

1 **Contents**

2	Notation	ii
3	1 Introduction	1
4	1.1 XXX motivation - why is it important	1
5	1.2 XXX problembaum / fragestellungen	1
6	1.3 XXX State-of-the-art	1
7	1.4 Roadmap	1
8	2 Problem Description	2
9	2.1 Available Data	2
10	2.1.1 Sentinel 2 Satellite Image Data	2
11	2.1.2 Yieldmapping Data	4
12	3 Interpolation Methods	6
13	3.1 Setting	6
14	3.2 XXX DAS vs GDD	6
15	3.3 Robustify	6
16	3.3.1 XXX Our Adjustment:	8
17	3.4 Parametric Regression	8
18	3.4.1 Double Logistic	8
19	3.4.2 Fourier Approximation	9
20	3.5 Non-Parametric Regression	9
21	3.5.1 Kernel Regression	10
22	3.5.2 Kriging	10
23	3.5.3 Savitzky-Golay Filter (SG Filter)	11
24	3.5.4 Locally Weighted Regression (LOESS)	13
25	3.5.5 B-splines	14
26	3.5.6 Natural Smoothing Splines	15
27	3.5.7 XXX Whittaker Smoother	15
28	3.6 Tuning parameter estimation	15
29	3.7 Robustification – Recap	16
30	3.7.1 Upper Envelope Approach - Penalty for negative resiudals	16
31	3.8 Performance Assecement	16
32	4 NDVI Correction / Improve NDVI Data	19
33	4.1 Considering other SCL Classes	19
34	4.2 XXX Correction	20
35	4.2.1 XXX idea -and- stepwise plots	20
36	4.2.2 XXX data-table-construction	20
37	4.2.3 XXX ml-methods	20
38	4.2.4 XXX Uncertainty	20
39	4.3 XXX Evaluation Method	20
40	4.3.1 yield estimation	20
41	5 Results	22
42	5.1 XXX small recap from “Interpolation Methods”	22
43	5.2 Robustification and NDVI-Correction	22

44	6 Discussion	23
45	7 Outlook	24
46	7.1 Data	24
47	7.2 Interpolation	24
48	7.3 NDVI Correction	24
49	7.4 NDVI Correction + +	24
50	Bibliography	25
51	A XXX Appendix	26

52 Notation

- 53 c : a (vector of) constant(s)
- 54 $\lambda \in \mathbb{R}$: a scalar
- 55 $n \in \mathcal{N}$: sample size
- 56 i, j are indices in $\{1, \dots, n\}$
- 57 $x \in \mathbb{R}^n$: covariate in 1-dim interpolation setting
- 58 $w \in \mathbb{R}^n$: a vector of weights for each location x
- 59 $y \in \mathbb{R}^n$: response in 1-dim interpolation setting
- 60 $\hat{y} \in \mathbb{R}^n$: estimate of y
- 61 $r \in \mathbb{R}^n$: residuals given by $y - \hat{y}$
- 62 Pixel: A pixel describes a specific location in a field. It has the size of 10 x 10 meters
63 and coincides with the resolution (and location) of the sentinel-2 pixels. Such pixels are
64 illustrated in figure ??.
- 65 P_t : this describes the observed data (weather and spectral bands) at time t and the location
66 of one pixel.
- 67 P : a pixel. We see it as a collection of all the observations at the specified location within
68 one season. More formally, $P := \{P_t | t \text{ is a valid sample time within a defined season}\}$
- 69 SCL: scene classification layer. This indicates what one can expect at a pixel at a sampled
70 time. For an overview c.f. table 2.2
- 71 P^{SCL45} : similar to P but we only consider observations which belong to the classes 4 and
72 5. This is used done to get a subset of observations which are less contaminated by clouds
73 and shadows.
- 74 NDVI: normalized vegetation difference index
- 75 DAS: days after sowing
- 76 GDD: growing degree days – cumulative sum of (temperature – threshold)⁺

77 **Chapter 1**

78 **Introduction**

79 **1.1 XXX motivation - why is it important**

- 80 - NDVI-timeseries is very simple and widely used. Examples are: - Plant Models REF -
81 Season Start (start of spring) (community name: land-surface-plant-phenology) -
82 Since satellite images are “for free” researchers extract

83 **1.2 XXX problebaum / fragestellungen**

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

85 **1.3 XXX State-of-the-art**

- 86 zusammenfassung mit literaturrecherche hier:
87 — Doublelogistic (winter-ndvi)
88 — parametric / non-parametric approaches
89 — spatio-temporal approaches

90 **1.4 Roadmap**

91 In chapter

92 **Chapter 2**

93 **Problem Description**

94 **2.1 Available Data**

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

97 **2.1.1 Sentinel 2 Satellite Image Data**

98 **General Information**

99 The European Space Agency (ESA)¹ freely distributes the high quality images of the two
100 Sentinel satellites 2 (S2). Together, both satellites have a revisit time of 5 days at the
101 equator and 2-3 at mid-latitudes. However, at our study region we only receive an image
102 every 5 days. In order to decrease the effect of atmospheric conditions like reflections
103 and scattering, we will not work with the raw data but with the results of the Level-2A
104 processing²³.

105 **Data Description**

106 The Level-2A processed images we use contain 12 spectral bands with local resolutions up
107 to 10 meters (see 2.1). Bands which have a lower resolution (20 and 60 meters) will be
108 scaled up to 10 meters using cubic interpolation (REF gregor perich). Additional to the
109 spectral bands the ESA also supplies a **Scene Classification Layer (SCL)** where for each
110 location the observed subject is assigned to an *SCL-class* (c.f. table 2.2). In chapter 3 we
111 will use this classification to filter out unreliable data points considering only SCL-classes
112 4 and 5.

113 **Data Illustration**

114 xxx plot beschreiben

115 In fig. 2.1

¹REF: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

²REF <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithms>

³XXXREF gregor perich “Data prior to March 2018 was only available in the top-of-atmosphere L1C format and was downloaded as such [...] L1C data was processed to L2A product level using the ‘Sen2Cor’ processor provided by ESA”

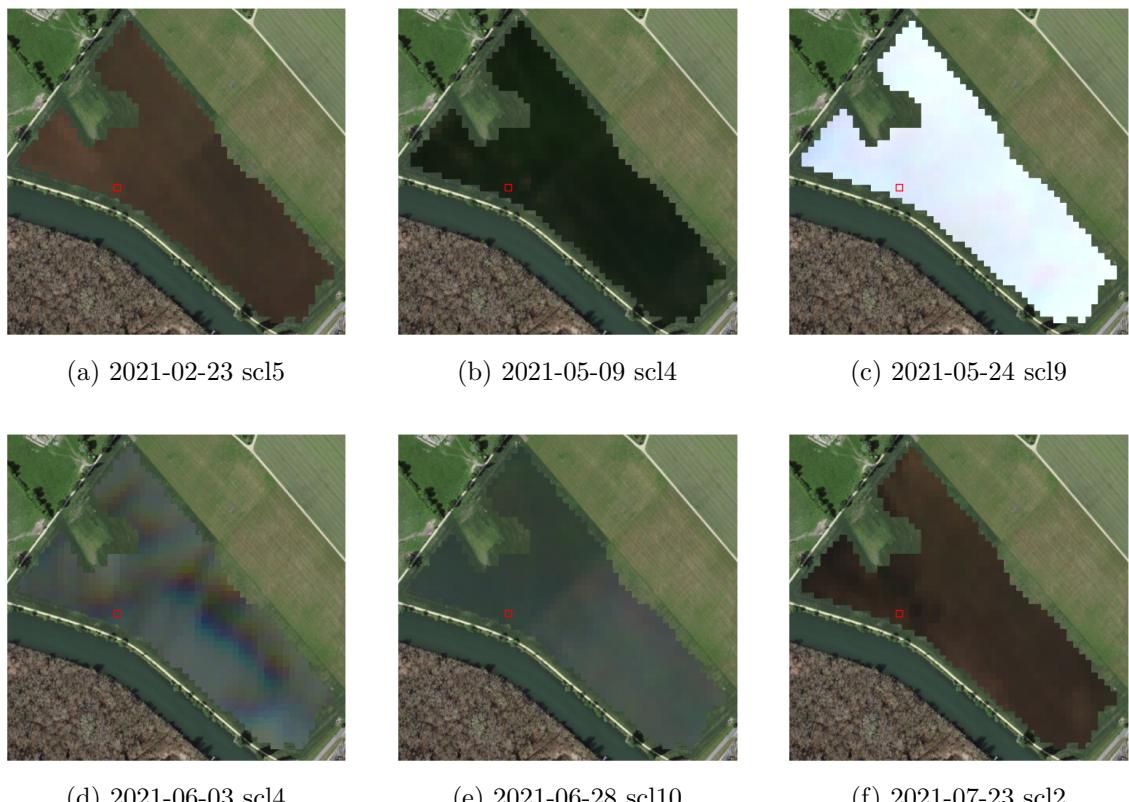


Figure 2.1: Satellite images of a field at selected times with a static background for orientation. The SCL-class of the highlighted pixel is provided in the respective subtitle. (???xxx include scl legend?)

Table 2.1: Jaramaz, Perović, Belanovic Simic, Saljnikov, Cakmak, Mrvić, and Zivotic (Jaramaz et al.) List of spectral bands of the S2-satellites. Each band has its center at the wavelength λ in nm with the spectral width $\Delta\lambda$ in nm with a spatial resolution SR in m.

Band	λ	$\Delta\lambda$	SR	Purpose
1	443	20	60	Atmospheric correction (aerosol scattering)
2	490	65	10	Sensitive to vegetation senescing, carotenoid, browning and soil background; atmospheric correction (aerosol scattering)
3	560	35	10	Green peak, sensitive to total chlorophyll in vegetation
4	665	30	10	Maximum chlorophyll absorption
5	705	15	20	Position of red edge; consolidation of atmospheric corrections / fluorescence baseline.
6	740	15	20	Position of red edge, atmospheric correction, retrieval of aerosol load.
7	783	20	20	Leaf Area Index (LAI), edge of the Near-Infrared (NIR) plateau.
8	842	115	10	LAI
8a	865	20	20	NIR plateau, sensitive to total chlorophyll, biomass, LAI and protein; water vapor absorption reference; retrieval of aerosol load and type.
9	945	20	60	Water vapor absorption, atmospheric correction.
10	1375	30	60	Detection of thin cirrus for atmospheric correction.
11	1610	90	20	Sensitive to lignin, starch and forest above ground biomass. Snow/ice/-cloud separation.
12	2190	180	20	Assessment of Mediterranean vegetation conditions. Distinction of clay soils for the monitoring of soil erosion. Distinction between live biomass, dead biomass and soil, e.g. for burn scars mapping.

116 DE: Die Abb. 2.1 zeigt eine Auswahl von 6 Satelitenbildern von einer Parzelle, welche
 117 unsere Herausforderungen aufzeigen. Im Februar (Bild(a)) sehen wir wie erwartet keine
 118 Vegetation, sondern nackte Erde. Anfang Mai beobachten wir ein wolkenfreies dunkel-
 119 grünes feld. In (c) wird ersichtlich, dass wir bei starker Bewölkung keine Hoffnung haben
 120 nützliche information zu erhalten. Bild (d) zeigt auf, dass die SCL-Klassifizierung nicht
 121 zuverlässig ist. In (e) sehen wir ein blasses Grün. Vermutlich sehen wir durch zirrus wolken
 122 hindurch.

123 2.1.2 Yieldmapping Data

124 XXX description of how harvester gets data, knn-interpolation and rasterization (using lin-
 125 ear interpolation), reference to gregors paper note: discrepancy between sum of estimated
 126 raster and manually weighted yield (per field per year)

127 Gather Data: define Pixel

128 scale of x axis? define DAS ,define GDD -> compare (plot) (note that gdd can be non-
 129 unique) get GDD

130 get NDVI

131 DE Mit den Bändern $B4$ und $B8$ berechnen wir den bekannten **Normalized Difference**
 132 **Vegetation Index (NDVI)** anhand der Formel:

$$NDVI = \frac{B8 - B4}{B8 + B4} \quad (2.1.2.1)$$

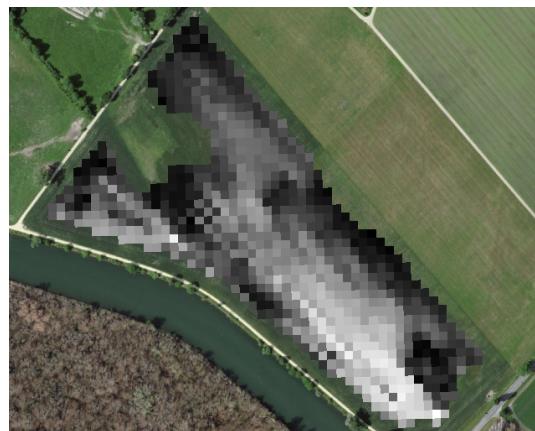
133 Bemerke, dass wir die berechneten Werte nur den *observed NDVI* nennen, da wir aufgrund

Table 2.2: Overview: Scene Classification Layers (SCL)

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)	



(a) A subfigure XXX



(b) A subfigure xxx

Figure 2.2: xxx

134 von wolken und schatten auf ungenauigkeiten gefasst seien müssen.

135 **Chapter 3**

136 **Interpolation Methods**

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

139 First, we give a brief overview in table [3.1](#).

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

142 Later, we introduce and discuss each method.

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

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

146 **3.1 Setting**

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

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

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

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

147 Different assumptions on m will lead to the following methods:

148 **3.2 XXX DAS vs GDD**

149 **3.3 Robustify**

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

Table 3.1: A short summary of the studied interpolation methods. Important assumptions are stated, pros/cons are listed and it is indicated whether the method supports weighted observations (w) and if the resulting interpolation is bounded w.r.t. a fixed interval (b).

	assumptions	pros	cons	w	b
Savitzky-Golay filter	<ul style="list-style-type: none"> - high frequencies are noise (low.pass filter) - equidistant points - local polynomials 	<ul style="list-style-type: none"> - computationally very fast 	<ul style="list-style-type: none"> - cannot deal natively with missing data (need some interpolation) 	no	(yes)
SG + NDVI	<ul style="list-style-type: none"> - upper envelope - vegetation cannot grow faster than some slope 	<ul style="list-style-type: none"> - biological knowledge 	<ul style="list-style-type: none"> - bad “upper envelope” since weights are not used for the estimation itself 	(no)	(yes)
Loess	<ul style="list-style-type: none"> - local polynomial with points closer to the estimated point are more important 	<ul style="list-style-type: none"> - flexible - generalization of SG - weighting function makes intuitive sense 	<ul style="list-style-type: none"> - computationally expensive 	yes	(yes)
Smoothing Splines	<ul style="list-style-type: none"> - 2cd derivative of function is integrable 	<ul style="list-style-type: none"> - intuitive meaning of penalty - general assumptions - flexible shape 	<ul style="list-style-type: none"> - unbounded 	yes	no
B-Splines (Smoothed)	<ul style="list-style-type: none"> - function can be approximated by a linear combination of B-splines basis functions 	<ul style="list-style-type: none"> - general assumption - flexible shape 	<ul style="list-style-type: none"> - unbounded - no intuitive meaning for smoothing 	yes	no
(Gaussian) Kernel Smoothing		<ul style="list-style-type: none"> - simple - general assumptions 	<ul style="list-style-type: none"> - bandwidth: fails if there are big data-gaps 	yes	yes
Double-Logistic	<ul style="list-style-type: none"> - function first increases then decreases - ndvi has a minimal value 	<ul style="list-style-type: none"> - good for evergreen plants (if snow masks ndvi) - upper envelope 	<ul style="list-style-type: none"> - parameterestimation can go seriously wrong - strange behaviour for long data-gaps 	yes	(yes)
Universal Kriging	<ul style="list-style-type: none"> - function is a realization of a stationary gaussian process 	<ul style="list-style-type: none"> - informative parameters - flexible 	<ul style="list-style-type: none"> - regression to the mean - assumptions clearly not met 	yes	(yes)

153 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.

156 XXX¹

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

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

weight. That is the bisquare function B:

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

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

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

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

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

162 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

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

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

164 3.4 Parametric Regression

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

168 Optimization Issues

169 Since we aim to minimize the residuals sum of squares over 5 (or 6) parameters, we try
 170 to solve a non-convex optimization problem. Thus, the algorithm² either struggles to find
 171 the global minimum or fails to converge. This was fixed by providing for each parameter
 172 reasonable initial values and generous bounds (which match our experience).

173 3.4.1 Double Logistic

174 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
 175 NDVI time series).

176 Assumptions:

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

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

- 181 — The growth cycle can be divided into an increase and a decrease period, where
 182 the time series follows a logistic function. The maximum increase (or decrease) is
 183 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:

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

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

186 Similar as for the Savitzky-Golay Filter (c.f. section 3.5.3) we reestimate (only once) the
 187 parameters by giving less weight to the overestimated observations and more weight to
 188 the underestimated observations³.

Pros	Cons
<ul style="list-style-type: none"> — Incorporates subject specific knowledge in the case of evergreen plants covered in snow. — Optimized parameters have an intuitive meaning. 	<ul style="list-style-type: none"> — 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)

189 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:

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

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

Pros	Cons
<ul style="list-style-type: none"> — Assumption of periodicity can be helpful if we are modelling multiyear grow cycles — Flexible curve shape 	<ul style="list-style-type: none"> — Bad behavior in regions with little data (cf. figure 3.1) — Hard to interpret estimated parameters — Parameter estimation can go wrong. Introducing bounds can help.

191 3.5 Non-Parametric Regression

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

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

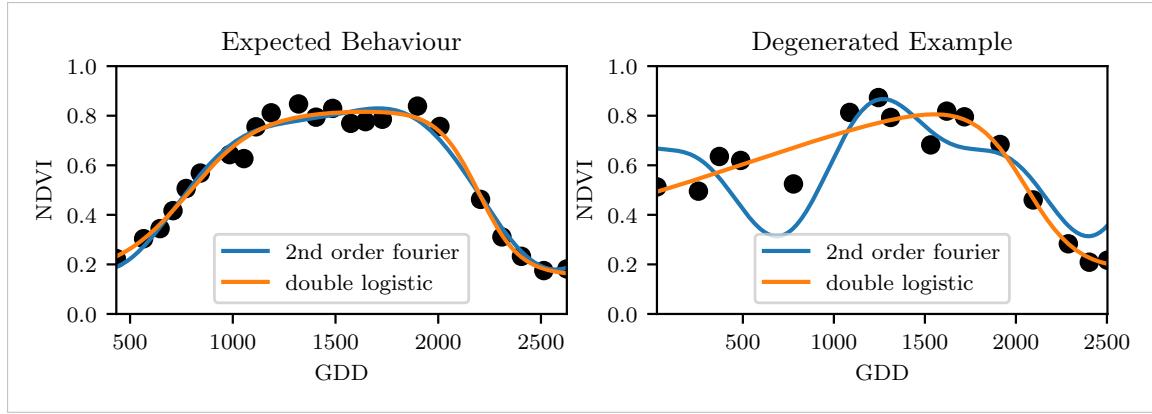


Figure 3.1: Here we observe the nice fitting possibilities of the two parametric methods but notice also some misbehavior

195 **3.5.1 Kernel Regression**

196 As described previously, we would like to estimate

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

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

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

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

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

Pros

- 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)

Cons

- if the $x \mapsto K(x)$ is not continuous, \hat{m} isn't either
- choice of bandwidth, especially if x_i are not equidistant.

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

198 XXX

199 **3.5.2 Kriging**

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

205 **Definitions and Assumptions**

- 206 A *Gaussian Process* $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \dots, S(t_k))$ has a multi-
 207 variate Gaussian distribution for every collection of times t_1, \dots, t_k . S can be fully charac-
 208 terized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') = \text{Cov}(S(t), S(t'))$
 209 Assumption: We will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be
 210 constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following
 211 only $\gamma(h)$.⁴

We also define the variogram of a Gaussian process as

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

And decide to use a Gaussian Variogram defined by

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

- 212 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

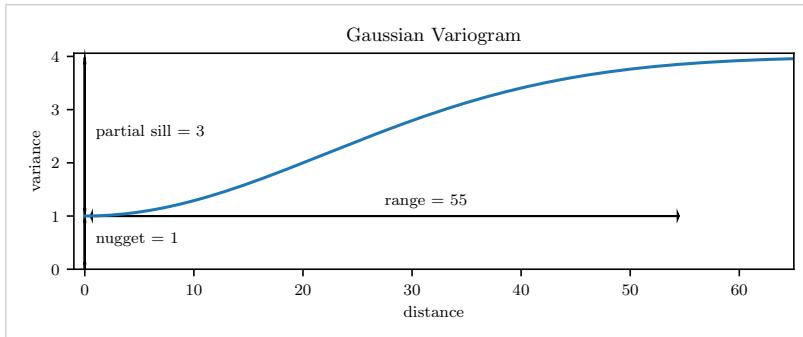


Figure 3.2: Gaussian Variogram with nugget=1, partial sill=3, range=55

213

Pros	Cons
<ul style="list-style-type: none"> — It is a well-studied method. — Parameters have an intuitive meaning. — Flexible covariance structure. 	<ul style="list-style-type: none"> — 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.

214 **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

⁴Note 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.

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

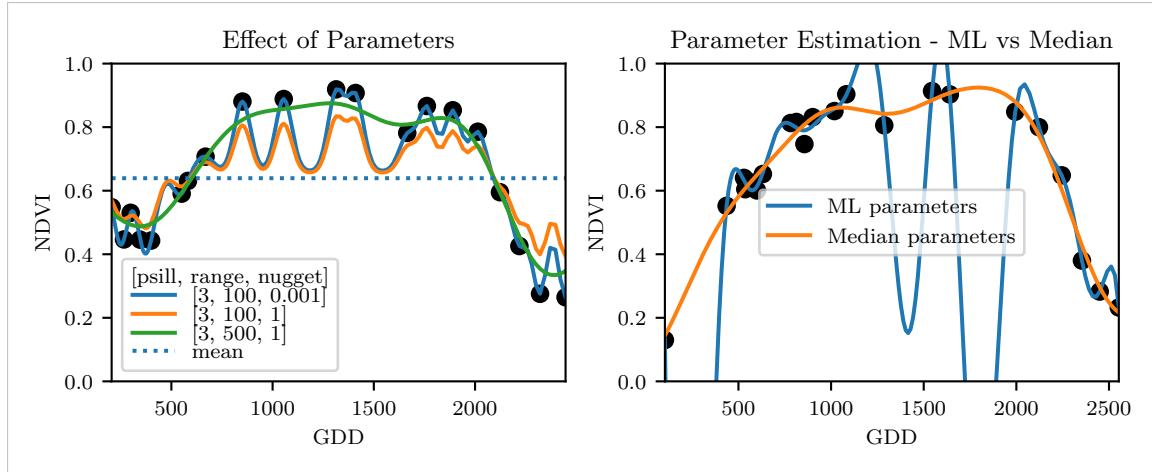


Figure 3.3: On the left, we see how the interpolation change if we increase the nugget and the range parameter. On the right we compare two kriging interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such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

filtering high frequency noise while keeping the low frequency signal. First, we choose a window size m . Then, for each point, $j \in \{m, m+1, \dots, n-m\}$ we fit a polynomial of degree k by:

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

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

For equidistant points this can efficiently be calculated by

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

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 quickly⁶.

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 .

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

227 v.) Update X^1 by applying again a Savitzky-Golay Filter. Repeat this until $w^T |X^1 - X^0|$
 228 stops decreasing, where w is a weight vector with $w_i = \min\left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|}\right)$.
 229 This reduces the penalty introduced by outliers⁷ and by repeating this step we ap-
 230 proach the “upper NDVI envelope”.

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

231 **Extension: Spatial-Temporal-Savitzky-Golay Filter**

232 One notable adaptation of the Savitzky-Golay is the presented by Cao, Chen, Shen, Chen,
 233 Zhou, Wang, and Yang (Cao et al.). The key difference is the additional assumption of the
 234 cloud cover being discontinuous and that we can improve by looking at adjacent pixels⁸.
 235 Because we are working with rather high resolution satellite data, and we need the variance
 236 in the predictors, we will waive this extension.

237 **3.5.4 Locally Weighted Regression (LOESS)**

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

242 The Locally Weighted Regression (LOESS) can be understood as a generalization of the
 243 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

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

244 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
 245 for each y_i we only consider a proportion α of the observations.

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

⁸Here, 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

⁹If 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.

246 **How does the Robust LOESS differ from the SG Filter?**

247 The LOESS smoother takes a fraction of points instead of a fixed number and therefore
 248 automatically adapts to the size of the data we wish to interpolate. However, we run
 249 into the danger of considering too little observations, since the estimation breaks down if
 250 $[an] < d + 1$. Furthermore, LOESS gives less weight to points further away. This yields a
 251 "smoother" estimate, since when we slide the window (e.g. for estimating the next value)
 252 an influential point at the border does not suddenly get zero weight from being weighted
 253 equally before. Finally, the LOESS also can be used for non-equidistant data and allows
 254 for arbitrary interpolation.

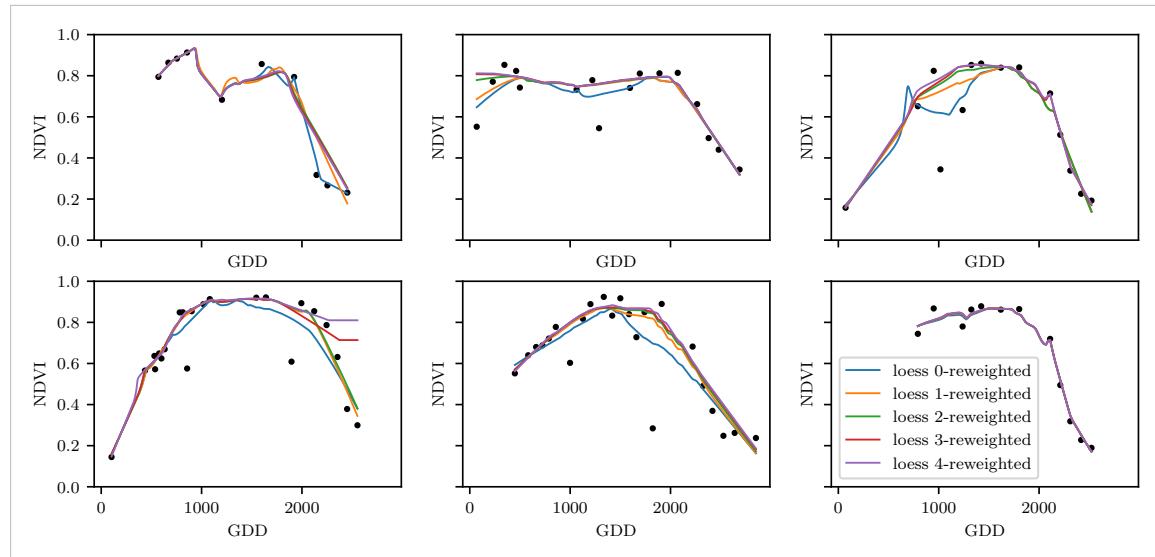


Figure 3.4: The LOESS smoother fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.3) are also displayed

Pros	Cons
<ul style="list-style-type: none"> — Flexible generalization of Savitzky-Golay — arbitrary interpolation possible — Intuitive parameters 	<ul style="list-style-type: none"> — The nature of local regression might lead to surprising estimates (no smoothness guarantees for the second derivative) — Multiple XXXXXXx

255 **3.5.5 B-splines**

from [Lyche and Mørken](#) ([Lyche and Mørken](#))

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

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

$$B_{i,k}(x) = \frac{x-t_i}{t_{i+k}-t_i} B_{i,k-1}(x) + \frac{t_{i+k+1}-x}{t_{i+k+1}-t_{i+1}} B_{i+1,k-1}(x)$$

256 ****Smoothing:**** We can relax the constraint that we have to perfectly interpolate. Thus,
 257 we use the minimum number of knots¹⁰ such that: $\sum_{i=1}^n (w(y_i - \hat{y}_i))^2 \leq s$

¹⁰SciPy uses FITPACK and DFITPACK, the documentation suggests that smoothness is achieved by

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

258 **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 unique¹¹ minimizer

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

259 is a natural¹² cubic spline (i.e. a piecewise cubic polynomial function). The objective
 260 function has an intuitive meaning, as to avoid lateral acceleration it is desirable to move
 261 the steering wheel as little as possible, when driving a car.

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)	

262 **3.5.7 XXX Whittaker Smoother**

263 XXX

264 **3.6 Tuning parameter estimation**

265 lots of cross validation

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

268 ?plot with different densities for each statistic

reducing the number knots used

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

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

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

269 3.7 Robustification – Recap

- 270 introduced in section ?? we want to review it
 271 robustifieng from loess -> lets try it with all. Result in figures ...
 272 issues when reiterating often (we lose some points completely)
 273 from pictures ... we get that one

274 3.7.1 Upper Envelope Approach - Penalty for negative resiudals

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

276 3.8 Performance Assecement

277 TEMP — Figures

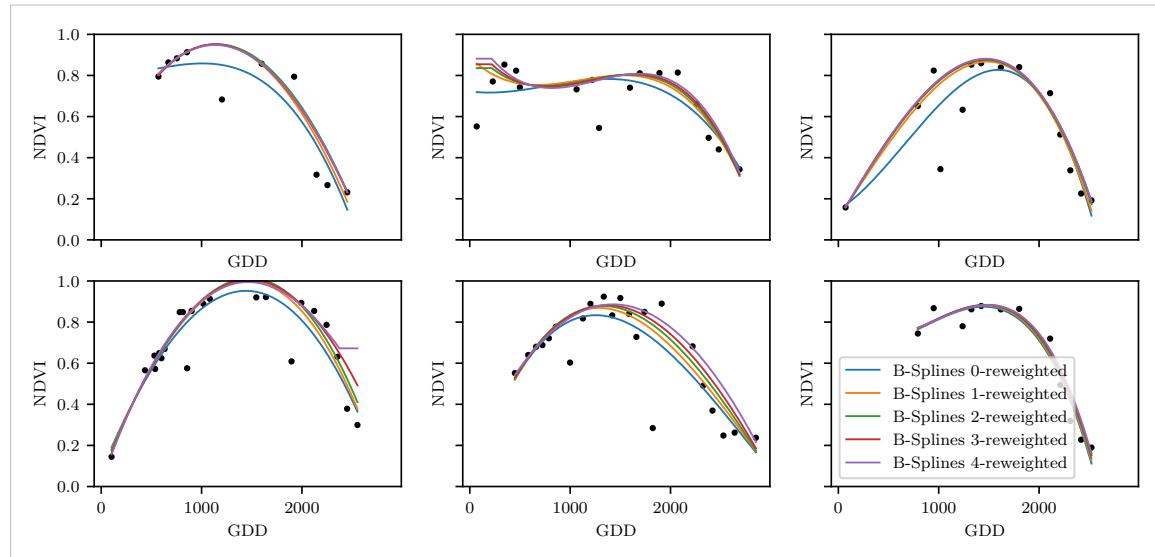


Figure 3.5: B-Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.3) are also displayed

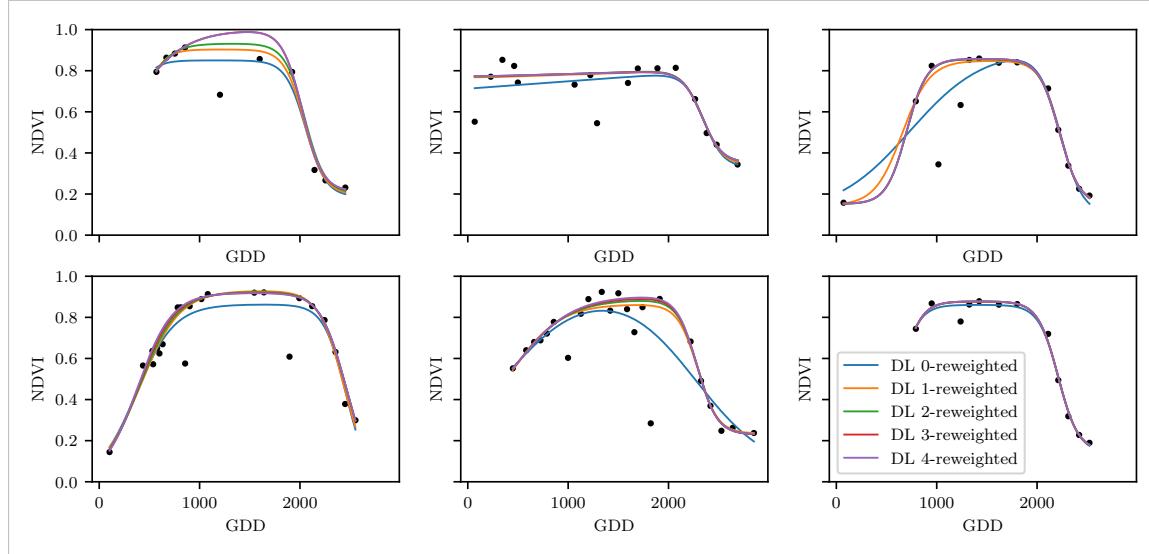


Figure 3.6: A Double Logistic curve fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.3) are also displayed

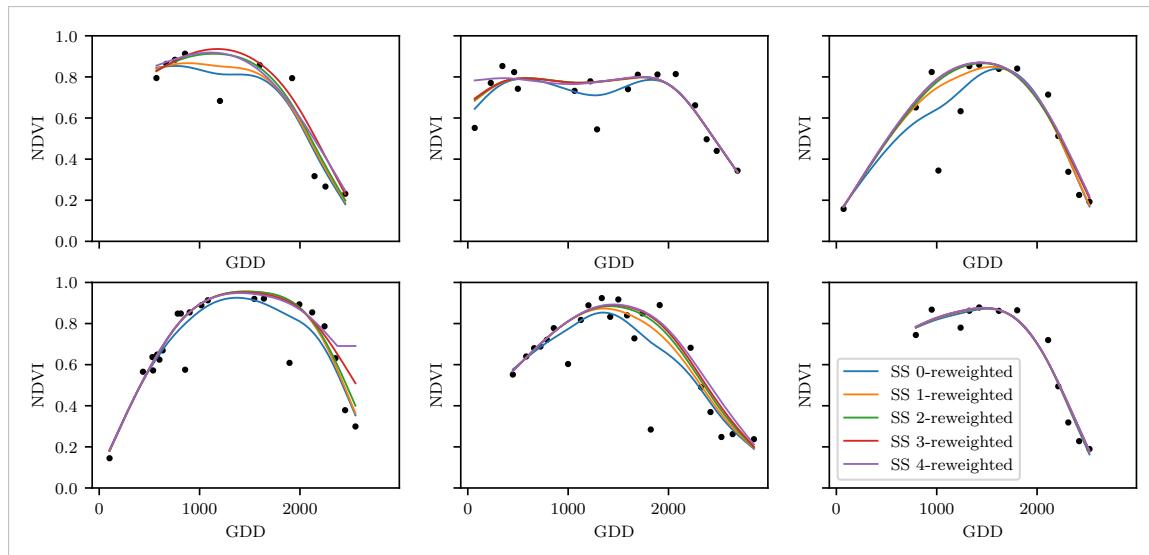


Figure 3.7: Smoothing Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.3) are also displayed

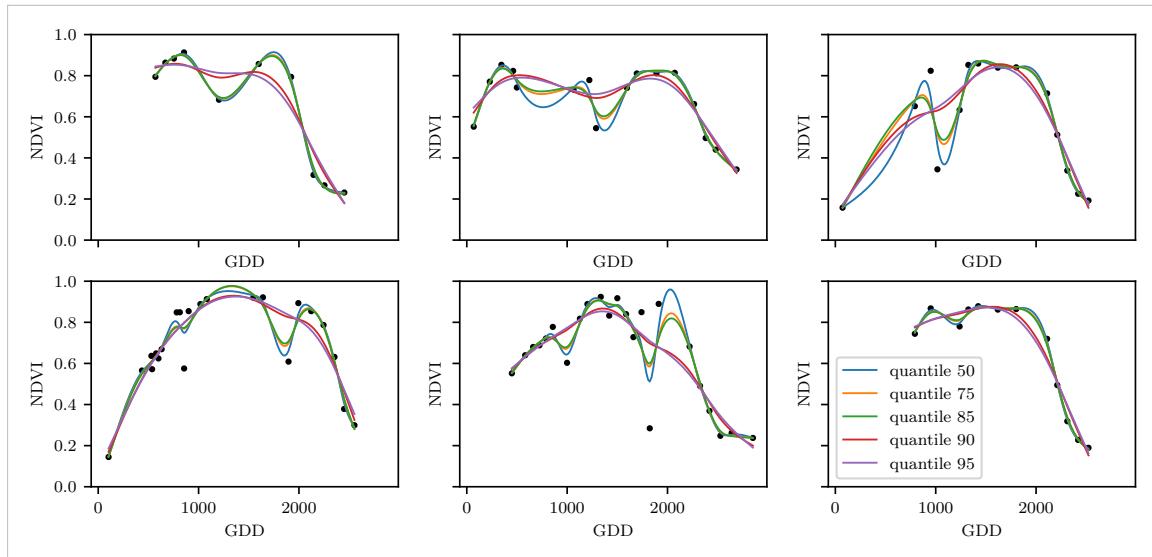


Figure 3.8: Smoothing splines fit with smoothing parameter optimized by minimizing the “...”-quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

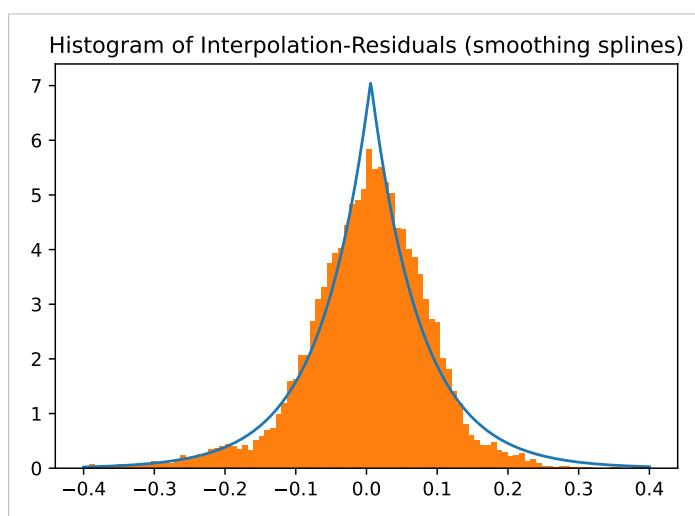


Figure 3.9: XXX caption XXX

278 **Chapter 4**

279 **NDVI Correction / Improve NDVI
280 Data**

281 Let's remind ourselves that the data from the Sentinel-2 is equipped with a scene classi-
282 fication layer (SCL) and we therefore have some information of what is observed at each
283 pixel for each sampled time (c.f. table 2.2). In this chapter we would like to improve
284 the observed NDVI values by using more information than just the two bands used to
285 calculate the NDVI (B4 and B8).

286 **4.1 Considering other SCL Classes**

287 In figure 4.1 we see for example that some blue points¹ follow the interpolated line closely
288 and that they might be useful in improving an interpolation fit.

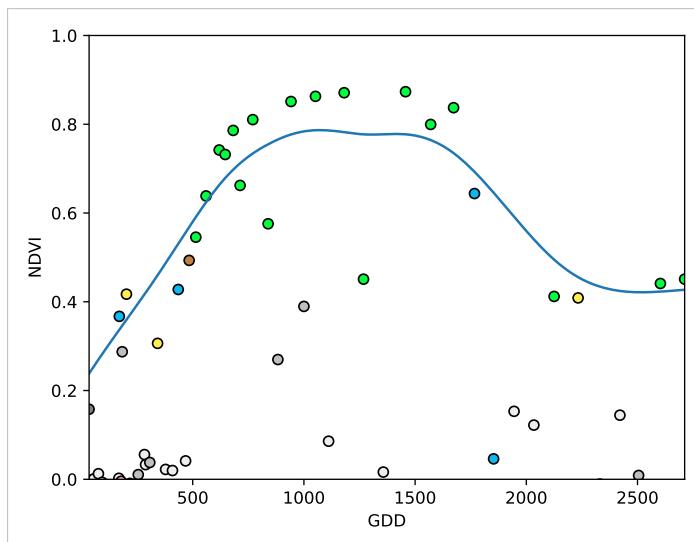


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45)

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

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

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

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

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

- 299 i.) For each pixel and for each observation (every SCL-class):
 300 estimate the NDVI value (via out-of-the-box interpolation²)
- 301 ii.)

302 4.2 XXX Correction

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

304 4.2.1 XXX idea -and- stepwise plots

305 4.2.2 XXX data-table-construction

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

308 4.2.3 XXX ml-methods

309 4.2.4 XXX Uncertainty

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

315 4.3 XXX Evaluation Method

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

320 4.3.1 yield estimation

- 321 problem: high dimensionality and unequal duration/length -> use features
- 322 name approaches for yield estimation (we will use a simple but flexible one)
- 323 random forest ■ for evaluation out-of-bag estimates

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

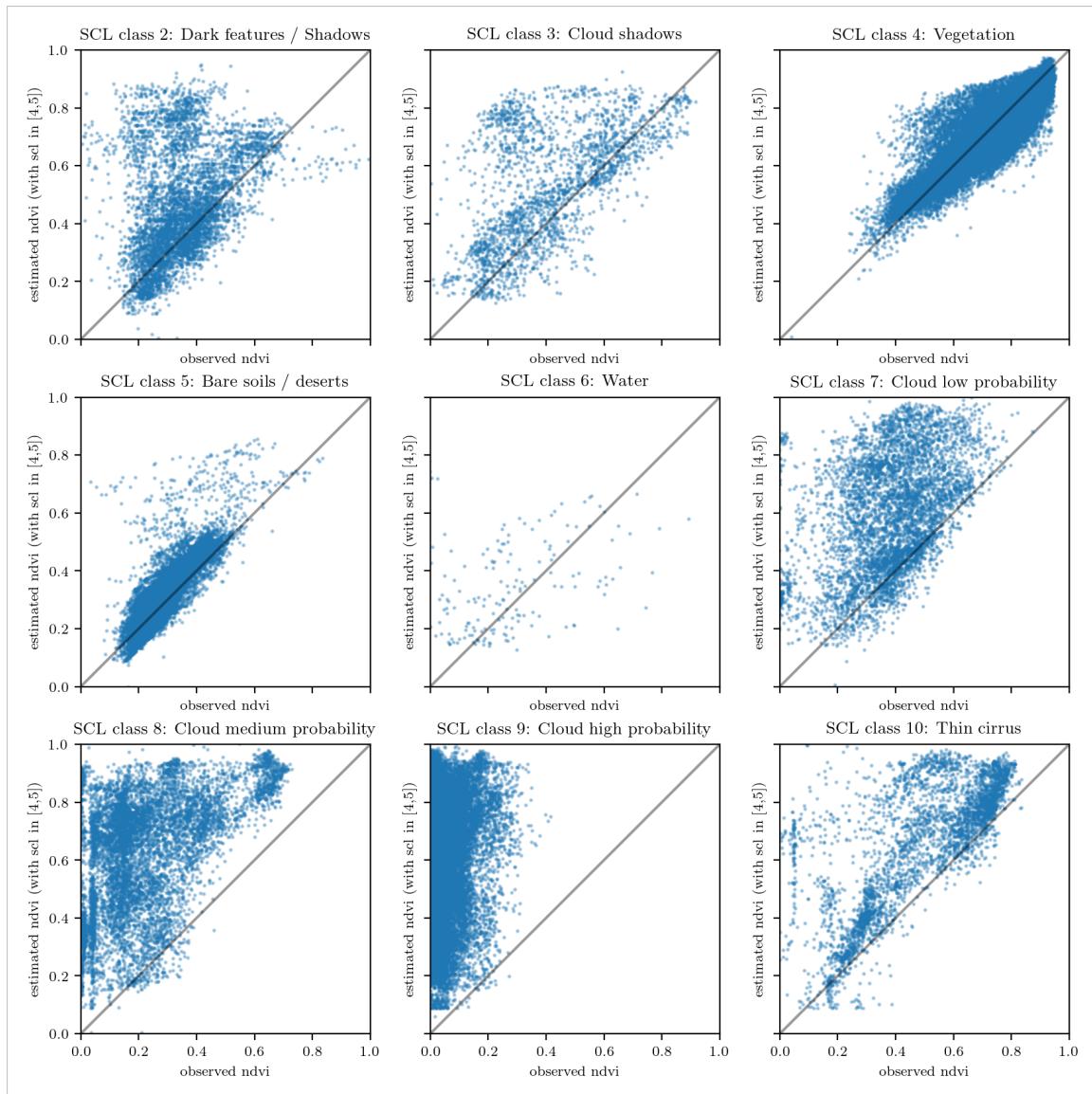


Figure 4.2: XXX caption XXX

324 Covariates used

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

326 **Chapter 5**

327 **Results**

328 **5.1 XXX small recap from “Interpolation Methods”**

329 **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

330 **Chapter 6**

331 **Discussion**

332 **High RMSE in ...:** How much can we expect to get? We have multiple sources of uncer-
333 tainty in the data: 1. Uncertainty in Yield data collected by the combine harvester 2.
334 Uncertainty in Yield data through rasterization 3. Uncertainty in satellite images through
335 “measurement errors” introduced via clouds and other atmospheric effects 4. Uncertainty
336 introduced by interpolating (especially when long data-gaps are present)

337 **Chapter 7**

338 **Outlook**

339 **7.1 Data**

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

344 **7.2 Interpolation**

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

347 **7.3 NDVI Correction**

- 348 — try different link functions in section ... between estimated absolute residuals and
349 weights

350 **7.4 NDVI Correction + +**

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

357

Bibliography

- 358 Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.), *Model-
359 Based Geostatistics*, pp. 46–78. Springer.
- 360 Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. Improved
361 monitoring of vegetation dynamics at very high latitudes: A new method using MODIS
362 NDVI. *100*(3), 321–334.
- 363 Cao, R., Y. Chen, M. Shen, J. Chen, J. Zhou, C. Wang, and W. Yang. A simple method to
364 improve the quality of NDVI time-series data by integrating spatiotemporal information
365 with the Savitzky-Golay filter. *217*, 244–257.
- 366 Cappellari, M., R. M. McDermid, K. Alatalo, L. Blitz, M. Bois, F. Bournaud, M. Bu-
367 reau, A. F. Crocker, R. L. Davies, T. A. Davis, P. T. de Zeeuw, P.-A. Duc, E. Em-
368 sellem, S. Khochfar, D. Krajnović, H. Kuntschner, R. Morganti, T. Naab, T. Oosterloo,
369 M. Sarzi, N. Scott, P. Serra, A.-M. Weijmans, and L. M. Young. The ATLAS3D project -
370 XX. Mass-size and mass-sigma distributions of early-type galaxies: Bulge fraction drives
371 kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function.
372 *432*, 1862–1893.
- 373 Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh. A simple method
374 for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay
375 filter. *91*(3), 332–344.
- 376 Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots.
377 *74*(368), 829–836.
- 378 Jaramaz, D., V. Perović, S. Belanovic Simic, E. Saljnikov, D. Cakmak, V. Mrvić, and
379 L. Zivotic. The ESA Sentinel-2 mission Vegetation variables for Remote sensing of
380 Plant monitoring.
- 381 Lyche, T. and K. Mørken. Spline Methods.
- 382 Savitzky, A. and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified
383 Least Squares Procedures. *36*(8), 1627–1639.
- 384 Schafer, R. W. What Is a Savitzky-Golay Filter? [Lecture Notes]. *28*(4), 111–117.

385 **Appendix A**

386 **XXX Appendix**

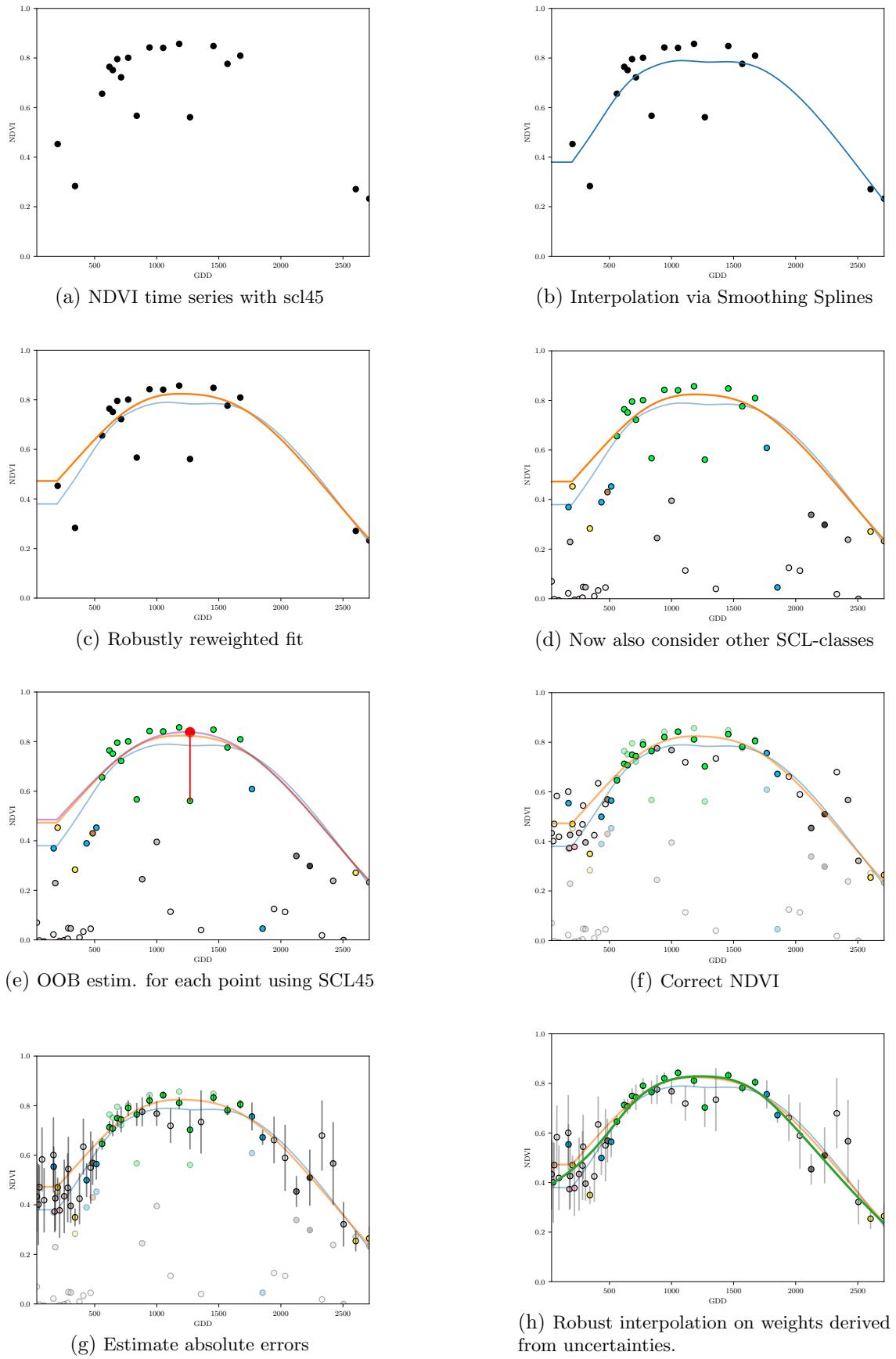


Figure A.1: Stepwise illustration of robust NDVI-Correction