp_var_stack.df_BS <-
as.data.frame(Exp_var_stack@data[,c(SPC_SEBAL_number, SMU_number)],
drop=FALSE)
Exp_var_stack.df_BS <- lapply(Exp_var_stack.df_BS, function(x){ifelse((x ==
0 |

NDVI_dry$band1 > NDVI_thrh_high |

NDVI_wet$band1 > NDVI_thrh_high |

NDVI_dry$band1 <= NDVI_thrh_low |

NDVI_wet$band1 <= NDVI_thrh_low |

is.na(mask$band1)), yes = NA, no = x)})
Exp_var_stack.df_BS <- as.data.frame(Exp_var_stack.df_BS)
Exp_var_stack@data[,c(SPC_SEBAL_number, SMU_number)] <- Exp_var_stack.df_BS

## Remove outliers use only SEBAL variables
Exp_var_stack.df_outlier <-
as.data.frame(Exp_var_stack@data[,c(SPC_SEBAL_number)])
Outlier_indices <- unique(do.call(c,lapply(Exp_var_stack.df_outlier,
function(x){which(x %in% boxplot.stats(x)$out)})))
Exp_var_stack@data[Outlier_indices,] <- NA

## Remove data without variation
```

```
Exp_var_stack.df_variation <- as.data.frame(Exp_var_stack@data)
sd_list <- do.call(c,lapply(Exp_var_stack.df_variation,
function(x){ifelse(sd(x, na.rm = TRUE)==0, yes = 1, no = NA)}))
No_variation_col <- names(which(sd_list == 1))
No_variation_col <- list.append(No_variation_col, ifelse(grep("SEBAL",
No_variation_col)>0, yes=sub("SEBAL","diff",names(which(sd_list == 1)))))
No_variation_col <- unique(do.call(c,lapply(No_variation_col,
                                    function(x){grep(substr(x, 1,
nchar(x)-4), names(Exp_var_stack))})))
Exp_var_stack@data <- Exp_var_stack.df_variation[,!names(Exp_var_stack)
%in% names(Exp_var_stack[No_variation_col]), drop = FALSE]

## Count different type of predictors
SMU_number <- grep("SMU", names(Exp_var_stack))

## Standardize data drop Soil Map Units (SMU)
Exp_var_stack.df_standardized <- as.data.frame(Exp_var_stack@data[,-
c(SMU_number)])
Exp_var_stack.df_standardized <- lapply(Exp_var_stack.df_standardized,
function(x){(x - mean(x, na.rm=TRUE)) / sd(x, na.rm=TRUE)})
Exp_var_stack.df_standardized <-
as.data.frame(Exp_var_stack.df_standardized)
Exp_var_stack@data[,-c(SMU_number)] <- Exp_var_stack.df_standardized

#Save multiple objects
save(Exp_var_stack, Resolution_input, Projection, mask,
     file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Module_1/1B_Data_prep.RData
")
```

Appendix B. SoilGrids30m R-script module 2

Module 2A: data correlation

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(sp)
library(rgdal)
library(corrplot)

## Set working directory
setwd("~/Thesis/Data")

## Load data to environment
load(file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Module_1/1B_Data_prep.RData
")

## Create observation location dataset
# Read-in pointsamples
SoilSamples <-
readOGR(paste0("Bodemkaart/SoilMapping_kriging/SoilSamples/SoilSamples/Soil
Samples_",Projection,".shp"))
# Remove spatial duplicates
SoilSamples <- remove.duplicates(SoilSamples)
SoilProperty <- SoilSamples["Lutum"]
names(SoilProperty)[1] <- "Clay"

# Add explenatory data to predictor object and inspect correlation
SoilProperty_df <- over(SoilProperty, Exp_var_stack)
SoilProperty@data[,2:(length(SoilProperty_df)+1)] <- SoilProperty_df

# Remove all rows which contain NA values
row.has.na <- apply(SoilProperty@data, 1, function(x){any(is.na(x))})
SoilProperty <- SoilProperty[!row.has.na,]

## Check correlation between target variable and explanatory variable
Corr_thrh = 0.15
Percentile = "15p"

# Create correlation matrix
correlation <- cor(SoilProperty@data)
# Adjust correlation row soilproperty to obtain only positvie values from
low to high
cor_adj <-
as.data.frame(abs(abs(correlation[1,2:length(correlation[1,])])))
# Extract rownmaes based on correlation threshold
perc <- quantile(cor_adj[,1], Corr_thrh)
cor_rownames <- rownames(cor_adj)[cor_adj >= perc]

# Create correlation matrix
correlation_name <- cor(SoilProperty_name@data)

png(file =
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_2/Corrpl
ot.png"),
```

```
    width = 750, height = 150)
corrplot(correlation_name[length(correlation_name[,1]),1:(length(correlatio
n_name[,1])-1), drop=FALSE], method = "circle", cl.pos='n', tl.col="black")
dev.off()

##Save data to environment
save(cor_rownames, Percentile, file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_2/2A_Data_cor_C
lay.RData")
```

Module 2B: data preparation

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(rlist)

## Set working directory
setwd("~/Thesis/Data")

## Load object
load("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Module_1/1A_Data_colle
ction.RData")
load("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_2/2A_Data_
cor_Clay.RData")

## Eliminate all explanatory variables based on correlation threshold
Exp_var_stack <- Exp_var_stack[which(names(Exp_var_stack) %in%
cor_rownames)]

## Count different type of predictors
SPC_SEBAL_number <- grep("SEBAL", names(Exp_var_stack), fixed = TRUE)
SMU_number <- grep("SMU", names(Exp_var_stack), fixed = TRUE)

## Remove vegetation, NA and 0 valued pixels use only SEBAL variables and
SMU
NDVI_thrh_high = 0.2 #Bare soil threshold
NDVI_thrh_low = 0.0 #Waterbodies threshold
Exp_var_stack.df_BS <-
as.data.frame(Exp_var_stack@data[,c(SPC_SEBAL_number, SMU_number)],
drop=FALSE)
Exp_var_stack.df_BS <- lapply(Exp_var_stack.df_BS, function(x){ifelse((x ==
0 |

NDVI_dry$band1 > NDVI_thrh_high |

NDVI_wet$band1 > NDVI_thrh_high |

NDVI_dry$band1 <= NDVI_thrh_low |

NDVI_wet$band1 <= NDVI_thrh_low |

is.na(mask$band1)), yes = NA, no = x)})
Exp_var_stack.df_BS <- as.data.frame(Exp_var_stack.df_BS)
Exp_var_stack@data[,c(SPC_SEBAL_number, SMU_number)] <- Exp_var_stack.df_BS
```

```
## Remove outliers use only SEBAL variables
Exp_var_stack.df_outlier <-
as.data.frame(Exp_var_stack@data[,c(SPC_SEBAL_number)])
Outlier_indices <- unique(do.call(c,lapply(Exp_var_stack.df_outlier,
function(x){which(x %in% boxplot.stats(x)$out)})))
Exp_var_stack@data[Outlier_indices,] <- NA

## Standardize data drop Soil Map Units (SMU)
Exp_var_stack.df_standardized <- as.data.frame(Exp_var_stack@data[,-
c(SMU_number)])
Exp_var_stack.df_standardized <- lapply(Exp_var_stack.df_standardized,
function(x){(x - mean(x, na.rm=TRUE)) / sd(x, na.rm=TRUE)})
Exp_var_stack.df_standardized <-
as.data.frame(Exp_var_stack.df_standardized)
Exp_var_stack@data[,-c(SMU_number)] <- Exp_var_stack.df_standardized

## Create observation location dataset
# Read-in pointsamples
SoilSamples <-
readOGR(paste0("Bodemkaart/SoilMapping_kriging/SoilSamples/SoilSamples/Soil
Samples_",Projection,".shp"))
# Remove spatial duplicates
SoilSamples <- remove.duplicates(SoilSamples)
SoilProperty <- SoilSamples["Lutum"]
names(SoilProperty)[1] <- "Clay"

# Add explenatory data to predictor object and inspect correlation
SoilProperty_df <- over(SoilProperty, Exp_var_stack)
SoilProperty@data[,2:(length(SoilProperty_df)+1)] <- SoilProperty_df

# Remove all rows which contain NA values
row.has.na <- apply(SoilProperty@data, 1, function(x){any(is.na(x))})
SoilProperty <- SoilProperty[!row.has.na,]

#Save multiple objects
save(Exp_var_stack, SoilProperty, Resolution_input, mask, Percentile,
     file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_2/2B_Data_prep_
cor_Clay.RData")
```

Appendix C. SoilGrids30m R-script module 3

Module 3: principal component analysis

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(FactoMineR)
library(factoextra)

## Set working directory
setwd("~/Thesis/Data")

# Load multiple objects
load(file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_2/2B_Data_prep_
cor_Clay.RData")

## Count different type of predictors
SPC_number <- grep("SPC", names(Exp_var_stack), fixed = TRUE)
SMU_number <- grep("SMU", names(Exp_var_stack), fixed = TRUE)

# Selecet Soil Predictive Components (SPC) and Soil Map Units (SMU)
SPC <- SoilProperty[,c(SPC_number+1)]
SMU <- SoilProperty[,c(SMU_number+1)]

# PCA
PCA <- prcomp(SPC@data, scale. = FALSE)
PCA_var <- PCA$sdev^2
Contribution_expvar <- PCA_var/sum(PCA_var)
cum_contribution_expvar <- cumsum(Contribution_expvar)
var <- get_pca_var(PCA)

# Minimum variance explained threshold
Var_exp_thrh = 0.99
sdev_trh <- length(which(PCA$sdev > 1))
var_trh <- length(which(cum_contribution_expvar <= Var_exp_thrh)) + 1
PC_number <- max(sdev_trh, var_trh)
Exp_var_contrib.df <- as.data.frame(var$contrib[,1:PC_number])

# Scree plot
png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/Sc
reeplot_Clay_",Percentile,".png"),
    width = 750, height = 500)
fviz_eig(PCA, addlabels = TRUE, ncp=PC_number, main = paste0("Scree plot
percentile ",

substr(Percentile, 1, nchar(Percentile)-1)),
         xlab="Principal components") +
  theme(text = element_text(size = 15),
        axis.title = element_text(size = 15),
        axis.text = element_text(size = 15),
        plot.title = element_text(hjust = 0.5))
dev.off()

# Total contribution per explanatory variable
```

```
png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/To
tal_contrib_expvar_Clay_",Percentile,".png"),
    width = 750, height = 500)
fviz_contrib(PCA, choice="var", axes = 1:PC_number, top = 19) +
  ggtitle(paste0("Total contributions of significant explanatory variables
for all selected principal component")) +
  theme(text = element_text(size = 15),
        axis.title = element_text(size = 15),
        axis.text = element_text(size = 15),
        plot.title = element_text(hjust = 0.5))
dev.off()

# Contribution plot per principal component
plot_list = list()
for (i in 1:PC_number) {
  top_number <- length(which(var$contrib[,i] >
(100/length(Contribution_expvar))))
  p = fviz_contrib(PCA, choice = "var", axes = i, top = (top_number+1)) +
    ggtitle(paste0("Contributions of significant explanatory variables to
principal component ", i)) +
    theme(text = element_text(size = 15),
          axis.title = element_text(size = 15),
          axis.text = element_text(size = 15),
          plot.title = element_text(hjust = 0.5))
  plot_list[[i]] = p
}

# Save contribution plot per principal component
for (i in 1:PC_number) {

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/Co
ntrib_dim_",i,"_Clay_",Percentile,".png"),
      width = 750, height = 500)
  print(plot_list[[i]])
  dev.off()
}

# Create dataframe with all variables for regression kriging
PCA_dataset <- SoilProperty["Clay"]
PCA_dataset@data <- data.frame(Clay = SoilProperty$Clay,
PCA$x[,c(1:PC_number)], SMU@data)
colnames(PCA_dataset@data)[2:(PC_number+1)] <- colnames(PCA$x)[1:PC_number]

PCA_eigenvectors_SPC <- data.frame(PCA$rotation[,1:PC_number])
colnames(PCA_eigenvectors_SPC)[1:PC_number] <-
colnames(PCA$rotation)[1:PC_number]

# Create mask stack
mask_SPC <- mask["band1"]
mask_SPC@data[2:(length(Exp_var_stack@data)+1)] <- Exp_var_stack@data
mask_SPC.df <- as.data.frame(mask_SPC@data)

mask_RK <- mask["band1"]
for(i in 2:length(PCA_dataset@data)){
  ifelse(grepl("PC",names(PCA_dataset@data[i]), fixed=TRUE),
         yes = mask_RK@data <- cbind(mask_RK@data,

data.frame(rowSums(data.frame(mapply(`*`,PCA_eigenvectors_SPC[,(i-1)],

mask_SPC.df[,2:(length(mask_SPC.df)-
```

```
length(SMU_number))])))))),
         no = mask_RK@data <- cbind(mask_RK@data,
data.frame(eval(parse(text =
paste0("Exp_var_stack$",names(PCA_dataset[i])))))))
}
colnames(mask_RK@data) <- c(names(PCA_dataset@data))
names(mask_RK)[1] <- "band1"

# Save multiple objects
save(PCA_dataset, mask_RK, Resolution_input, Percentile,
     file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/3_PCA_Clay.RD
ata")
```

Appendix D. SoilGrids30m R-script module 4-5

Module 4-5: Regression-kriging

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(gstat)
library(raster)
library(rgdal)
library(caret)

## Set working directory
setwd("~/Thesis/Data")

## Load object
load(file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/3_PCA_Clay.RD
ata")

source("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/RK_function_Clay.R")

## Regression kriging
# Properties semi variogram
VG_nugget = 5
VG_psill = 13
VG_range = 2000
VG_model = "Exp"

# Run for final result
mask_RK.df <- as.data.frame(mask_RK@data)

# Regression kriging function
RK_output <- RK(PCA_dataset, mask_RK, mask_RK.df, VG_nugget, VG_psill,
VG_range, VG_model)

TrainData.rk <- RK_output[[1]]
Variogram_exp <- RK_output[[2]]
Variogram_fit <- RK_output[[3]]

# Plot semivariogram
png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/Semivariance_Clay_",Percentile,".png"),
    width = 750, height = 500)
plot.new()
plot(Variogram_exp, Variogram_fit, main=list(paste0("Semivariogram
percentile ",
                                                    substr(Percentile, 1,
nchar(Percentile)-1)) , cex=1.5),
     xlab = list("Distance [m]", cex=1.5),
     ylab=list("Semivariance [%^2]", cex=1.5),
     pch=16, cex=1.5, col="black")
legend("center", bty="n", legend=paste("Nugget =",
format(Variogram_fit[1,2], digits=2), "[%^2]",
                                       "\nSill =",
format(Variogram_fit[2,2], digits=2), "[%^2]",
```

```
                                        "\nRange =",
format(Variogram_fit[2,3], digits=2)), "[m]", cex=1.5)
dev.off()

# Write result to tif file
TrainData_RK_raster_predict = raster(TrainData.rk, layer=4)
writeRaster(TrainData_RK_raster_predict,
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/predict_RK_"

,Resolution_input,"m_Clay_",Percentile,".tif"),
            overwrite = FALSE)

writeOGR(obj=PCA_dataset,
dsn=paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5"),
         layer=paste0("PCA_dataset_Clay_",Percentile), driver="ESRI
Shapefile")

# Validation result
PCA_dataset$predict <- over(PCA_dataset, TrainData.rk)$predict

Performance <- postResample(PCA_dataset$predict, PCA_dataset$Clay)

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/Predict_result_Clay_",Percentile,".png"),
    width = 750, height = 500)
plot(PCA_dataset$Clay, PCA_dataset$predict, main=paste0("Regression kriging
results percentile ",
     substr(Percentile, 1, nchar(Percentile)-1)),
     xlab="Measured clay content [%]", ylab="Predicted clay content [%]",
     cex.lab=1.5, cex.axis=1.5, cex.main=1.5, pch=3)
abline(fit <- lm(predict ~ Clay, data=PCA_dataset@data), col='red')
abline(0,1, col='black')
legend("topleft", bty="n", legend=paste("Number of observations =",
format(length(PCA_dataset), digits=2),
                                        "\nR2 =",
format(summary(fit)$adj.r.squared, digits=2),
                                        "\nRMSE =", format(Performance[1],
digits=2),
                                        "\nMPE =", format(Performance[3],
digits=2)), cex=1.5)
dev.off()

# Analysis with SoilGrids250m
Clay_SoilGrids250m <-
readGDAL("Bodemkaart/SoilGrids250m_Clay_0m_UTM32N.tif")
PCA_dataset$Clay250m <- over(PCA_dataset, Clay_SoilGrids250m)$band1
Performance250m <- postResample(PCA_dataset$Clay250m, PCA_dataset$Clay)

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/SoilGrids250m_result_Clay_",Percentile,".png"),
    width = 750, height = 500)
plot.new()
par(mar=c(5,5,5,1))
plot(PCA_dataset$Clay, PCA_dataset$Clay250m, main=paste0("Regression
kriging results percentile ",
                                                         substr(Percentile,
1, nchar(Percentile)-1), " SoilGrids250m"),
     xlab="Measured clay content [%]", ylab="Predicted clay content [%]",
     cex.lab=2, cex.axis=2, cex.main=2, pch=3)
```

```
abline(fit <- lm(predict ~ Clay250m, data=PCA_dataset@data), col='red')
abline(0,1, col='black')
legend(x=min(PCA_dataset$Clay) - 2.5, y=max(PCA_dataset$Clay250m) + 2.75,
bty="n",
       legend=paste("\nR2 =", format(Performance250m[2], digits=2),
                                        "\nRMSE =",
format(Performance250m[1], digits=2),
                                        "\nMPE =",
format(Performance250m[3], digits=2)), cex=2)
dev.off()

# Analysis with SoilGrids1000m
Clay_SoilGrids1000m <-
readGDAL("Bodemkaart/SoilGrids1000m_Clay_0m_UTM32N.tif")
PCA_dataset$Clay1000m <- over(PCA_dataset, Clay_SoilGrids1000m)$band1
Performance1000m <- postResample(PCA_dataset$Clay1000m, PCA_dataset$Clay)

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/SoilGrids1000m_result_Clay_",Percentile,".png"),
    width = 750, height = 500)
plot.new()
par(mar=c(5,5,5,1))
plot(PCA_dataset$Clay, PCA_dataset$Clay1000m, main=paste0("Regression
kriging results percentile ",
                                                    substr(Percentile,
1, nchar(Percentile)-1), " SoilGrids1000m"),
     xlab="Measured clay content [%]", ylab="Predicted clay content [%]",
     cex.lab=2, cex.axis=2, cex.main=2, pch=3)
abline(fit <- lm(predict ~ Clay1000m, data=PCA_dataset@data), col='red')
abline(0,1, col='black')
legend(x=min(PCA_dataset$Clay) - 2.5, y=max(PCA_dataset$Clay1000m) + 2.25,
bty="n",
       legend=paste("\nR2 =", format(Performance1000m[2], digits=2),
                                        "\nRMSE =",
format(Performance1000m[1], digits=2),
                                        "\nMPE =",
format(Performance1000m[3], digits=2)), cex=2)
dev.off()
```

Regression-kriging function

```
# Function for regression kriging
RK <- function(RK.data, mask_RK, mask_RK.df, VG_nugget, VG_psill, VG_range,
VG_model) {
  # Fit a linear regression model and inspect the results
  TrainData.lm <- lm(Clay~., data = RK.data@data)

  # Append residuals to dataset
  RK.data$residuals <- TrainData.lm$residuals

  # Regression prediction
  TrainData.trend <- predict(TrainData.lm, newdata = mask_RK.df)

  # Set prediction outside mask to NA
  TrainData.trend <- ifelse(test = is.na(mask_RK$band1), yes = NA, no =
TrainData.trend)

  # Set predictions smaller than 0 to 0
  TrainData.trend <- ifelse(TrainData.trend<0, yes = 0, no =
TrainData.trend)
```

```
  # Define gstat object and compute experimental semivariogram
  gpb <- gstat(formula = residuals~1, data = RK.data)
  vgpb <- variogram(gpb)

  # Define initial semivariogram model
  vgmpb <- vgm(nugget = VG_nugget, psill = VG_psill, range = VG_range,
model = VG_model)
  show(plot(vgpb,vgmpb))

  # Fit semivariogram model
  vgmpb <- fit.variogram(vgpb, vgmpb, fit.method = 7)
  show(plot(vgpb,vgmpb))

  # Krige the residuals
  TrainData.rk <- krige(formula = residuals~1, locations = RK.data, newdata
= mask_RK, model = vgmpb, beta = 0)

  names(TrainData.rk)[1] <- "resid"
  TrainData.rk$trend <- TrainData.trend

  # Set kriged residuals and variance outside mask to NA
  TrainData.rk$resid <- ifelse(test = is.na(mask_RK$band1), yes = NA, no =
TrainData.rk$resid)

  TrainData.rk$var1.var <- ifelse(test = is.na(mask_RK$band1), yes = NA, no
= TrainData.rk$var1.var)

  # Obtain RK prediction
  TrainData.rk$predict <- TrainData.rk$trend + TrainData.rk$resid

  # Set predictions smaller than 0 to 0
  TrainData.rk$predict <- ifelse(TrainData.rk$predict<0, yes = 0, no =
TrainData.rk$predict)

  RK_output <- list(TrainData.rk, vgpb, vgmpb)

  return(RK_output)
}
```

Appendix E. SoilGrids30m R-script module 6

Module 6: validation

```
## clean environment
rm(list = ls()) #clean memory

## Set libraries
library(gstat)
library(rlist)
library(caret)
library(raster)

## Set working directory
setwd("~/Thesis/Data")

## Load object
load(file =
"Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_3/3_PCA_Clay.RD
ata")

source("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_4-
5/RK_function_Clay.R")

## Split data for validation run
Data_split = 0.7
smp_size <- floor(Data_split * nrow(PCA_dataset))

mask_RK.df <- as.data.frame(mask_RK@data)

TrainData_RK <- list()
Rsq_list <- list()
MAE_list <- list()
RMSE_list <- list()
Variogram_nugget <- list()
Variogram_sill <- list()
Variogram_range <- list()
Calc_num <- 100
for(i in 1:Calc_num){
  # Create validationset and trainingsset
  validateIndexes <- sample(seq_len(nrow(PCA_dataset)), size = smp_size)
  TrainData <- PCA_dataset[validateIndexes, ]
  ValidateData <- PCA_dataset[-validateIndexes, ]

  RK.data <- PCA_dataset[validateIndexes,]

  #Properties semi variogram
  VG_nugget = 5
  VG_psill = 13
  VG_range = 2000
  VG_model = "Exp"

  # Run regression kriging function
  RK_output <- RK(RK.data, mask_RK, mask_RK.df, VG_nugget, VG_psill,
VG_range, VG_model)

  TrainData.rk <- RK_output[[1]]
  Variogram_fit <- RK_output[[3]]
```

```
  Variogram_nugget <- list.append(Variogram_nugget, Variogram_fit[1,2])
  Variogram_sill <- list.append(Variogram_sill, Variogram_fit[2,2])
  Variogram_range <- list.append(Variogram_range, Variogram_fit[2,3])

  # Add Trainingsdata to list
  TrainData_RK[[paste0("Set_",i)]] <- TrainData.rk

  # Validate result
  ValidateData$predict <- over(ValidateData, TrainData.rk)$predict

  Performance <- postResample(ValidateData$predict, ValidateData$Clay)

  Rsq_list <- list.append(Rsq_list, Performance[2])
  MAE_list <- list.append(MAE_list, Performance[3])
  RMSE_list <- list.append(RMSE_list, Performance[1])
}

# Plot validation results
png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_6/Perf_MAE_Clay_",Percentile,".png"),
    width = 750, height = 500)
hist(unlist(MAE_list), breaks=100, main="Histogram of mean absolute estimation error",
     xlab="MAE [%]", ylab="Frequency [-]", cex.lab=1.5, cex.axis=1.5, cex.main=1.5)
legend("topright", bty="n", legend=paste("Number of runs = ",Calc_num,
                                         "\nMAE mean =",
format(mean(unlist(MAE_list)), digits=3),"%"), cex=1.5)
dev.off()

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_6/Perf_RMSE_Clay_",Percentile,".png"),
    width = 750, height = 500)
hist(unlist(RMSE_list), breaks=100, main="Histogram of root mean square error",
     xlab="RMSE [%]", ylab="Frequency [-]", cex.lab=1.5, cex.axis=1.5, cex.main=1.5)
legend("topright", bty="n", legend=paste("Number of runs = ",Calc_num,
                                         "\nRMSE mean =",
format(mean(unlist(RMSE_list)), digits=3),"%"), cex=1.5)
dev.off()

png(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_6/Perf_R2_Clay_",Percentile,".png"),
    width = 750, height = 500)
hist(unlist(Rsq_list), breaks=100, main="Histogram of coefficient of determination",
     xlab="R2 [-]", ylab="Frequency [-]", cex.lab=1.5, cex.axis=1.5, cex.main=1.5)
legend("topright", bty="n", legend=paste("Number of runs = ",Calc_num,
                                         "\nR2 mean =",
format(mean(unlist(Rsq_list)), digits=3)), cex=1.5)
dev.off()

#Save multiple objects
save(TrainData_RK, Variogram_nugget, Variogram_range, Variogram_sill,
Rsq_list, MAE_list, RMSE_list,
     file =
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module_6/6_validation_Clay_",Percentile,".RData"))
```

Appendix F. Pedotransfer function R-script

Pedotransfer functions

```
rm(list = ls()) #clean memory

# Load libraries
library(rgdal)
library(raster)

# Set working directory
setwd("~/Thesis/Data")

## Parameters
Resolution <- "30"
Projection <- "EPSG32632"

## Read in data
SoilProperty <-
readGDAL(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/Clay/Module
_4-5/predict_RK_",
                                    Resolution,"m_Clay_15p.tif"))
names(SoilProperty)[1] <- "Clay"
OM <-
readGDAL(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/OM/Module_4
-5/predict_RK_",
                Resolution,"m_OM_70p.tif"))
SoilProperty$OM <- OM$band1

#Remove values without information of Clay and or OM
SoilProperty$OM <- ifelse(test=is.na(SoilProperty$Clay), yes=NA,
no=SoilProperty$OM)
SoilProperty$Clay <- ifelse(test=is.na(SoilProperty$OM), yes=NA,
no=SoilProperty$Clay)

#PTFs
SoilProperty$BD <- 1 / (0.6117 + 0.003601*SoilProperty$Clay +
0.002172*SoilProperty$OM^2 +
                          0.01715*log(SoilProperty$OM)) #g/cm3, input is in
percentages

SoilProperty$Theta_s <- 0.85*(1-SoilProperty$BD/2.65) +
0.13*(SoilProperty$Clay/100) #cm3/cm3, input: BD=g/cm3, clay=g/g

SoilProperty$Theta_r <- 0.51*(SoilProperty$Clay/100) +
0.0017*((SoilProperty$OM/100*1000)/2) #cm3/cm3; factor 2 see Pribyl, 2010,
input: clay=g/g, OM=g/kg

SoilProperty$alpha = exp(-19.13 + 0.812*SoilProperty$OM +
23.4*SoilProperty$BD - 8.16*SoilProperty$BD^2 +
  0.423*SoilProperty$OM^(-1) + 2.388*log(SoilProperty$OM) -
1.338*SoilProperty$BD*SoilProperty$OM) #1/cm, input in percentages

SoilProperty$n <- exp(-0.235 + 0.972*SoilProperty$BD^(-1) -
0.7743*log(SoilProperty$Clay) - 0.3154*log(SoilProperty$OM) +
  0.0678*SoilProperty$BD*SoilProperty$OM) + 1 #input in percentages
```

```
SoilProperty$Theta_fc <- SoilProperty$Theta_r + (SoilProperty$Theta_s -
SoilProperty$Theta_r) /
  ((1+abs(SoilProperty$alpha*-100)^SoilProperty$n)^(1-1/SoilProperty$n))
#cm3/cm3

SoilProperty$Theta_wp <- SoilProperty$Theta_r + (SoilProperty$Theta_s -
SoilProperty$Theta_r) /
  ((1+abs(SoilProperty$alpha*-16000)^SoilProperty$n)^(1-1/SoilProperty$n))
#cm3/cm3

SoilProperty$WHC <- 1000*(SoilProperty$Theta_fc - SoilProperty$Theta_wp)
#mm/m

## Write output to a .tif file
WHC_tif <- raster(SoilProperty, layer=10)
writeRaster(WHC_tif,
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/WHC/WHC_",Resolution
,"m_",Projection,".tif"),
            overwrite = TRUE)

## Write output to a .tif file
BD_tif <- raster(SoilProperty, layer=3)
writeRaster(BD_tif,
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/WHC/BD_",Resolution,
"m_",Projection,".tif"),
            overwrite = TRUE)
```

Visualization

```
rm(list = ls()) #clean memory

# Load libraries
library(rgdal)
library(rlist)
library(ggplot2)
library(viridis)
library(lubridate)
library(outliers)

# Set working directory
setwd("~/Thesis/Data")

Projection = "EPSG32632"
Resolution = 30
Croptype = "Sugarbeet"
Croptype_adj = "sugar beet"

# Read in data crop type
SoilProperty <-
readGDAL(paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/WHC/WHC_buf
fer/WHC_merged/WHC_",
                                  Croptype,"_",Resolution,"m_merged.tif"))

names(SoilProperty)[1] <- "WHC"

Outliers <- which(SoilProperty$WHC %in%
boxplot.stats(SoilProperty$WHC)$out)
SoilProperty$WHC[Outliers] <- NA
```

```
SOM_date <- c("2018_03_20", "2018_04_21", "2018_05_07", "2018_07_03",
"2018_07_26")
SOM_date_adj <- c("20-03-2018", "21-04-2018", "07-05-2018", "03-07-2018",
"26-07-2018")

for(i in 1:length(SOM_date)){
  date <- dmy(SOM_date_adj[i])
  Day <- yday(date)
  Year <- year(date)
  SEBAL_moisture <-
readGDAL(paste0("SEBAL/SEBAL_output/NOP/",Projection,"/Soil_Moisture_SEBAL/
",

"LS8_LS8_Total_soil_moisture_",Resolution,"m_",Year,"_",Day,".tif"))
  SEBAL_moisture$band1 <- ifelse(test = is.na(SoilProperty$WHC), yes = NA,
no = SEBAL_moisture$band1)
  SEBAL_moisture$band1 <- SEBAL_moisture$band1 / 0.45
  SoilProperty$WHC <- ifelse(test = is.na(SEBAL_moisture$band1), yes = NA,
no = SoilProperty$WHC)

  SoilProperty$SEBAL_moisture <- SEBAL_moisture$band1

  # Increment data
  df <- data.frame(SoilProperty$SEBAL_moisture, SoilProperty$WHC)

  # Remove rows with NA
  row.has.na <- apply(df, 1, function(x){any(is.na(x))})
  df <- df[!row.has.na,]

  increment = 0.05
  WHC_subset <- list()
  WHC_mean <- list()
  WHC_median <- list()
  WHC_std <- list()
  WHC_increment_start <- list()
  WHC_increment_end <- list()
  WHC_count <- list()
  for(j in 1:(1/increment)){
    count = j*increment
    WHC_subset <- list.append(WHC_subset,
                              subset(df$SoilProperty.WHC
,df$SoilProperty.SEBAL_moisture > (count-increment) &
                                       df$SoilProperty.SEBAL_moisture <
count))
  }

  # Boxplot
  png(file =
paste0("Bodemkaart/SoilMapping_kriging/SoilGrids_model/WHC/Boxplot_WHC_SM_"
,
                    Croptype,"_",Resolution,"m_",Year,"_",Day,".png"),
      width = 750, height = 500)
  bp <- boxplot(WHC_subset, cex.lab=1.5, cex.axis=1.5, cex.main=1.5,
                xlab = "SEBAL total soil moisture [-]",
                ylab = "Soil water holding capacity [mm/m]",
                main = c(paste("Total soil moisture content vs. soil water
holding capacity"), paste(SOM_date_adj[i], Croptype_adj)),
                varwidth = TRUE,
                notch = TRUE,
                col = "lightgray",
```

```
                names = c("0-0.05","0.05-0.1","0.1-0.15","0.15-0.2","0.2-
0.25","0.25-0.3","0.3-0.35","0.35-0.4","0.4-0.45","0.45-0.5",
                          "0.5-0.55","0.55-0.6","0.6-0.65","0.65-0.7","0.7-
0.75","0.75-0.8","0.8-0.85","0.85-0.9","0.9-0.95","0.95-1.0"))
  dev.off()
}
```

Appendix G. SWI module 1

Module 1A: buffer zone

```
import geopandas as gpd
from Trapezoid.Module_1.Shapefile_buffer import buffer

Projection = ["EPSG32631", "EPSG32632"]
CropType = ["Sugarbeet", "Winterwheat"]
bufferDist = -60

for i in range(len(Projection)):
    for j in range(len(CropType)):
        inputfn =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\BRP\brpgewaspercelen_2018\%s\BR
P_%s_2018_NOP_%s.shp" % (Projection[i], CropType[j], Projection[i])
        outputBufferfn =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\BRP\brpgewaspercelen_2018\%s\Bu
ffer\%s" % (Projection[i], CropType[j])

        FileName = "BRP_%s_2018_NOP_%i" %(CropType[j], bufferDist)

        # Create bufferzone
        schema, Output_shp, geomBuffer_list = buffer(inputfn,
outputBufferfn, bufferDist, FileName)

        # Copy attribute columns from original shapefile to new shapefile
        gdf_in = gpd.read_file(inputfn)
        gdf_out = gpd.read_file(Output_shp)
        for k in range(len(schema)):
            gdf_out[schema[k]] = gdf_in[schema[k]]

        index_out = []
        for l in range(len(geomBuffer_list)):
            if geomBuffer_list[l].IsEmpty():
                index_out.append(l)
            else:
                continue

        # Write to raster file
        gdf_out.drop(gdf_out.index[index_out], inplace=True)
        gdf_out.to_file(Output_shp)
```

Module 1B: buffer zone function

```
import os
from osgeo import ogr
from shutil import copyfile

def buffer(inputfn, outputBufferfn, bufferDist, FileName):
    inputds = ogr.Open(inputfn)
    inputlyr = inputds.GetLayer()

    shpdriver = ogr.GetDriverByName('ESRI Shapefile')
    outputBufferds = shpdriver.CreateDataSource(outputBufferfn)
    bufferlyr = outputBufferds.CreateLayer(FileName,
geom_type=ogr.wkbPolygon)
    featureDefn = bufferlyr.GetLayerDefn()

    schema = []
```

```
    ldefn = inputlyr.GetLayerDefn()
    for n in range(ldefn.GetFieldCount()):
        fdefn = ldefn.GetFieldDefn(n)
        schema.append(fdefn.name)

    geomBuffer_list = []
    for feature in inputlyr:
        ingeom = feature.GetGeometryRef()
        geomBuffer = ingeom.Buffer(bufferDist)
        geomBuffer_list.append(geomBuffer)

        outFeature = ogr.Feature(featureDefn)
        outFeature.SetGeometry(geomBuffer)
        bufferlyr.CreateFeature(outFeature)

    FileName = FileName + ".shp"
    Output_shp = os.path.join(outputBufferfn, FileName)

    copyfile(inputfn.replace('.shp', '.prj'), Output_shp.replace('.shp',
'.prj'))
    copyfile(inputfn.replace('.shp', '.qpj'), Output_shp.replace('.shp',
'.qpj'))

    return schema, Output_shp, geomBuffer_list
```

Appendix H. SWI module 2

Module 2A: extract LST and NDVI data per field

```
import time
from datetime import timedelta
from datetime import datetime
from Trapezoid.Module_2.get_raster_per_polygon import
get_raster_per_polygon
import os
import geopandas as gpd
from Trapezoid.Module_2.Merging_tif_files import Merging_tiff_files
from Trapezoid.Metadatafile_reader import build_data
import glob
start_time = time.monotonic()

Image_date = ["2018_03_20", "2018_04_21", "2018_05_07", "2018_07_03",
"2018_07_26"]
T_inst = [6.23, 17.92, 22.21, 22.15, 30.28] #See Excel input file SEBAL
Crop_type = ["Sugarbeet", "Winterwheat"]
buffdist = -60
Property_name = "LST"
Property_folder = "Output_vegetation"
Raster_inputfile = "LS8_LS8_surface_temp_sharpened"

for i in range(len(Crop_type)):
    for j in range(len(Image_date)):
        # Determining the input parameters from metadatafile
        LANDSAT8_metadata_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_input\Landsat8\%s"
%(Image_date[j])
        file_name = os.path.join(LANDSAT8_metadata_dir, "*[_MTL].txt")
        file_path = glob.glob(file_name)
        f = open(file_path[0], 'r')  # open file for reading
        data = build_data(f)
        UTM_zone = int(data["UTM_ZONE"])
        Projection = "EPSG326%s" % (UTM_zone)
        Resolution = int(data["GRID_CELL_SIZE_REFLECTIVE"][0:2])
        Date = data["DATE_ACQUIRED"]
        adate = datetime.strptime(Date, "%Y-%m-%d")
        DOY = adate.timetuple().tm_yday
        Year = adate.timetuple().tm_year

        BRP_root_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\BRP\brpgewaspercelen_%i\%s\Buff
er" %(Year, Projection)
        output_root_dir_property_buffer =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\%s
" %(Projection, Image_date[j], Property_folder)
        property_file = "%s_%im_%i_%i.tif" %(Raster_inputfile, Resolution,
Year, DOY)

        # """
        # --Combining property with BRP--
        # """
        # Reading shapefile
        FileName = "%s\BRP_%s_%i_NOP_%i.shp" %(Crop_type[i], Crop_type[i],
Year, buffdist)
        BRP_path_out = os.path.join(BRP_root_dir, FileName)

        shapefile_gpd = gpd.read_file(BRP_path_out)
```

```
        # Creating new folder for property
        dir_name = "%s_buffer" %(Property_name)
        Output_property_field =
os.path.join(output_root_dir_property_buffer, dir_name)
        if not os.path.exists(Output_property_field):
            os.makedirs(Output_property_field)

        # Creating new folder for croptype
        dir_name = "%s" %(Crop_type[i])
        Output_property_field2 = os.path.join(Output_property_field,
dir_name)
        if not os.path.exists(Output_property_field2):
            os.makedirs(Output_property_field2)

        field_list = range(len(shapefile_gpd))

        property_path = os.path.join(output_root_dir_property_buffer,
property_file)

        if Property_name == "LST":
            T_inst_K = 273.15 + T_inst[j]
        else:
            T_inst_K = 0

        # Combining BRP and property
        get_raster_per_polygon(property_path, shapefile_gpd, field_list,
Output_property_field2, Image_date[j], T_inst_K)

        """
        ---Merging all property.tif files---
        """
        # File and folder paths
        dir_name = "Merged_%s" %(Property_name)
        output_root_dir_merged = os.path.join(Output_property_field,
dir_name)
        if not os.path.exists(output_root_dir_merged):
            os.makedirs(output_root_dir_merged)

        FileName = "%s_merged_%s_buffer_%i.tif" %(Crop_type[i],
Property_name, buffdist)
        Output_merged = os.path.join(output_root_dir_merged, FileName)

        # Make a search criteria to select the .tif files
        search_criteria = "*.tif"

        Merging_tiff_files(Output_property_field2, search_criteria,
Projection[4:], Output_merged)

end_time = time.monotonic()
print("Execution time processing: %s" %timedelta(seconds=end_time -
start_time))
```

Module 2B: extract LST and NDVI data per field function

```
import os
import rasterio
import rasterio.mask

def get_raster_per_polygon(raster_path, shapefile_gpd, field_list,
output_root_dir, Image_date, T_inst_K):
    """
    Get the raster data per polygon
    :param raster_path: raster path (.tif)
    :param plot_data: shapefile path or geopandas dataframe containing the
plot polys (.shp)
    :param field_list: list of field_id numbers in format: [],
    to obtain all fields use range(len(shapefile_data)) without []
    :param output_root_dir: root dir output path raster files
    :return: .tif file per requested shapefile feature
    """
    shp_gpd = shapefile_gpd.mask(shapefile_gpd.eq('None')).dropna()
    for index, row in shp_gpd.iterrows():
        with rasterio.open(raster_path) as img:
            raster_data, out_t = rasterio.mask.mask(img, [row.geometry],
pad=True, crop=True)
            if T_inst_K == 0:
                raster_data = raster_data
            else:
                for i in range(raster_data.shape[1]):
                    for j in range(raster_data.shape[2]):
                        if raster_data[0][i][j] == -9999:
                            raster_data[0][i][j] == -9999
                        else:
                            raster_data[0][i][j] = raster_data[0][i][j] -
T_inst_K
            raster_meta = img.meta.copy()

            raster_meta.update({"driver": "GTiff",
                        "height": raster_data.shape[1],
                        "width": raster_data.shape[2],
                        "transform": out_t})

        for i in range(len(field_list)):
            if field_list[i] == index:
                field_name = "%s_LST_%s_FIELDid_%s.tif" %(Image_date,
shp_gpd.gewas[index], field_list[i])
                output_path = os.path.join(output_root_dir, field_name)
                with rasterio.open(output_path, "w", **raster_meta) as
dest:
                    dest.write(raster_data)
```

Module 2C: merge all fields into one tif file

```
def Merging_tiff_files(dir_path, search_criteria, Projection, Output_tif):
    import rasterio
    from rasterio.merge import merge
    import glob
    import os
    from Thermal_images_SelfMade.EPSG_to_Proj4 import from_epsg_code

    q = os.path.join(dir_path, search_criteria)
    dem_fps = glob.glob(q)
```

```
    src_files_to_mosaic = []
    for fp in dem_fps:
        src = rasterio.open(fp)
        src_files_to_mosaic.append(src)

    Proj4 = from_epsg_code(Projection)

    mosaic, out_trans = merge(src_files_to_mosaic)

    out_meta = src.meta.copy()

    out_meta.update({"driver": "GTiff", "height": mosaic.shape[1], "width":
mosaic.shape[2], "transform": out_trans, "crs": Proj4})

    with rasterio.open(Output_tif, "w", **out_meta) as dest:
        dest.write(mosaic)
```

Appendix I. SWI module 3

Module 3A: Pixel envelope boundaries

```
from Trapezoid.Module_3.Pixel_envelope_boundaries import Class_boundaries
from Trapezoid.Metadatafile_reader import build_data
from datetime import datetime
import os
import glob
import rasterio
import numpy as np
import pickle

Image_date = ["2018_03_20", "2018_04_21", "2018_05_07", "2018_07_03",
"2018_07_26"]
T_inst = [6.23, 17.92, 22.21, 22.15, 30.28] #See Excel input file SEBAL
# Remove outliers manually detected boundary points
high_thrh_delete = [[0],[],[],[],[0,1,2,3,8],[0,1,8]]
low_thrh_delete = [[0],[0],[0],[0],[0,1,2,8],[0]]

for i in range(len(Image_date)):
    # Determining the input parameters from metadatafile
    LANDSAT8_metadata_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_input\Landsat8\%s"
% (Image_date[i])
    file_name = os.path.join(LANDSAT8_metadata_dir, "*[_MTL].txt")
    file_path = glob.glob(file_name)
    f = open(file_path[0], 'r')  # open file for reading
    data = build_data(f)
    UTM_zone = int(data["UTM_ZONE"])
    Projection = "EPSG326%s" % (UTM_zone)
    Resolution = int(data["GRID_CELL_SIZE_REFLECTIVE"][0:2])
    Date = data["DATE_ACQUIRED"]
    adate = datetime.strptime(Date, "%Y-%m-%d")
    DOY = adate.timetuple().tm_yday
    Year = adate.timetuple().tm_year

    # Read in all LST and NDVI data
    NDVI_path =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL output\NOP\%s\%s\Ou
tput_vegetation\LS8_ndvi_%im_%i_%i.tif" %(Projection, Image_date[i],
Resolution, Year, DOY)
    with rasterio.open(NDVI_path) as img:
        NDVI = img.read(1)

    LST_path =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LS8_LS8_surface_temp_sharpened_%im_%i_%i.tif" %(Projection,
Image_date[i], Resolution, Year, DOY)
    with rasterio.open(LST_path) as img:
        LST = img.read(1)

    # Correct for instantaneous air temperature
    LST = LST - (273.15 + T_inst[i])

    #Remove values which are not representative for analysis
    NDVI_min = 0.0
    NDVI_max = 0.9
    NDVI[NDVI>NDVI_max] = np.nan
    NDVI[NDVI<NDVI_min] = np.nan
```

```
    #Determine boundary points per bin based on percentile (low is given
threshold, high is 100 minus given threshold)
    Percentile_thrh = 2

    #Number of classes
    classes = 20

    #Number of bins for density plot
    nbins = 100

    #Determine the polygons
    polygons, LST, xi, yi, zi = Class_boundaries(NDVI, LST, Image_date[i],
low_thrh_delete[i], high_thrh_delete[i], Percentile_thrh, classes,
NDVI_min, NDVI_max, nbins)

    # Saving the objects:
    with
open('C:\Python_projects\Thesis\Trapezoid\Module_3\Pixel_envelope_boundarie
s_%s.pkl' %(Image_date[i]), 'wb') as f:
        pickle.dump([Projection, LST, NDVI, polygons, nbins, xi, yi, zi],
f)
```

Module 3B: Pixel envelope boundaries function

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
from shapely.geometry.polygon import Polygon
from Trapezoid.Module_3.Boxplot_outlier import Boxplot_outlier
from scipy.stats import kde

def Class_boundaries(NDVI, LST, Image_date, low_thrh_delete,
high_thrh_delete, Percentile_thrh, classes, NDVI_min, NDVI_max, nbins):
    """
    Function to determine the boundary lines for each class
    :param Image_date: Date of image (2018_07_26)
    :param Median_thrh: Number of points to determine the median value of
top and bottom boundary per bin
    :param low_thrh_delete: Delete indices from points lower boundary in
list form []
    :param high_thrh_delete: Delete indices from points upper boundary in
list form []
    :param classes: Number of classes available for the data
    :return: Set of y-coordinates (y_start,y_end) for a line, x-coordinates
are generally (0,1) for NDVI
    """

    # Determine high and low percentile values per bin
    NDVI_split = []
    LST_split = []
    NDVI_LST = []
    NDVI_LST_low = []
    NDVI_LST_high = []
    LST_min_list = []
    LST_max_list = []
    for i in range(0, 10):
        NDVI_idx = np.where((NDVI >= i / 10.0) & (NDVI < i / 10.0 + 0.1))
        NDVI_split.append(NDVI[NDVI_idx[0], NDVI_idx[1]])
        LST_split.append(LST[NDVI_idx[0], NDVI_idx[1]])

        # Determine outliers based on boxplot theory
```

```
        if len(LST_split[i]) != 0:
            LST_min, LST_max = Boxplot_outlier(LST_split[i])
            LST_min_list.append(LST_min)
            LST_max_list.append(LST_max)
            LST_split[i][LST_split[i] < LST_min] = np.nan
            LST_split[i][LST_split[i] > LST_max] = np.nan
        else:
            LST_min_list.append(np.nan)
            LST_max_list.append(np.nan)

        NDVI_LST.append((np.array(NDVI_split[i]), np.array(LST_split[i])))
        NDVI_LST[i] = np.asarray(NDVI_LST[i])
        NDVI_LST[i] = np.ndarray.transpose(NDVI_LST[i])
        mask = ~np.isnan(NDVI_LST[i][:, 0]) & ~np.isnan(NDVI_LST[i][:, 1])
        NDVI_LST[i] = NDVI_LST[i][mask]

        # Calculate percentile value of top/bottom values
        if len(NDVI_LST[i]) > 0:
            NDVI_LST_high.append(np.percentile(NDVI_LST[i], (100-
Percentile_thrh), axis=0))
            NDVI_LST_low.append(np.percentile(NDVI_LST[i], Percentile_thrh,
axis=0))

    # Remove boxplot outlier values
    LST_min = np.nanmin(LST_min_list)
    LST[LST < LST_min] = np.nan
    LST_max = np.nanmax(LST_max_list)
    LST[LST > LST_max] = np.nan

    # Remove nan values and transform to array
    NDVI_LST_low = np.asarray(NDVI_LST_low)
    mask = ~np.isnan(NDVI_LST_low[:, 0]) & ~np.isnan(NDVI_LST_low[:, 1])
    NDVI_LST_low = NDVI_LST_low[mask]

    NDVI_LST_high = np.asarray(NDVI_LST_high)
    mask = ~np.isnan(NDVI_LST_high[:, 0]) & ~np.isnan(NDVI_LST_high[:, 1])
    NDVI_LST_high = NDVI_LST_high[mask]

    # fit line through points left after threshold
    NDVI_LST_low_thrh = np.delete(NDVI_LST_low, low_thrh_delete, axis=0)
    slope_low, intercept_low, r_value_low, p_value_low, std_err_low =
stats.linregress(NDVI_LST_low_thrh[:, 0], NDVI_LST_low_thrh[:, 1])
    y2_low = slope_low * NDVI_max + intercept_low

    NDVI_LST_high_thrh = np.delete(NDVI_LST_high, high_thrh_delete, axis=0)
    slope_high, intercept_high, r_value_high, p_value_high, std_err_high =
stats.linregress(NDVI_LST_high_thrh[:, 0], NDVI_LST_high_thrh[:, 1])
    y2_high = slope_high * NDVI_max + intercept_high

    LST1_high = slope_high * NDVI_min + intercept_high
    LST2_high = slope_high * NDVI_max + intercept_high

    LST1_low = slope_low * NDVI_min + intercept_low
    LST2_low = slope_low * NDVI_max + intercept_low

    lineset = []
    lineset.append([LST1_low, LST2_low])
    line_start_incr = (LST1_high - LST1_low) / (classes - 2)
    line_end_incr = (LST2_high - LST2_low) / (classes - 2)
    for i in range(classes - 2):
        line_start = lineset[i][0] + line_start_incr
```

```
        line_end = lineset[i][1] + line_end_incr
        lineset.append([line_start, line_end])
    lineset = np.asarray(lineset)

    polygons = []
    for i in range(len(lineset)):
        if i == 0:
            polygons.append(Polygon(
                [(NDVI_min, lineset[i, 0]), (NDVI_min, LST_min), (NDVI_max,
LST_min), (NDVI_max, lineset[i, 1])]))
        elif i == len(lineset) - 1:
            polygons.append(Polygon(
                [(NDVI_min, lineset[i - 1, 0]), (NDVI_min, lineset[i, 0]),
(NDVI_max, lineset[i, 1]),
                 (NDVI_max, lineset[i - 1, 1])]))
            polygons.append(Polygon(
                [(NDVI_min, LST_max), (NDVI_min, lineset[i, 0]), (NDVI_max,
lineset[i, 1]), (NDVI_max, LST_max)]))
        else:
            polygons.append(Polygon(
                [(NDVI_min, lineset[i - 1, 0]), (NDVI_min, lineset[i, 0]),
(NDVI_max, lineset[i, 1]),
                 (NDVI_max, lineset[i - 1, 1])]))

    # Create arrays
    NDVI_array = np.reshape(NDVI, (NDVI.size, 1))
    LST_array = np.reshape(LST, (LST.size, 1))

    # Create dataset
    data = np.hstack((NDVI_array, LST_array))
    mask = ~np.isnan(data[:, 0]) & ~np.isnan(data[:, 1])
    data = data[mask]
    x, y = data.T

    # Evaluate a gaussian kde on a regular grid of nbins x nbins over data
extents
    k = kde.gaussian_kde(data.T)
    xi, yi = np.mgrid[x.min():x.max():nbins * 1j, y.min():y.max():nbins *
1j]
    zi = k(np.vstack([xi.flatten(), yi.flatten()]))

    #Make density plot of LST-NDVI field along with class polygons
    fig, ax = plt.subplots()
    ax.pcolormesh(xi, yi, zi.reshape(xi.shape), shading='gouraud',
cmap=plt.cm.GnBu_r)
    ax.contour(xi, yi, zi.reshape(xi.shape))
    plt.scatter(NDVI_LST_low_thrh[:, 0], NDVI_LST_low_thrh[:, 1],
color='r')
    plt.scatter(NDVI_LST_high_thrh[:, 0], NDVI_LST_high_thrh[:, 1],
color='r')
    for j in range(len(polygons)):
        x, y = polygons[j].exterior.xy
        plt.plot(x, y, color='k', alpha=0.25)
    plt.title("NDVI-LST plot\n%s" % (Image_date))
    plt.xlabel("NDVI")
    plt.ylabel("LST (K)")
    ax.legend(loc = 'upper right')

    return polygons, LST, xi, yi, zi
```

Module 3C: boxplot outlier function

```
import numpy as np

def Boxplot_outlier(Data):
    Data_array = np.array(Data)
    Data_array = Data_array[~np.isnan(Data_array)]
    Data_sort = sorted(Data_array)
    q1, q3 = np.percentile(Data_sort, [25, 75])
    iqr = q3 - q1
    lower_bound = q1 - (1.5 * iqr)
    upper_bound = q3 + (1.5 * iqr)
    return lower_bound, upper_bound
```

Appendix J. SWI module 4

Module 4A: Assign classes to pixels

```
from Trapezoid.Module_4.Pixel_class_assign import Pixel_class_assign
import rasterio
import matplotlib.pyplot as plt
import pickle
import os
import numpy as np

#Input parameters
Image_date = ["2018_03_20", "2018_04_21", "2018_05_07", "2018_07_03",
"2018_07_26"]
Image_date_adj = ["20-03-2018", "21-04-2018", "07-05-2018", "03-07-2018",
"26-07-2018"]
CropType = ["Sugarbeet", "Winterwheat"]
CropType_adj = ["sugar beet", "winter wheat"]
buff_dist = -60

for i in range(len(Image_date)):
    # Open dataset from module 3
    with
open('C:\Python_projects\Thesis\Trapezoid\Module_3\Pixel_envelope_boundarie
s_%s.pkl' % (Image_date[i]),'rb') as f:
        Projection, LST, NDVI, polygons, nbins, xi, yi, zi = pickle.load(f)

    for j in range(len(CropType)):
        # Read in land use data
        NDVI_crop_path =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\NDVI_buffer\Merged_NDVI\%s_merged_NDVI_buffer_%i.tif"
%(Projection, Image_date[i], CropType[j], buff_dist)
        with rasterio.open(NDVI_crop_path) as img:
            NDVI_crop = img.read(1, masked=True)

        LST_crop_path =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LST_buffer\Merged_LST\%s_merged_LST_buffer_%i.tif"
%(Projection, Image_date[i], CropType[j], buff_dist)
        with rasterio.open(LST_crop_path) as img:
            LST_crop = img.read(1, masked=True)
            raster_meta = img.meta.copy()

        # Creating new folder for croptype
        dir_name = "LST_NDVI"
        Output_property_field =
os.path.join(r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output
\NOP\%s\%s\Output_vegetation"  %(Projection, Image_date[i]), dir_name)
        if not os.path.exists(Output_property_field):
            os.makedirs(Output_property_field)

        #Apply pixel class assign function
        Class_data = Pixel_class_assign(LST_crop, NDVI_crop, polygons)

        #Write results to .tif file
        Output_tif =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LST_NDVI\%s_LST_NDVI_classes_%s.tif" %(Projection,
Image_date[i], CropType[j], Image_date[i])
        with rasterio.open(Output_tif, 'w', **raster_meta) as dst:
```

```
            dst.write(Class_data, 1)

        #Make density plot of LST-NDVI field along with class polygons
        fig, ax = plt.subplots()
        ax.pcolormesh(xi, yi, zi.reshape(xi.shape), shading='gouraud',
cmap=plt.cm.GnBu_r)
        ax.contour(xi, yi, zi.reshape(xi.shape))
        ax.scatter(NDVI_crop, LST_crop, color='r', alpha=0.1, label="%s
pixels" %(CropType_adj[j]))
        for k in range(len(polygons)):
            x, y = polygons[k].exterior.xy
            plt.plot(x, y, color='k', alpha=0.25)
        plt.title("NDVI-RCT plot %s\n%s" % (CropType_adj[j],
Image_date_adj[i]))
        plt.xlabel("NDVI")
        plt.ylabel("RCT (K)")
        ax.legend(loc = 'upper right')

        #Save figure to .png file
        figure_output =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LST_NDVI\%s_NDVI_LST_space.png" %(Projection,
Image_date[i], CropType[j])
        plt.savefig(figure_output)
```

Module 4A: Assign classes to pixels function

```
from shapely.geometry import Point
import numpy.ma as ma
import numpy as np

def Pixel_class_assign(LST_crop,NDVI_crop, polygons):
    """
    Assign each pixel to class of wetness
    :param LST_crop: relative land surface temperature compared to the
instantaneous air temperature of crop specific pixels
    :param NDVI_crop: normalized difference vegetation index of crop
specific pixels
    :param polygons: classes of wetness
    :return: masked array of classvalue per pixel and density plot
parameters
    """

    # Assign pixels to a class based on polygons from Module 3
    NDVI_data = ma.getdata(NDVI_crop)
    LST_data = ma.getdata(LST_crop)
    class_data = np.copy(LST_data)
    for i in range(NDVI_data.shape[0]):
        for j in range(NDVI_data.shape[1]):
            if NDVI_data[i,j] == -9999 or LST_data[i,j] == -9999:
                class_data[i, j] == np.nan
            else:
                for k in range(len(polygons)):
                    point = Point(NDVI_data[i,j], LST_data[i,j])
                    if polygons[k].contains(point):
                        class_data[i,j] = k+1

    class_data = ma.masked_values(class_data, -9999)

    return class_data
```

Appendix K. SWI module 5

Module 5: Validation

```
import rasterio
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import kde
from Trapezoid.Metadatafile_reader import build_data
from datetime import datetime
import os
import glob

Image_date = ["2018_04_21", "2018_05_07", "2018_07_03", "2018_07_26"]
Image_date_adj = ["21-04-2018", "07-05-2018", "03-07-2018", "26-07-2018"]
CropType = ["Sugarbeet", "Winterwheat"]
CropType_adj = ["sugar beet", "winter wheat"]
number_classes = 20
nbins = 100

for i in range(len(Image_date)):
    # Determining the input parameters from metadatafile
    LANDSAT8_metadata_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_input\Landsat8\%s"
% (Image_date[i])
    file_name = os.path.join(LANDSAT8_metadata_dir, "*[_MTL].txt")
    file_path = glob.glob(file_name)
    f = open(file_path[0], 'r')  # open file for reading
    data = build_data(f)
    UTM_zone = int(data["UTM_ZONE"])
    Projection = "EPSG326%s" % (UTM_zone)
    Resolution = int(data["GRID_CELL_SIZE_REFLECTIVE"][0:2])
    Date = data["DATE_ACQUIRED"]
    adate = datetime.strptime(Date, "%Y-%m-%d")
    DOY = adate.timetuple().tm_yday
    Year = adate.timetuple().tm_year
    for j in range(len(CropType)):
        Classes_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LST_NDVI\%s_LST_NDVI_classes_%s.tif" %(Projection,
Image_date[i], CropType[j], Image_date[i])
        with rasterio.open(Classes_dir) as img:
            Classes = img.read(1)

        Classes[Classes>number_classes] = np.nan
        Classes[Classes<0] = np.nan

        SM_dir =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_soil_moisture\SOM_buffer\Merged_SOM\%s_merged_SOM_buffer_-60.tif"
%(Projection, Image_date[i], CropType[j])
        with rasterio.open(SM_dir) as img:
            SM = img.read(1)

        SM[SM>1] = np.nan
        SM[SM<0] = np.nan

        # Create arrays
        Classes_array = np.reshape(Classes, (Classes.size, 1))
        # Classes_array = abs(Classes_array - number_classes) + 1
        SM_array = np.reshape(SM, (SM.size, 1))
```

```
        # Create dataset
        data = np.hstack((SM_array, Classes_array))
        mask = ~np.isnan(data[:, 0]) & ~np.isnan(data[:, 1])
        data = data[mask]
        x, y = data.T

        # Evaluate a gaussian kde on a regular grid of nbins x nbins over
data extents
        k = kde.gaussian_kde(data.T)
        xi, yi = np.mgrid[x.min():x.max():nbins * 1j, y.min():y.max():nbins
* 1j]
        zi = k(np.vstack([xi.flatten(), yi.flatten()]))

        # Make density plot of LST-NDVI field along with class polygons
        fig, ax = plt.subplots()
        ax.pcolormesh(xi, yi, zi.reshape(xi.shape), shading='gouraud',
cmap=plt.cm.GnBu_r)
        ax.contour(xi, yi, zi.reshape(xi.shape))
        plt.title("Soil moisture vs. soil wetness indicator\n%s %s" %
(CropType_adj[j], Image_date_adj[i]))
        plt.xlabel("Soil moisture content [-]")
        plt.ylabel("Soil wetness indicator [-]")

        #Save figure to .png file
        figure_output =
r"C:\Users\TUDelftSID\Documents\Thesis\Data\SEBAL\SEBAL_output\NOP\%s\%s\Ou
tput_vegetation\LST_NDVI\%s_SM_SWI_space.png" %(Projection, Image_date[i],
CropType[j])
        plt.savefig(figure_output)
```

Appendix L. SoilGrids30m principal component results

Contribution explanatory variables to each principal component for clay content

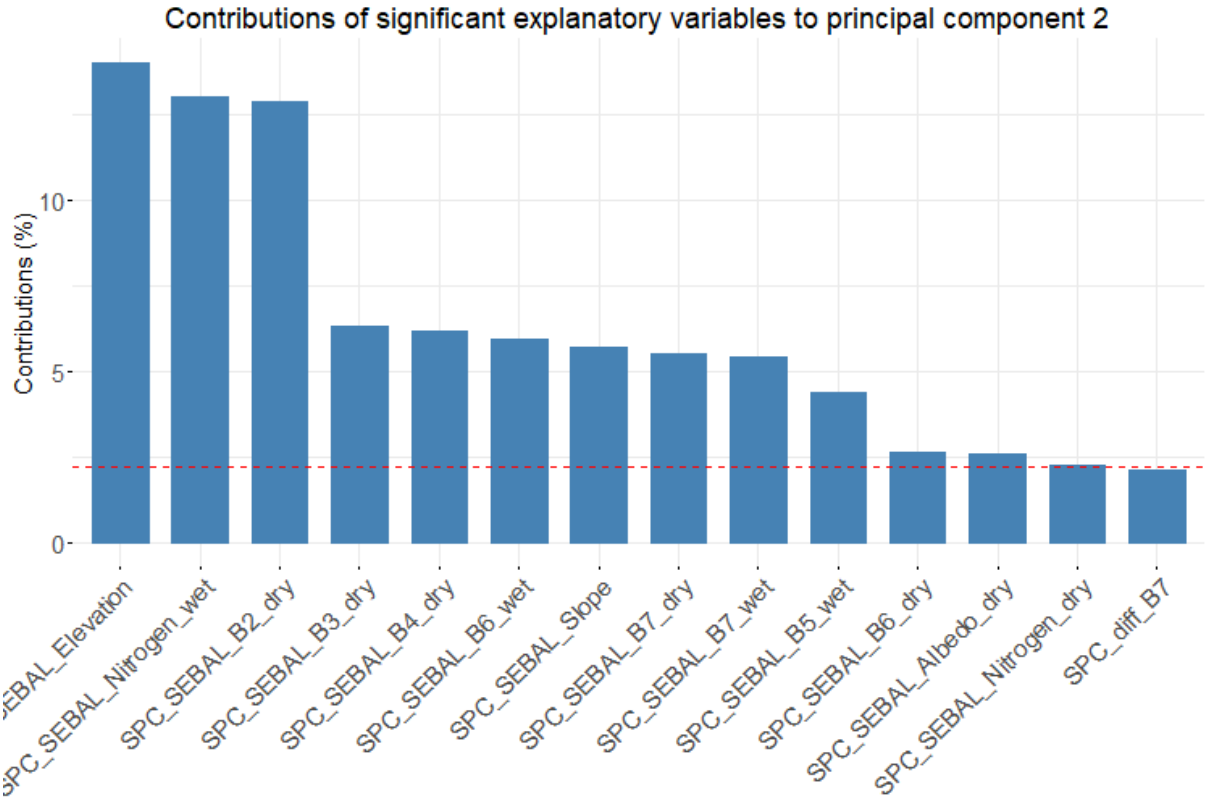


Figure L.1 Contribution significant explanatory variables to principal component 2 for clay content

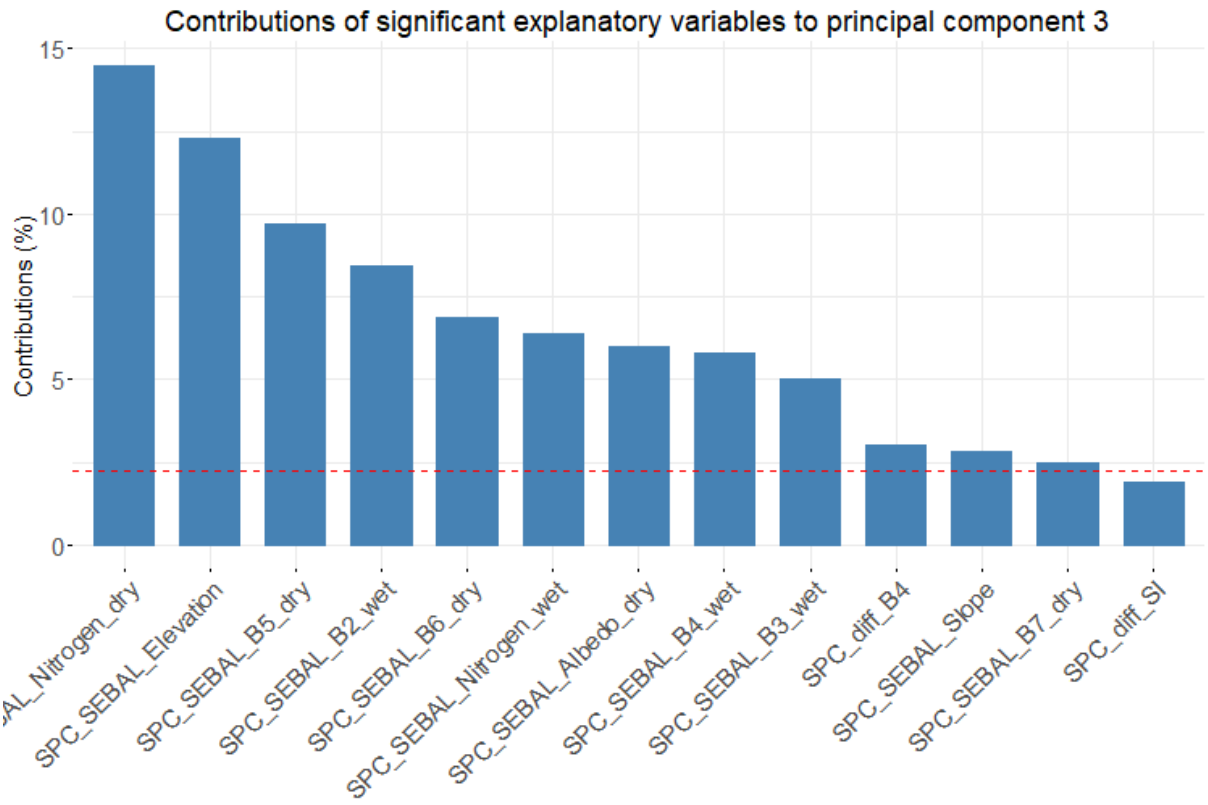


Figure L.2 Contribution significant explanatory variables to principal component 3 for clay content

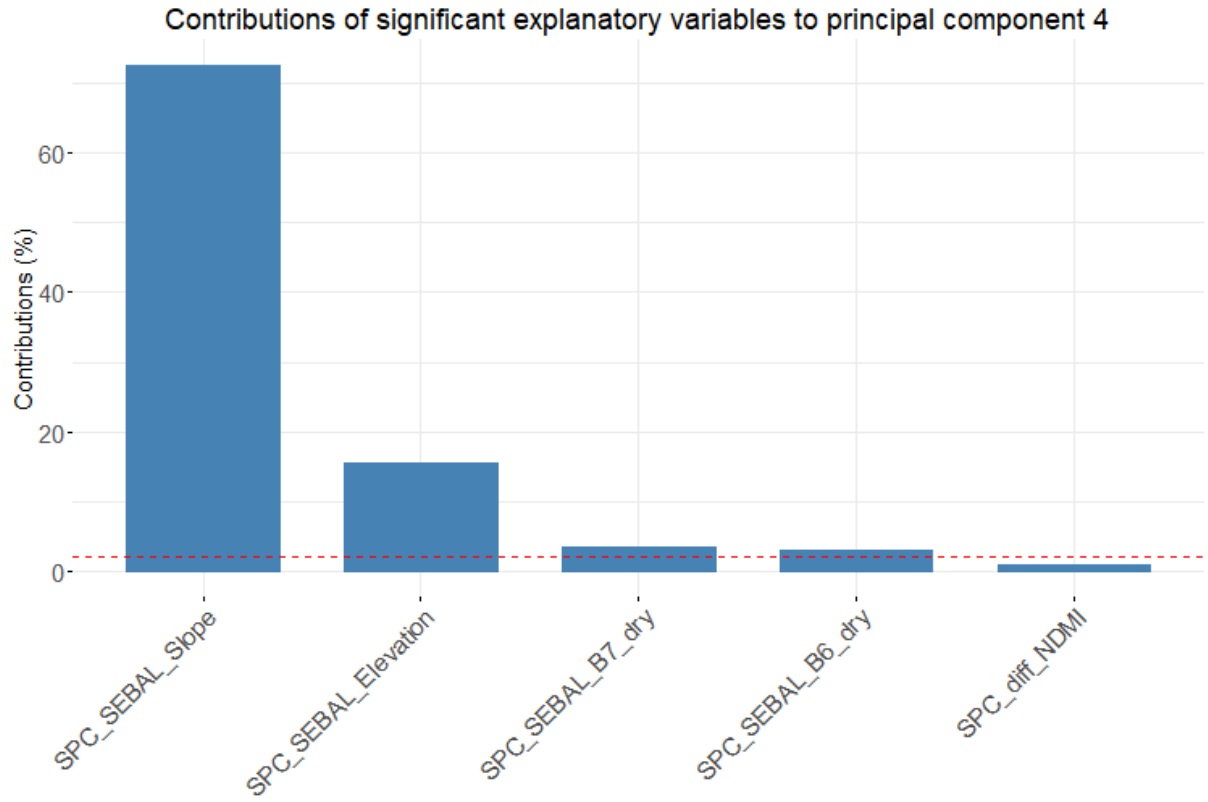


Figure L.3 Contribution significant explanatory variables to principal component 4 for clay content

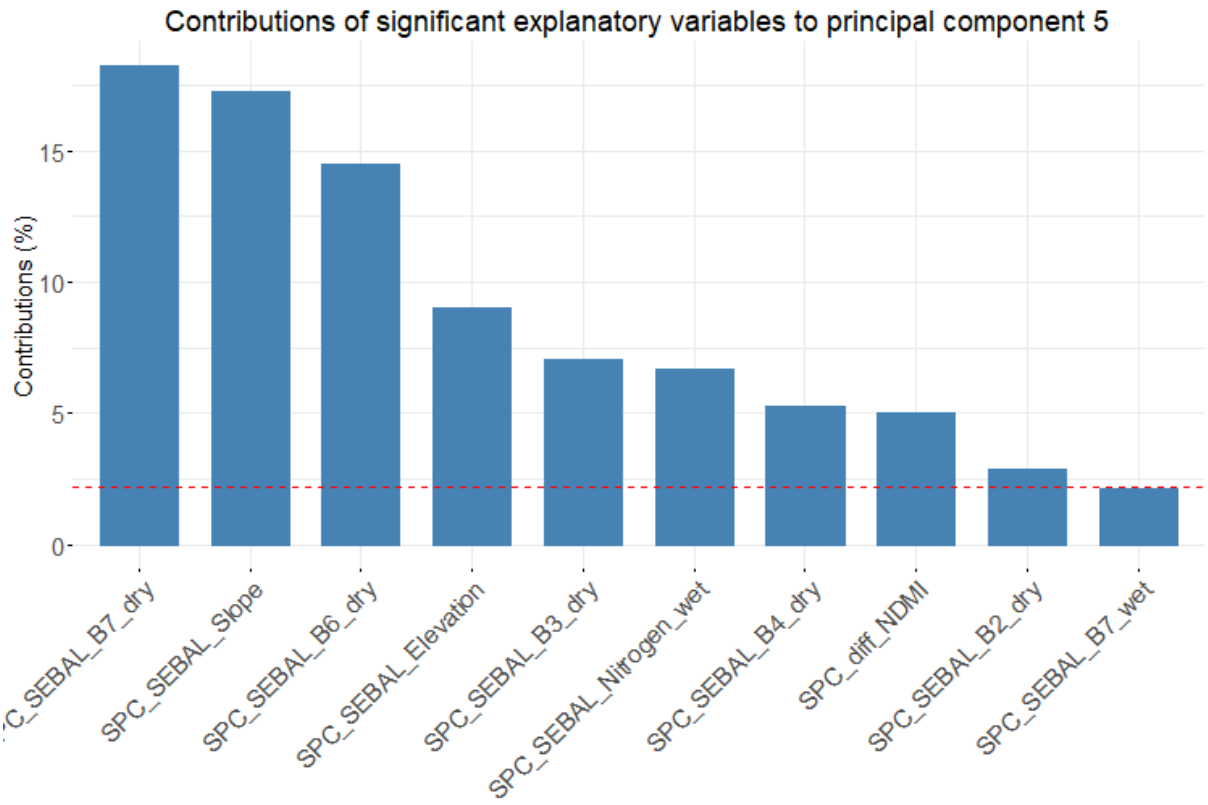


Figure L.4 Contribution significant explanatory variables to principal component 5 for clay content

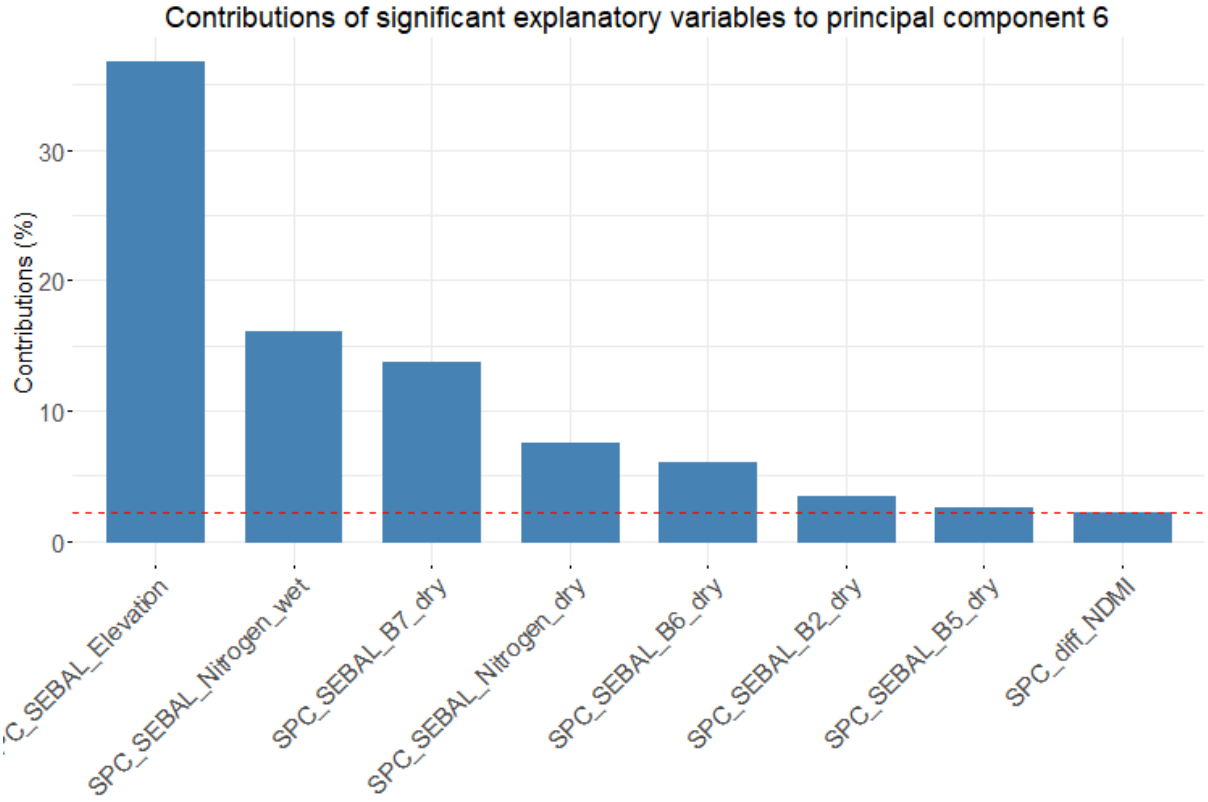


Figure L.5 Contribution significant explanatory variables to principal component 6 for clay content

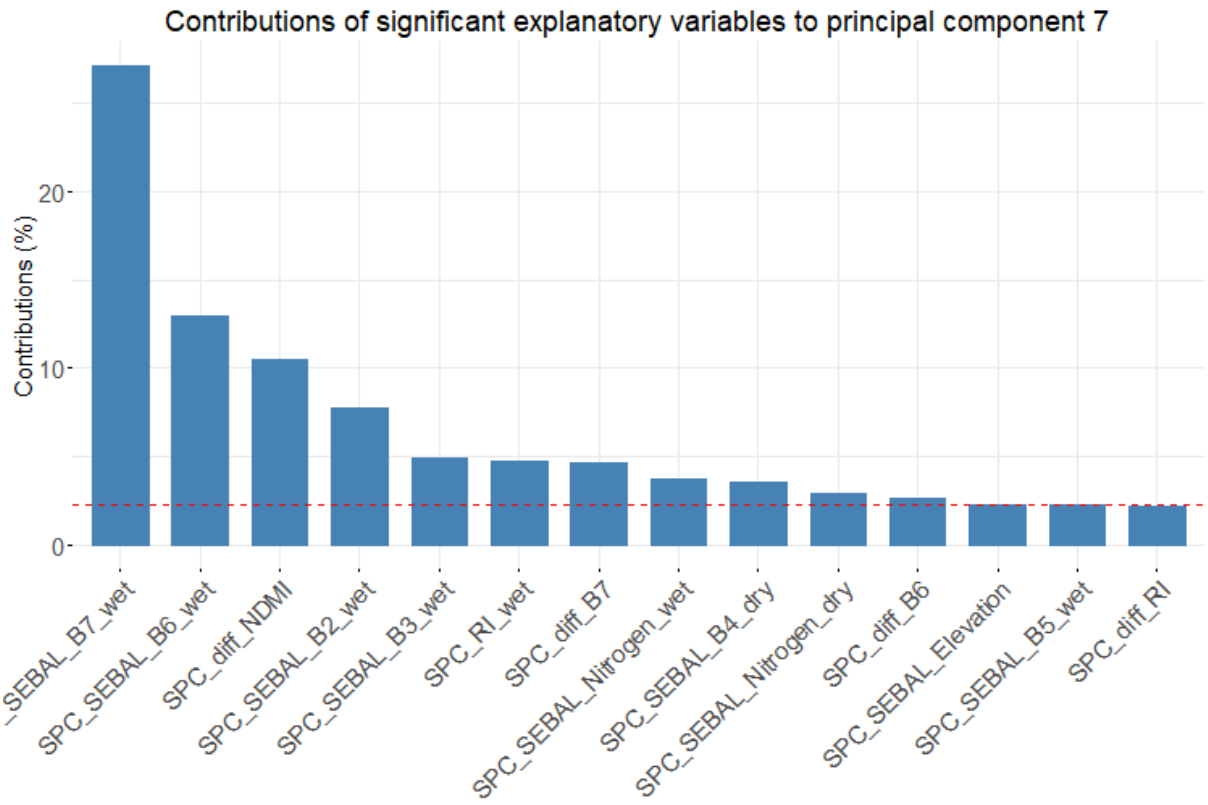


Figure L.6 Contribution significant explanatory variables to principal component 7 for clay content

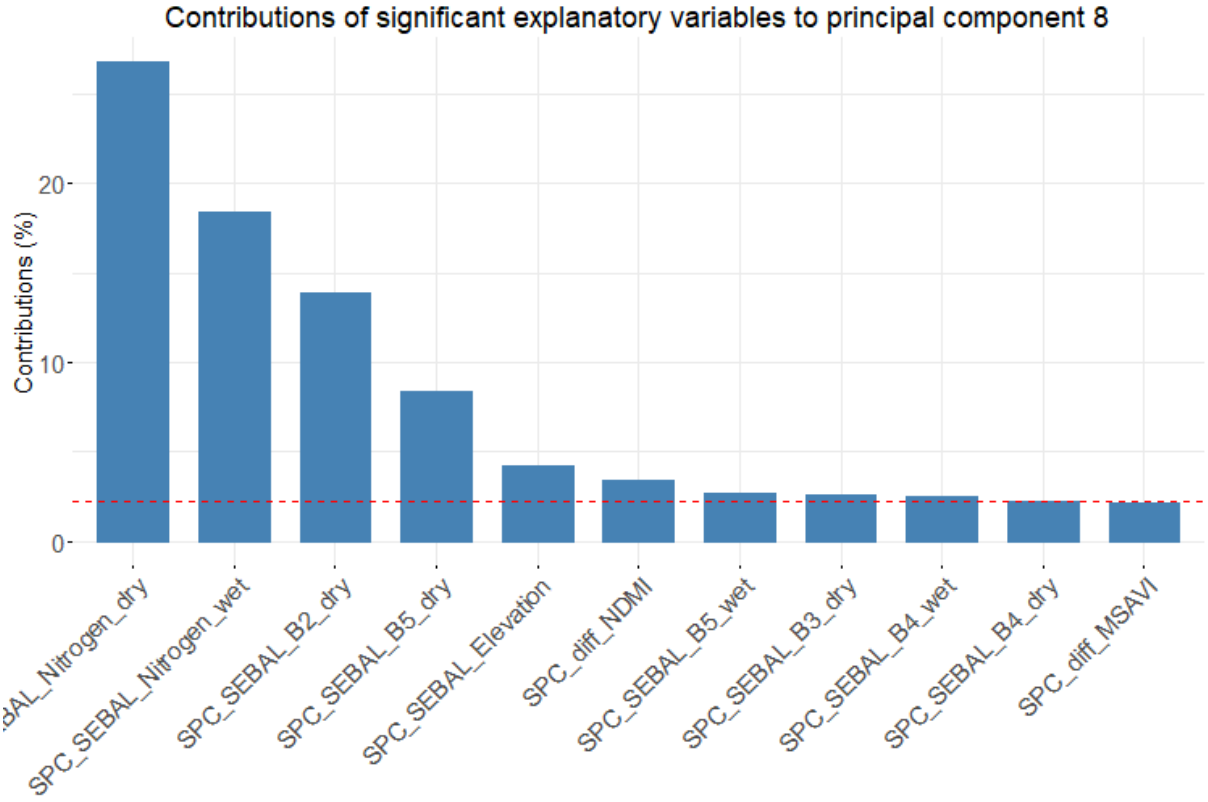


Figure L.7 Contribution significant explanatory variables to principal component 8 for clay content

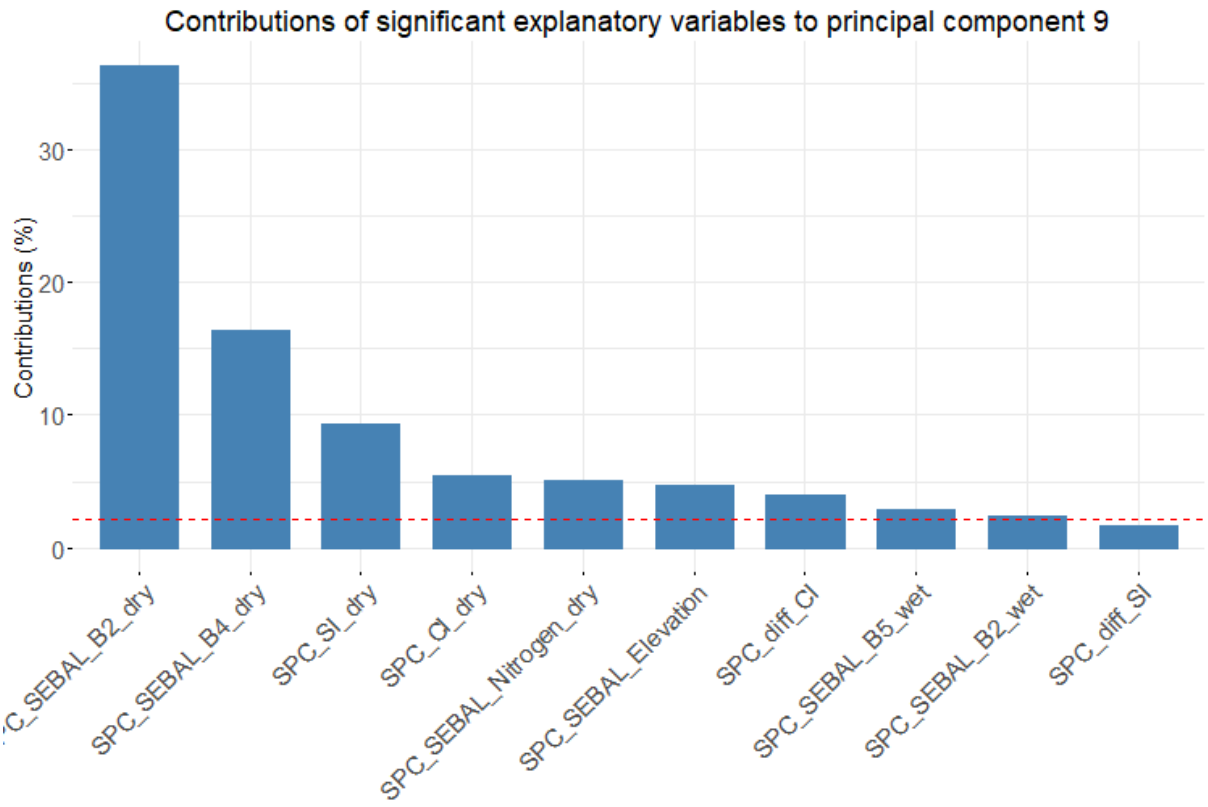


Figure L.8 Contribution significant explanatory variables to principal component 9 for clay content

Contribution explanatory variables to each principal component for organic matter content

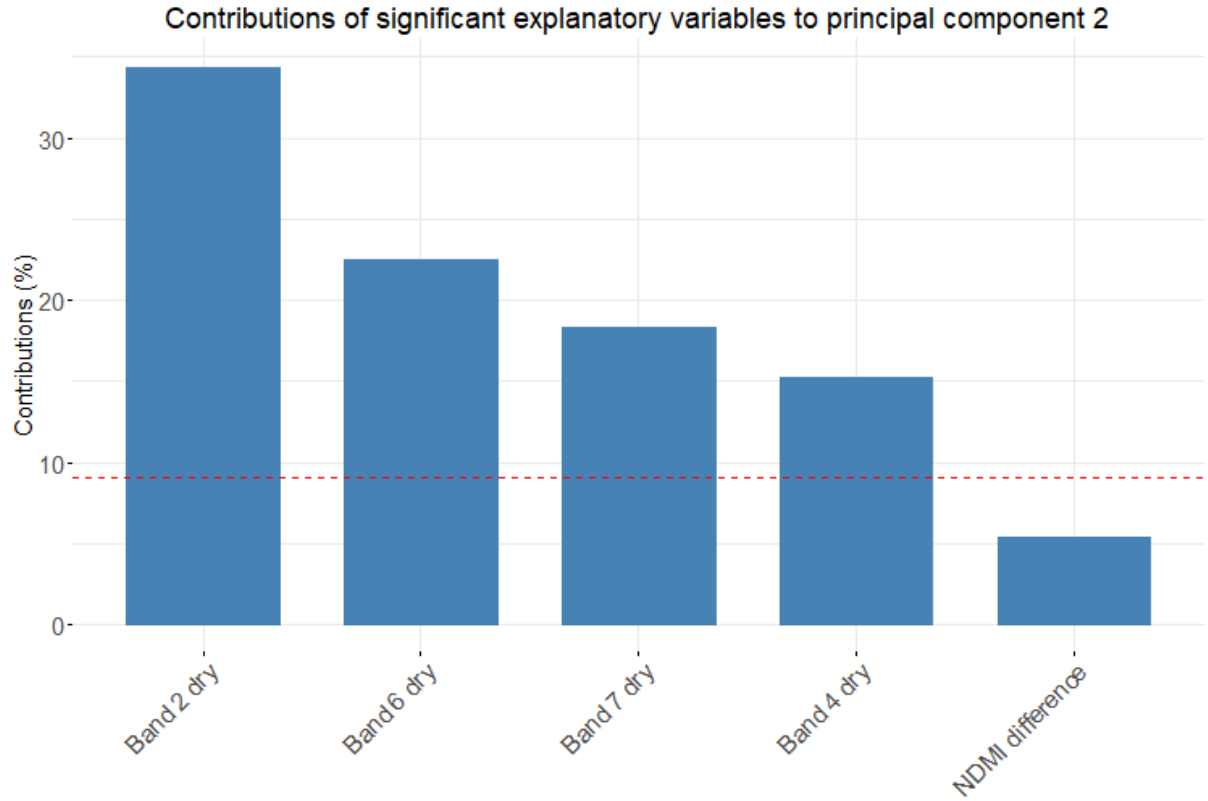


Figure L.9 Contribution significant explanatory variables to principal component 2 for organic matter content

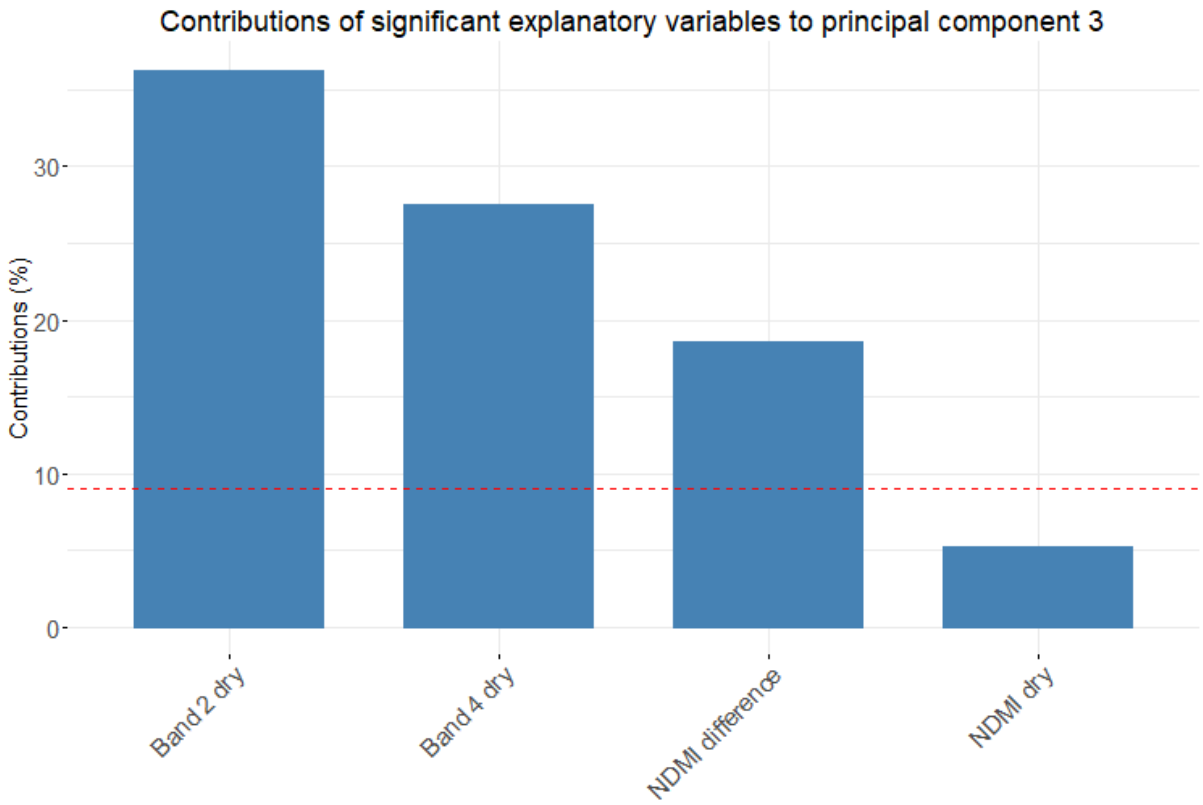


Figure L.10 Contribution significant explanatory variables to principal component 3 for organic matter content

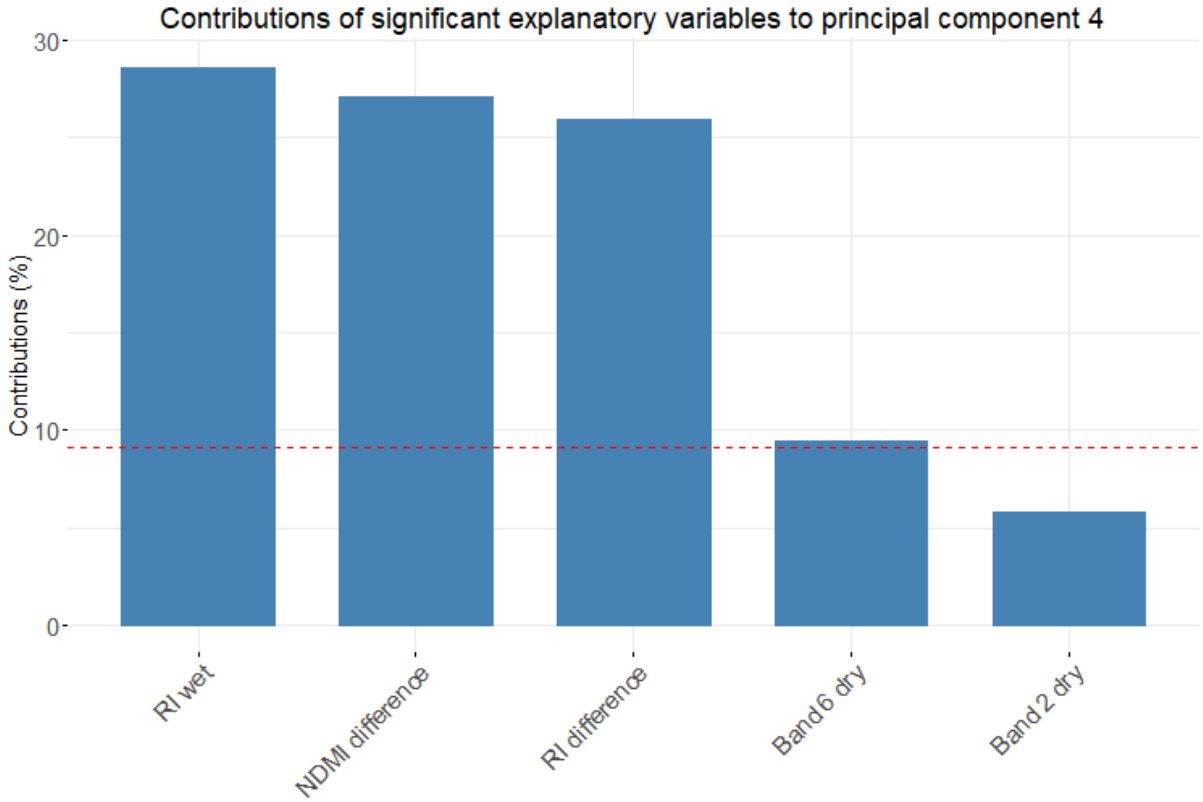


Figure L.11 Contribution significant explanatory variables to principal component 4 for organic matter content

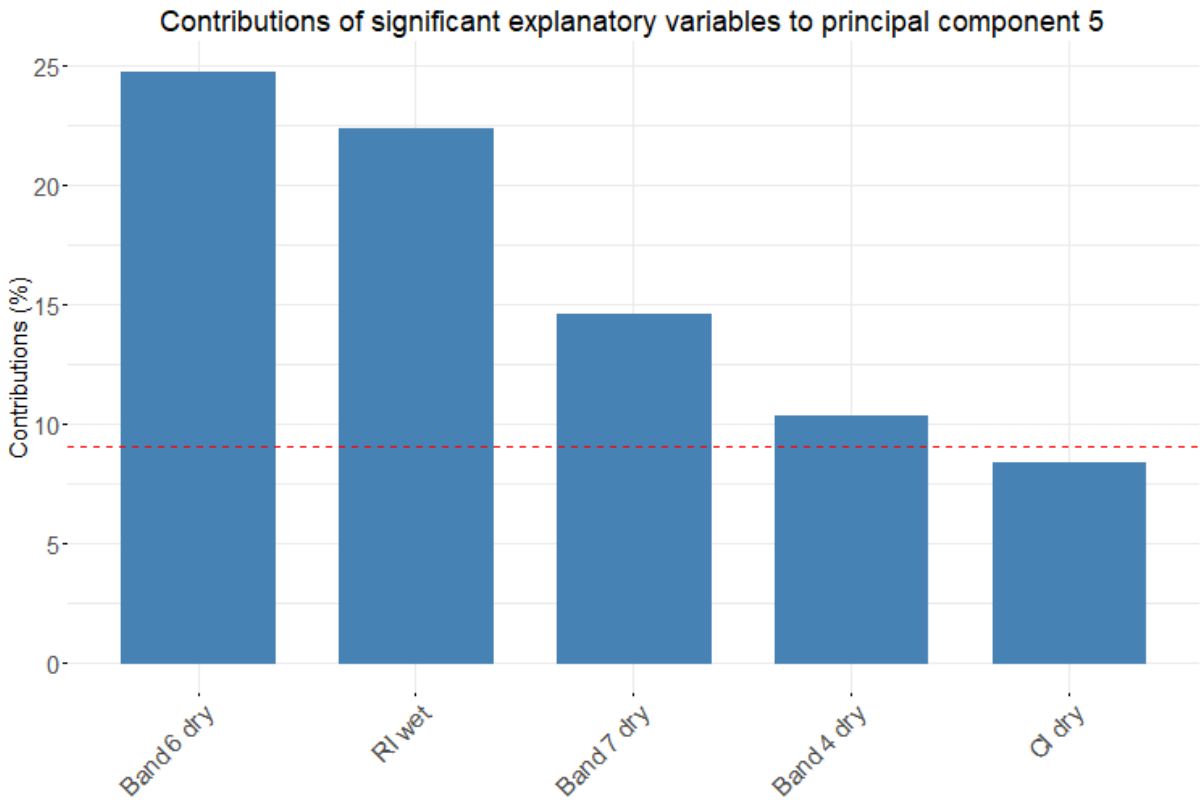


Figure L.12 Contribution significant explanatory variables to principal component 5 for organic matter content

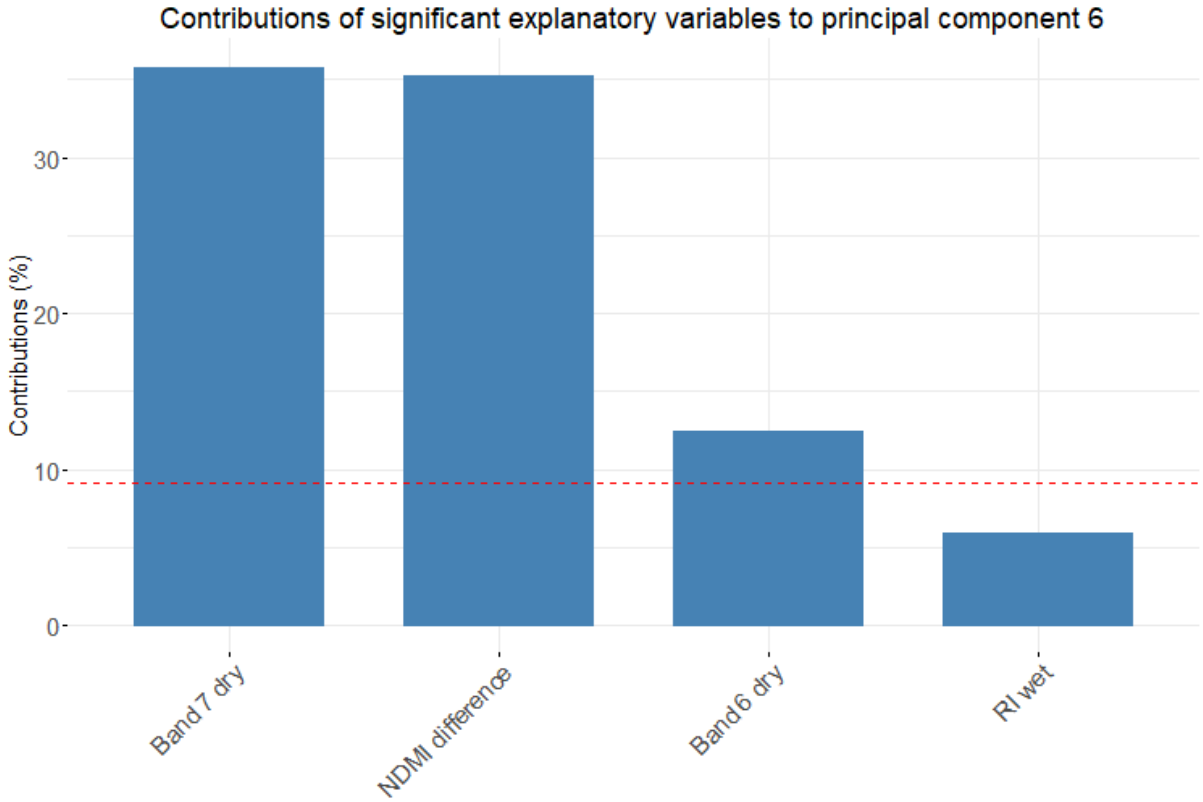


Figure L.13 Contribution significant explanatory variables to principal component 6 for organic matter content

Appendix M. SWHC-SM boxplots

Boxplot sugar beet

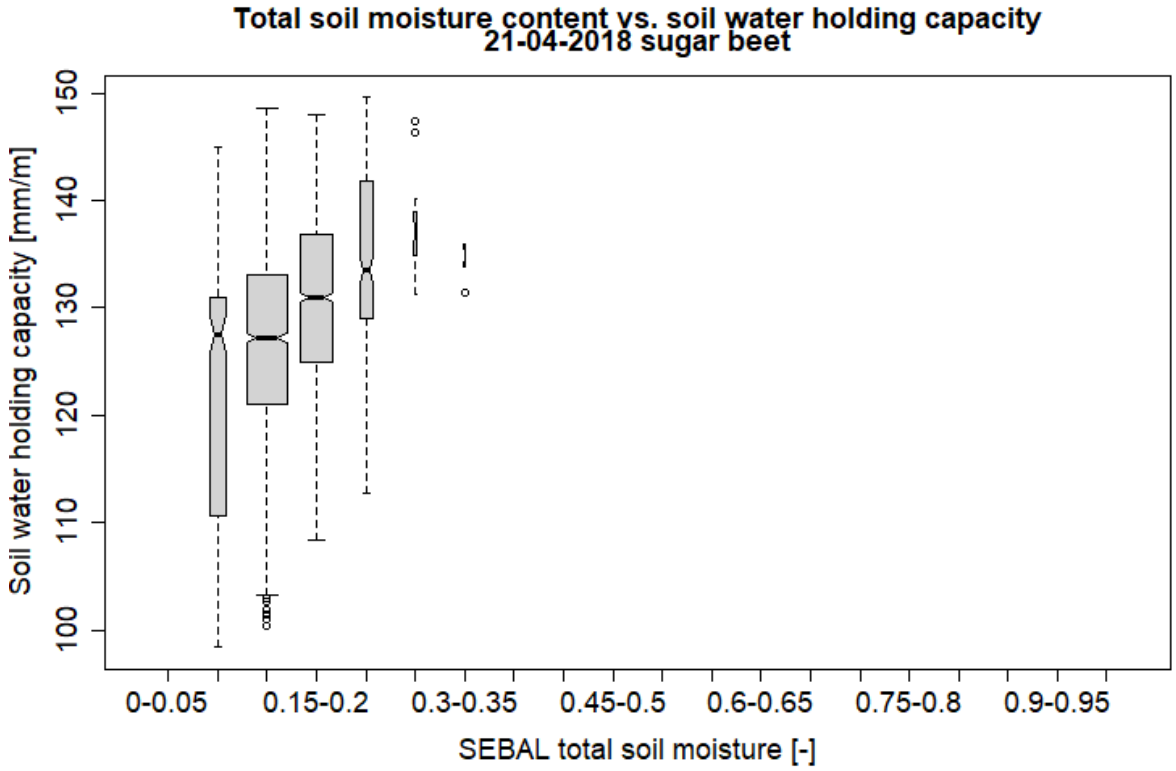


Figure M.1 Boxplot soil moisture vs. soil water holding capacity, 21-04-2018 sugar beet

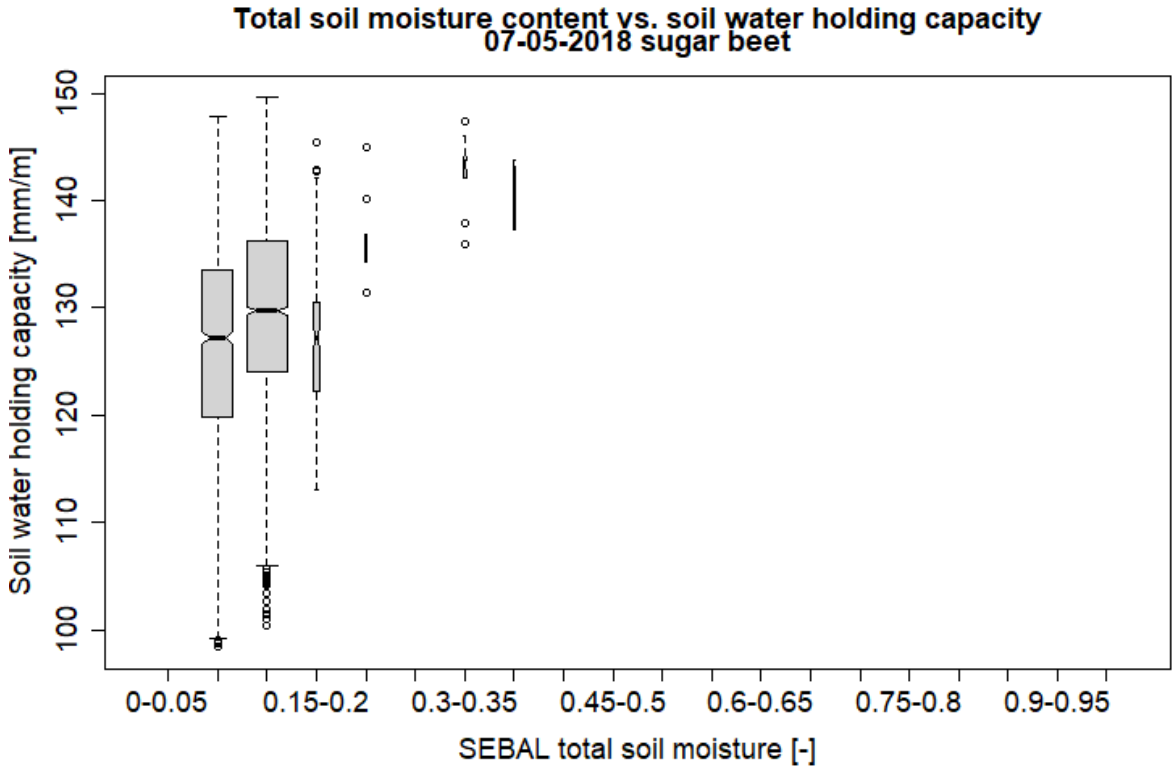


Figure M.2 Boxplot soil moisture vs. soil water holding capacity, 07-05-2018 sugar beet

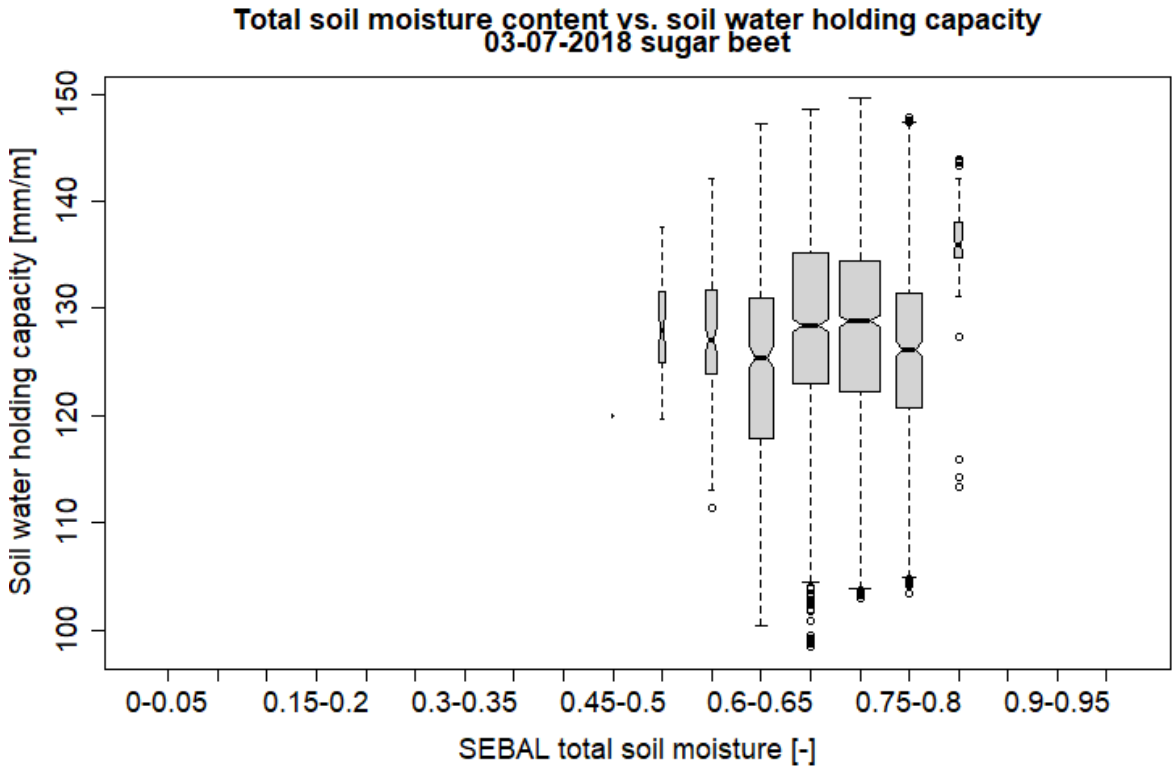


Figure M.3 Boxplot soil moisture vs. soil water holding capacity, 03-07-2018 sugar beet

Boxplot winter wheat

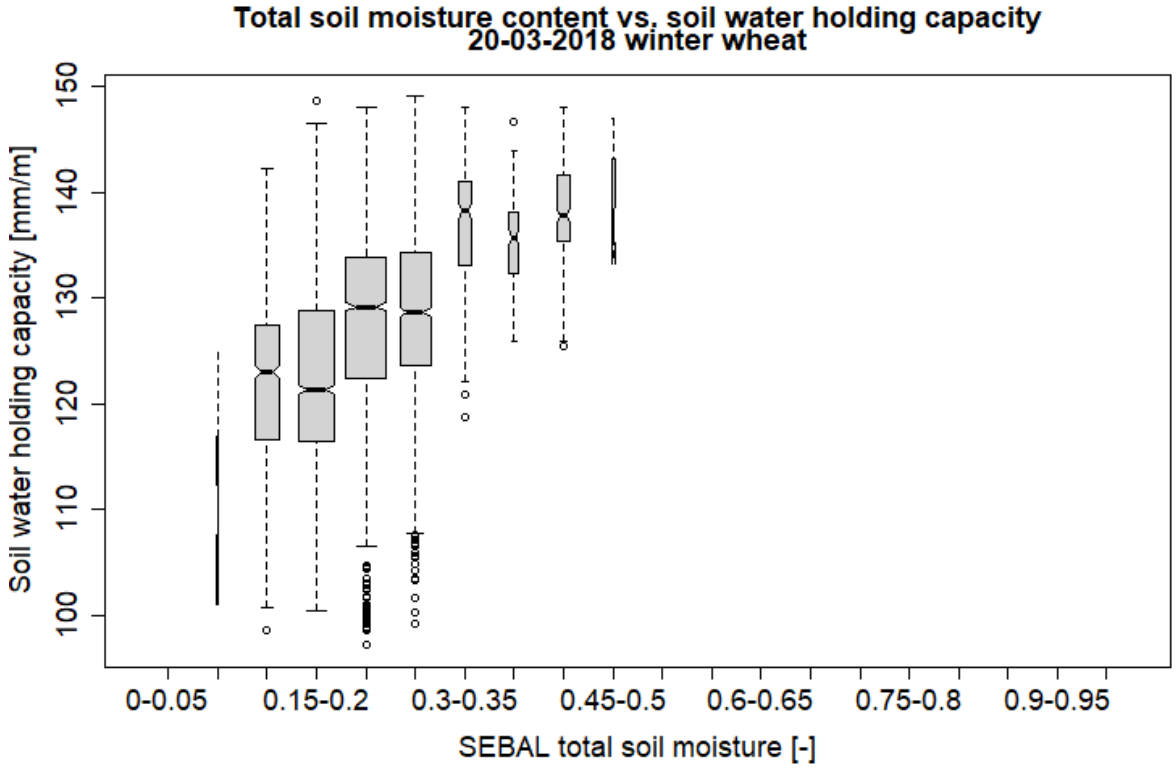


Figure M.4 Boxplot soil moisture vs. soil water holding capacity, 20-03-2018 winter wheat

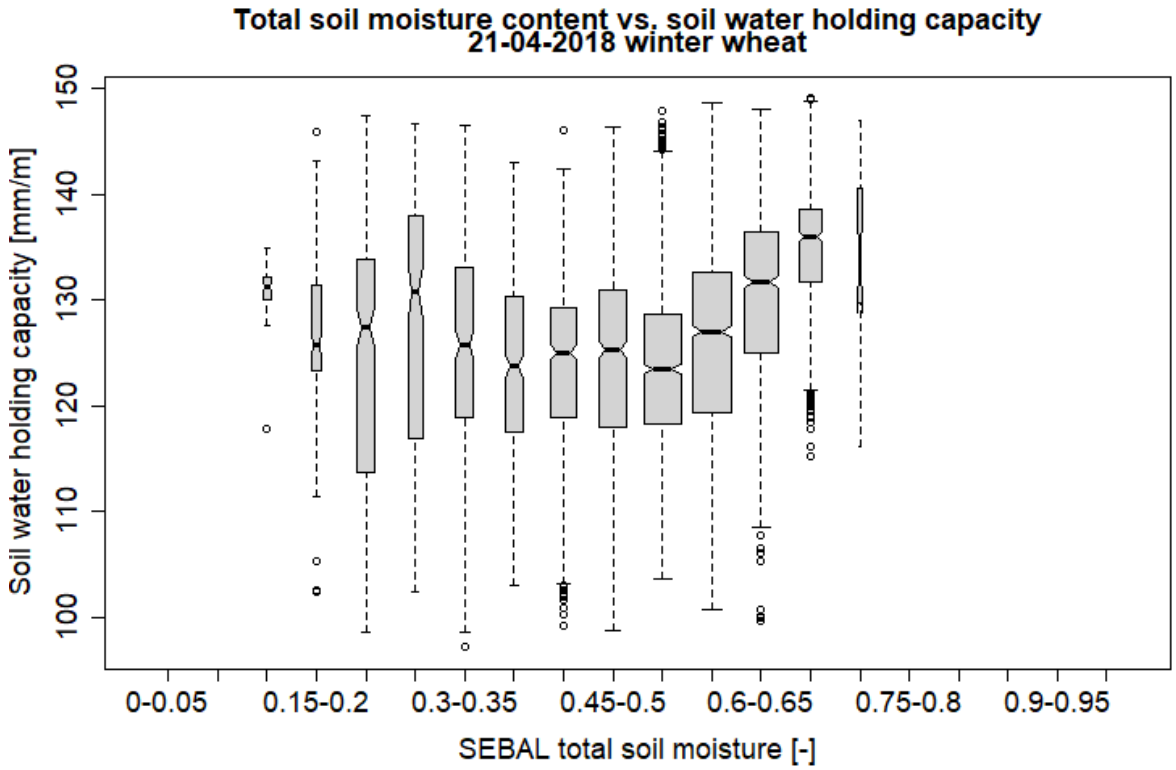


Figure M.5 Boxplot soil moisture vs. soil water holding capacity, 21-04-2018 winter wheat

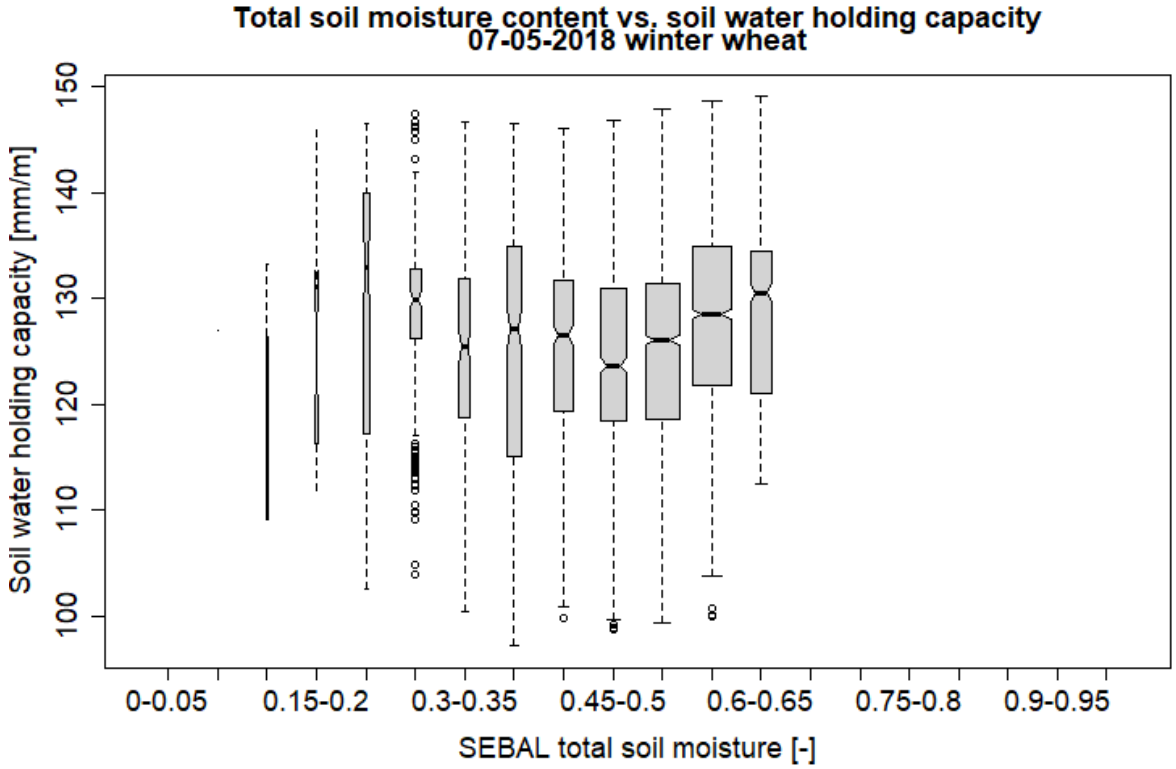


Figure M.6 Boxplot soil moisture vs. soil water holding capacity, 07-05-2018 winter wheat

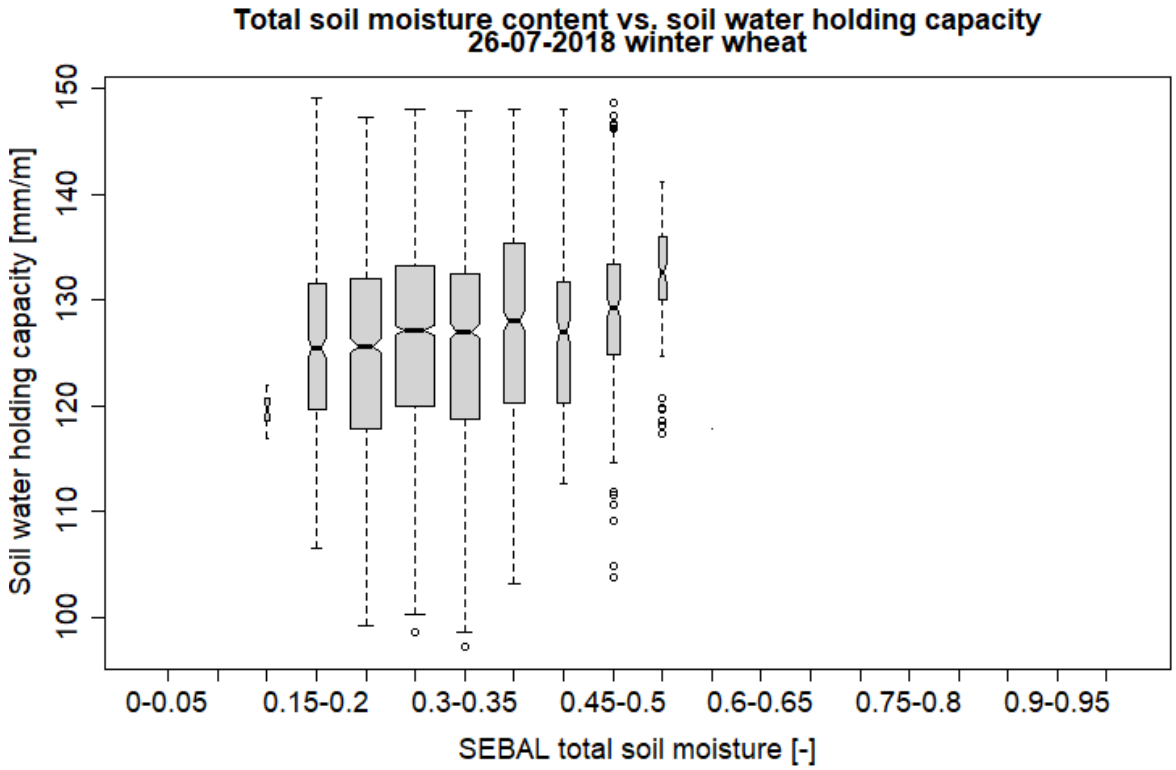


Figure M.7 Boxplot soil moisture vs. soil water holding capacity, 26-07-2018 winter wheat

Appendix N. SWI NDVI-RCT plot

NDVI-RCT plot sugar beet

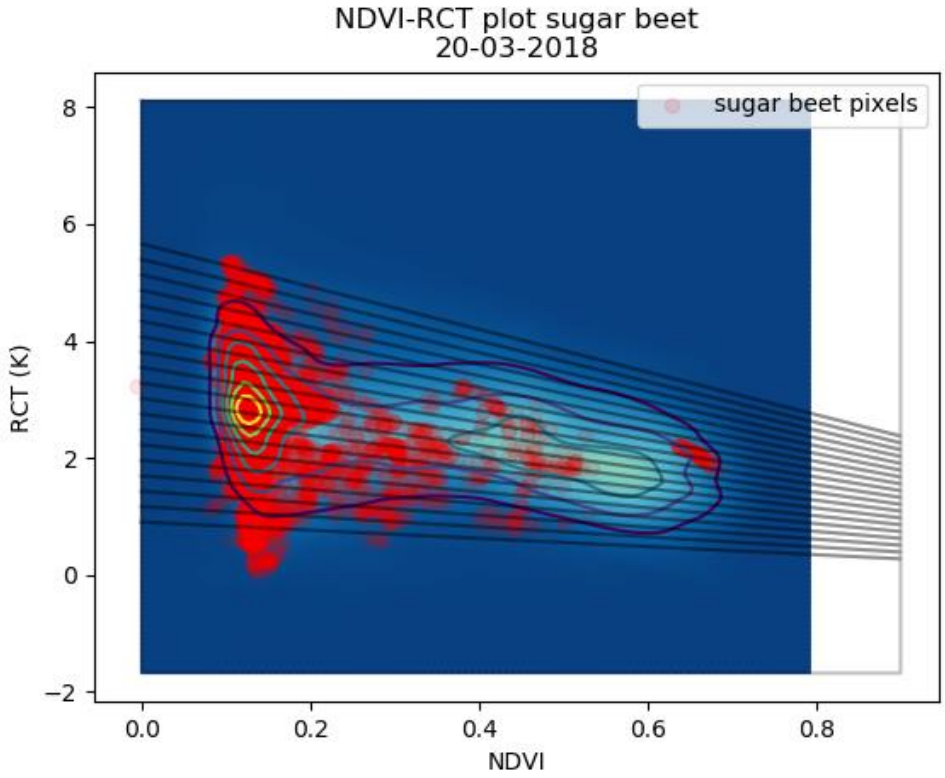


Figure N.1 NDVI-RCT plot sugar beet 20-03-2018

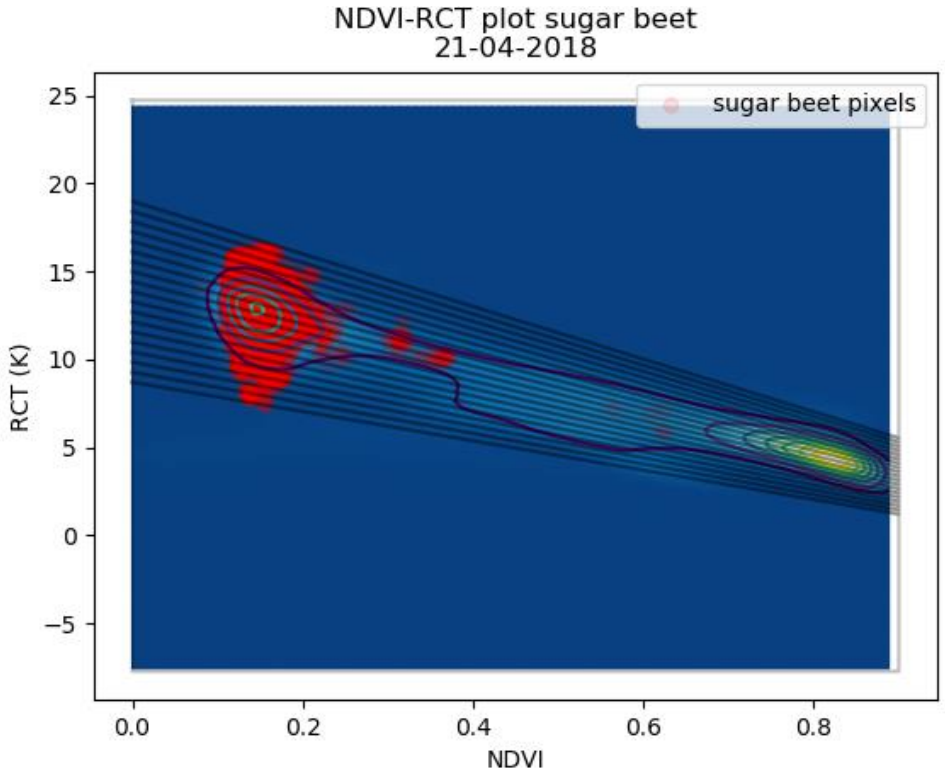


Figure N.2 NDVI-RCT plot sugar beet 21-04-2018

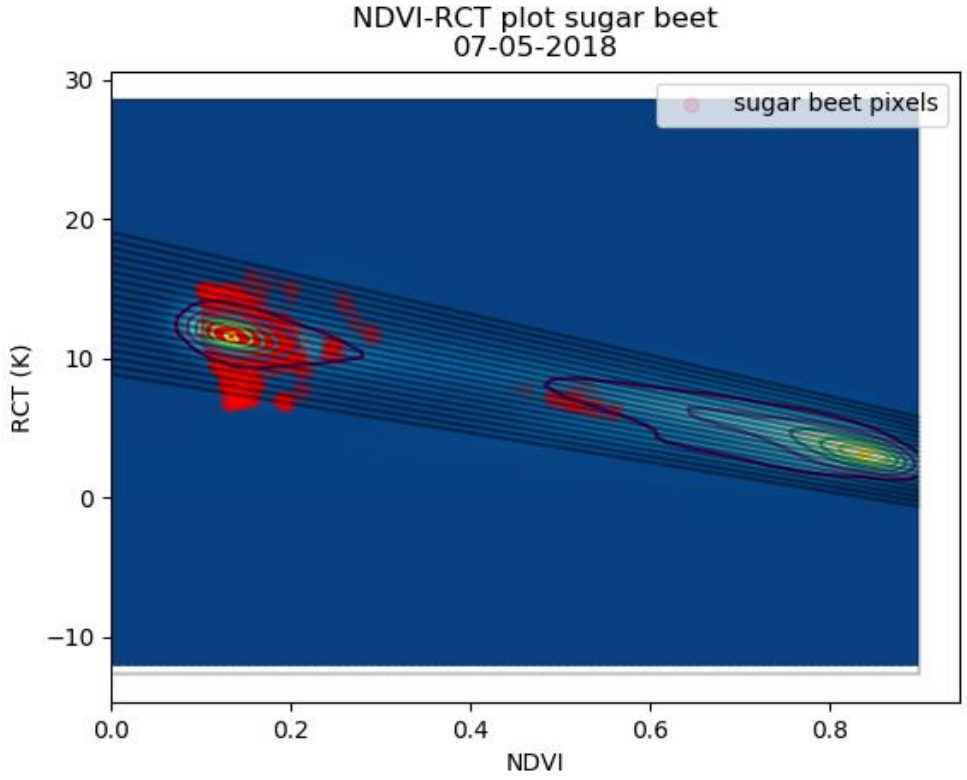


Figure N.3 NDVI-RCT plot sugar beet 07-05-2018

NDVI-RCT plot winter wheat

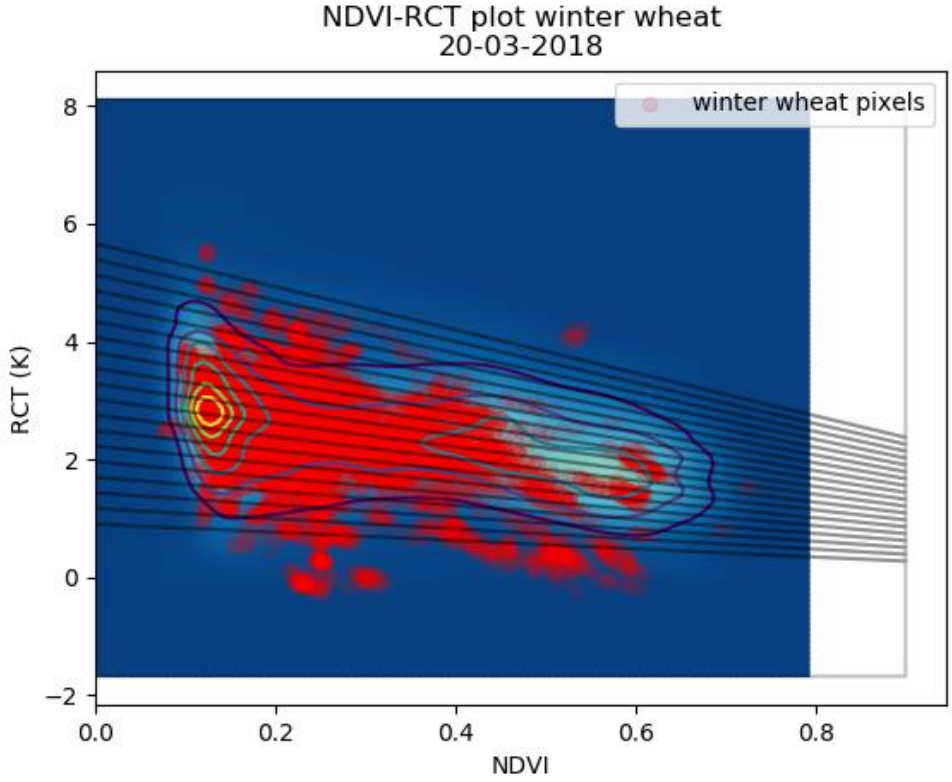


Figure N.4 NDVI-RCT plot winter wheat 20-03-2018

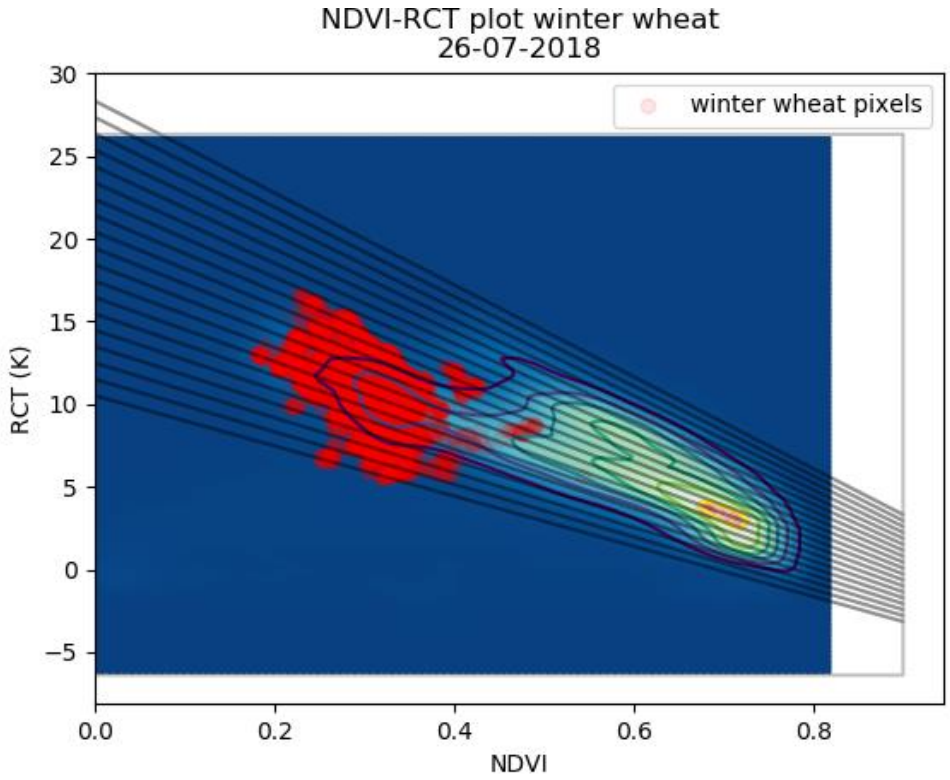


Figure N.5 NDVI-RCT plot winter wheat 26-07-2018