Contents

Notation

c: a (vector of) constant(s)

$\lambda \in \mathbb{R}$: a scalar

$n \in \mathcal{N}$: sample size

i, j are indices in $\{1, \ldots, n\}$

$x \in \mathbb{R}^n$: covariable in 1-dim interpolation setting

$w \in \mathbb{R}^n$: a vector of weights for each location x

$y \in \mathbb{R}^n$: response in 1-dim interpolation setting

$\hat{y} \in \mathbb{R}^n$: estimate of y

$r \in \mathbb{R}^n$: residuals given by $y - \hat{y}$

Pixel: A pixel describes a specific location in a field. It has the size of 10 x 10 meters and coincides with the resolution (and location) of the sentinel-2 pixels. Such pixels are illustrated in figure **??**.

P_t: this describes the observed data (weather and spectral bands) at time t and the location of one pixel.

P: a pixel. We see it as a collection of all the observations at the specified location within one season. More formally, $P := \{P_t | t \text{ is a valid sample time within a defined season}\}$

SCL: scene classification layer. This indicates what one can expect at a pixel at a sampled time. For an overview c.f. table 2.1

P^{SCL45}: similar to P but we only consider observations which belong to the classes 4 and 5. This is used done to get a subset of observations which are less contaminated by clouds and shadows.

NDVI: normalized vegetation difference index

DAS: days after sowing

GDD: growing degree days – cumulative sum of (temperature – threshold)$^+$

Chapter 1

Introduction

1.1 XXX motivation - why is it important

- NDVI-timeseries is very simple and widely used. Examples are: - Plant Models REF - Season Start (start of spring) (community name: land-surface-plant-phenology) -

Since satellite images are “for free” researchers extract

1.2 XXX problembaum / fragestellungen

problem schilderung anhand des Leitfadens: **pictures?**

1.3 XXX State-of-the-art

zusammenfassung mit literaturrecherche hier:

— Doublelogistic (winter-ndvi)

— parametric / non-parametric approaches

— spatio-temporal approaches

1.4 Roadmap

In chapter ...

Chapter 2

Problem Description

2.1 Available Data

XXX field region Witzwil, Data from gregor perich (ref xxx) fields over 5 years cereals (not other cultures)

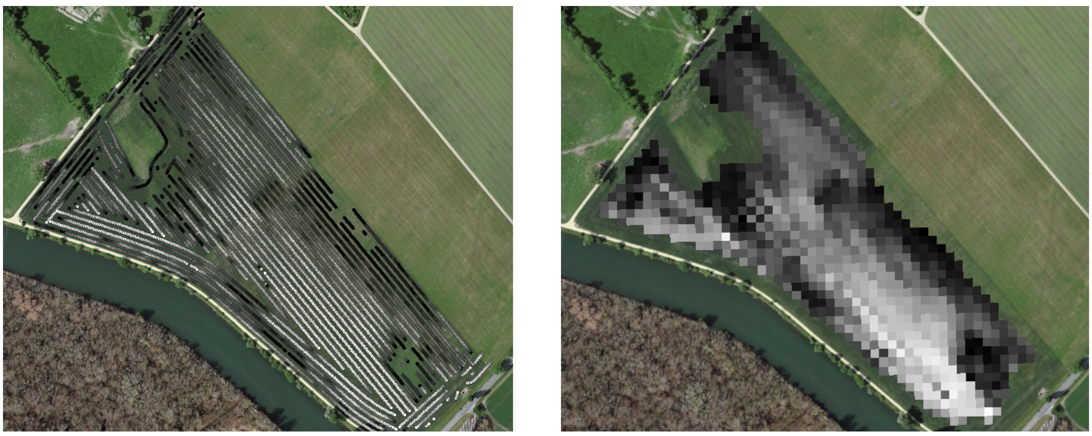
2.1.1 Yieldmapping Data

XXX description of how harvester gets data, knn-interpolation and rsterization, reference to gregors paper

2.1.2 Sentinel 2 Satellite Image Data

General Information

DE: Die ESA [1] revisit times = 5 days am Äquator mittlere breitengrade 2-3 tage XXX CH-grafik von LukasGraf? witzwil befindet



(a) A subfigure XXX (b) A subfigure xxx

Figure 2.1: xxx

Table 2.1: XXX SCL classes

No.	Class	Color
0	No Data (Missing data on projected tiles) (black)	
1	Saturated or defective pixel (red)	
2	Dark features / Shadows (very dark gray)	
3	Cloud shadows (dark brown)	
4	Vegetation (green)	
5	Bare soils / deserts (dark yellow)	
6	Water (dark and bright) (blue)	
7	Cloud low probability (dark gray)	
8	Cloud medium probability (gray)	
9	Cloud high probability (white)	
10	Thin cirrus (very bright blue)	
11	Snow or ice (very bright pink)	

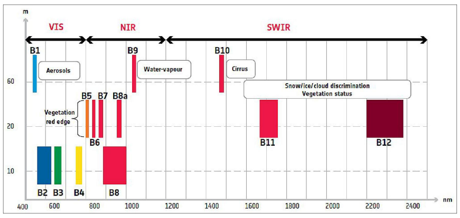


Figure 2.2: XXX Sentinel 2 bands

Data Description

Data Illustration

xxx plot beschreiben

In fig. 2.3

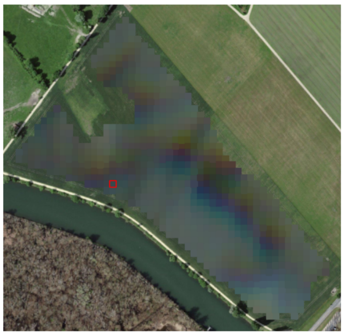
DE: Die Abb. 2.3 zeigt eine Auswahl von 6 Satelitenbildern von einer Parzelle, welche unsere Herrausvorderungen aufzeigen. Im Februar (Bild(a)) sehen wir wie erwartet keine Vegetation, sondern nackte Erde. Anfang Mai beobachten wir ein wolkenfreies dunkelgrünes feld. In (c) wird ersichtlich, dass wir bei starker Bewölkung keine Hoffnung haben nützliche information zu erhalten. Bild (d) zeigt auf, dass die SCL-Klassifizierung nicht zuverlässig ist. In (e) sehen wir ein blasses Grün. Vermutlich sehen wir durch zirrus wolken hindurch.

[1]XXX reference: https://sentinel.esa.int/web/sentinel/missions/sentinel-2

(a) 2021-02-23 scl5

(b) 2021-05-09 scl4

(c) 2021-05-24 scl9



(d) 2021-06-03 scl4

(e) 2021-06-28 scl10

(f) 2021-07-23 scl2

Figure 2.3: Satellite images of a field at selected times with a static background for orientation. The SCL-class of the highlighted pixel is provided in the respective subtitle. (???xxx include scl legend?)

Chapter 3

Interpolation Methods

In this section, we take a closer look at several interpolation methods, which will be used to interpolate and smooth the NDVI time series.

First, we give a brief overview in table 3.1.

Second, we define the general setting and discuss a general approach to make the interpolation more robust (i.e. reduce the impact of outliers).

Later, we introduce and discuss each method.

Then, we try to extract the main ingredients of each method to forge our own one.

Finally, using leave-one-out cross validation, we tune the parameters (where necessary) and get a first idea of the performance of each method.

3.1 Setting

We are given data in the form of (x_i, Y_i) for $i = 1, \ldots, n$). Assume that it can be represented by

$$Y_i = m(x_i) + \varepsilon_i,$$

where ε_i is some noise and $m : \mathbb{R} \to \mathbb{R}$ being some (parametric or non-parametric) function. If we assume that $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. with $\mathbb{E}[\varepsilon_i] = 0$ then

$$m(x) = \mathbb{E}[Y \mid x]$$

Different assumptions on m will lead to the following methods:

3.2 XXX DAS vs GDD

3.3 Robustify

Now we discuss a general approach of how to robustify an interpolation. The main idea is to give less weight to observations which have high residuals after the initial (or if we reiterate, the last) fit.

Table 3.1: A short summary of the studied interpolation methods

	assumtpions	**pros**	**cons**	weights	bounded
Savitzky-Golay filter	– high frequencies are noise (low.pass filter) – equidistant points – local polynomials	– computationally very fast	– cannot deal natively with missing data (need some interpolation)	no	mostly
SG + NDVI	– upper envelope – vegetation cannot grow faster than some slope	– biological knowledge	– bad “upper envelope” since weights are not used for the estimation itselfe	(no)	mostly
Loess	– local polynomial with points closer to the estimated point are more important	– flexible – generalization of SG – weighting function makes intuitive sense	– computationally expensive	yes	mostly
Smoothing-Splines	– 2cd derivative of function is integrable	– intuitive meaning of penalty – general assumptions – flexible shape	– unbounded	yes	no
B-Splines (Smoothed)	– function can be approximated by a linear combination of B-splines basis functions	– general assumtion – flexible shape	– unbounded – no intuitive meaning for smoothing		no
(Gaussian) Kernel Smoothing		– simple – general assumptions	– bandwidh: failes if there are big data-gaps	yes	yes
Double-Logistic	– function first increases then decreases – ndvi has a minimal value	– good for evergreen plants (if snow masks ndvi) – upper envelope	– parameterestimation can go seriously wrong – strange behaviour for long data-gaps	yes	mostly
Universal Kriging	– function is a realization of a stationary gaussian process	– informative parameters – flexible	– regression to the mean – assumptions clearly not met	yes	mostly

Even though the procedure is taken from the robust version of the LOESS smoother (c.f. section 3.5.4 and Cleveland (Cleveland)), we discuss it now because we will apply it also to other interpolation methods.

XXX[1]

Before we describe the procedure, we define a function which will determine the weight given to each observation such that observations with large scaled residuals will have less weight. That is the bisquare function B:

$$B(x) := \begin{cases} \left(1 - x^2\right)^2, & \text{if } |x| < 1 \\ 0, & \text{else} \end{cases}$$

Now, we do something similar to what is done in iteratively reweighted least squares. After an initial interpolation, update the weights of each observation with

$$w_i^{\text{new}} := w_i^{\text{old}} \, \mathrm{B}\left(\frac{|r_i|}{6 \, \mathrm{mad}\,(r_1, \dots, r_n)}\right) \tag{3.3.0.1}$$

where $r_i = y_i - \hat{y}_i$ denotes the residuals. We can iterate this reweighting and stop after several steps or when the change of the values is smaller than some tolerance.

Examples of such iterative fits are illustrated in the figures 3.4 3.5, 3.6, 3.4 and 3.7.

3.3.1 XXX Our Adjustment:

Since we usually observe outliers with negative residuals we decide to divide the negative residuals by two(XXX) before updating the weights. Furthermore, we want to prevent low-weighted observations to corrupt our estimation of scale (the median) and thus we use the weighted median. This can be defined as

$$\mathrm{med}_{\text{weighted}}(r, w) := \operatorname*{arg\,min}_{\lambda \in \mathbb{R}} \sum_{i=1}^{n} |r_i w_i - \lambda|$$

for $r, w \in \mathbb{R}^n$

3.4 Parametric Regression

Parametric Curve estimation tries to fit a parametric function (e.g. a Gaussian function with parameter μ and σ) to a dataset. In the following, we introduce 2 such parametric approaches.

Optimization Issues

Since we aim to minimize the residuals sum of squares over 5 (or 6) parameters, we try to solve a non-convex optimization problem. Thus, the algorithm[2] either struggles to find the global minimum or fails to converge. This was fixed by providing for each parameter reasonable initial values and generous bounds (which match our experience).

[1] Note that due to using the median for the normalization, we gain a breakdown point of 50% for outliers in y.

[2] We used the python function `scipy.optimize.curve_fit`

3.4.1 Double Logistic

The Double Logistic smoothing as described in Beck, Atzberger, Høgda, Johansen, and Skidmore (Beck et al.) heavily relies on shape assumptions of the fitted curve (i.e. the NDVI time series).

Assumptions:

— There is a minimum NDVI level Y_{min} in the winter (e.g. due to evergreen plants), which might be masked by snow. This can be estimated beforehand, taking into several years into account.

— The growth cycle can be divided into an increase and a decrease period, where the time series follows a logistic function. The maximum increase (or decrease) is observed at t_0 (or t_1) with a slope of d_0 (or d_1).

The equation of the double-logistic fit is given by:

$$Y(t) = Y_{\text{min}} + (Y_{\text{max}} - Y_{\text{min}}) \left(\frac{1}{1 + e^{-d_0(t-t_0)}} + \frac{1}{1 + e^{-d_1(t-t_1)}} - 1 \right)$$

Where the five free parameters: Y_{max}, d_0, d_1, t_0, t_1 are initially estimated by least squares. Such fit can be seen in figure 3.1.

Similar as for the Savitzky-Golay Filter (c.f. section 3.5.3) we reestimate (only once) the parameters by giving less weight to the overestimated observations and more weight to the underestimated observations[3].

Pros	**Cons**
— Incorporates subject specific knowledge in the case of evergreen plants covered in snow. — Optimized parameters have an intuitive meaning.	— Strong shape assumptions on the NDVI curve. — Parameter optimization might go wrong. This can be mitigated to some extent to provide bounds for the parameters — Strange behavior in regions with little observations. (cf. figure 3.1)

3.4.2 Fourier Approximation

Similar as in section 3.4.1 we fit a parametric curve to the data by least squares. Here we take the second order Fourier series:

$$\text{NDVI}(t) = \sum_{j=0}^{2} a_j \times \cos(j \times \Phi_t) + b_j \times \sin(j \times \Phi_t)$$

where $\Phi = 2\pi \times (t-1)/n$.

[3] For the details on the weights we refer to Beck, Atzberger, Høgda, Johansen, and Skidmore (Beck et al.)

Pros	Cons
— Assumption of periodicity can be helpful if we are modelling multiyear grow cycles	— Bad behavior in regions with little data (cf. figure 3.1)
— Flexible curve shape	— Hard to interpret estimated parameters
	— Parameter estimation can go wrong. Introducing bounds can help.

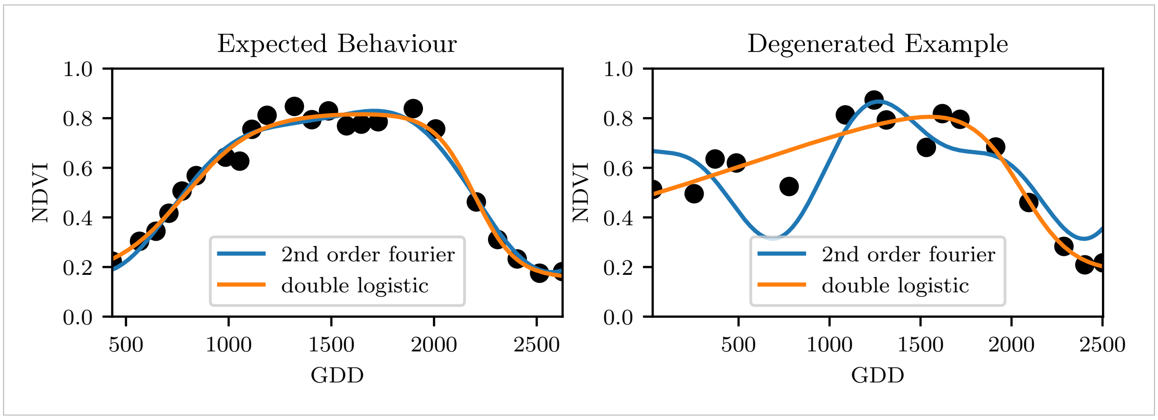


Figure 3.1: Here we observe the nice fitting possibilities of the two parametric methods but notice also some misbehavior

3.5 Non-Parametric Regression

In non-parametric curve estimation, we no longer demand our curve to be fully determined by several parameters, but we allow it to also dependent on the data. That said, we might still use some tuning-parameters sometimes.

3.5.1 Kernel Regression

As described previously, we would like to estimate

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f_X(x)}, \tag{3.5.1.1}$$

where $f_{Y|X}, f_{X,Y}, f_X$ denote the conditional, joint and marginal densities. This can be done with a kernel K:

$$\hat{f}_X(x) = \frac{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)}{nh}, \hat{f}_{X,Y}(x, y) = \frac{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) K\left(\frac{y-Y_i}{h}\right)}{nh^2}$$

By plugging the above into equation 3.5.1.1 we arrive at the *Nadaraya-Watson* kernel estimator:

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left((x - x_i)/h\right) Y_i}{\sum_{i=1}^n K\left((x - x_i)/h\right)}$$

Examples: Normal, Box For local bandwidth selection see Brockmann et al. (1993) XXX

Pros	Cons
— flexible due to different possible kernels — can be assigned degrees of freedom (trace of the hat-matrix) — estimation of the noise variance $\hat{\sigma}_\varepsilon^2$ (XXX c.f. CompStat 3.2.2)	— if the $x \mapsto K(x)$ is not continuous, $\hat{m}$ isn't either — choice of bandwidth, especially if x_i are not equidistant.

3.5.2 Kriging

Kriging was developed in geostatistics to deal with autocorrelation of the response variable at nearby points. By applying the notion that two spectral indices which are (timewise) close should also take similar values, we justify the application of Kriging. In the end, we would like to fit a smooth Gaussian process to the data. For this subsection, we will follow Diggle and Ribeiro (dig).

Definitions and Assumptions

A *Gaussian Process* $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \dots, S(t_k))$ has a multivariate Gaussian distribution for every collection of times $t_1, \dots, t_k$. S can be fully characterized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') = \operatorname{Cov}(S(t), S(t'))$

Assumption: We will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following only $\gamma(h)$.[4]

We also define the variogram of a Gaussian process as

$$V(h) := V(t, t+h) := \frac{1}{2} \operatorname{Var}(S(t) - S(t+h)) = (\gamma(0))^2 (1 - \operatorname{corr}(S(t), S(t+h)))$$

And decide to use a Gaussian Variogram defined by

$$V(h) = p \cdot \left(1 - e^{-\frac{h^2}{\left(\frac{4}{7} r\right)^2}} \right) + n,$$

where h is the distance, n is the nugget, r is the range and p is the partial sill visualized in figure 3.2.[5]

Pros	Cons
— It is a well-studied method. — Parameters have an intuitive meaning. — Flexible covariance structure.	— Regression to the mean. — Violated assumption of constant mean and constant variance. Thus, the NDVI is not a stationary process. — Skewness of errors is not taken into account.

[4] Note that the process is also *isotropic* (i.e. $\gamma(h) = \gamma(\|h\|)$ since we are in a one-dimensional setting and the covariance is symmetric.

[5] Strictly speaking we use a scaled version of the variogram. Thus, only the ratio of p/n matters.

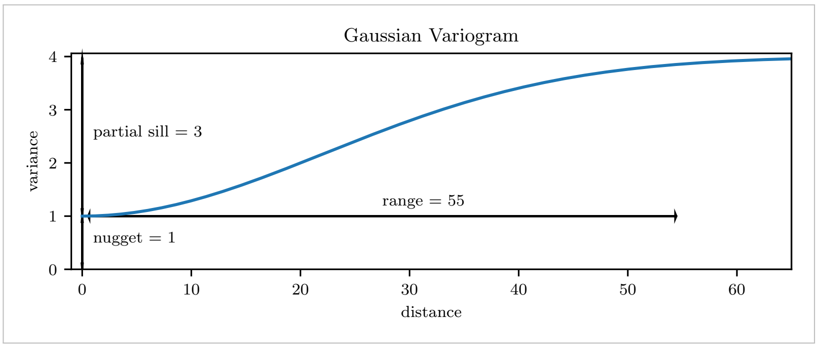


Figure 3.2: Gaussian Variogram with nugget=1, partial sill=3, range=55

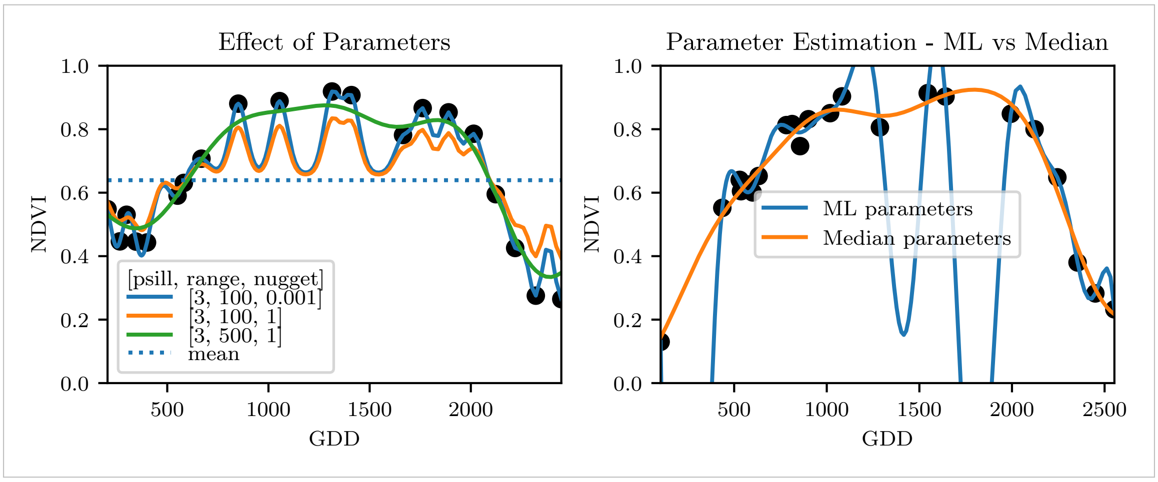


Figure 3.3: On the left, we see how the interpolation change if we increase the nugget and the range parameter. On the right we compare two kriging interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such numerical optimizations.

194 ### 3.5.3 Savitzky-Golay Filter (SG Filter)

The *Savitzky-Golay Filter*, introduced in Savitzky and Golay (Savitzky and Golay) is a technique in signal processing and can be used to filter out high frequencies (low-pass filter) as argued in Schafer (Schafer). Furthermore, it also can be used for smoothing by filtering high frequency noise while keeping the low frequency signal. First, we choose a window size m. Then, for each point, $j \in \{m, m+1, \ldots, n-m\}$ we fit a polynomial of degree k by:

$$\hat{y}_j = \min_{p \in P_k} \sum_{i=-m}^{m} (p(x_{j+i}) - y_{i+j})^2,$$

195 where P_k denotes the Polynomials of degree k over $\mathbb{R}$.

For equidistant points this can efficiently be calculated by

$$\hat{y}_j = \sum_{i=-m}^{m} c_i y_{j+i},$$

196 where the c_i are only dependent on the m and k and are tabulated in the original paper.

Adaptation to the NDVI

In a rather famous paper Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh (Chen et al.) a "robust" method based on the Savitzky-Golay has been used. The method is based on the assumption that due to atmospheric effects the observed NDVI tends to be underestimated and that it cannot increase too quickly[6].

Algorithm:

i.) Remove points which are labeled as cloudy.

ii.) Remove points which would indicate an increase greater than 0.4 within 20 days.

iii.) Linearly interpolate to obtain an equidistant time series X^0.

iv.) Apply the Savitzky-Golay Filter to obtain a new time series X^1.

v.) Update X^1 by applying again a Savitzky-Golay Filter. Repeat this until $w^T|X^1 - X^0|$ stops decreasing, where w is a weight vector with $w_i = \min\left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|}\right)$. This reduces the penalty introduced by outliers[7] and by repeating this step we approach the "upper NDVI envelope".

Pros	Cons
— Popular technique in signal processing. — Efficient calculation for equidistant points. — Upper envelope matches intuition for the NDVI. Therefore, it is robust against outliers with small values.	— No natural way of how to estimate points which are not in the data. — Not generalizable to other spectral indices. — Linear interpolation to account for missing data might be not appropriate. — No smooth interpolation between two measurements.

Extension: Spatial-Temporal-Savitzky-Golay Filter

One notable adaptation of the Savitzky-Golay is the presented by Cao, Chen, Shen, Chen, Zhou, Wang, and Yang (Cao et al.). The key difference is the additional assumption of the cloud cover being discontinuous and that we can improve by looking at adjacent pixels[8]. Because we are working with rather high resolution satellite data, and we need the variance in the predictors, we will waive this extension.

3.5.4 Locally Weighted Regression (LOESS)

Introduced by : Cleveland (Cleveland) implemented here Cappellari, McDermid, Alatalo, Blitz, Bois, Bournaud, Bureau, Crocker, Davies, Davis, de Zeeuw, Duc, Emsellem, Khochfar, Krajnović, Kuntschner, Morganti, Naab, Oosterloo, Sarzi, Scott, Serra, Weijmans, and Young (Cappellari et al.)

[6] The latter is argued by the biological impossibility of such fast vegetation changes

[7] Here we call a point i an outlier if $X_i^0 < X_i^1$.

[8] Here, we say that a pixel is adjacent if it is the same pixel but from a different year (keeping the same day of the year) or (if not enough of such temporal-adjacent pixel are found) it is spatially adjacent

The Locally Weighted Regression (LOESS) can be understood as a generalization of the Savitzky-Golay Filter (c.f. sec. 3.5.3).

Given a proportion $\alpha \in (0,1]$, we estimate each y_i separately by fitting a polynomial of order d by weighted least squares. The weights are (usually) defined by

$$w_i(x_j) = \begin{cases} \left(1 - \left(\frac{x_j}{h_i}\right)^3\right)^3, & \text{for } |x_j| < h_i \\ 0, & \text{for } |x_j| \geqslant h_i \end{cases},$$

where h_i is the minimal distance such that $\lceil \alpha n \rceil$ observations are in the ball $B_{h_i}(x_i)$.[9] So for each y_i we only consider a proportion α of the observations.

How does the Robust LOESS differ from the SG Filter?

The LOESS smoother takes a fraction of points instead of a fixed number and therefore automatically adapts to the size of the data we wish to interpolate. However, we run into the danger of considering too little observations, since the estimation breaks down if $\lceil \alpha n \rceil < d+1$. Furthermore, LOESS gives less weight to points further away. This yields a "smoother" estimate, since when we slide the window (e.g. for estimating the next value) an influential point at the border does not suddenly get zero weight from being weighted equally before. Finally, the LOESS also can be used for non-equidistant data and allows for arbitrary interpolation.

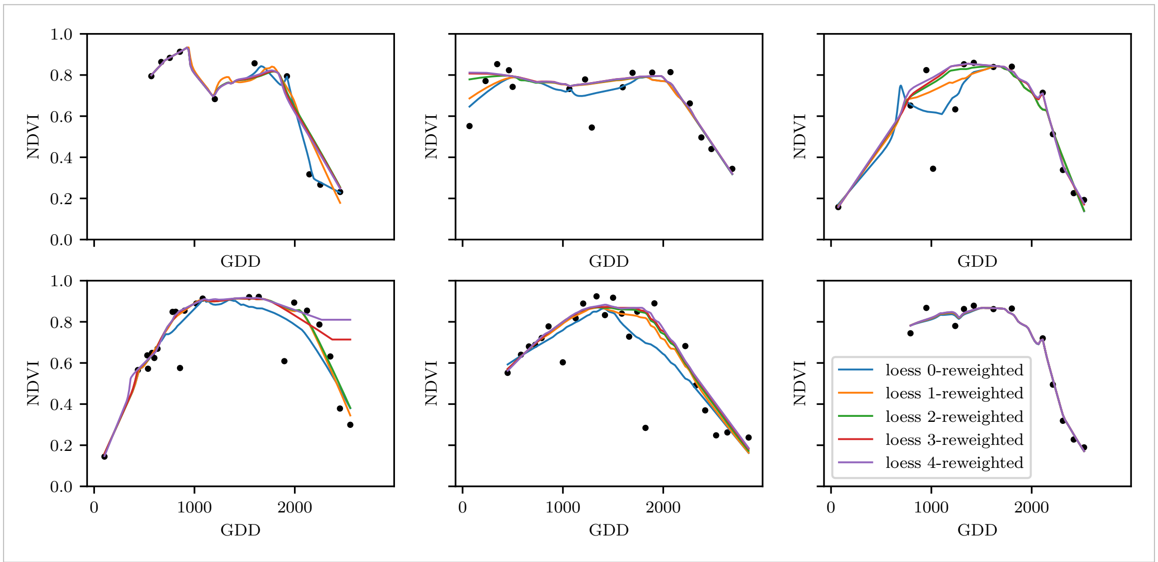


Figure 3.4: The LOESS smoother fitted to different (SCL45) NDVI time series. Iterations of a robustifing refit (as indicated in section 3.3) are also displayed

[9]If too many weights are set to zero, we might end up considering not enough observations and thus get a singular design-matrix (for the least squares estimation). Therefore, we substitute h_i with $1.01h_i$, so that the observation on the boundary of $B_{h_i}(x_i)$ does not get completely ignored.

Pros	Cons
— Flexible generalization of Savitzky-Golay — arbitrary interpolation possible — Intuitive parameters	— The nature of local regression might lead to surprising estimates (no smoothness guarantees for the second derivative) — Multiple XXXXXXx

3.5.5 B-splines

from Lyche and Morken (Lyche and Morken)

$$S(x) = \sum_{j=0}^{n-1} c_j B_{j,k;t}(x)$$

$$B_{i,0}(x) = 1, \text{ if } t_i \leq x < t_{i+1}, \text{ otherwise } 0$$
$$B_{i,k}(x) = \tfrac{x-t_i}{t_{i+k}-t_i} B_{i,k-1}(x) + \tfrac{t_{i+k+1}-x}{t_{i+k+1}-t_{i+1}} B_{i+1,k-1}(x)$$

Smoothing: We can relax the constraint that we have to perfectly interpolate. Thus, we use the minimum number of knots[10] such that: $\sum_{i=1}^{n}(w(y_i - \hat{y}_i))^2 \leq s$

Pros	Cons
— can be assigned degrees of freedom — extendable to ”smooth” version — performs also well if points are not equidistant	— smoothing process does not translate well to a interpretation (unlike smoothing splines) — choice of smoothing parameter s

3.5.6 Natural Smoothing Splines

Let $\mathcal{F}$ be the Sobolev space (the space of functions of which the second derivative is integrable). Then the unique[11] minimizer

$$\hat{m} := \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (Y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

is a natural[12] cubic spline (i.e. a piecewise cubic polynomial function). The objective function has an intuitive meaning, as to avoid lateral acceleration it is desirable to move the steering wheel as little as possible, when driving a car.

3.5.7 XXX Whittaker Smoother

XXX

[10] SciPy uses FITPACK and DFITPACK, the documentation suggests that smoothness is achieved by reducing the number knots used

[11] Strictly speaking it is only unique for $\lambda > 0$

[12] It is called natural since it is affine outside the data range ($\forall x \notin [x_1, x_n] : \hat{m}''(x) = 0$)

Pros	Cons
— can be assigned degrees of freedom (trace of the hat-matrix)	— choose λ
— efficient estimation (closed form solution)	
— intuitive penalty (we don't want the function to be too "wobbly" — change slopes)	
— performs also well if points are not equidistant	
— fixes the Runge's phenomenon (fluctuation of high degree polynomial interpolation)	

3.6 Tuning parameter estimation

lots of cross validation

what is the best? RMSE is bad, since we know that outliers are present optimizing w.r.t different statistics

?plot with different denisties for each statistic

3.7 Robustification – Recap

introduced in section **??** we want to review it

robustifieng from loess -> lets try it with all. Result in figures ...

issues when reiterating often (we lose some points completely)

from pictures ... we get that one

3.7.1 Upper Envelope Approach - Penalty for negative resiudals

discussion of idea, and explenation why we did no use it (arbitrary choice)

3.8 Performance Assecement

TEMP — Figures

Table 3.2: Performance comparison of different interpolation methods measured with various statistics. Considering only SCL45 points, we get the out-of-bag estimates using the given interpolation method. Consequently, we compute the absolute (value of the) residuals and apply the given statistic to it.

	ss	loess	dl	bspl	fourier	ss rob	loess rob	dl rob	bspl rob	fourier rob
rmse	0.063	0.061	0.061	0.074	0.075	0.070	0.065	0.065	0.079	0.208
qtile50	0.036	0.034	0.027	0.043	0.031	0.032	0.031	0.022	0.037	0.049
qtile75	0.063	0.061	0.051	0.077	0.058	0.061	0.057	0.044	0.070	0.099
qtile85	0.080	0.079	0.070	0.098	0.083	0.081	0.076	0.063	0.094	0.158
qtile90	0.092	0.092	0.088	0.112	0.108	0.097	0.090	0.082	0.113	0.226
qtile95	0.119	0.115	0.122	0.142	0.161	0.132	0.115	0.124	0.157	0.375

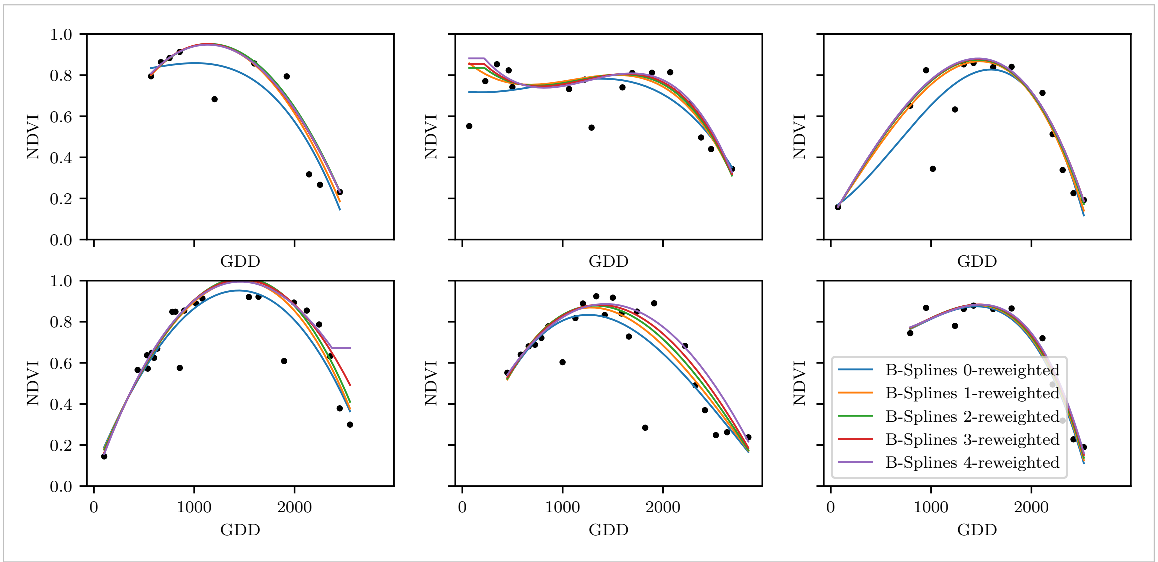


Figure 3.5: B-Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifing refit (as indicated in section 3.3) are also displayed

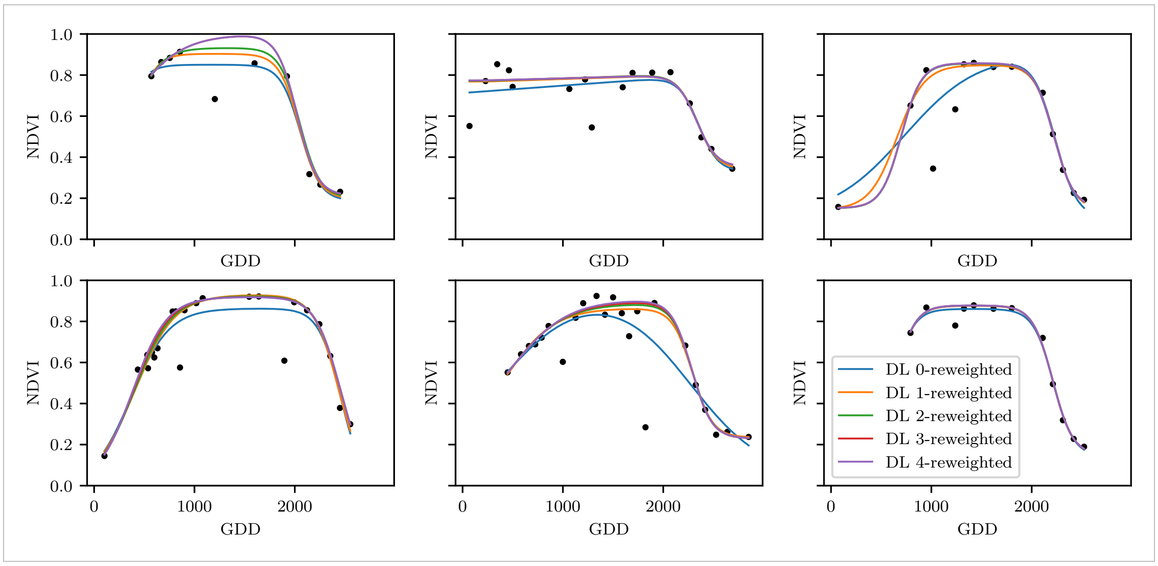


Figure 3.6: A Double Logistic curve fitted to different (SCL45) NDVI time series. Iterations of a robustifing refit (as indicated in section 3.3) are also displayed

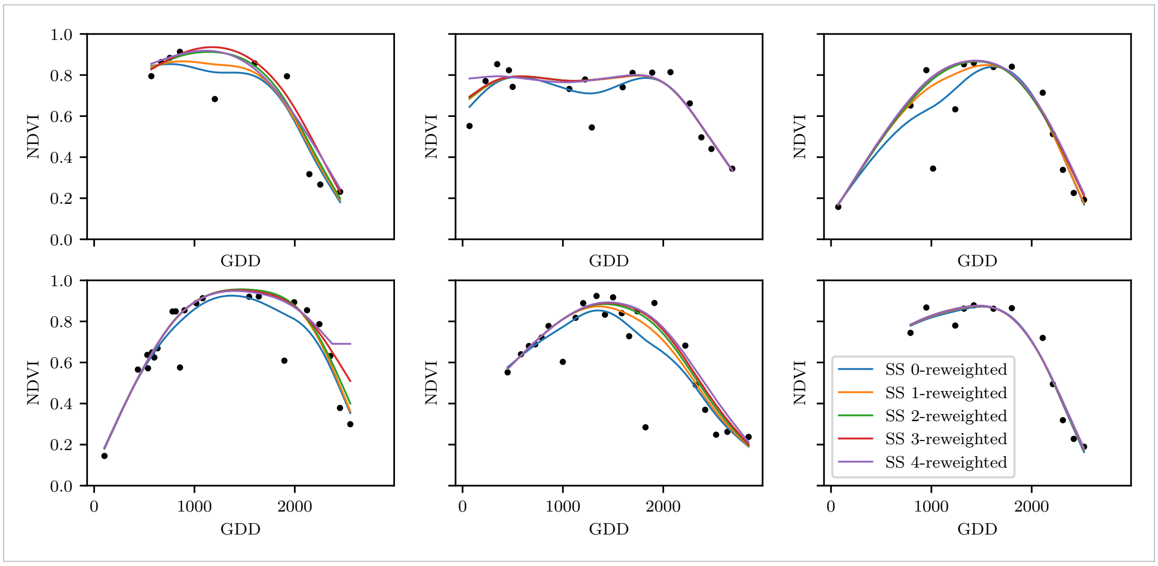


Figure 3.7: Smoothing Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifing refit (as indicated in section 3.3) are also displayed

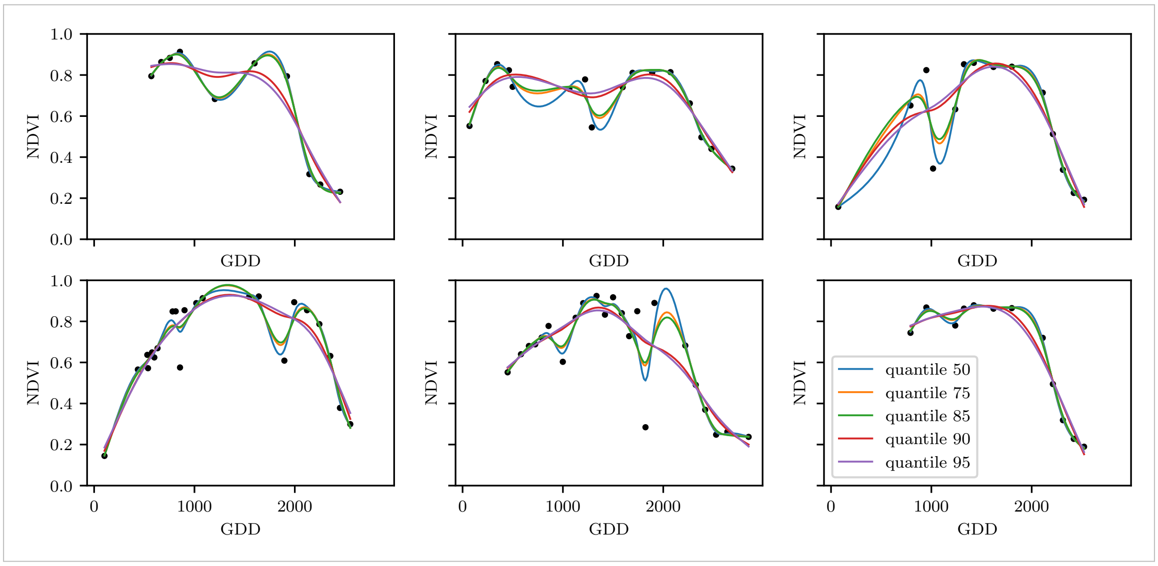


Figure 3.8: Smoothing splines fit with smoothing parameter optimized by minimizing the "…"-quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

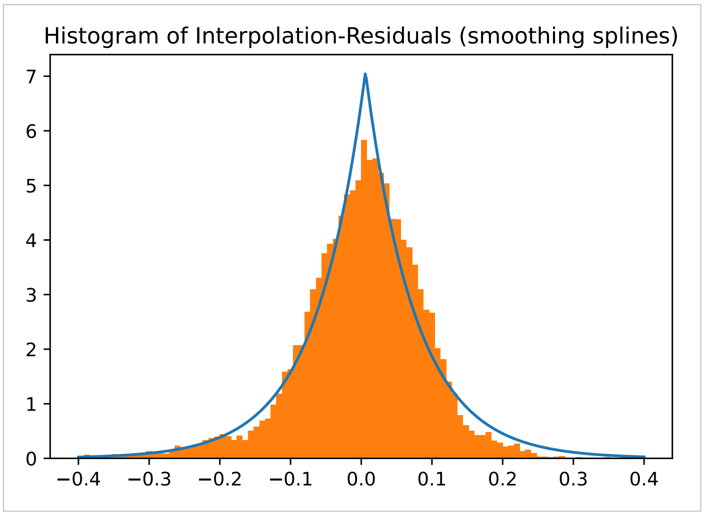


Figure 3.9: XXX caption XXX

Chapter 4

NDVI Correction / Improve NDVI Data

Let's remind ourselves that the data from the Sentinel-2 is equipped with a scene classification layer (SCL) and we therefore have some information of what is observed at each pixel for each sampled time (c.f. table 2.1). In this chapter we would like to improve the observed NDVI values by using more information than just the two bands used to calculate the NDVI (B4 and B8).

4.1 Considering other SCL Classes

In figure 4.1 we see for example that some blue points[1] follow the interpolated line closely and that they might be useful in improving an interpolation fit.

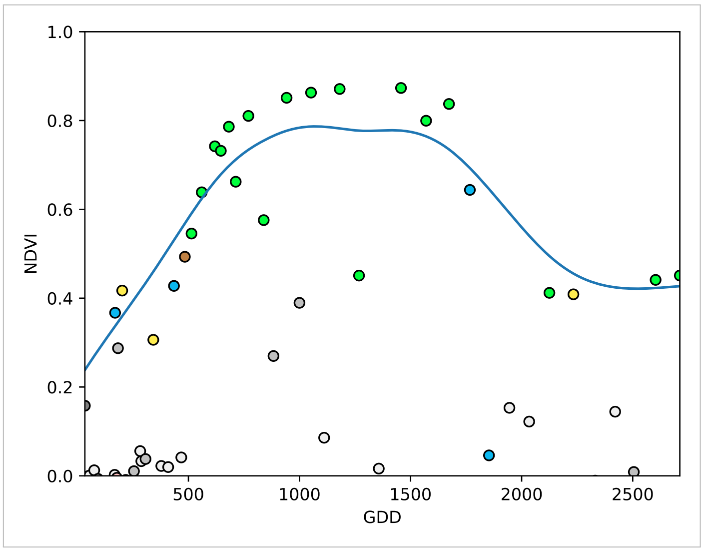


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45)

To get an impression whether there is some useful information contained in the remaining SCL-classes (all except 4 and 5) we would like to compare the observed NDVI with the

[1] The blue points correspond to the SCL-class 10: Thin cirrus clouds

true NDVI. But since we do not have any ground truth data, we will make the following assumption:

Definition 4.1.0.1. *XXXAssumption (true NDVI) The true NDVI value at time t can be successfully estimated by out-of-bag interpolation using high quality observations. That is the interpolated value (using XXX) considering the points $P^{SCL45} \setminus P_t$. In the following, we will call this estimate the "true"-NDVI*

shall pair every observed NDVI value with its out-of-bag-estimate. Then for each category we collect all pairs and create a scatter plot in fig 4.2XXXXXXXXXXXXXX

i.) For each pixel and for each observation (every SCL-class):
estimate the NDVI value (via out-of-the-box interpolation[2])

ii.)

4.2 XXX Correction

roadmap ... (intuition, data-table, ml-methods, uncertainty, refit and evaluation)

4.2.1 XXX idea -and- stepwise plots

4.2.2 XXX data-table-construction

XXX discussion about choosen covariates: list of things we considered but rejected + reasoning –> no weather to keep it general even though we have it implemeted

4.2.3 XXX ml-methods

4.2.4 XXX Uncertainty

abs(residuals), train models for uncertainty, estimate residuals, get weights (via weight-function) (problem of weight function -> we should norm the weights somehow since smoothing parameters are "dependent" on weights -> then, some outer points get really low weights (just because others in the middle have very little residuals and thus very high weight))

4.3 XXX Evaluation Method

yield estimation is a main goal. Claim that yield-estimation-accuracy is a objective measure : - we have not looked at the yield so far - if the one NDVI-time-series predicts the yield better than a different one, we conclude that the first time-series carries more true information about the plants Now: "yield NDVI-TS / derived-covariates"

4.3.1 yield estimation

problem: high dimensionality and unequal duration/length -> use features

name approaches for yield estimation (we will use a simple but flexible one)

random forest ■ for evaluation out-of-bag estimates

[2] That is, we use all observations (in SCL45) but the current one.

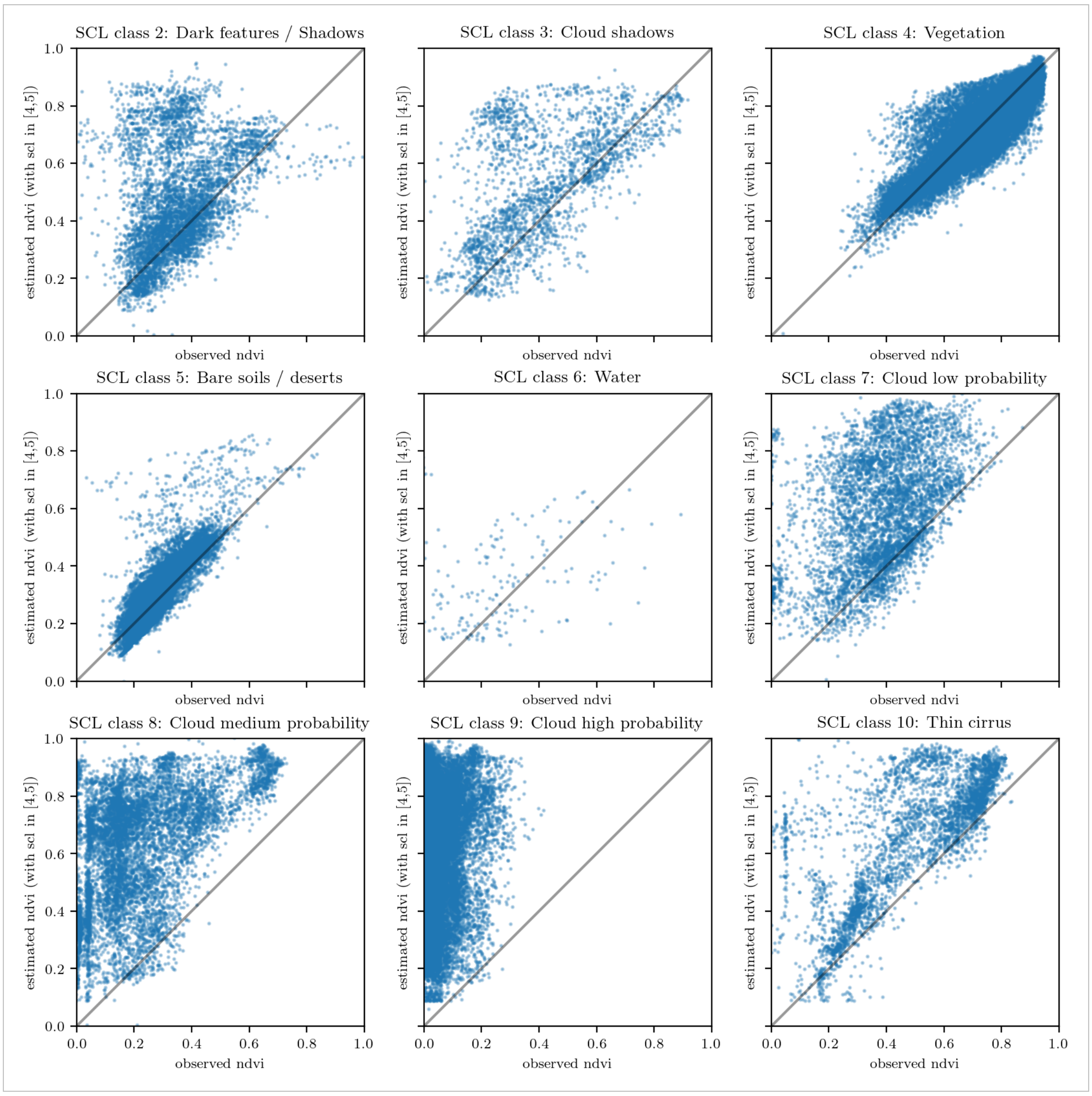


Figure 4.2: XXX caption XXX

Covariates used

reference to kamir et al, why we did choosed some and not others

Chapter 5

Results

5.1 XXX small recap from "Interpolation Methods"

5.2 Robustification and NDVI-Correction

Table 5.1: XXX RMSE of yield prediction

	rf	lm-scl	lm-all	mars	gam	lasso	no-correction
ss	1.999	1.872	1.829	2.055	2.047	2.033	1.941
dl	1.873	1.886	1.896	1.988	1.898	1.833	2.018
ss-rob	1.895	2.010	2.037	1.970	1.874	1.928	1.880
dl-rob	1.865	1.884	2.002	1.996	1.808	1.875	2.005

Chapter 6

Discussion

High RMSE in ...: How much can we expect to get? We have multiple sources of uncertainty in the data: 1. Uncertainty in Yield data collected by the combine harverster 2. Uncertainty in Yield data through rasterization 3. Uncertainty in satellite images through "measurement errors" introduced via clouds and other atmospheric effects 4. Uncertainty introduced by interpolating (especially when long data-gaps are present)

Chapter 7

Outlook

7.1 Data

- Method how data has been extrapolated to the grid could possibly be improved
- For computational reasons we mostly considered all years and split the data (on the pixel level) randomly into a train/test set. A cross Validation with leaving one year out would be

7.2 Interpolation

- Penalized Regressions as described in ... are similar to smoothing splines (c.f. ...) but different. Better?

7.3 NDVI Correction

- try different link functions in section ... between estimated absolute resiudals and weights

7.4 NDVI Correction + +

- NDVI Correction can be applied to all sorts of land observed via. satellites (without the need of ground truth data)
- The idea of NDVI Correction could be applied to other spectral indices like the Green Leaf Area Index.
- Yield is not the only target variable of interest. Other variables like protein content could also be used in section ... for the method evaluation.

Bibliography

Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.), *Model-Based Geostatistics*, pp. 46–78. Springer.

Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *100*(3), 321–334.

Cao, R., Y. Chen, M. Shen, J. Chen, J. Zhou, C. Wang, and W. Yang. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *217*, 244–257.

Cappellari, M., R. M. McDermid, K. Alatalo, L. Blitz, M. Bois, F. Bournaud, M. Bureau, A. F. Crocker, R. L. Davies, T. A. Davis, P. T. de Zeeuw, P.-A. Duc, E. Emsellem, S. Khochfar, D. Krajnović, H. Kuntschner, R. Morganti, T. Naab, T. Oosterloo, M. Sarzi, N. Scott, P. Serra, A.-M. Weijmans, and L. M. Young. The ATLAS3D project - XX. Mass-size and mass-sigma distributions of early-type galaxies: Bulge fraction drives kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function. *432*, 1862–1893.

Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *91*, 332–344.

Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *74*(368), 829–836.

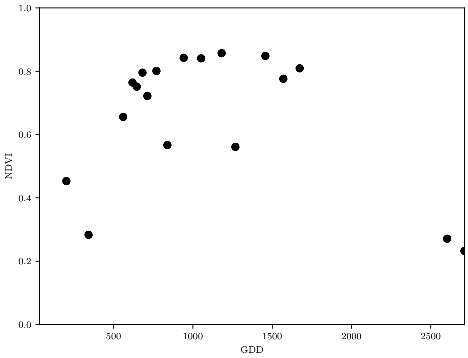
Lyche, T. and K. Morken. Spline Methods.

Savitzky, A. and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *36*(8), 1627–1639.

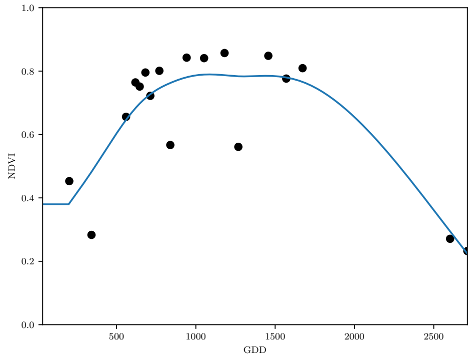
Schafer, R. W. What Is a Savitzky-Golay Filter? [Lecture Notes]. *28*(4), 111–117.

Appendix A

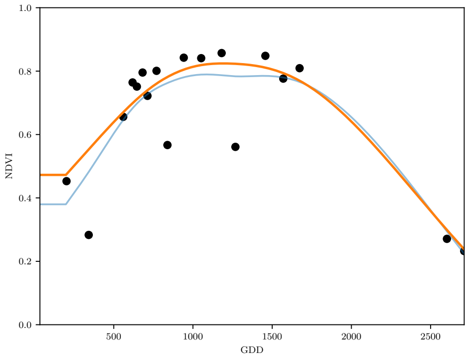
XXX Appendix



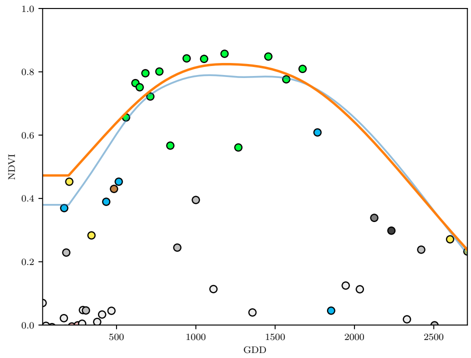
(a) NDVI time series with scl45



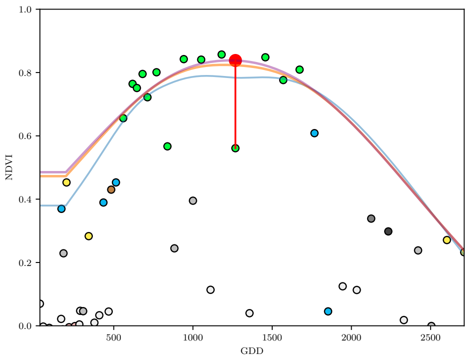
(b) Interpolation via Smoothing Splines



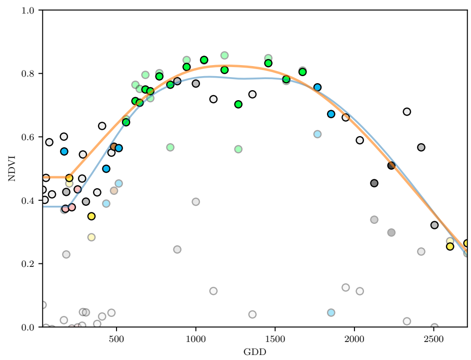
(c) Robustly reweighted fit



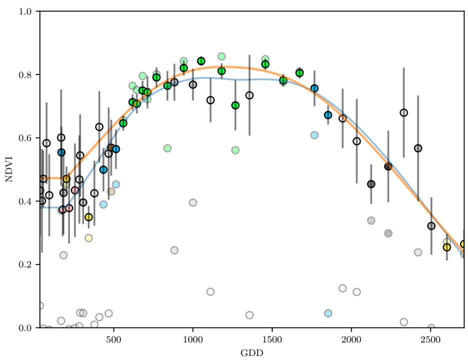
(d) Now also consider other SCL-classes



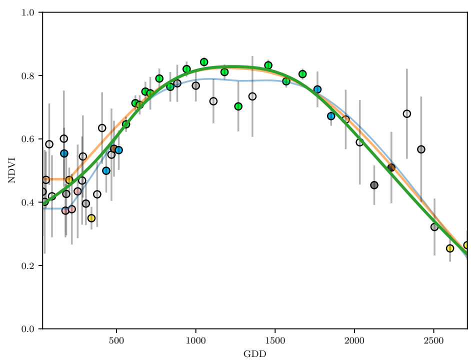
(e) OOB estim. for each point using SCL45



(f) Correct NDVI



(g) Estimate absolute errors



(h) Robust interpolation on weights derived from uncertainties.

Figure A.1: Stepwise illustration of robust NDVI-Correction