



Swiss Federal Institute of Technology Zurich

Seminar for
Statistics

1 **Department of Mathematics**

2

3

4

5 Master Thesis

Spring 2022

6

7

Lukas Graz

8

Interpolation and Correction

9

of

10

Multispectral Satellite Image Time Series

11

12

Submission Date: September 18th 2022

13

14

Co-Adviser: Gregor Perich
Adviser: Prof. Dr. Nicolai Meinshausen

15 Preface

16 Supplementary Material

17 GitHub: <https://github.com/LGraz/MasterThesis-Code>

18 R package: <https://github.com/LGraz/CorrectTimeSeries>

19 Acknowledgements

20 First, I wish to express my sincere gratitude to my supervisor Prof. Dr. Nicolai Mein-
21 shausen who took the responsibility for my work and happily took the time to discuss
22 conceptual and guiding questions and to inspire me with new ideas.

23 It is necessary to highlight that without Gregor Perich this project would not have been
24 possible. His high personal commitment, reliability as well as the weekly instructive su-
25 pervision meetings were, without question, essential for this work.

26 It was a real pleasure for me to be part of the *Crop Science* group for this time. Enjoying
27 everyday company, a two-day excursion, and harvesting wheat together have made this
28 time truly remarkable. In particular, I would like to thank Prof. Dr. Achim Walter, who
29 supported this collaboration at its core.

30 Last but not least, I would like to express my gratitude to the *Seminar for Statistics*,
31 which created the framework conditions for this work and did everything to help me with
32 conceptional and administrative questions. I should also mention the computing resources
33 provided by them, without which my computations would not have been feasible.

34 Abstract

35 Die Kern-Resultate müssen auch in den Abstract. Ebenso würde ich die vollständige
Reproduzierbarkeit und die R-Package erwähnen.

- 36 Kurze problemerläuterung (NDVI-ts im Zentrum)
- 37 NDVI Interpolation gewinner
- 38 erforscht Robusification
- 39 NDVI Correction + yield-based evaluation

40 Contents

41	Notation	vi
42	1 Introduction	1
43	1.1 XXX motivation - why is it important	1
44	1.2 XXX Problembaum / Fragestellungen	1
45	1.3 XXX State-of-the-art	1
46	1.4 Research Questions	2
47	1.5 Roadmap – anderer Name XXX	2
48	2 Data and Methods	3
49	2.1 Sentinel 2 Data	3
50	2.2 Crop Yield Data	3
51	2.3 Normalized Difference Vegetation Index (NDVI)	5
52	2.4 Timescale Transformation	6
53	2.5 The Concept of a ‘Pixel’	6
54	2.6 Challenges in S2 Data	6
55	2.7 General Methods	8
56	2.7.1 Root Mean Square Error (RMSE)	8
57	2.7.2 Out-Of-Bag (<i>OOB</i>) and Leave-One-Out-Cross-Validation (<i>LOOCV</i>)	8
58	3 Interpolation Methods	9
59	3.1 Interpolation Setup	9
60	3.2 Parametric Regression	11
61	3.2.1 Double Logistic	11
62	3.2.2 Fourier Approximation	11
63	3.2.3 Optimization Issues	12
64	3.3 Non-Parametric Regression	12
65	3.3.1 Kernel Regression	12
66	3.3.2 Kriging	13
67	3.3.3 Savitzky-Golay Filter (SG Filter)	14
68	3.3.4 Locally Weighted Regression (LOESS)	16
69	3.3.5 B-splines	17
70	3.3.6 Natural Smoothing Splines	17
71	3.4 Tuning Parameter Estimation	18
72	3.5 Robustification	18
73	3.5.1 Our Adjustment:	19
74	3.5.2 Examples and Conclusions	20
75	3.5.3 Upper Envelope Approach - Penalty for Negative Residuals	20
76	3.6 Performance Assessment	20
77	4 NDVI Correction	21
78	4.1 Considering other SCL Classes	21
79	4.2 Response and Covariates	22
80	4.3 Correction Methods	22
81	4.3.1 Ordinary Least Squares (<i>OLS</i>)	22
82	4.3.2 Least Absolute Shrinkage and Selection Operator (LASSO)	23
83	4.3.3 Random Forest (<i>RF</i>)	24

84	4.3.4 Multivariate Adaptive Regression Splines (<i>MARS</i>)	24
85	4.3.5 General Additive Model (<i>GAM</i>)	25
86	4.4 Uncertainty Estimation	26
87	4.5 Interpolation	26
88	4.6 Resulting Interpolation Strategies	26
89	4.7 Evaluation Method	27
90	4.7.1 Yield Estimation	27
91	5 Results	30
92	5.1 Goodness of Fit for Selected Interpolation Methods	30
93	5.2 Robustification and NDVI-Correction	30
94	6 Discussion	32
95	6.1 Data Gaps	32
96	6.2 Interpolation Methods	32
97	6.3 NDVI Correction	33
98	6.3.1 Bootstrap	33
99	6.3.2 Using Additional Covariates	33
100	6.3.3 High RMSE in Yield Prediction	33
101	7 Conclusion	34
102	7.1 Future Work	34
103	7.1.1 Time Series Correction-Interpolation as a General Method	34
104	7.1.2 Minor Improvements	35
105	Bibliography	36
106	A Reproducibility	38
107	A.1 Reproduce Results	38
108	A.2 R-Package	38
109	B Further Material	40
110	B.1 Data and Methods	40
111	B.1.1 GDD	40
112	B.2 Interpolation	40
113	B.3 NDVI correction	40

114 Todo list

115 Die Kern-Resultate müssen auch in den Abstract. Ebenso würde ich die vollständige Reproduzierbarkeit und die R-Package erwähnen.	iii
117 Why do we do interpolation in NDVI (and other indices) time series? What are possible shortcomings thereof?	1
119 verdeutliche dem leser, dass ein auftrag das findne von interpolationmethoden war . .	9
120 Paper zitieren wo eingeführt oder wo benutzt (falls einführung fast schon trivial) . .	9
121 Ähnliche struktur sich überlegen	9
122 TODO: include Weighted versions	12
123 figure / tabelle / pseudocode anstatt aufzählung	15
124 consider naming the sub-plots	20
125 Ich finde die sections, in denen Du die Modelle erklärt, gut. Allerdings fehlt mir die Überleitung/Einleitung, warum die Modelle gebraucht werden	25
127 Here in the discussion, you should take up the points you mentioned in the introduction . .	32
128 where does this section belong to? Chapter ‘NDVI Correction’ or ‘Further Work’? .	33
129 You already capture the ”main” structure of your thesis with the interpolation and the NDVi correction sections. Can you combine them both in a ”synthesis”	
131 subsection at the end of the discussion?	33
132 which data? I assume the combine harvester point data?	35
133 page breaks	40
134 replace space before ref by tilda	44
135 check quantile definitions	44

¹³⁶ Notations

¹³⁷ Variables

c	a (vector of) constant(s)
$\lambda \in \mathbb{R}$	a scalar
$n \in \mathbb{N}$	sample size
i, j	are indices in $\{1, \dots, n\}$
$x \in \mathbb{R}^n$	time, usually in GDD
¹³⁸ $x \in \mathbb{R}^n$	covariate 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
$\bar{y} \in \mathbb{R}$	sample mean of y
$r \in \mathbb{R}^n$	residuals given by $y - \hat{y}$

¹³⁹ Abbreviations and Objects

Pixel	A pixel originates of an image pixel and describes a square of 10 x 10 meters in the field which coincides with the resolution (and location) of the Sentinel-2 pixels. Such pixels are illustrated in figure 2.1b. Additional information like yield is also attached.
P_t	describes the observed data (weather and spectral bands) at time t and the location of one pixel.
P	is 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 provided by the European Space Agency (ESA) that gives an estimation of the land cover class of each pixel. It indicates what one can expect at a pixel at a sampled time. For an overview, c.f. table 2.2
P^{SCL45}	is 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 Difference Vegetation Index (Rouse, 1974)

DAS	Days After Sowing
GDD	Growing Degree Days – cumulative sum of “ $\max(0, \text{temperature} - \text{threshold})$ ”
RYEA	Relative Yield-Estimation-Accuracy. Definition 4.7.0.1
OOB	Out Of the Box. Describes the procedure of estimating the value for a point but not consider the point itself (c.f. section 2.7.2)

140 XXX ML models and their shortnames

141 European Space Agency (ESA)

142 **MATLAB Matrix Notation**

143 We will use the MATLAB ‘:’ notation to indicate rows and columns of a matrix. That is
144 if $X \in \mathbb{R}^{n \times p}$ is a matrix, then $X_{[:,3]}$ is the 3rd column of X and $X_{[2,:]}$ is the second row of
145 X .

146 XXX only equations that are referenced are equipped with a number

¹⁴⁷ **Chapter 1**

¹⁴⁸ **Introduction**

¹⁴⁹ Research Questions:

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

¹⁵¹ - NDVI-timeseries is simple and widely used. Examples are: - Plant Models REF - Season
¹⁵² Start (start of spring) (community name: land-surface-plant-phenology) - Yield prediction
¹⁵³ - crop classification

¹⁵⁴ - NDVI is not only of interest to researchers but also public agents and insurance companies

¹⁵⁵ Since satellite images are “for free” researchers extract it (only S2 for free)

¹⁵⁶ Please also add some words on the S2 satellites of ESA in the introduction.

¹⁵⁷ “Similarly, smoothing the time series of satellite data is helpful to address inconsistency
¹⁵⁸ in observation frequency and timing due to clouds and other sensor artefacts Skakun,
¹⁵⁹ Vermote, Franch, Roger, Kussul, Ju, and Masek (2019)”

¹⁶⁰ **1.2 XXX problembaum / fragestellungen**

¹⁶¹ problem schilderung anhand referenzen und evtl. eines bileds:

¹⁶² **1.3 XXX State-of-the-art**

¹⁶³ Why do we do interpolation in NDVI (and other indices) time series? What are possible shortcomings thereof?

¹⁶⁴ zusammenfassung mit literaturrecherche hier (jetzige antowrt auf problemstellung):

¹⁶⁵ — Doublelogistic (winter-ndvi)

¹⁶⁶ — parametric / non-parametric approaches

¹⁶⁷ — spatio-temporal approaches

168 1.4 Research Questions

169 XXX

170 1.5 Roadmap – anderer name XXX

171 This thesis is structured as follows: XXX

172 **Chapter 2**

173 **Data and Methods**

174 We will start by describing the available data and the challenges associated with it. Our
175 study region is a farm of over 800ha, which is located in western Switzerland. From
176 [Perich, Turkoglu, Graf, Wegner, Aasen, Walter, and Liebisch \(2022\)](#) we acquire satellite
177 image data (section 2.1), yield maps of several cereals from 2017 to 2021 (section 2.2),
178 and meteorological data (section 2.5). Afterwards, we will introduce general methods in
179 section 2.7, which will be used in the remaining chapters.

180 **2.1 Sentinel 2 Data**

181 The European Space Agency (ESA)¹ freely distributes the high-quality images of the two
182 Sentinel satellites (S2). Together, both satellites have a revisit time of 5 days at the
183 Equator and 2-3 days at mid-latitudes. However, in our study region, we only receive an
184 image every 5 days.

185 The S2 images contain 12 spectral bands with spatial resolutions up to 10 meters (see
186 2.1). Bands with a lower resolution (20 and 60 meters) were upscaled to 10 meter reso-
187 lution using cubic interpolation ([Perich et al. \(2022\)](#)). In order to decrease the effect of
188 atmospheric conditions like reflections and scattering, bottom-of-atmosphere, radiometric
189 corrected Level-2A data was used². The ESA also supplies an algorithm³ produces Scene
190 Classification Layer (*SCL*) where for each location the observed subject is assigned to one
191 of 11 *SCL*-classes (c.f. table 2.2). In this thesis, we will use this classification to filter out
192 data points, which we believe to be less informative. That are all observations which *SCL*-
193 class does not correspond to vegetation or bare soils (classes 4 and 5). For convenience,
194 we define the set *SCL45* as the observations which belong to *SCL*-class 4 or 5.

195 **2.2 Crop Yield Data**

196 The crop yield data were collected using a combine harvester. Equipped with GPS, the
197 harvester drives over the fields and continuously estimates the dry crop yield density in

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

²According to [Perich et al. \(2022\)](#): “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”

³<https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/>
algorithm

Table 2.1: 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 ([Jaramaz et al. \(2013\)](#)).

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.

Table 2.2: Overview: Scene Classification Layers (SCL)

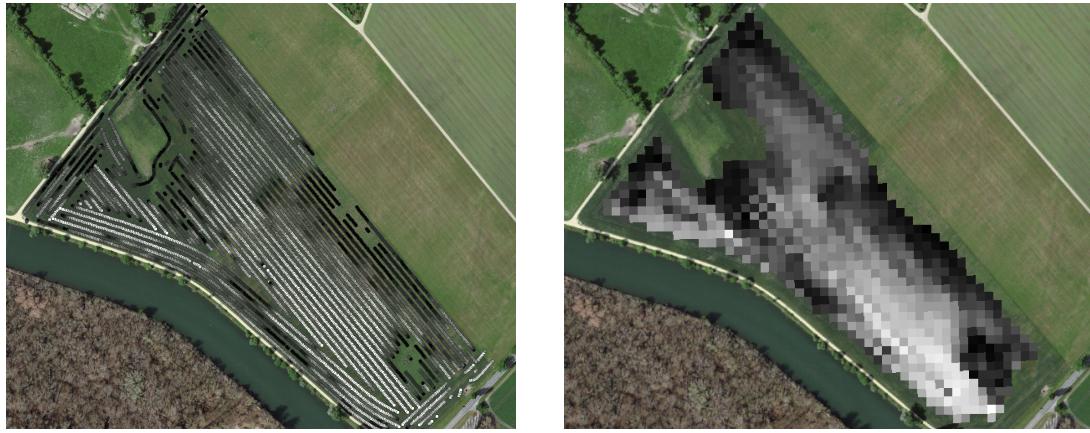
Color	No.	Class	Color	No.	Class
	0:	Missing Data		6:	Water
	1:	Saturated or defective pixel		7:	Cloud low probability
	2:	Dark features / Shadows		8:	Cloud medium probability
	3:	Cloud shadows		9:	Cloud high probability
	4:	Vegetation		10:	Thin cirrus cloud
	5:	Bare soils		11:	Snow or ice

198 t/ha (see fig. [2.1a](#)). We take the data set derived in [Perich et al. \(2022\)](#), where error-prone measurement points (such as during a tight curve of the combine harvester) were removed and then the yield map was rasterized using linear interpolation (c.f. fig. [2.1b](#)).

200 We summarize the rasterized dry-yield values by the following statistics:

202 Minimum 1st Quartile Median Mean 3rd Quartile Maximum Variance
0.107 6.186 7.560 7.359 8.756 13.35 4.035

203 Comparing the average per-field crop yield reported by the farmer with the yield estimated by the combine harvester shows that the latter overestimates crop yield by ca. 10% (c.f. [Perich et al. \(2022\)](#)). Since the relative estimation error is approximately constant and we do not aim for an accurate yield prediction, we will not consider this deviation.



(a) Raw combine harvester data (cleaned)

(b) rasterized to Sentinel 2 resolution.

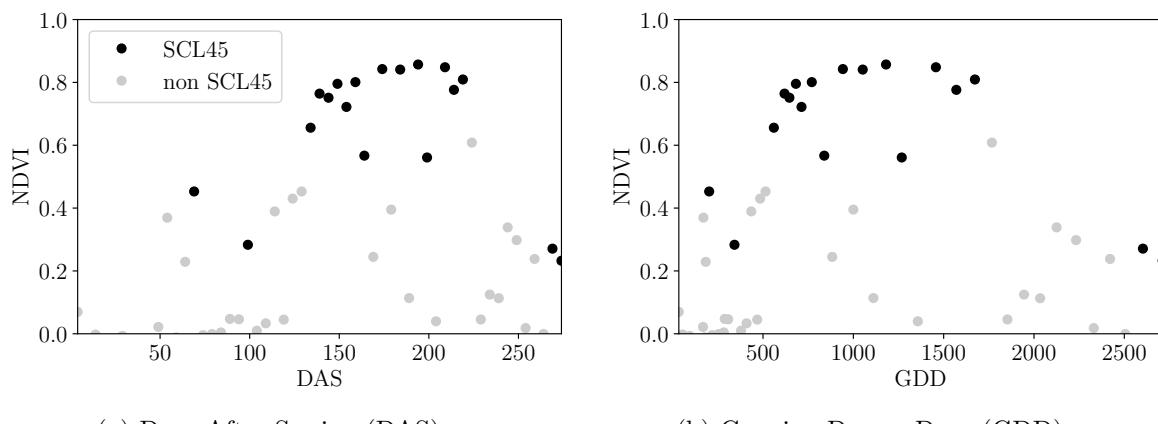
Figure 2.1: Crop yield density map of a field. Ranges from 0.1 t/ha (black) to 5.35 t/ha (white)

207 2.3 Normalized Difference Vegetation Index (NDVI)

208 The well-known (*NDVI*) introduced in [Rouse \(1974\)](#) is used to measure vegetation in
 209 remote sensing. It utilizes a large jump of reflectancy between red and infrared and can
 210 be calculated using the bands *B4* and *B8* (table 2.1) by:

$$NDVI = \frac{B8 - B4}{B8 + B4}$$

211 Since we measure the NDVI via the S2 satellites from space we can not expect to measure
 212 the true NDVI. This is especially true if we do not see the ground because of clouds or the
 213 ground signal is disturbed by cloud shadows. Even if we only use SCL45 observations we
 214 still encounter issues as will be described in section 2.6. Therefore, we call the calculated
 215 values merely the *observed NDVI*. In the following chapters, we will study the resulting
 216 NDVI time series (for one location and one season) extensively. Such a time series is shown
 in figure 2.2a.



(a) Days After Sowing (DAS)

(b) Growing Degree Days (GDD)

Figure 2.2: NDVI time series plotted against DAS and GDD. GDD are introduced in
 section 2.4.

218 2.4 Timescale Transformation

219 Regarding the Days After Sowing (DAS) time scale shown in fig. 2.2a, we detect two
 220 drawbacks. First, this scale makes it difficult to compare two NDVI time series because
 221 wheat is not always sown on the same day of the year and in some years plants begin
 222 to emerge earlier. Second, because there are only few SCL45 observations in the winter,
 223 we face significant data gaps in this period. The time scale transformation introduced in
 224 McMaster and Wilhelm (1997) fixes both problems. The resulting Growing Degree Days
 225 (*GDD*) are defined as the cumulative sum since sowing of temperature above a given base
 226 temperature T_{base} . For cereals, we use $T_{base} = 0$ (Perich et al. (2022)). Thus, the GGD
 227 for n days after sowing will be equal to:

$$GDD_n := \sum_{i=0}^n \max(T_i - T_{base}, 0).$$

228 Important plant growth stages and their corresponding GDD values are tabultaed in B.1.1
 229 In figure 2.2 we see an example for comparison of the DAS and GDD timescale. Here
 230 we see that the first 120 DAS are compressed to just 500 GDD and hence the gap in
 231 observations was succesfully compressed. Due to the reasons mentioned above, from now
 232 on we will only consider GDD.

233 2.5 The Concept of a ‘Pixel’

234 Now we create a new data structure that we call Pixel. This originates from the pixels of
 235 the S2 satellite images. It will contain all the information needed to confront the tasks in
 236 the following chapters.

237 Consider a 10 by 10 meter square that coinsides with a S2 image pixel and T the GDD
 238 values for which S2 images are avialable in a given season. For $t \in T$ let P_t be a tupel of
 239 all the spectral bands, the observed NDVI and the SCL class (at the considered location
 240 at time t). Then, define P as the collection of all the P_t and the estimated dry-yield for
 241 this square. Analogously to P , define P^{SCL45} by only considering P_t with SCL-class 4 or
 242 5 (vegetation and soil).

243 2.6 Challenges in S2 Data

244 Now, we shall illustrate with an example pixel the challenges, we will confront in the
 245 coming chapters. The figure 2.3 shows a selection of 6 satellite images of a field, one
 246 selected Pixel and the NDVI time series of that pixel. In February (image a), we see
 247 no vegetation but bare soil and thus also a low NDVI. At the beginning of May (b), we
 248 observe a cloudless dark green field with a high NDVI. In (c) heavy cloud cover (SCL class
 249 9) leads to a complete loss of plant information in this S2 observation. Figure (d) shows
 250 that the SCL classification is not reliable, since we evidently observe clouds which is also
 251 reflected in a sudden NDVI drop. Even though SCL indicates that (e) are thin cirrus
 252 clouds, we see a pale green and we also note a NDVI.

253 So in conclusion, we remark that some SCL45 observations are not accurate and even
 254 though a few non-SCL45 observations contain useful information, most of them are too
 255 unreliable (e.g. all SCL 9 observations). Thus, we aim to substitute the unreliable ones
 256 with interpolated versions and correct corrupt ones.

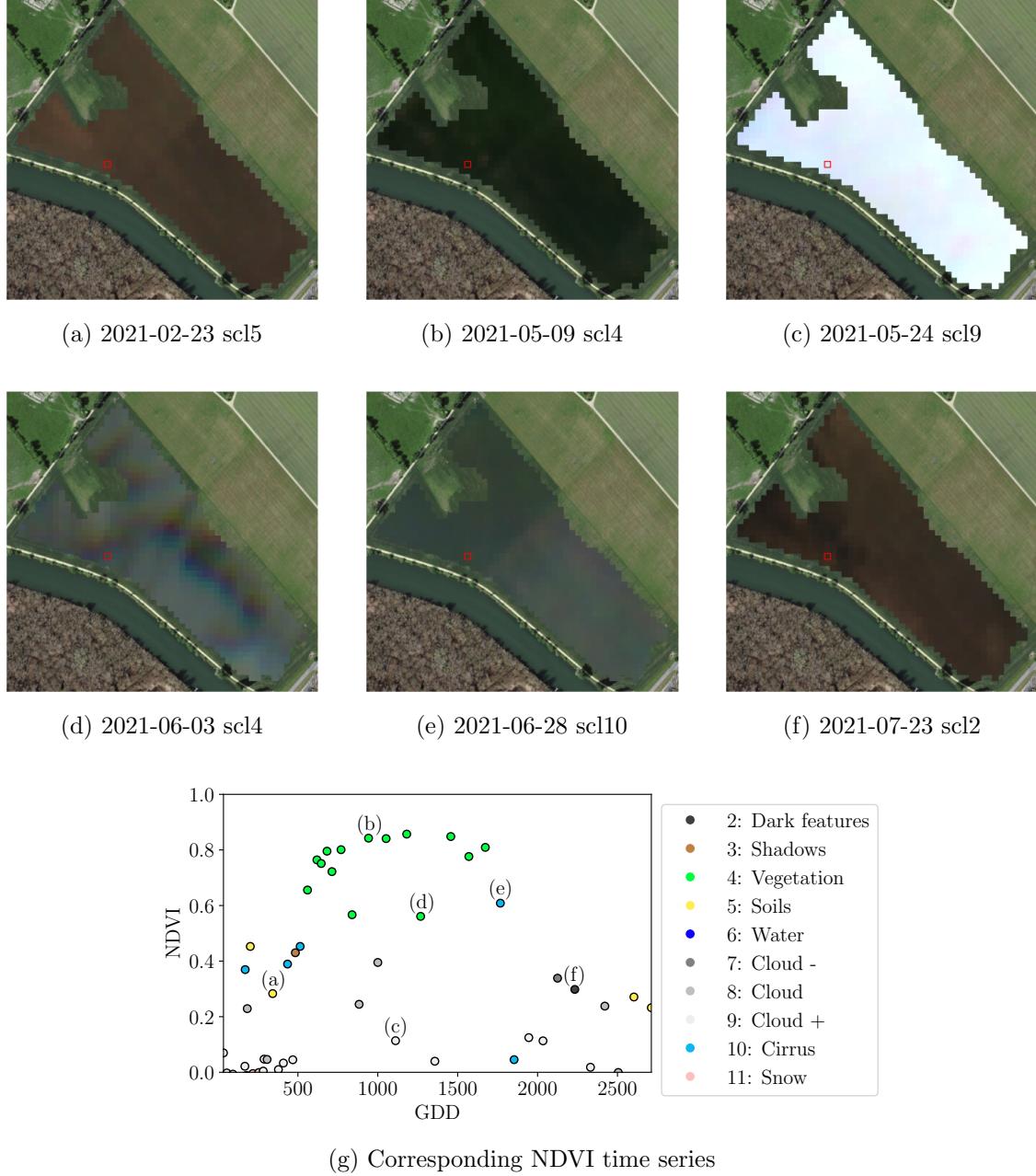


Figure 2.3: Satellite images of a field at selected times with a static background for orientation. Moreover, the NDVI time series of the red-highlighted pixel is shown in (g) colored by the SCL labels.

257 2.7 General Methods

258 Here we will only introduce Methods which will accure in several places. For interpolation
 259 methods we refer to sections 3.2 and 3.3, for a robust interpolation strategy to section 3.5.
 260 In section 3.4 we describe a method to objectively determine the quality of an interpolation,
 261 and in chapter 4 we present the NDVI correction together with an adapted interpolation
 262 strategy.

263 2.7.1 Root Mean Square Error (RMSE)

264 In this section we describe different criteria to evaluate models. Hence, given a vector
 265 $y \in \mathbb{R}^n$ and its estimator \hat{y} (estimated using the model), we define the RMSE as:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

266 2.7.2 Out-Of-Bag (OOB) and Leave-One-Out-Cross-Validation (LOOCV)

267 The rationale for OOB and LOOCV is that we intend to evaluate a model M with unseen
 268 data. That is, if D describes the entire dataset and we train a model on a subset of D , we
 269 can use the remaining data to evaluate the model.

To formally introduce this, let:

$$D = \{(X_{[j,:]}, y_j) \mid X \in \mathbb{R}^{n \times p}, y \in \mathbb{R}^n, j = 1, \dots, n\}$$

270 be a dataset, $i \in \{1, \dots, n\}$ and $M^{(-i)}$ a model fitted on a subset of $D \setminus \{(X_{[i,:]}, y_i)\}$. Then
 271 we call $\hat{y}_i := M^{(-i)}(X_{[i,:]})$ an OOB estimator of y_i . If we do this for all $i \in \{1, \dots, n\}$, we
 272 obtain $\hat{y} := (\hat{y}_1, \dots, \hat{y}_n)$ the OOB estimator for $y \in \mathbb{R}^n$.

273 In the bootstrap (e.g., random forest) framework, we define \hat{y}_i to be the average of all
 274 computed and admissible $M^{(-i)}$.

275 In the case that $M^{(-i)}$ was fitted on the set $D \setminus \{(X_i, y_i)\}$ (i.e., not a true subset), we call
 276 the corresponding \hat{y}_i also the LOOCV estimator.

277 If we optimize some parameter via OOB (or LOOCV) this means that we search for the
 278 parameter that minimizes some loss function which takes the OOB (or LOOCV) residuals.
 279 Usually we approximate this parameter by searching on a grid.

280 **Chapter 3**

281 **Interpolation Methods**

282

283 In section 2.6 we have established the need for interpolating the NDVI time series. In
284 this chapter we first specify a setting for the interpolation and divide the interpolation
285 methods into those that make fundamental shape assumptions (parametric) and those
286 that are more flexible (non-parametric). We give an introduction for each method with
287 an compact definition, highlight adjustments or give remarks where appropriate, and then
288 point out strengths and weaknesses of each method. Additionally, a brief overview of
289 the considered interpolation methods is provided in table 3.1. Afterwards, we extract an
290 robustification strategy from the one interpolation method and generalize it so we can use
291 it for all methods that allow for a priori weighted observations. Finally, using LOOCV,
292 we tune the parameters (where necessary) and get a first idea of the performance of each
293 method.

verdeutliche
dem
leser,
dass ein
auftrag
das
findne
von
interpo
lation-
metho
den war

294 **3.1 Interpolation Setup**

In this chapter we will only consider SCL45 observations, since they are more reliably. Hence, data in the form of (t_i, y_i) for $i = 1, \dots, n$ is given, where t_i is the time in GDD and y_i denotes the NDVI at time t_i . Assume that it can be represented by

$$y_i = m(t_i) + \varepsilon_i,$$

where ε_i is some noise and $m : \mathbb{R} \rightarrow \mathbb{R}$ is 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

$$m(t) = \mathbb{E}[y | t]$$

295 We will introduce parametric and non-parametric approaches to estimate m in section 3.2
296 and 3.3 Furthermore, in the subsequent, we denote $w \in \mathbb{R}^n$ as the vector of weights such
297 that w_i corresponds to the weight that (t_i, y_i) should have in the interpolation.

298 Paper zitieren wo eingeführt oder wo benutzt (falls einföhrung fast schon trivial)

299 Ähnliche struktur sich überlegen

Table 3.1: Summary of the studied interpolation methods containing important assumptions, advantages and disadvantages and whether the method supports weighted observations (w) and if the resulting interpolation is bounded w.r.t. a fixed interval (b).

	Assumptions	Advantages	Disadvantages	w	b
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"> - Parameter estimation can be very difficult - Strange behavior for long data-gaps 	Yes	(Yes)
Fourier Approximation	<ul style="list-style-type: none"> - NDVI can be approximated by a 2cd order fourier series. 	<ul style="list-style-type: none"> - Incorporates periodical growth-cycles 	<ul style="list-style-type: none"> - Parameter estimation can be very difficult - Curve easily exceeds bounds of the NDVI 	Yes	No
(Gaussian) Kernel Smooth- ing	<ul style="list-style-type: none"> - Close points are related to each other via a kernel function 	<ul style="list-style-type: none"> - Simple - Computationally very fast 	<ul style="list-style-type: none"> - Biased, especially at ‘peaks’ and ‘valleys’ - Bandwidth: fails if there are big 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)
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)
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
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"> - Choice of smoothing parameter 	Yes	No

300 **3.2 Parametric Regression**

301 Parametric Curve estimation tries to fit a parametric function, such as, for example, a
 302 Gaussian function with parameters μ and σ , to a dataset. In the following, we introduce
 303 two parametric approaches.

304 **3.2.1 Double Logistic**

The Double Logistic smoothing as described in [Beck, Atzberger, Høgda, Johansen, and Skidmore \(2006\)](#)REF heavily relies on shape assumptions of the fitted curve (i.e. the NDVI time series). First, we assume that 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 several years into account. Second, we assume that 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:

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

305 Where the five free parameters: y_{\max} , d_0 , d_1 , t_0 , t_1 are initially estimated by least squares.
 306 Such fit can be seen in figure [3.1](#).

307 **Robustification**

308 Similar as for the Savitzky-Golay Filter (c.f. section ??) one can reestimate (only once) the
 309 parameters by giving less weight to the overestimated observations and more weight to the
 310 underestimated observations. For the details on the choice of the weights we refer to [Beck
311 et al. \(2006\)](#). We wont apply this reestimation but rather the robustification introduced
 312 later in section [3.5](#).

Advantages	Disadvantages
<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. (c.f. figure 3.1)

313 **3.2.2 Fourier Approximation**

Analogous to section [3.2.1](#) we fit a parametric curve to the data by least squares. Here we take the second order Fourier series:

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

314 where $\Phi = 2\pi \times (t - 1)/n$. Thus, we assume an periodacal behavior. If we would set the
 315 period to match one year this would coinced with the nothion that plans grow every year.
 316 Example fits can be seen in figure [3.1](#)

Advantages	Disadvantages
<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 (c.f. figure 3.1) — Hard to interpret estimated parameters — Parameter estimation can go wrong. Introducing bounds can help.

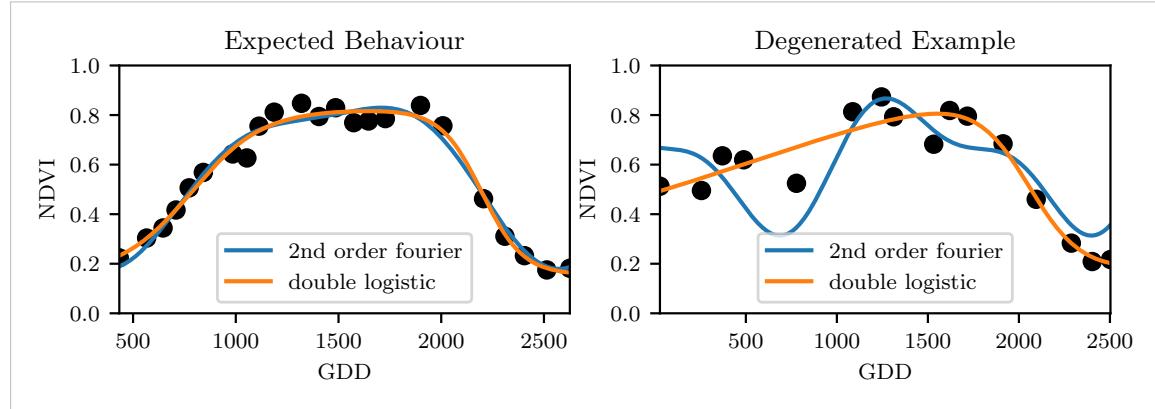


Figure 3.1: Here we observe the possibilities of a precise fit for the two parametric methods but notice also some misbehavior

317 3.2.3 Optimization Issues

318 We shall mention some optimization issues we countered during implementation. Since we
 319 aim to minimize the residual sum of squares over 5 (or 6) parameters, we try to solve a
 320 non-convex optimization problem. Thus, the algorithm¹ either struggles to find the global
 321 minimum or fails to converge. This was fixed by providing for each parameter reasonable
 322 initial values and generous bounds (which match our experience).

323 3.3 Non-Parametric Regression

325 In non-parametric curve estimation, the curve does no longer have to be fully determined
 326 by parameters, but we allow it to flexibly approximate the data. Note, that we do not
 327 exclude the use of tuning-parameters.

TODO:
include
Weighted
versions

328 3.3.1 Kernel Regression

329 As described in section 3.1, we aim to estimate

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x,y) dy}{f_X(x)}, \quad (3.3.1.1)$$

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

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

¹We used the python function `scipy.optimize.curve_fit`.

where h , the bandwidth, symbolizes the windowsize of to consider. By using the above function in equation (3.3.1.1) we arrive at the *Nadaraya-Watson* kernel estimator:

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

- 330 Common choices for the kernel are the normal function or a uniform function (also called
 331 ‘box’ function).

332 **Choose Bandwidth**

- 333 Note that we still need to choose the bandwidth of the function. This can be done with
 334 the help of LOOCV while optimizing the RMSE. For non-equidistant data we refere to
 335 [Brockmann, Gasser, and Herrmann \(1993\)](#) where a local adaptive bandwidth selection is
 336 presented.

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

337 **3.3.2 Kriging**

- 338 Kriging as described in [Diggle and Ribeiro \(2007\)](#) was developed in geostatistics to deal
 339 with autocorrelation of the response variable at locations which are spatially close. By
 340 applying the notion that two spectral indices which are timewise close should also take
 341 similar values, we justify the application of Kriging. In the end, we would like to fit a
 342 smooth Gaussian process to the data.

- 343 A Gaussian Process $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \dots, S(t_k))$ has a multi-
 344 variate Gaussian distribution for every collection of times t_1, \dots, t_k . S can be fully charac-
 345 terized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') := \text{Cov}(S(t), S(t'))$.
 346 Furthermore, we will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be
 347 constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following
 348 only $\gamma(h)$.²

Now, we need to make some assumption on the covariance function. For this we introduce the variogram of a Gaussian process as

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

and define γ via the above equation by choosing the Gaussian Variogram defined by

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

²Note 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.

349 Here h denotes the distance, n is the nugget, r is the range and p is the partial sill. The
 350 influence of the parameters is visualized in figure 3.2.³

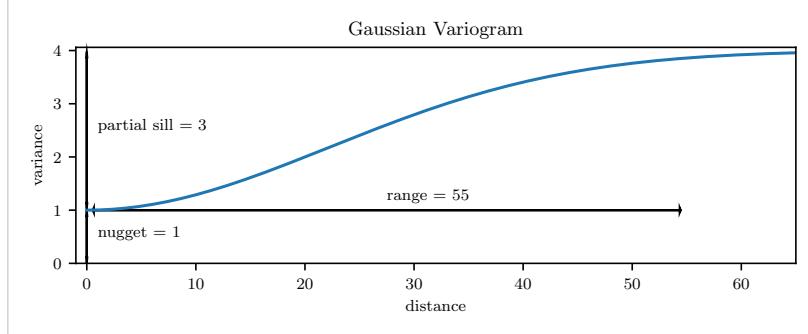


Figure 3.2: Gaussian Variogram with nugget=1, partial sill=3, range=55

351 Finally, we consider a one-dimensional Gaussian process G_γ with variogram γ and tune
 352 the variogram parameters using maximum likelihood⁴. Let z be a vector with the new
 353 values to extrapolate, then we can determine the values $m(z) = \mathbb{E}[G_\gamma(z)|(x, y)]$ using
 354 Bayes rule⁵. For an example fit, we refer to figure 3.3.

355 Violated Assumption

356 Since we observe a clear pattern of a growth period in spring and harvest in the end
 357 of summer, we have to admit that our stationarity assumption with the constant mean
 358 is structurally violated. This is also the reason why we observe (for every variogram
 359 parameter) a tendency to the mean, as indicated in figure 3.3.

Advantages	Disadvantages
<ul style="list-style-type: none"> — It is a well-studied method. — Variogram 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. — Pure maximum likelihood can result in overfitting.

360 3.3.3 Savitzky-Golay Filter (SG Filter)

361 The *Savitzky-Golay Filter*, introduced in [Savitzky and Golay \(1964\)](#) is a technique in signal
 362 processing and can be used to filter out high frequencies (low-pass filter) ([Schafer, 2011](#)).
 363 Furthermore, it can also be used for smoothing by filtering high frequency noise while
 364 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

³Strictly speaking we use a scaled version of the variogram. Thus, only the ratio of p/n matters.

⁴As illustrated in figure 3.3 maximum likelihood estimation can lead to overfitting. Thus, we will in practice sample several such optimized parameters and use their median in the end.

⁵Bayes rule generally claims, that for two random variables A and B we have that $P(A|B) = P(B|A)/P(B)$

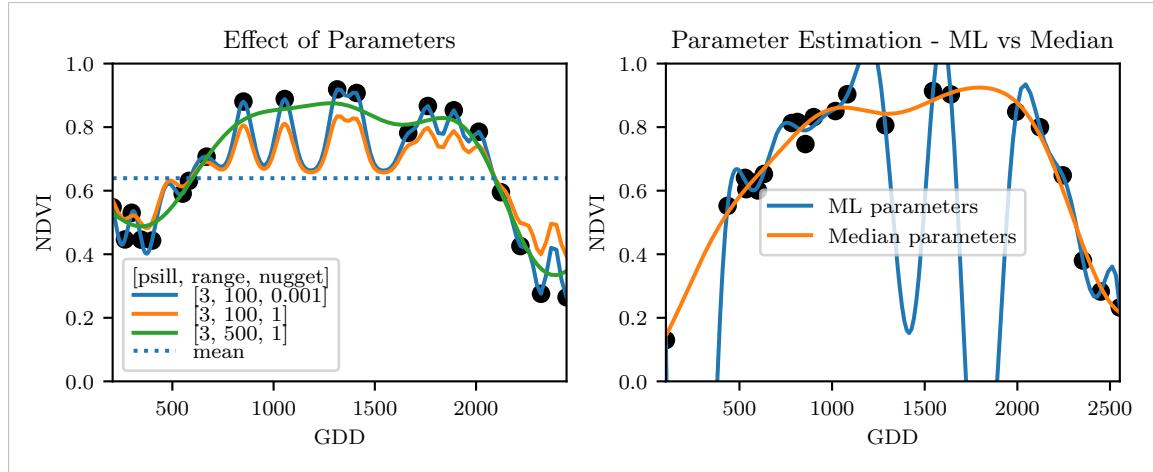


Figure 3.3: On the left, we see how the interpolation change if we increase the nugget and the range parameter. On the right, we compare two kriging interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such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

a polynomial of degree k by:

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

where P_k denotes the Polynomials of degree k over \mathbb{R} . For equidistant points this can efficiently be calculated by

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

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

366 Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh (2004) developed a ‘robust’ 367 interpolation method for the NDVI based on the SG Filter. The method is based on the 368 assumption that due to atmospheric effects the observed NDVI tends to be underestimated 369 and that it cannot increase too quickly. The latter is argued by the biological impossibility 370 of such fast vegetation changes. Their proposed algorithm is:

- 371 i.) Remove non-SCL45 points.
- 372 ii.) Remove points which would indicate an increase greater than 0.4 within 20 days.
- 373 iii.) Linearly interpolate to obtain an equidistant time series X^0 .
- 374 iv.) Apply the SG Filter to obtain a new time series X^1 .
- 375 v.) Update X^1 by applying again a SG Filter. Repeat this until $w^T |X^1 - X^0|$ stops 376 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 377 reduces the penalty introduced by outliers⁶ and by repeating this step we approach 378 the “upper NDVI envelope”.

figure /
tabelle /
pseu-
doode
anstatt
aufzäh-
lung

⁶Here we call a point i an outlier if $X_i^0 < X_i^1$.

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

380 One notable adaptation of the SG Filter is the presented by [Cao, Chen, Shen, Chen, Zhou, Wang, and Yang \(2018\)](#). The key difference is the additional assumption of the cloud cover
 381 being discontinuous and that we can improve by looking at adjacent pixels⁷. Because we
 382 are working with rather high resolution satellite data, and we need the variance in the
 383 predictors, we will waive this extension.

Advantages	Disadvantages
— 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.

385 **3.3.4 Locally Weighted Regression (LOESS)**

386 The LOESS introduced by [Cleveland \(1979\)](#) can be understood as a generalization of the
 387 SG Filter (c.f. sec. [3.3.3](#)).

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

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

388 where h_i is the minimal distance such that $\lceil \alpha n \rceil$ observations are in the ball $B_{h_i}(x_i)$.⁸ So
 389 for each y_i we only consider a proportion α of the observations.

390 **Differences between the Robust LOESS and the SG Filter?**

391 The LOESS smoother takes a fraction of points instead of a fixed number and therefore
 392 automatically adapts to the size of the data we wish to interpolate. However, we run
 393 into the danger of considering too little observations, since the estimation breaks down if
 394 $\lceil \alpha n \rceil < d + 1$.⁸ Furthermore, LOESS gives less weight to points further away. This yields
 395 a "smoother" estimate, since when we slide the window (e.g. for estimating the next value)
 396 an influential point at the border does not suddenly get zero weight from being weighted
 397 equally before. Finally, the LOESS also can be used for non-equidistant data and allows
 398 for arbitrary interpolation.

⁷Here, 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

⁸If 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. But we also have to assure that α is big enough.

Advantages	Disadvantages
<ul style="list-style-type: none"> — Flexible generalization of SG Filter — 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)

399 **3.3.5 B-splines**

B-splines as discussed in [Lyche and Mørken \(2005\)](#) are piecewise cubic polynomials defined by

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

where B are basis functions and recursively defined by:

$$\begin{aligned} B_{i,0}(z) &= 1, \text{ if } t_i \leq z < t_{i+1}, \text{ otherwise } 0 \\ B_{i,k}(z) &= \frac{z-x_i}{x_{i+k}-x_i} B_{i,k-1}(z) + \frac{x_{i+k+1}-z}{x_{i+k+1}-x_{i+1}} B_{i+1,k-1}(z). \end{aligned}$$

Assuming that all x_i are distinct, this yields an interpolation which fits the data perfectly. To reduce the amount of overfitting and increase the smoothness, we relax the constraint that we have to perfectly interpolate. Thus, we use the minimum number of basis functions⁹ such that:

$$\sum_{i=1}^n (w_i(y_i - \hat{y}_i))^2 \leq s$$

Advantages	Disadvantages
<ul style="list-style-type: none"> — can be assigned degrees of freedom — extendable to "smooth" version — performs also well if points are not equidistant 	<ul style="list-style-type: none"> — smoothing process does not translate well to a interpretation (unlike smoothing splines) — choice of smoothing parameter s

400 **3.3.6 Natural Smoothing Splines**

401 Let \mathcal{F} be the Sobolev space (the space of functions of which the second derivative is
402 integrable). Then the unique¹⁰ minimizer

$$\hat{m} := \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n w_i (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx \quad (3.3.6.1)$$

403 is a natural¹¹ cubic spline (i.e. a piecewise cubic polynomial function). The objective
404 function ensures that we decrease the curvature while keeping the RMSE low.

⁹So we do not require one basis function for each neighboring pair of knots. SciPy uses FITPACK and DFITPACK, the documentation suggests that smoothness is achieved by reducing the number of knots used

¹⁰Strictly speaking it is only unique for $\lambda > 0$

¹¹It is called natural since it is affine outside the data range ($\forall x \notin [x_1, x_n] : \hat{m}''(x) = 0$)

Advantages	Disadvantages
<ul style="list-style-type: none"> — Can be assigned degrees of freedom (trace of the hat-matrix). — Efficient estimation (closed form solution). — Intuitive penalty (we don't want the function to be too "wobbly" — change slopes). — Also performs well if points are not equidistant. — Fixes the Runge's phenomenon (fluctuation of high degree polynomial interpolation). 	<ul style="list-style-type: none"> — The tuning parameter λ must be chosen. This can be done via cross validation and optimizing a score function (e.g. the RMSE).

405 3.4 Tuning Parameter Estimation

406 Many of the interpolation methods introduced in section 3.2 and 3.3 include a free parameter.
 407 To determine this parameter for a specific interpolation method, we will estimate the
 408 absolute residuals using OOB estimation and then optimize the parameter using a score
 409 function. We clarify the procedure step by step:

- 410 i.) Construct a set Λ of candidate parameters that generously covers the parameter
 411 space.
- 412 ii.) Consider \mathcal{P} , a set of Pixels.
- 413 iii.) For each parameter $\lambda \in \Lambda$ consider the individual pixels and compute the LOOCV¹²
 414 for the absolute residuals of the specific NDVI-interpolation method for all Pixels in
 415 \mathcal{P} and store them in the set R_λ .
- 416 iv.) Determine $\lambda_{optimal} = \arg \min_{\lambda \in \Lambda} q_{90}(R_\lambda)$, where we describe the 90% quantile with
 417 q_{90} .

418 We choose quantile(90) as our optimization function because we want to allow 10% of
 419 outliers (corrupt points) but also aim for an accurate fit in 90% of the cases.

420 Figure 3.4 exemplifies the effect of the optimization function (different quantiles). To
 421 summarize, we may say that the higher the quantile, the stronger the smoothing.

422 3.5 Robustification

423 Now we discuss a general approach of how to make an interpolation more robust against
 424 outliers. The main idea is to give less weight to observations that have high residuals after
 425 the initial (or if we reiterate, the previous) fit.

426 Even though the procedure is taken from the robust version of the LOESS smoother (c.f.
 427 section 3.3.4 and Cleveland (1979)), we can apply it to every interpolation method that
 428 allows for prior weighting of observations.

¹²For a definition of the leave-one-out-cross-validation we refer to section 2.7.2

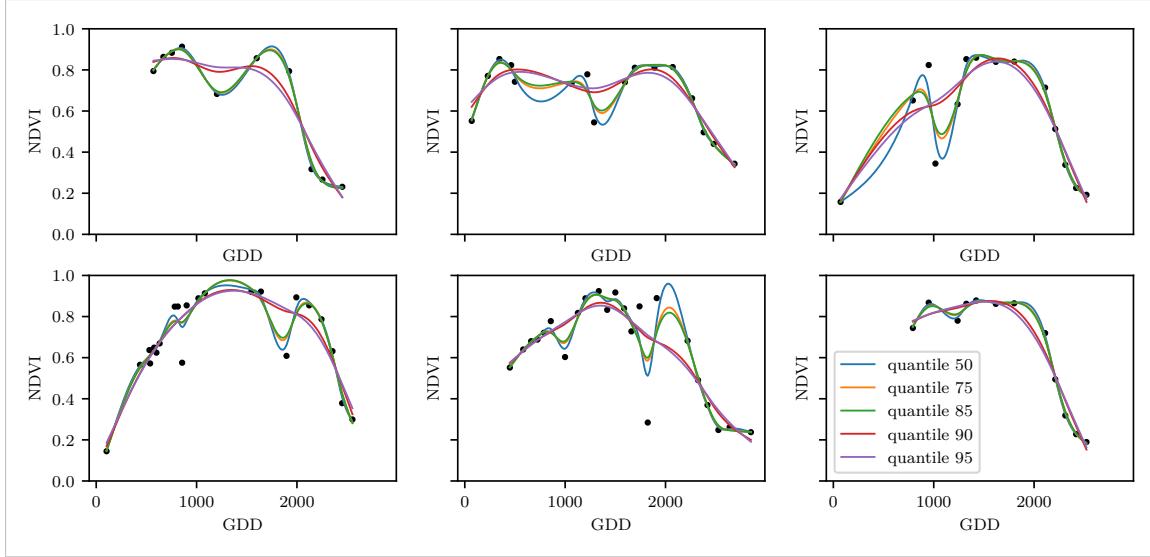


Figure 3.4: Smoothing splines fit with smoothing parameter optimized by minimizing the given quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

429 After an initial fit we calculate the residuals $r_i := y_i - \hat{y}_i$ and obtain \tilde{r}_i by scaling with the
430 median of the absolute residuals:

$$\tilde{r}_i := \frac{r_i}{6 \text{ med}(|r_1|, \dots, |r_n|)}$$

431 Next, we compute new weights by

$$w_i^{\text{new}} := w_i^{\text{old}} \begin{cases} (1 - \tilde{r}_i^2)^2, & \text{if } |\tilde{r}_i| < 1 \\ 0, & \text{else} \end{cases}; \quad (3.5.0.1)$$

432 Using the new weights, we can re-interpolate. This reweighting can be iterate for several
433 steps or till the change of the values is smaller than some tolerance.

434 Note that this procedure is indeed robust since we use the median for the normalization
435 which has a breakdown point¹³ of 50%.¹⁴

436 3.5.1 Our Adjustment:

During the iterations or when supplying prior weights low-weighted observations can corrupt our estimation of scale (the median of absolute residuals). Thus, we introduce the weighted median as

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

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

¹³Intuitively, the breakdown point denotes the fraction of observations a “vicious” player can replace without breaking the estimator. For example, the median has a breakdown point of 50%.

¹⁴The breakdown point relates only to outliers in the y values. Note that we do not require the interpolation methods to be robust, since the residual for an outlier will still be larger than for non-outliers and thus will be down weighted more and more in each iteration (because for the next iteration the residual of the outlier will be even larger, since we gave less weight to it).

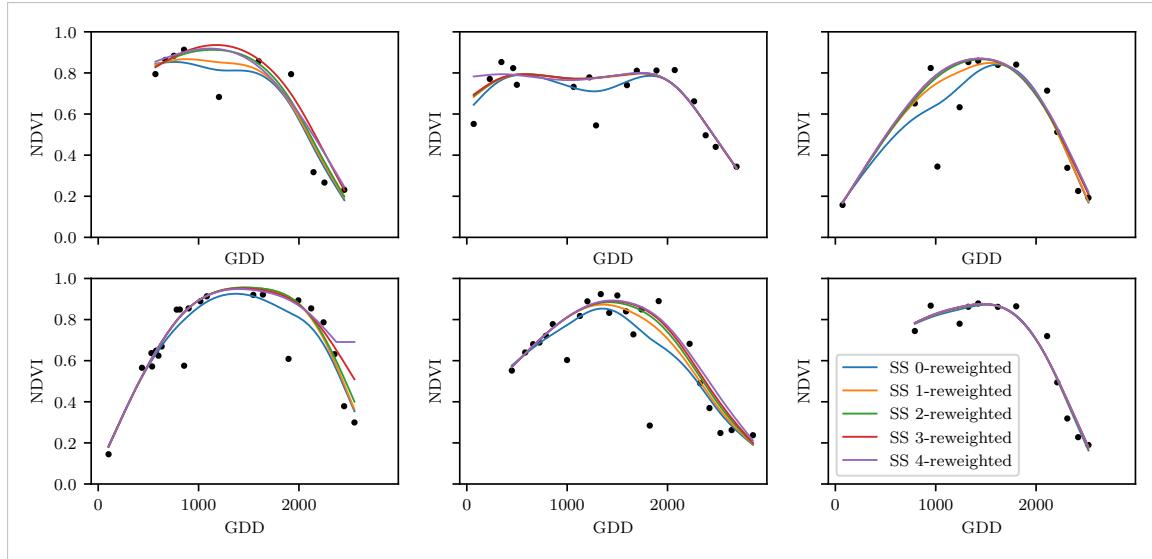
438 **3.5.2 Examples and Conclusions**

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

439 Examples of the first four iterative fits using smoothing splines are shown in figure 3.5 for
 440 six pixels. For the analogous figures of the other interpolation methods c.f. figures B.1, B.2,
 441 B.3 and B.1. Indeed, we observe how the interpolated time series is less affected by outliers
 442 after each iteration. We notice the biggest difference in the first iteration. Furthermore, in
 443 the plot at the bottom left we see how the interpolation ‘escapes’ from the right endpoint
 444 with each successive iteration, even though our intuition does not necessarily identify this
 445 point as an outlier. Therefore, in the following, we will always stop after one iteration.

consider
naming
the sub-
plots

446 **3.5.3 Upper Envelope Approach - Penalty for Negative Residuals**

447 If we artificially increase the negative residuals in 3.5.0.1 by multiplying (e.g. factor 2),
 448 the corresponding points will get less weight in the next iteration. This allows us to create
 449 an interpolation that resembles an upper envelope. Intuitively, this upper envelope can be
 450 thought of as a sheet that is laid on top of the points.

451 This approach is based on the premise that we tend to underestimate the NDVI (as argued
 452 in Cao et al. (2018)). Since we want to develop a general method that is in principle not
 453 related to the NDVI, we will not pursue this approach further.

454 **3.6 Performance Assessment**

455 Next, we will benchmark the different interpolation methods with and without robustifi-
 456 cation. For this, we will use the same technique as we did for the parameter determination
 457 in section 3.4. On B_λ we apply the RMSE and different quantiles.

458 The results are presented in section 5.1 and are discussed in section 6.2. The double logistic
 459 turns out to be the best convincing parametric method and form the non-parametric
 460 methods we choose the smoothing splines.

461 **Chapter 4**

462 **NDVI Correction**

463 Let's remind ourselves that the data from the S2 satellites is distributed with an SCL and
464 we therefore have some evidence about what is observed at each pixel for each sampled
465 time (c.f. table 2.2). So far, we have only considered points, labeled as cloud- and shadow-
466 free (SCL45). However, we remind ourselves that the satellite images in figure 2.3d, where
467 we had cloudy images despite the 'vegetation' label and see fragments in figure 2.3e even
468 though we are supposed to see cirrus clouds.

469 In this chapter we will try to improve our NDVI interpolation by not relying only on the
470 observed NDVI, but by training our own model to correct the NDVI using all S2 bands.
471 For this, we introduce several statistical modelling approaches and discuss the strengths
472 and weaknesses for each of them. After correcting the observed NDVI, we will assess the
473 uncertainties of our corrections and translate them into weights (in section 4.4). These will
474 be used for the subsequent interpolation. This step-by-step procedure is illustrated by the
475 figure B.4 in the appendix. Finally, we will evaluate which combination of interpolation
476 methods and correction method performs the best.

477 **4.1 Considering other SCL Classes**

478 In figure 4.1 we plot the observed NDVI and notice that some blue points which correspond
479 to the SCL-class 10 (thin cirrus clouds) follow the interpolated line closely. Hence, they
480 might be useful in improving an interpolation fit.

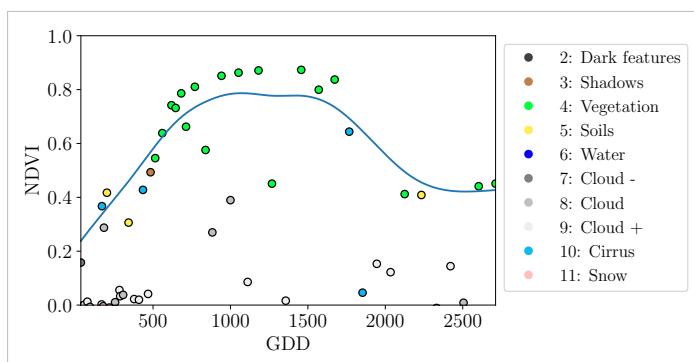


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45)

481 To get an impression of whether there is some useful information contained non-SCL45

482 observations we would like to compare the observed NDVI with the true NDVI. But since,
 483 we do not have any ground truth data, we will make the following assumption:

484 **Assumption 1.** The “true” NDVI value at time t can be successfully estimated by robustified
 485 LOOCV interpolation using high-quality observations. That is, the interpolated value
 486 (using an robustified interpolation method from chapter 3) considering the points $P^{SCL45} \setminus$
 487 P_t . In the following, we will call this estimate the “true”-NDVI.

488 We would like to get an idea if there is any information that can be recovered from non-
 489 SCL45 observations. For that, we will check for the other SCL-classes if there is a relation
 490 between the “true” NDVI (derived with robustified Smoothing Splines) and the observed
 491 NDVI. Thus, we pair each “true” NDVI with its observed one, collect all pairs, and create
 492 a scatter plot for each SCL-class in fig 4.2. As expected, the “true” and the observed
 493 NDVI seem to be highly correlated for SCL45. But we can also detect some patterns of
 494 correlation in the SCL-classes 2, 3, 7, 8 and 10.

495 It might be tempting to include some of the above SCL classes (for interpolation). But
 496 on the one hand, the choice would not be objective and on the other hand, the correlation
 497 seems to be weaker than for SCL45. Therefore, in the following section, we will correct
 498 the observed NDVI and estimate the uncertainty of each correction.

499 4.2 Response and Covariates

500 For training an NDVI correction model, we need ground-truth (response) and informative
 501 covariates. We organize those in a table, where each row corresponds to a P_t (i.e., a pixel at
 502 a time t). Since ground-truth NDVI data is not available, we will again use the assumption
 503 1 and use the “true” NDVI instead. There is no canonical answer to the question of
 504 which covariates we should use. It is a tradeoff between simplicity, generalizability and
 505 performance (with the danger of overfitting). Our desire with the NDVI correction is
 506 to develop a product that is simple for others to understand and use. Therefore, in the
 507 subsequent, we will only take the spectral data of the satellite (i.e. all the bands) and the
 508 observed NDVI derived from it as covariates.

509 4.3 Correction Methods

510 In the following, we will introduce different modelling approaches, which we will use to
 511 model the relation between the response $y = y_{\text{true OOB NDVI}} \in \mathbb{R}^n$ and the covariates
 512 encoded in the design matrix $X \in \mathbb{R}^{n \times p}$ which contains all covariates.

513 Note that in order to reduce computation time, only 10% of the training data has been
 514 used to fit the subsequent models which are still more than 120'000 observations.

515 4.3.1 Ordinary Least Squares (OLS)

516 The OLS is a linear model which aims to minimize the sum of the squared residuals. Let
 517 $y \in \mathbb{R}^n$ be the vector of responses and $X \in \mathbb{R}^{n \times p}$ be the design matrix, where each row
 518 corresponds to one pixel and each column consist of one covariate¹. We assume a linear
 519 relationship between y and X and allow for Gaussian noise. That is:

$$y = X\beta + \epsilon \quad \text{where } \epsilon \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

¹Strictly speaking, since SCL-classes are dummy variables

520 Assuming that X is regular, we can estimate the regression coefficients β by

$$\hat{\beta} = (X^T X)^{-1} X^T y = \arg \min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2$$

521 We will train two models, one using only the SCL-classes as covariates and the other one
 522 using all covariates (which are discussed in section 4.2).

Advantages	Disadvantages
— Simple method with good interpretability of coefficients.	— Catches only linear relationships. — No integrated variable selection. ²
— Computationally cheap.	

523 4.3.2 Least Absolute Shrinkage and Selection Operator (LASSO)

524 The LASSO can be similarly expressed than the OLS but adds a penalty to the minimization
 525 problem:

$$\hat{\beta}_\lambda = \arg \min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 = \arg \min_{\beta \in \mathbb{R}^p \text{ and } \|\beta\|_1 < \lambda} \|y - X\beta\|_2^2. \quad (4.3.2.1)$$

526 Even though we do not have a closed form solution for equation (4.3.2.1) we can solve
 527 it easily via optimization, since the function $\beta \in \{\beta \in \mathbb{R}^p | \|\beta\|_1 < \lambda\} \mapsto \|y - X\beta\|_2^2$ is
 528 continuous and convex.

529 Tibshirani (2011) shows that the LASSO solution tends to be sparse (for moderate λ).
 530 That is $\beta_i = 0$ for most $i = 1, \dots, p$

531 In order to know which λ to choose, we try a huge range of possible values. For each
 532 β_λ , we calculate the cross-validated $RMSE_\lambda$ ⁴ (and its standard deviation σ_λ using the k
 533 folds) and define the λ with the smallest corresponding $RMSE_\lambda$ as λ_{min} . From here we
 534 choose the largest λ for which the $RMSE_\lambda$ is smaller than $RMSE_{\lambda_{min}} + \sigma_\lambda$. This yields
 535 a simpler model while keeping the $RMSE$ reasonable model.

536 We will apply the LASSO using the selected covariates in section 4.2 and their second
 537 degree of interactions.⁵

Advantages	Disadvantages
— Usually yields a sparse solution. This tends to give better generalizability (prediction performance on unseen data).	— Estimate is biased. — Computationally expensive.
— Successfully deals with correlation in covariates.	
— Interpretable results.	

³The last two terms are equivalent by lagrangian optimization

⁴The cross validated Root Mean Square Error is the mean of the RMSE's obtained for each fold (using the model trained on the remaining folds). We use the following definition of the $RMSE$: $\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}$

⁵This is if our covariates are $\{a, b\}$, then we will now use $\{a, b, ab, a^2, b^2\}$.

538 **4.3.3 Random Forest (*RF*)**

539 To define a random Forest introduced by Breiman (2001) we will first define what a Tree
 540 is. A (*decision*) Tree is a graph (V, E) without circles, a distinct root node, every node
 541 has at most two children and every leaf has a value assigned to it. At each node there
 542 is a boolean condition testing if one variable is greater than some value and a pointer to
 543 one child depending on the boolean value. To evaluate a tree we start at the root node,
 544 test the boolean expression and go to the node indicated by the resulting pointer. This
 545 we repeat until we end up at a leaf-node, where we return the value assigned to it.

546 To build such a Tree, we will recursively partition the covariate space using greedy splits⁶
 547 decreasing the RMSE⁷ each time. If the set we want to split contains less than a certain
 548 amount of training points, we stop.

549 To build a *Random Forest* we will bootstrap-aggregate⁸ many such Trees⁹. The prediction
 550 of the Random Forest for a new point x is then the mean of the predictions from all the
 551 Trees.

Advantages	Disadvantages
— Captures non-linear relationships.	— The resulting (prediction) function is not continuous but locally constant.
— Captures all interactions and performs automatic variable selection.	— Computationally expensive.
— Can deal with missing data.	— No interpretability.

552 **4.3.4 Multivariate Adaptive Regression Splines (*MARS*)**

553 A MARS model as introduced in Friedman (1991) can be described by

$$g(x) = \sum_{m=0}^M \beta_m h_m(x),$$

554 where the h_m are simple functions (explained later) and the β_m are estimated via Least
 555 Squares.

556 In the building procedure of a MARS model, we first select many of those simple functions
 557 and later drop some of them to avoid overfitting. For the construction of those simple
 558 functions, define \mathcal{B} be the set of pairs of ‘hockystick functions’

$$\mathcal{B} := \left\{ (b_1, b_2) \mid (b_1(x), b_2(x)) = \left((x_j - d)_+, (d - x_j)_+ \right), d = X_{1,j}, \dots, X_{n,j}, j = 1, \dots, p \right\}$$

559 and the set $\mathcal{M} = \{1\}$ of all functions currently in the model. Now, consider \mathcal{C} the set of
 560 candidate functions-pairs

$$\mathcal{C} := \{(h(\cdot)b_1(\cdot), h(\cdot)b_2(\cdot)) \mid h \in \mathcal{M}, (b_1, b_2) \in \mathcal{B}\} \quad (4.3.4.1)$$

⁶For computational reasons, we will only use splits along one covariate. So we ‘cut’ our covariate space into rectangles.

⁷To calculate the RMSE, we need a prediction. Let P be the current partition, then the predicted value for some $x \in A \in P$ is the mean of the responses of all the points in A (included in the training data).

⁸That is we will sample (with replacement) several times n observations from our original data and fit a Tree to each such sample.

⁹Building the Tree, this time we will not test every covariate at each node (for the RMSE minimization) but a node-specific subsample of the covariates. Thus, also the “second best split” can be selected.

561 and select the pair (which when added to \mathcal{M} and the coefficients refitted) reduces the
 562 RMSE the most. Add the selected pair to \mathcal{M} and repeat until the RMSE reduction
 563 becomes insignificant.

564 Finally, to avoid overfitting, we prune the set \mathcal{M} by optimizing a LOOCV score.¹⁰

565 To reduce computational complexity, we follow the recommendation from REFStephen
 566 (2021) and restrict h in equation (4.3.4.1) to be of degree one (so it is also in a pair of \mathcal{B}).
 567 Consequently, \mathcal{C} contains functions with a degree of at most 2.

Advantages	Disadvantages
<ul style="list-style-type: none"> — Catches non-linear relationships. — Interpretability via functions in \mathcal{M} and their coefficients. — Allows for interactions with variable selection. 	<ul style="list-style-type: none"> — Computationally expensive (can be reduced by restricting the degree of interactions).

568 4.3.5 General Additive Model (*GAM*)

569 GAMs as described in Hastie and Tibshirani (1987) are a special case of Projection Pursuit
 570 Regression, where only the p directions parallel to the coordinate axes are considered. The
 571 result is different to a linear model since the coordinate functions are not restricted to be
 572 linear but are assumed to be non-parametric functions. The model can be written as:

$$g_{add}(x) = \mu + \sum_{i=1}^p g_j(x_j).^{11}$$

573 To estimate the non-parametric functions, we can use smoothing splines (ref sec. 3.3.6).
 574 For this let \mathcal{S}_j be the function which takes some $z \in \mathbb{R}^n$ and returns the smoothing splines
 575 fitted to $(X_{:,j}, z)$ where the smoothing parameter is optimized by GCV. Since we cannot
 576 fit all g_j simultaneously, we will use a strategy named Backfitting. We basically cycle
 577 through the indices $1, \dots, p$ and refit \hat{g}_j each time. The following illustrates the procedure:

- 1) $\hat{g}_1 = \mathcal{S}_1(y - \mu)$
 - 2) $\hat{g}_j = \mathcal{S}_j(y - \mu - \hat{g}_1(X_{:,1}) - \dots - \hat{g}_{j-1}(X_{:,j-1}))$ for $j = 2, \dots, p$
 - 3) $\hat{g}_1 = \mathcal{S}_1(y - \mu - \hat{g}_2(X_{:,2}) - \dots - \hat{g}_p(X_{:,p}))$
 - 4) $\hat{g}_j = \mathcal{S}_j(y - \mu - \sum_{k \neq j} \hat{g}_k(X_{:,k}))$ for $j = 2, \dots, p$
- \vdots

578 We repeat step 3) and 4) until the change falls below some tolerance.

Advantages	Disadvantages
<ul style="list-style-type: none"> — Captures non-linearity. — Good interpretability. 	<ul style="list-style-type: none"> — No automatic variable selection. — Computationally expensive.

Ich
finde die
sections,
in denen
Du die
Modelle
erklärest,
gut.
Allerd-

¹⁰This means that we perform an iterative procedure to reduce the number of functions in \mathcal{M} . For every function h in \mathcal{M} , we compute the model using $\mathcal{M} \setminus \{h\}$. We discard the function which – when excluding from \mathcal{M} – leads to the best LOOCV score.

¹¹where g_j is a real-valued function. For identifiability we also demand $\mathbb{E}[g_j(X_{:,j})] = 0$ for $j = 1, \dots, p$.

580 **4.4 Uncertainty Estimation**

581 Once we correct the NDVI using the models described in the previous section, we are left
 582 with the problem that not every correction is equally reliable.¹². Hence, we are interested
 583 in a measure of how uncertain an estimate is.

584 We do this by replacing the response with the absolute residuals $v := |y - \hat{y}|$ and modeling
 585 their relationship with the covariates defined by X . In this way, we obtain a model for
 586 the absolute residuals v and the estimator \hat{v} .

587 **4.5 Interpolation**

588 Consider now a pixel P , $\hat{y}^{(P)}$ its corrected NDVI and $\hat{v}^{(P)}$ the estimated uncertainties of
 589 $\hat{y}^{(P)}$. In order to interpolate $\hat{y}^{(P)}$, we will give less weight to unreliable observations. Thus,
 590 we define the weight function:

$$w_\tau^{(P)} := \frac{1}{R} \frac{1}{\hat{v}_\tau^{(P)}}, \quad \text{for } \tau = 1, \dots, n_P$$

591 where τ is an index over the satellite images and $R := \frac{\sum_i^{n_P} \hat{v}_i^{(P)}}{n_P}$ a normalization constant.
 592 The normalization is needed since for some interpolation methods, inflating the sum of
 593 weights would decrease the effect of the smoothing.

594 **4.6 Resulting Interpolation Strategies**

595 We have developed the following procedure to obtain a new interpolation (keyword-wise):

- 596 i.) OOB Interpolation (+ robustify?)
- 597 ii.) Correction
- 598 iii.) Uncertainty estimation
- 599 iv.) Interpolation (+ robustify?)

600 At each step we have a choice, more precisely:

- 601 — Interpolation: Smoothing Splines / Double Logistic
- 602 — Robustify: Yes / No
- 603 — Correction & uncertainty estimation: RF / OLS – considering only SCL-classes /
- 604 — OLS – considering all selected covariates / MARS / GAM / LASSO / no correction.

605 As it is not feasible to try every possible combination, we make the following restrictions
 606 on which combinations we will consider:

- 607 — We use the same interpolation method each time.
- 608 — Either we robustify both times, or we do not robustify at all.
- 609 — We use the same underlying method for correction and uncertainty estimation.

610 In this fashion, we obtain 28 distinct interpolation strategies, which we will benchmark in
 611 the next section.

¹²One correction is illustrated in the figure B.4f. In this figure, the outer points (labeled as clouds) have a large scatter.

612 4.7 Evaluation Method

613 In this section, we introduce the relative yield-estimation-accuracy (*RYEA*) and utilize it to
 614 evaluate the interpolation strategies from section 4.6. The fundamental assumption is that
 615 the closer the interpolated NDVI time series is to the true one, the better it can be used to
 616 determine crop yield. Implicitly, we believe that an NDVI time series which better models
 617 yield will incorporate more true information about the underlying vegetation. Therefore,
 618 we want to determine a comparable RYEA for each interpolation strategy and choose it
 619 as a benchmark criterion. This is an objective measure, since we have not considered crop
 620 yield in any of our previous steps. Moreover, this criterion is justified by the fact that
 621 yield estimation has been a motivation for the interpolation.

622 **Definition 4.7.0.1.** (*RYEA*) Let $y \in \mathbb{R}^n$ be the yield, M be a model for estimating y , and
 623 $\hat{y} = M(X)$ where X describes the data¹³. We define the *RYEA* as the relative RMSE in
 624 yield estimation. Formally expressed:

$$\text{RYEA} = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{y}},$$

625 where \bar{y} denotes the sample mean.

626 4.7.1 Yield Estimation

627 For all the pixels, we will interpolate the NDVI time series with every interpolation strat-
 628 egy. From the interpolated NDVI time series, we would like to estimate the yield. However,
 629 given the high dimensionality and different lengths of the interpolation (not every time
 630 series has the same start and end point), we must first map each NDVI time series into a
 631 low-dimensional vector space of covariates. For this, we will use the following statistics:

- Maximum slope
- Minimum slope
- Integral¹⁴ over all
- Peak (i.e. maximal NDVI)
- GDD for the Peak
- Integral¹⁴ up to the peak
- Integral¹⁴ after peak
- Integral¹⁴ from 0-685 GDD
- Integral¹⁴ from 685-1075 GDD

632 For the choice we were inspired by (c.f. table 2 in Kamir, Waldner, and Hochman (2020)).
 633 However, we deliberately omit any statistic that involves the minimum (e.g. the NDVI-
 634 range), since we regard the minimum as a very error-prone measure due to the large
 635 influence of clouds in the time series.

636 As a result, for each interpolation strategy, a matrix is obtained in which each row corre-
 637 sponds to a pixel and both the yield and the covariates (computed by applying the above
 638 statistics) are contained. Using this matrix, we train a random forest for yield estimation,
 639 and compute the integrated OOB estimates¹⁵ \hat{y} . Note that the choice of the modeling
 640 approach does not matter much, as long as it is general enough (i.e. able to approximate

¹³We will use the matrixes derived in section 4.7.1

¹⁴We will only consider the integral of the function $\max(0, NDVI - 0.3)$, where 0.3 is assumed to be a minimal NDVI value. REF

¹⁵By the integrated OOB estimates, we denote the predictions for each pixel where only trees are used, where the pixel has not been used (as n_{tree} , the number of Trees, grows the fraction of trees which do not contain a certain pixel converges to $\frac{1}{e}$).

any function) and we use the same one for each interpolation strategy. Finally, for each interpolation strategy, we calculate the RYEA and describe the results in section 5.2.

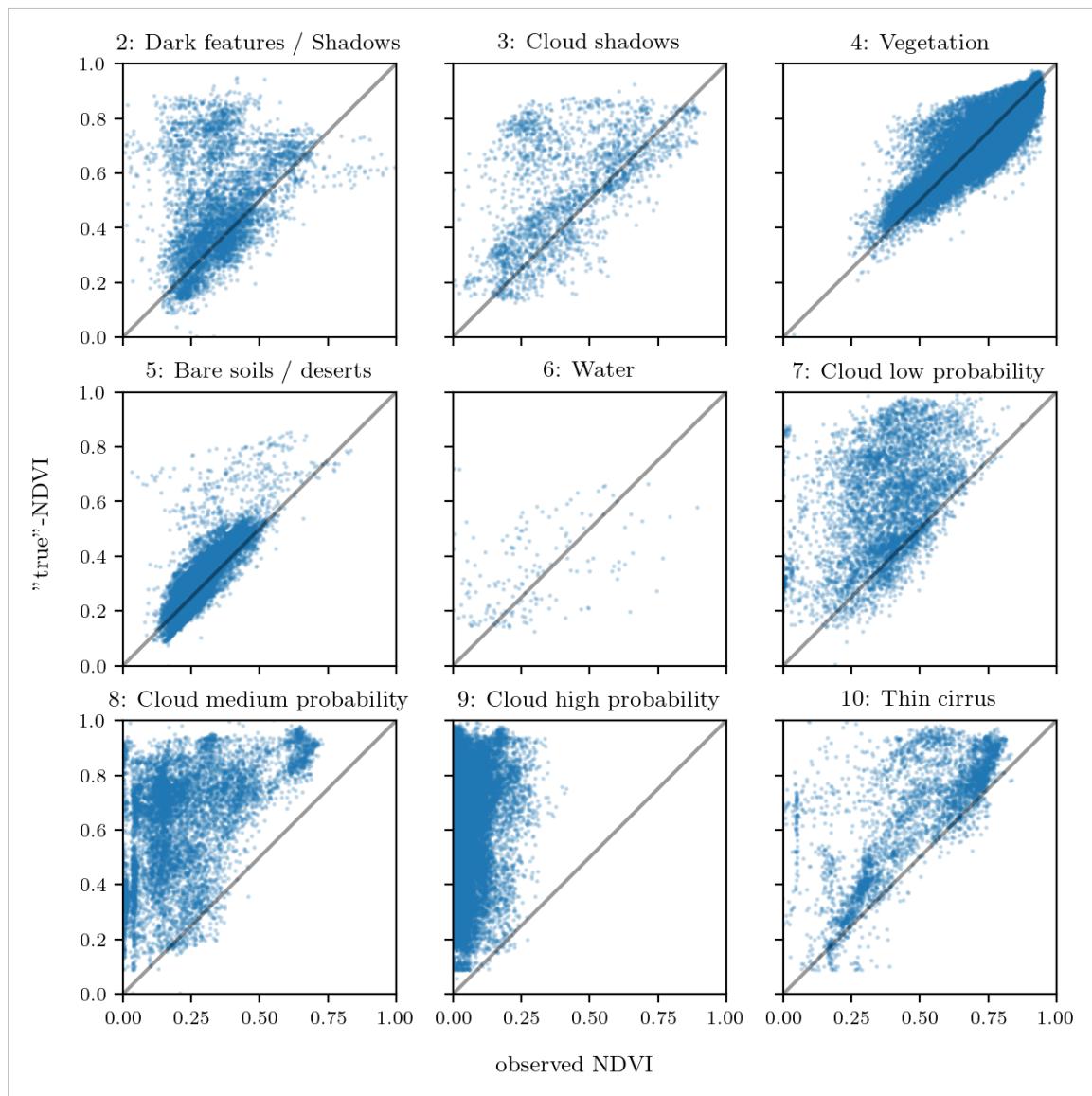


Figure 4.2: For each SCL class, we compare the true NDVI with the observed NDVI. (The true NDVI was estimated with OOB smoothing splines, and we used all observations of 10% of the total training pixels.)

643 **Chapter 5**

644 **Results**

645 **5.1 Goodness of Fit for Selected Interpolation Methods**

646 Table 5.1 benchmarks the interpolation methods (on P^{SCL45}) with respect to various
647 score functions. The score functions take the absolute values of the LOOCV residuals
648 and summarize them in a number (the smaller the better). For each of the 5 selected
649 interpolation methods, we consider the basic version and the robustified (see section 3.5)
650 version.

Table 5.1: Comparing the goodness of fit for different interpolation methods (on P^{SCL45}) measured with the score functions (which take the LOOCV residuals as input) listed in the left column. q_X denotes here the $X\%$ quantile.

	SS	LOESS	DL	BSPL	FR	SS^{rob}	$\text{LOESS}^{\text{rob}}$	DL^{rob}	$BSPL^{\text{rob}}$	FR^{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

651 DL is the best among both robustified and non-robustified with respect to most of the score
652 functions used (all except q95) and is especially superior to the other parametric approach,
653 which is Fourier interpolation. Especially the robust Fourier interpolation performs poorly.
654 The LOESS dominates (i.e. is superior on every score function) all other non-parametric
655 methods, but is closely followed by the SS. The BSPL, on the other hand, is the worst
656 non-parametric method tested here.

657 **5.2 Robustification and NDVI-Correction**

$$\begin{aligned}\widehat{\text{NDVI}}_{\text{corr}} = & 0.711 \text{NDVI}_{\text{observed}} + \mathbb{1}_{SCL=2} 0.215 + \mathbb{1}_{SCL=3} 0.237 + \mathbb{1}_{SCL=4} 0.210 \\ & + \mathbb{1}_{SCL=5} 0.116 + \mathbb{1}_{SCL=6} 0.162 + \mathbb{1}_{SCL=7} 0.327 + \mathbb{1}_{SCL=8} 0.474 \quad (5.2.0.1) \\ & + \mathbb{1}_{SCL=9} 0.575 + \mathbb{1}_{SCL=10} 0.306 + \mathbb{1}_{SCL=11} 0.512\end{aligned}$$

658 - strong upwards correction for SCL classes 8, 9 and 11 (correspond to ‘medium probability
659 clouds’, ‘high probability clouds’ and ‘thin cirrus clouds’).

Table 5.2: XXX RMSE of yield prediction. For the relative RMSE and the coefficient of determination (R^2) see table B.1 and B.2

	RF	OLS-SCL	OLS-all	MARS	GAM	LASSO	no-correction
ss	1.144	1.033	1.051	1.042	1.046	1.042	1.095
dl	1.150	1.115	1.116	1.116	1.097	1.098	1.159
ss-rob	1.144	1.054	1.084	1.094	1.072	1.071	1.091
dl-rob	1.159	1.128	1.117	1.064	1.093	1.105	1.156

$$\widehat{\text{abs}}(\text{NDVI}_{\text{true}} - \text{NDVI}_{\text{corr}}) = -0.133 \text{NDVI}_{\text{observed}} + \mathbb{1}_{SCL=2}0.186 + \mathbb{1}_{SCL=3}0.185 \\ + \mathbb{1}_{SCL=4}0.146 + \mathbb{1}_{SCL=5}0.089 + \mathbb{1}_{SCL=6}0.167 \\ + \mathbb{1}_{SCL=7}0.203 + \mathbb{1}_{SCL=8}0.181 + \mathbb{1}_{SCL=9}0.173 \\ + \mathbb{1}_{SCL=10}0.180 + \mathbb{1}_{SCL=11}0.172 \quad (5.2.0.2)$$

660 - the higher the observed NDVI the lower the estimated absolute residual. - estimated
 661 absolute residuals are the smalles for SCL classes 4 and 5.

662 **Chapter 6**

663 **Discussion**

664 Here in the discussion, you should take up the points you mentioned in the introduction

665 SCL is prone to errors as can be seen in figure 2.3. A machine learning approach like the
666 one developed in [Raiyani, Gonçalves, Rato, Salgueiro, and Marques da Silva \(2021\)](#) could
667 be used instead.

668 **6.1 Data Gaps**

669 fourier schlecht bei wenigen beobachtungen, siehe rob in [5.1](#)

670 Strange behaviour of LOESS prefer ss because of the smoothness guarantees (compare the
671 figures [B.1](#) and [3.5](#))

672 **6.2 Interpolation Methods**

673 Given that DL convinces regarding most of the selected score functions in table [5.1](#) we will
674 certainly investigate this method in chapter [4](#). Moreover, we see that the robustification
675 mostly improved the score regarding the 50, 75, 85, and 90 % Quantiles. Only for the
676 outlier-sensitive score functions (RMSE and q95)¹ we notice significant worsening (we
677 consider the robust fourier separately in section [6.1](#)). Consequently, we will also use the
678 robustification in section [4](#). Not wanting to rely on the form assumptions of the DL, we
679 further choose a non-parametric method for further consideration. Despite the LOESS
680 slightly dominating the SS in table [5.1](#), we choose the SS. This is due to the strange
681 behavior of the LOESS in case of data gaps (see section [6.1](#)) and the good interpretability
682 of the SS using the minimization function [3.3.6.1](#).

683 XXX discuss results from table [5.2](#)

¹For the RMSE one outlier is enough to take away the usefulness of the statics, in the case of q95 it is enough if 5% of the data are corrupt to break the statics.

684 **6.3 NDVI Correction**685 **6.3.1 Bootstrap**

686 The question arises if we can build the correction model on the same year as we want to
 687 apply it on. Usually, a similar approach might carry the danger of overfitting. However, we
 688 have not used any ground truth at any point (until the evaluation). Instead, we estimated
 689 the “true” NDVI with the assumption 1 via OOB. Thus, we have bootstrapped our way
 690 out of the problem. Consequently, we reason that we can apply our method to a new
 691 (comparable) dataset and solve the correction again via this bootstrap.

693 **6.3.2 Using Additional Covariates**

694 In section 4.2 we have only used the spectral data (and the observational NDVI calculated
 695 from them) as covariates. Since we have the weather data available (c.f. REF-SEC), it
 696 would be a small effort to incorporate it, together with statistics collected from it (i.e.
 697 GDD or ‘rainfall in the last 30 days’).

698 We decided against using this data, because on the one hand we have the problem that
 699 we have practically too few observations (we observe only 5 years) and we expect the
 700 weather in our study region to be rather homogeneous which is suggested by the fact
 701 that the weather data published by Meteoswiss are for a grid with a resolution of 1 km.
 702 On the other hand, we want the underlying model not to learn improper relationships.
 703 For example, the model might automatically predict a high NDVI for a day in summer
 704 (detected by high GDD / many sunshine hours / high temperature) just because it is
 705 “used” to observing a lot of vegetation in summer. Including temporally (e.g., P_{t-1} and
 706 P_{t+1}) and geographically adjacent pixels would likely improve performance. However, for
 707 simplicity, we omit it here².

708 - weight/uncertainty function (problem of weight function -> some outer points get really
 709 low weights (just because others in the middle have very little residuals and thus very high
 710 weight))

where
does
this sec-
tion be-
long to?
Chapter
‘NDVI
Correc-
tion’ or
‘Further
Work’?

711 **6.3.3 High RMSE in Yield Prediction**

712 How much can we expect to get? We have multiple sources of uncertainty in the data:

- 713 i.) Uncertainty in Yield data collected by the combine harvester
- 714 ii.) Uncertainty in Yield data through rasterization
- 715 iii.) Uncertainty in satellite images through “measurement errors” introduced via clouds
 716 and other atmospheric effects
- 717 iv.) Uncertainty introduced by interpolating (especially when long data-gaps are present)

You already capture the ”main” structure of your thesis with the interpolation and the
 NDVi correction sections. Can you combine them both in a ”synthesis” subsection at
 the end of the discussion?

718

²This is done for simplicity of understanding and using the model, since one would need to adapt to some convention of how to supply the data of adjacent pixels without redundancy (i.e. supplying P_t multiple times).

719 **Chapter 7**

720 **Conclusion**

721

722

```
- itpl methods,  
  parametric dl  
  non-param  
  discarded  
  kernel methods because of strong bias  
  kriging because assumptions and highdim parameters  
  savitzky-golay filter since we will investigate the LOESS which can be thought a  
  loess slightly best performing itpl method but we notice non-smooth behaviour if  
  loess > ss > bsp  
  choose ss because of its meaningful definition (minimizing the integral of the second  
  - robustifying useful?
```

733

734 XXX draw your conclusion to which you came during this thesis

735 **7.1 Future Work**

736 **7.1.1 Time Series Correction-Interpolation as a General Method**

737 Throughout this thesis, we developed a correction and interpolation method for the NDVI.
738 However, we never used features of the NDVI. Only the parameter estimated via cross-
739 validation in chapter 3.4 depends on the scale of the time series. For simplicity, we could
740 thus determine the parameter using Generalized Cross Validation (as Ripley and Maechler
741 suggest). Therefore, our approach of interpolation and correction of time series can be
742 applied to arbitrary time series as long as additional information is available. However,
743 further research is required, to demonstrate the usefulness of this approach in general.

744 **Example: Cloud Correction with Uncertainty Estimation and Interpolation**

745 This generalization can be used in particular for cloud correction. In the same manner as
746 we corrected the NDVI time series in chapter 4, we can correct each spectral band and
747 reunite the corrected bands with the uncertainties. If desired, the time series can also be
748 interpolated before merging as in chapter 4.5. The resulting question would be how well
749 this approach performs.

750 **7.1.2 Minor Improvements**

751 During this project, we also noticed some minor issues that we would have liked to invest-
752igate further if more resources were available. The most relevant of these are:

- 753 — **Data:** Method how data has been extrapolated to the grid could possibly be improved
754 — **Data:** For computational reasons, we mostly considered all years and split the data
755 (on the pixel level) randomly into a train/test set. A leave one year out cross
756 validation might yield more accurate results.
757 — **Data:** We have not included the spectral bands which have a resolution of 60m. But
758 precisely these seem to be promising for cloud correction, since they are a proxy of
759 the water (content and form) in the atmosphere.
760 — **NDVI Correction:** Explore the effect of different link functions between the esti-
761 mated absolute residuals and the weights in section 4.5.
762 — **NDVI Correction:** Yield is not the only target variable of interest. Other variables
763 like protein content could also be used in section 4.7 for the method evaluation.

which
data? I
assume
the
com-
bine
har-
vester
point
data?

764 Bibliography

- 765 (2007). Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.),
766 *Model-Based Geostatistics*, pp. 46–78. New York, NY: Springer.
- 767 Bailey, S. J. (2018, July). Using Growing Degree Days to Predict Plant Stages. pp. 8.
- 768 Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore (2006,
769 February). Improved monitoring of vegetation dynamics at very high latitudes: A new
770 method using MODIS NDVI. *Remote Sensing of Environment* 100(3), 321–334.
- 771 Breiman, L. (2001, October). Random Forests. *Machine Learning* 45(1), 5–32.
- 772 Brockmann, M., T. Gasser, and E. Herrmann (1993, December). Locally Adaptive Band-
773 width Choice for Kernel Regression Estimators. *Journal of the American Statistical
774 Association* 88(424), 1302–1309.
- 775 Cao, R., Y. Chen, M. Shen, J. Chen, J. Zhou, C. Wang, and W. Yang (2018, November). A simple method to improve the quality of NDVI time-series data by integrating
776 spatiotemporal information with the Savitzky-Golay filter. *Remote Sensing of Environ-
777 ment* 217, 244–257.
- 778 Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004, June). A simple method for reconstructing a high-quality NDVI time-series data set based on the
779 Savitzky–Golay filter. *Remote Sensing of Environment* 91(3), 332–344.
- 780 Cleveland, W. S. (1979, December). Robust Locally Weighted Regression and Smoothing
781 Scatterplots. *Journal of the American Statistical Association* 74(368), 829–836.
- 782 Friedman, J. H. (1991, March). Multivariate Adaptive Regression Splines. *The Annals of
783 Statistics* 19(1), 1–67.
- 784 Hastie, T. and R. Tibshirani (1987, June). Generalized Additive Models: Some Applica-
785 tions. *Journal of the American Statistical Association* 82(398), 371–386.
- 786 Jaramaz, D., V. Perović, S. Belanovic Simic, E. Salnikov, D. Cakmak, V. Mrvić, and
787 L. Zivotic (2013, May). The ESA Sentinel-2 mission Vegetation variables for Remote
788 sensing of Plant monitoring.
- 789 Kamir, E., F. Waldner, and Z. Hochman (2020, February). Estimating wheat yields
790 in Australia using climate records, satellite image time series and machine learning
791 methods. *ISPRS Journal of Photogrammetry and Remote Sensing* 160, 124–135.
- 792 Lyche, T. and K. Mørken (2005, January). Spline Methods.
- 793 McMaster, G. S. and W. W. Wilhelm (1997, December). Growing degree-days: One
794 equation, two interpretations. *Agricultural and Forest Meteorology* 87(4), 291–300.

- 797 Perich, G., M. O. Turkoglu, L. V. Graf, J. D. Wegner, H. Aasen, A. Walter, and F. Liebisch
798 (2022, July). Pixel-based yield mapping and prediction from Sentinel-2 using spectral
799 indices and neural networks.
- 800 Raiyani, K., T. Gonçalves, L. Rato, P. Salgueiro, and J. R. Marques da Silva (2021,
801 January). Sentinel-2 Image Scene Classification: A Comparison between Sen2Cor and
802 a Machine Learning Approach. *Remote Sensing* 13(2), 300.
- 803 Ripley, B. D. and M. Maechler. R: Fit a Smoothing Spline. [https://stat.ethz.ch/R-
804 manual/R-patched/library/stats/html/smooth.spline.html](https://stat.ethz.ch/R-manual/R-patched/library/stats/html/smooth.spline.html).
- 805 Rouse, J. W. (1974, May). Monitoring the vernal advancement and retrogradation (green
806 wave effect) of natural vegetation. Technical Report NASA-CR-139243.
- 807 Savitzky, A. and M. J. E. Golay (1964, July). Smoothing and Differentiation of Data by
808 Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), 1627–1639.
- 809 Schafer, R. W. (2011, July). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE
810 Signal Processing Magazine* 28(4), 111–117.
- 811 Skakun, S., E. Vermote, B. Franch, J.-C. Roger, N. Kussul, J. Ju, and J. Masek (2019,
812 July). Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating
813 Surface Reflectance, Through Phenological Fitting, into Regression Yield Models.
814 *Remote Sensing* 11(15), 1768.
- 815 Stephen, M. (2021, July). Earth: Multivariate Adaptive Regression Splines.
- 816 Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective.
817 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(3), 273–
818 282.

819 **Appendix A**

820 **Reproducibility**

821 **A.1 Reproduce Results**

822 For reproducibility of the whole computations, we refer to our codebase at:

823 <https://github.com/LGraz/MasterThesis-Code>

824 In order to reproduce our computations and results, set up the directory as described
825 in the README and execute the computations via `./shell_scripts/reproduce.sh`
826 and do not execute the python and R scripts by hand (unless you follow the order in
827 `./shell_scripts/reproduce.sh`).

828 **A.2 R-Package**

829 We also provide an R package for a general time series correction and interpolation if
830 additional data is available at:

831 <https://github.com/LGraz/CorrectTimeSeries>

832 In our case we consider the NDVI time series and the additional data consists of the unused
833 spectral bands.

834 We recommend installing it via the `devtools` package by:

835 `devtools::install_github("LGraz/CorrectTimeSeries")`

836 In the following, we shall give a stand-alone example of how the R package can be used:

```
837 1 library(CorrectTimeSeries)
838 2
839 3 # load a list of dataframes, each one describes one pixel with the covariates and
840 4 # the response
841 5 data(timeseries_list)
842 6 str(timeseries_list[[1]])
843 7
844 8 # Train/Load RF
845 9 train_model_myself <- TRUE
846 10 if (train_model_myself){
847 11     # Add "true" NDVI (or generally the response), by Out-Of-Bag estimation
848 12     timeseries_list <- lapply(timeseries_list, function(df) {
849 13         df$oob_ndvi <- OOB_est(df$gdd, df$ndvi_observed) # gdd is the time-axis
850 14         df
851 15     })
852 16     # Train correction model
853 17     formula <- "oob_ndvi ~ B02+B03+B04+B05+B06+B07+B08+B8A+B11+B12+scl_class"
854 18     RF <- train_RF_with_fromula(formula, timeseries_list, robustify=TRUE)
855 19 } else {
```

```
857 19  data(RF_for_NDVI)
858 20  RF <- RF_for_NDVI
859 21 }
860 22
861 23 # ADD CORRECTION
862 24 timeseries_list <- lapply(timeseries_list, function(df) {
863 25   df$corrected_ndvi <- randomForest:::predict.randomForest(RF, df)
864 26   df
865 27 })
866 28
867 29 # Get interpolation for each timeseries
868 30 newx <- 1:1000
869 31 lapply(timeseries_list, function(df){
870 32   ss <- smoothing_spline(df$gdd, df$corrected_ndvi)
871 33   predict(ss, newx)$y
872 34 })
```

Example of how to use the `CorrectTimeSeries` package

874 **Appendix B**

875 **Further Material**

876 **B.1 Data and Methods**

877 **B.1.1 GDD**

878 Bailey (2018) tabulates the corresponding GDD for each stage of wheat.

Stage	Description	GDD
Emergence	Leaf tip just emerging from above-ground coleoptyle.	125 – 160
Leaf development	Two leaves unfolded.	169 – 208
Tillering	First tiller visible	369 – 421
Stem elongation	First node detectable.	592 – 659
Anthesis	Flowering commences; first anthers of cereals are visible.	807 – 901
Seed fill	Seed fill begins. Caryopsis of cereals watery ripe (first grains have reached half of their final size).	1068 – 1174
Dough stage	Soft dough stage, grain contents soft but dry, fingernail impression does not hold.	1434 – 1556
Maturity complete	Grain is fully mature and drydown begins. Ready for harvest when dry.	1538 – 1665

879 **B.2 Interpolation**

880 **B.3 NDVI correction**

881 page breaks

882
883 1 Call:
884 2 lm(formula = (paste(response, " ~ ", "ndvi_observed + scl_class")) ,

Table B.1: XXX RMSE of yield prediction

	RF	OLS-SCL	OLS-all	MARS	GAM	LASSO	no-correction
ss	0.155	0.140	0.143	0.142	0.142	0.142	0.149
dl	0.156	0.151	0.152	0.152	0.149	0.149	0.158
ss-rob	0.155	0.143	0.147	0.149	0.146	0.145	0.148
dl-rob	0.157	0.153	0.152	0.145	0.148	0.150	0.157

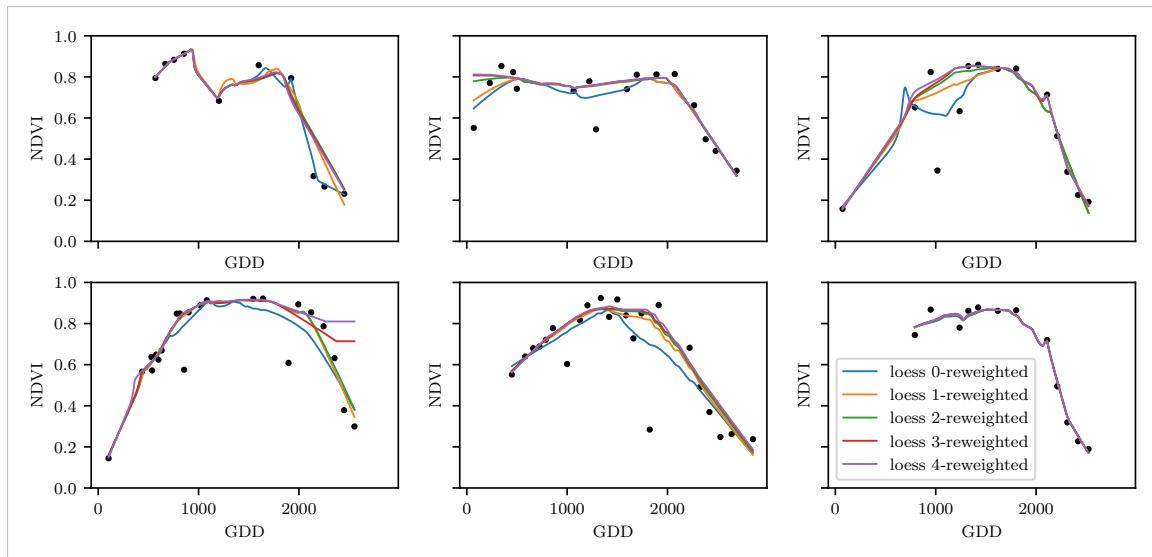


Figure B.1: The LOESS smoother fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed

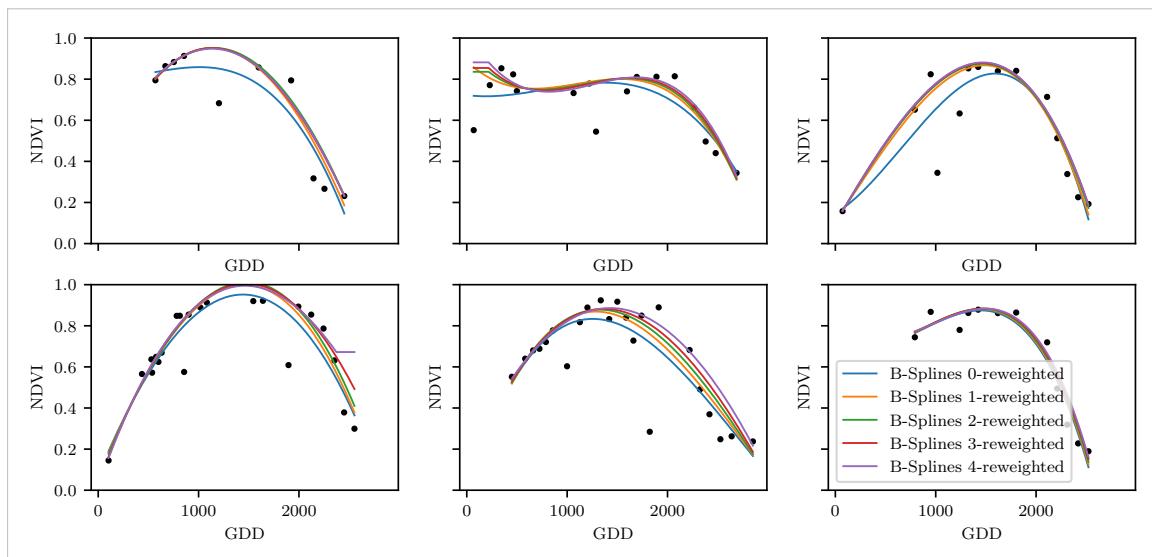


Figure B.2: B-Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed

```

885   3   data = ndvi_df)
886
887 Residuals:
888   Min     1Q   Median     3Q    Max
889 -0.7997 -0.0717  0.0039  0.0695  0.6632
890
891 Coefficients:
892             Estimate Std. Error t value Pr(>|t|)
893 (Intercept) 0.21465  0.00230  93.46 < 2e-16 ***
894 ndvi_observed 0.71116  0.00346 205.65 < 2e-16 ***
895 scl_class3   0.02205  0.00356   6.20 5.8e-10 ***
896 scl_class4   -0.00431  0.00251  -1.72  0.085 .
897 scl_class5   -0.09875  0.00234 -42.15 < 2e-16 ***
898 scl_class6   -0.05301  0.01104  -4.80 1.6e-06 ***

```

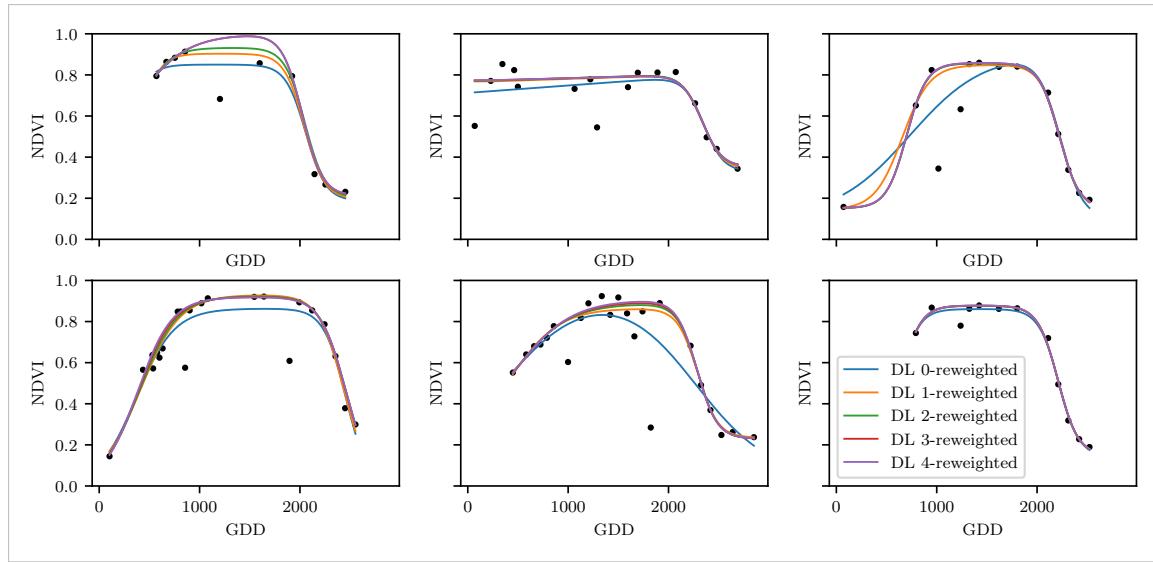


Figure B.3: A Double Logistic curve fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed

Table B.2: XXX RMSE of yield prediction

	RF	OLS-SCL	OLS-all	MARS	GAM	LASSO	no-correction
ss	0.431	0.486	0.477	0.481	0.479	0.481	0.455
dl	0.427	0.445	0.444	0.444	0.454	0.453	0.423
ss-rob	0.431	0.475	0.461	0.456	0.467	0.467	0.457
dl-rob	0.423	0.439	0.444	0.470	0.456	0.450	0.424

```

899 17 | scl_class7      0.11245    0.00274   41.09 < 2e-16 ***
900 18 | scl_class8      0.25963    0.00253   102.57 < 2e-16 ***
901 19 | scl_class9      0.35994    0.00236   152.47 < 2e-16 ***
902 20 | scl_class10     0.09091    0.00308   29.54 < 2e-16 ***
903 21 | scl_class11     0.29784    0.00392   76.06 < 2e-16 ***
904 22 |
905 23 | Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
906 24 |
907 25 | Residual standard error: 0.146 on 124978 degrees of freedom
908 26 | Multiple R-squared:  0.532,   Adjusted R-squared:  0.532
909 27 | F-statistic: 1.42e+04 on 10 and 124978 DF, p-value: <2e-16
910

```

R Summary of the NDVI correction model (c.f. equation 5.2.0.1)

```

911 1 | Call:
912 2 | lm(formula = (paste(get_res(), " ~ ", "ndvi_observed + scl_class"))),
913 3 |   data = ndvi_df)
914 4 |
915 5 | Residuals:
916 6 |   Min     1Q Median     3Q    Max
917 7 | -0.2051 -0.0427 -0.0074  0.0329  0.6589
918 8 |
919 9 | Coefficients:
920 10|            Estimate Std. Error t value Pr(>|t|)
921 11| (Intercept) 0.18647   0.00126 147.74 < 2e-16 ***
922 12| ndvi_observed -0.13265   0.00190 -69.80 < 2e-16 ***
923 13| scl_class3   -0.00180   0.00196 -0.92  0.3587
924 14| scl_class4   -0.04069   0.00138 -29.55 < 2e-16 ***
925 15| scl_class5   -0.09698   0.00129 -75.32 < 2e-16 ***
926 16| scl_class6   -0.01906   0.00606 -3.14  0.0017 **
927

```

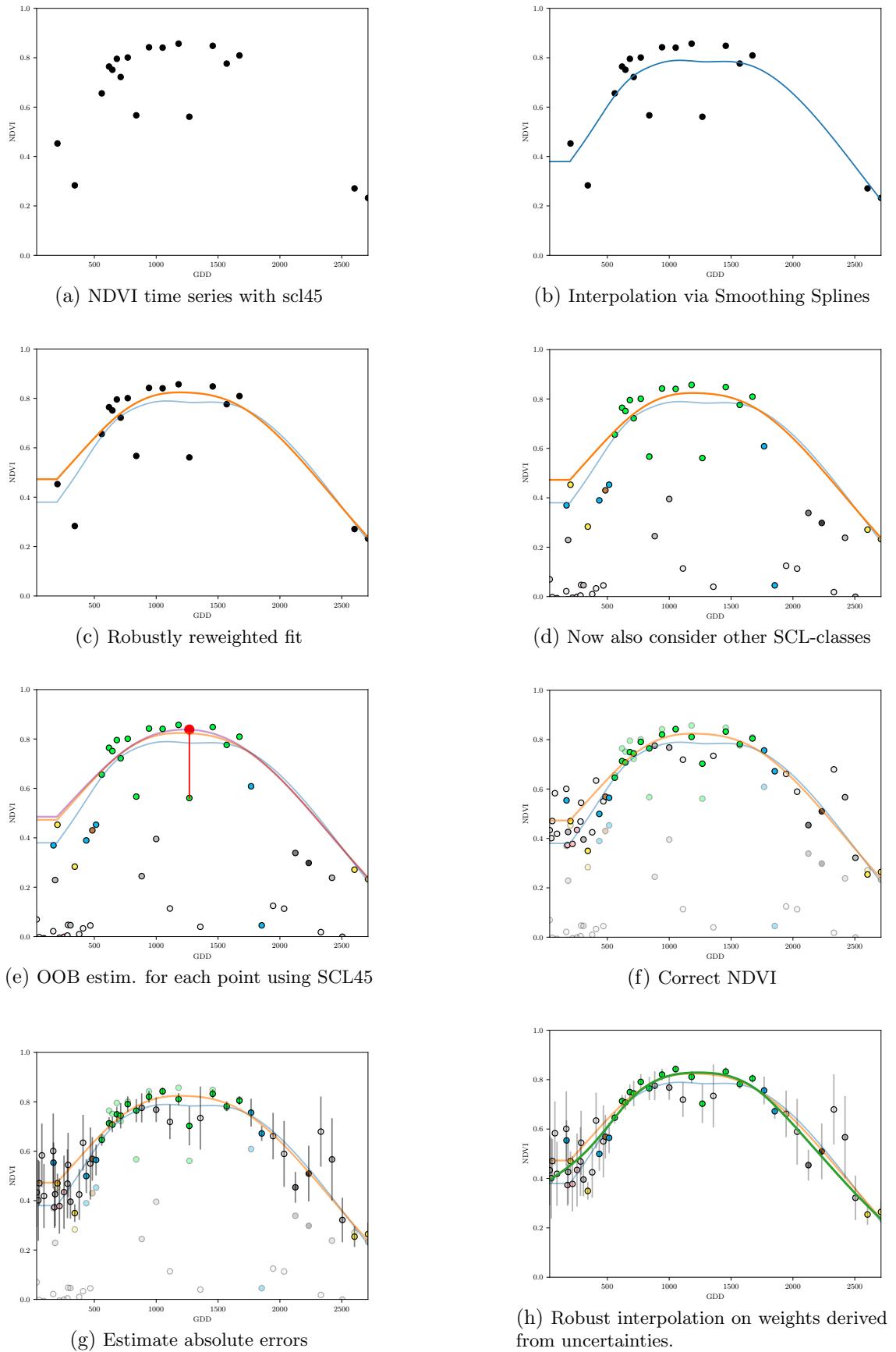


Figure B.4: Stepwise illustration of robust NDVI-Correction. For the color encoding of the SCL classes we refer to table 2.2.

```
928 17 | scl_class7      0.01641    0.00150   10.91  < 2e-16 ***  
929 18 | scl_class8     -0.00560    0.00139   -4.02  5.7e-05 ***  
930 19 | scl_class9     -0.01384    0.00130   -10.67 < 2e-16 ***  
931 20 | scl_class10    -0.00690    0.00169   -4.08  4.5e-05 ***  
932 21 | scl_class11    -0.01446    0.00215   -6.72  1.8e-11 ***  
933 22 | ---  
934 23 | Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
935 24 |  
936 25 | Residual standard error: 0.08 on 124978 degrees of freedom  
937 26 | Multiple R-squared:  0.352,    Adjusted R-squared:  0.352  
938 27 | F-statistic: 6.8e+03 on 10 and 124978 DF, p-value: <2e-16
```

R Summary of the NDVI correction model (c.f. equation 5.2.0.2)

940 replace space before ref by tilda

941 check quantile definitions