



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

9 **Interpolation and Correction**

10 **of**

11 **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	vii
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	1
47	1.5 Roadmap – anderer name XXX	2
48	2 Data and Methods	3
49	2.1 Available Data	3
50	2.1.1 Sentinel 2 Data	3
51	2.1.2 Crop Yield Data	5
52	2.1.3 The Concept of a ‘Pixel’	5
53	2.2 General Methods	7
54	2.2.1 (Relative) RMSE	7
55	2.2.2 Out-Of-Bag (<i>OOB</i>) and Leave-One-Out-Cross-Validation (<i>LOOCV</i>)	7
56	2.2.3 Generalized Cross Validation (GCV)	7
57	3 Interpolation Methods	8
58	3.1 DAS vs. GDD	8
59	3.2 Interpolation Setup	8
60	3.3 Parametric Regression	10
61	3.3.1 Double Logistic	10
62	3.3.2 Fourier Approximation	11
63	3.3.3 Optimization Issues	11
64	3.4 Non-Parametric Regression	11
65	3.4.1 Kernel Regression	12
66	3.4.2 Kriging	12
67	3.4.3 Savitzky-Golay Filter (SG Filter)	14
68	3.4.4 Locally Weighted Regression (LOESS)	15
69	3.4.5 B-splines	16
70	3.4.6 Natural Smoothing Splines	17
71	3.5 Tuning Parameter Estimation	17
72	3.6 Robustification	18
73	3.6.1 Our Adjustment:	19
74	3.6.2 Examples and Conclusions	19
75	3.6.3 Upper Envelope Approach - Penalty for Negative Residuals	19
76	3.7 Performance Assessment	20
77	3.8 XXX Evaluation	20
78	4 NDVI Correction	21
79	4.1 Considering other SCL Classes	21
80	4.2 Correction	23
81	4.2.1 Response and Covariates	23
82	4.2.2 Correction Methods	23
83	4.2.3 Uncertainty Estimation	27

84	4.2.4 Interpolation	27
85	4.3 Resulting Interpolation Strategies	27
86	4.4 Evaluation Method	28
87	4.4.1 Idea	28
88	4.4.2 Yield Estimation	28
89	5 Results	30
90	5.1 XXX small recap from “Interpolation Methods”	30
91	5.2 Robustification and NDVI-Correction	30
92	6 Discussion	32
93	6.1 Interpolation Methods	32
94	6.2 NDVI Correction	32
95	6.2.1 Bootstrap	32
96	6.2.2 Using Additional Covariates	32
97	6.2.3 High RMSE in Yield Prediction	33
98	7 Conclusion	34
99	7.1 Future Work	34
100	7.1.1 Time Series Correction-Interpolation as a General Method	34
101	7.1.2 Minor Improvements	35
102	Bibliography	36
103	A Reproducibility	38
104	A.1 Reproduce Results	38
105	A.2 R-Package	38
106	B Further Material	40
107	B.1 Interpolation	40
108	B.2 NDVI correction	40

109 Todo list

110 Die Kern-Resultate müssen auch in den Abstract. Ebenso würde ich die vollständige 111 Reproduzierbarkeit und die R-Package erwähnen.	iii
112 Hier u.U. nach dem Doppelpunkt ein Absatz/Tab einführen, s.d. die definition und 113 die Erklärung räumlich separiert sind?	viii
114 Why do we do interpolation in NDVI (and other indices) time series? What are 115 possible shortcomings thereof?	1
116 introduce SCL45	3
117 Hier noch erwähnen, dass die SCL das Resultat eines Algorithmus der ESA ist. 118 Das kann in der Diskussion dann auch wieder aufgenommen werden & kritisch 119 hinterfragt werden.	3
120 Hier noch eine NDVI Zeitreihe parallel dazu zeigen. Ansonsten wird nicht klar, 121 warum wir die Interpolation überhaupt machen.	3
122 which challenges? were they introduced earlier? E.g. in the introduction?	3
123 Hier bitte noch eine kleine Beschreibung der Crop-yields reinnehmen. Also was sind 124 die Ertragswerte der Daten.	5
125 Please clarify this in more detail. We used pixels flagged with SCL 4 & 5, but as 126 can be seen in Fig. 2.1 d), this can yield erroneous NDVI values, etc.	5
127 Für den Leser wäre es interessant, wenn Du noch kurz die wichtigsten GDD Werte 128 aus der Literatur beschreiben würdest (D.h. z.B. Sowing, Emergence of Plants, 129 Anthesis, Senescence, Harvest)	6
130 necessary info for what? To answer the research questions asked in section XXX . . 131 which?	6
132 which context? I would write: "... introduced in chapter/section XXX in more detail. 133 definitition here and also relative	7
134 Hier fehlt mir eine kurze Erklärung, was OOB und LOOCV sind.	7
135 definition here and explanation why (computational cheap)	7
136 verdeutliche dem leser, dass ein auftrag das findne von interpolationmethoden war .	8
137 put section in methods / data	8
138 Findet man hier noch Literatur, in welcher ähnliches diskutiert wurde, die man 139 zitieren kann?	8
140 add fourier and fix order to match chapters	9
141 Paper zitieren wo eingeführt oder wo benutzt (falls einföhrung fast schon trivial) .	9
142 Ähnliche struktur sich überlegen	10
143 Die Aufzählung ist hier m.M.n. nicht so passend für einen "Fliesstext".	10
144 TODO: include Weighted versions	11
145 Die Aufzählung ist hier m.M.n. nicht so passend für einen "Fliesstext".	12
146 ohne unterkapitel struktur	13
147 figure / tabelle / pseudocode anstatt aufzählung	15

148	consider naming the sub-plots	19
149	write out keywords, after final results	20
150	hier könnte man auch wieder bezug nehmen auf die originale Sektion, wo man SCL einführt. dort fehlt momentan auch die Erklärung, dass SCL ein model output ist	23
153	Ich finde die sections, in denen Du die Modelle erklärt, gut. Allerdings fehlt mir die Überleitung/Einleitung, warum die Modelle gebraucht werden	27
155	tabelle wäre sauberer	28
156	check reference	29
157	welche characterizing statistics genau?	29
158	shoud w write 1:1 the sam es in the end of section 3	30
159	Here in the discussion, you should take up the points you mentioned in the introduction	32
160	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?	32
163	where does this section belong to? Chapter 'NDVI Correction' or 'Further Work'? .	32
164	which data? I assume the combine harvester point data?	35
165	page breaks	40

166 Notations

167 Variables

Hier u.U. nach dem Doppelpunkt ein Absatz/Tab einführen, s.d. die definition und die Erklärung räumlich separiert sind?

168

169 c : a (vector of) constant(s)

170 $\lambda \in \mathbb{R}$: a scalar

171 $n \in \mathcal{N}$: sample size

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

173 $x \in \mathbb{R}^n$: covariate in 1-dim interpolation setting

174 $w \in \mathbb{R}^n$: a vector of weights for each location x

175 $y \in \mathbb{R}^n$: response in 1-dim interpolation setting

176 $\hat{y} \in \mathbb{R}^n$: estimate of y

177 $\bar{y} \in \mathbb{R}$: sample mean of y

178 $r \in \mathbb{R}^n$: residuals given by $y - \hat{y}$

179 Abbreviations and Objects

180 Pixel: A pixel originates of an image pixel and describes a square of 10 x 10 meters in
181 the field which coincides with the resolution (and location) of the Sentinel-2 pixels. Such
182 pixels are illustrated in figure 2.2b. Additional information like yield is also attached.

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

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

187 SCL: Scene Classification Layer provided by the European Space Agency (ESA) that gives
188 an estimation of the land cover class of each pixel. It indicates what one can expect at a
189 pixel at a sampled time. For an overview, c.f. table 2.2

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

- 193 NDVI: Normalized Difference Vegetation Index ([Rouse](#), [Rouse](#))
- 194 DAS: Days After Sowing
- 195 GDD: growing degree days – cumulative sum of (temperature – threshold)⁺
- 196 XXX ML models and their shortnames
- 197 RYEA : relative yield-estimation-accuracy. Definition [4.4.0.1](#)
- 198 OOB : out-of-box. Describes the procedure if we estimate the value for a point but not
199 consider the point itself (cf. section [2.2.2](#))

200 **MATLAB Matrix Notation**

- 201 We will use the MATLAB ‘:’ notation to indicate rows and columns of a matrix. That is
202 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
203 X .

204 **Chapter 1**

205 **Introduction**

206 **1.1 XXX motivation - why is it important**

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

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

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

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

213 “Similarly, smoothing the time series of satellite data is helpful to address inconsistency
214 in observation frequency and timing due to clouds and other sensor artefacts Skakun,
215 Vermote, Franch, Roger, Kussul, Ju, and Masek (Skakun et al.)”

216 **1.2 XXX problembaum / fragestellungen**

217 problem schilderung anhand referenzen und evtl. eines bileds:

218 **1.3 XXX State-of-the-art**

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

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

- 221 — Doublelogistic (winter-ndvi)
- 222 — parametric / non-parametric approaches
- 223 — spatio-temporal approaches

224 **1.4 Research Questions**

225 XXX

226 1.5 Roadmap – anderer name XXX

227 This thesis is structured as follows: XXX

228

Chapter 2

229

Data and Methods

230 introduce SCL45

231

2.1 Available Data

232 Our study region is a farm of over 800ha, which is located in western Switzerland. From
233 REF-gregor we acquire satellite image data (section 2.1.1), yield maps of several cereals
234 from 2017 to 2021 (section 2.1.2), and meteorological data (section 2.1.3).

235

2.1.1 Sentinel 2 Data

236 The European Space Agency (ESA)¹ freely distributes the high-quality images of the two
237 Sentinel satellites 2 (S2). Together, both satellites have a revisit time of 5 days at the
238 Equator and 2-3 days at mid-latitudes. However, in our study region, we only receive an
239 image every 5 days. In order to decrease the effect of atmospheric conditions like reflections
240 and scattering, bottom-of-atmosphere, radiometric corrected Level-2A data was used²³.

241 The S2 images contain 12 spectral bands with spatial resolutions up to 10 meters (see
242 2.1). Bands with a lower resolution (20 and 60 meters) were upscaled to 10 meter
243 resolution using cubic interpolation (REF gregor perich). Additional to the spectral bands,
244 the ESA also supplies a Scene Classification Layer (*SCL*) where for each location the
245 observed subject is assigned to an *SCL-class* (cf. table 2.2). In chapter 3 we will use this
246 classification to filter out unreliable data points, considering only SCL-classes 4 and 5.

248

Data Illustration XXXorXXX Challenges in S2 Data

249 The figure 2.1 shows a selection of 6 satellite images of a field, which display our challenges.
250 In February (image a), we see no vegetation but bare soil. At the beginning of May, we
251 observe a cloudless dark green field. In (c) heavy cloud cover (*SCL* class 9) leads to a
252 complete loss of plant information in this S2 observation. Figure (d) shows that the *SCL*

Hier noch erwähnen, dass die *SCL* das Resultat eines Algorithmus der ESA ist. Das kann in der Diskussion dann auch wieder aufgenommen werden

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

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

³XXXREF gregor perich "Data prior to March 2018 was only 145 available in the top-of-atmosphere L1C format and was downloaded as such [...] L1C data was processed to L2A product level using the 'Sen2Cor' processor provided by ESA"

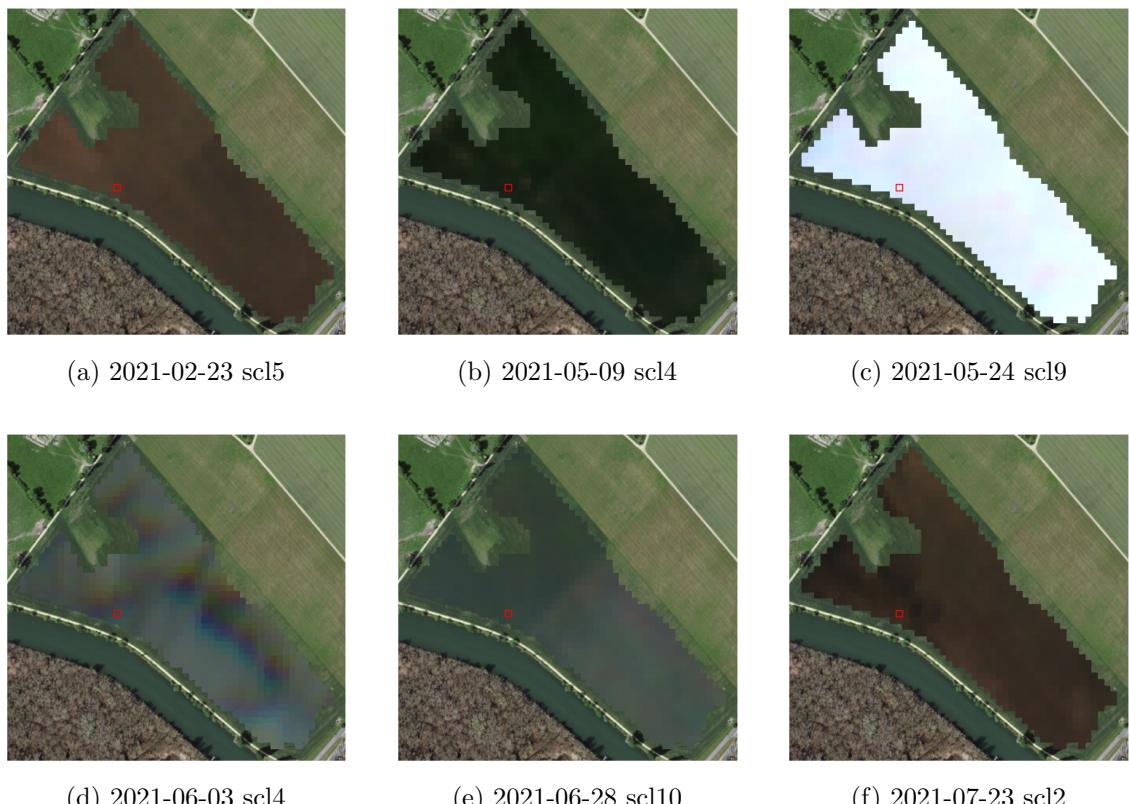


Figure 2.1: Satellite images of a field at selected times with a static background for orientation. The SCL-class of the highlighted pixel is provided in the respective subtitle. (???xxx include scl legend?)

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

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

253 classification is not reliable, since we evidently observe clouds. In (e) we see a pale green.
 254 This likely shimmers through cirrus clouds.

2.1.2 Crop Yield Data

257 The crop yield data were collected using a combine harvester. Equipped with GPS, the
 258 harvester drives over the fields and continuously estimates the crop density in t/ha (see fig.
 259 2.2a). We take the data set derived from this in REF-Gregor-Perich, where error-prone
 260 measurement points (such as during a tight curve of the combine harvester) were removed
 261 and then the yield map was rasterized using linear interpolation (cf. fig. 2.2b).

262 Comparing the average per-field crop yield reported by the farmer with the yield estimated
 263 by the combine harvester shows that the latter overestimates crop yield by ca. 10% (cf.
 264 REF-gregor). Since the relative estimation error is approximately constant and we do not
 265 aim to estimate the absolute yield, we will not consider this deviation.

Hier
bitte
noch
eine
kleine
Beschrei-
bung
der
Crop-
yields
rein-
nehmen.
Also
was sind
die Er-
tragswerte
der
Daten.

2.1.3 The Concept of a ‘Pixel’

267 Before we join all the data, we define a few concepts.

268 The well-known Normalized Difference Vegetation Index (*NDVI*) introduced in Rouse
 269 (Rouse) can be calculated using the bands *B4* and *B8* (table 2.1) by:

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

270 Note that we call the calculated values merely the *observed NDVI*, as we must be aware
 271 of imprecisions due to clouds and shadows.

Please
clarify
this in
more
detail.
We used

Table 2.2: Overview: Scene Classification Layers (SCL)

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

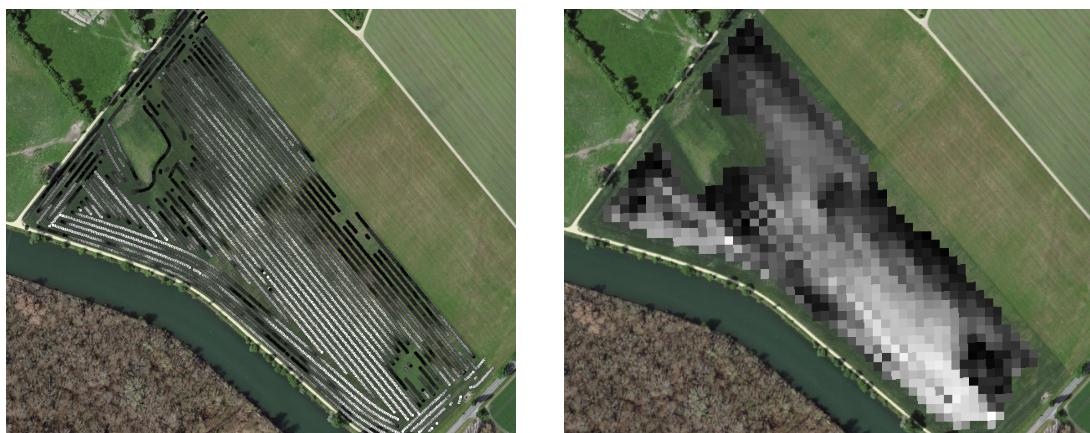


Figure 2.2: Crop yield density map of a field. Ranges from 0.1 t/ha (black) to 5.35 t/ha (white)

To define a timescale, we consider Days After Sowing (*DAS*) and a transformed timescale, Growing Degree Days (*GDD*) ([McMaster and Wilhelm \(McMaster and Wilhelm\)REF](#)). The latter are defined as the cumulative sum (since sowing) of temperature above a given base temperature T_{base} . For cereals, we use $T_{base} = 0$ (REF-Gregor). Thus, the GGD for n days after sowing will be equal to:

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

Now we create a data set, which will contain all the necessary information. Given that we have the spectral data at a $10m \times 10m$ resolution, we introduce the concept of a Pixel. A *Pixel* P is associated with a $10m \times 10m$ square defined by the S2 satellites and contains all relevant information for a season and this location. More precisely, P is a collection of general information (like yield and coordinates) and all associated P_t of a given season. Where P_t represents a tuple of the spectral data for time t , the NDVI calculated from it, and the associated GDD. We will call the resulting data set *PIXELS*, as it is the collection of all Pixels (over all seasons).

Für den Leser wäre es interessant, wenn Du noch kurz die wichtigsten GDD Werte aus der Literatur beschreiben würdest (D.h. z.B. ...)

285 **2.2 General Methods**

286 We will only introduce general methods within this section, whereas more specific methods
 287 will be introduced in their context. We discuss interpolation methods in sections 3.3 and
 288 3.4, a robustification strategy in section 3.6, a method how we can objectively determine
 289 the quality of an interpolation in section 3.5, and in section 4.2 we present the NDVI
 290 correction with an adapted interpolation strategy.

291 **2.2.1 (Relative) RMSE**

292 definition here and also relative

293 **2.2.2 Out-Of-Bag (*OOB*) and Leave-One-Out-Cross-Validation (*LOOCV*)**

294 Hier fehlt mir eine kurze Erklärung, was OOB und LOOCV sind.

which
context?
I would
write:
"... in-
tro-
duced
in chap-
ter/-
section
XXX
in more
detail."

Let

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

295 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
 296 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
 297 obtain $\hat{y} := (\hat{y}_1, \dots, \hat{y}_n)$ the *OOB* estimator for $y \in \mathbb{R}^n$.

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

300 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
 301 the corresponding \hat{y}_i also the *LOOCV* estimator.

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

305 **2.2.3 Generalized Cross Validation (GCV)**

306 definition here and explanation why (computational cheap)

307 **Chapter 3**

308 **Interpolation Methods**

309

310 In this section, we take a closer look at several interpolation methods, which will be
311 used to interpolate and smooth the NDVI time series, while considering only SCL45 in
312 this chapter. A brief overview of the considered interpolation methods can be found in
313 table 3.1.

314 First, we define the general setting and discuss a general approach to make the interpola-
315 tion more robust (i.e. reduce the impact of outliers).

316 Afterwards, we introduce and discuss each method.

317 Then, we try to extract the main ingredients of each method to construct a new one with
318 all benefits.

319 Finally, using LOOCV, we tune the parameters (where necessary) and get a first idea of
320 the performance of each method.

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

321 **3.1 DAS vs. GDD**

322 Prior to interpolating the NDVI time series, we should decide on a timescale. We can
323 choose between DAS and GDD (cf. section 2.1.3 and equation 2.1.3). In figure 3.1 we see
324 an example for comparison of the two. Here we see that the first 120 DAS are compressed
325 to just 500 GDD. This has several advantages. First, it makes the scales comparable (in
326 terms of plant growth) because the plants are not concerned with the month of the year
327 but the current temperature. Second, in winter we tend to have higher cloud cover and
328 thus fewer SCL45 observations. Hence, this gap in observations is compressed. Therefore,
329 we will only use GDD in the subsequent.

put sec-
tion in
methods
/ data

331 **3.2 Interpolation Setup**

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

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

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

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

Findet
man
hier
noch
Liter-
atur, in
welcher
ähn-
liches
disku-
tiert
wurde,
die man
zitieren
kann?

add fourier and fix order to match chapters

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
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 edge 	<ul style="list-style-type: none"> - Bad “upper envelope” since weights are not used for the estimation itself 	(No)	(Yes)
LOESS	<ul style="list-style-type: none"> - Local polynomial with points closer to the estimated point are more important 	<ul style="list-style-type: none"> - Flexible - Generalization of SG - Weighting function makes intuitive sense 	<ul style="list-style-type: none"> - Computationally expensive 	Yes	(Yes)
Smoothing Splines	<ul style="list-style-type: none"> - 2cd derivative of function is integrable 	<ul style="list-style-type: none"> - Intuitive meaning of penalty - General assumptions - Flexible shape 	<ul style="list-style-type: none"> - Unbounded 	Yes	No
B-Splines (Smoothed)	<ul style="list-style-type: none"> - Function can be approximated by a linear combination of B-splines basis functions 	<ul style="list-style-type: none"> - General assumption - Flexible shape 	<ul style="list-style-type: none"> - Unbounded - No intuitive meaning for smoothing 	Yes	No
(Gaussian) Kernel 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 - General assumptions 	<ul style="list-style-type: none"> - Bandwidth: fails if there are big data-gaps 	Yes	Yes
Double-Logistic	<ul style="list-style-type: none"> - Function first increases then decreases - Ndvi has a minimal value 	<ul style="list-style-type: none"> - Good for evergreen plants (if snow masks NDVI) - Upper envelope 	<ul style="list-style-type: none"> - Parameter estimation can go seriously wrong - Strange behavior for long data-gaps 	Yes	(Yes)
Universal Kriging	<ul style="list-style-type: none"> - Function is a realization of a stationary Gaussian process 	<ul style="list-style-type: none"> - Informative parameters - Flexible 	<ul style="list-style-type: none"> - Regression to the mean - Assumptions clearly not met 	Yes	(Yes)

³³² We will introduce parametric and non-parametric approaches to estimate m in section 3.3 and 3.4.

³³³ Furthermore, in the subsequent, we denote $w \in \mathbb{R}^n$ as the vector of weights such that w_i corresponds to the weight that (x_i, Y_i) should have in the interpolation.

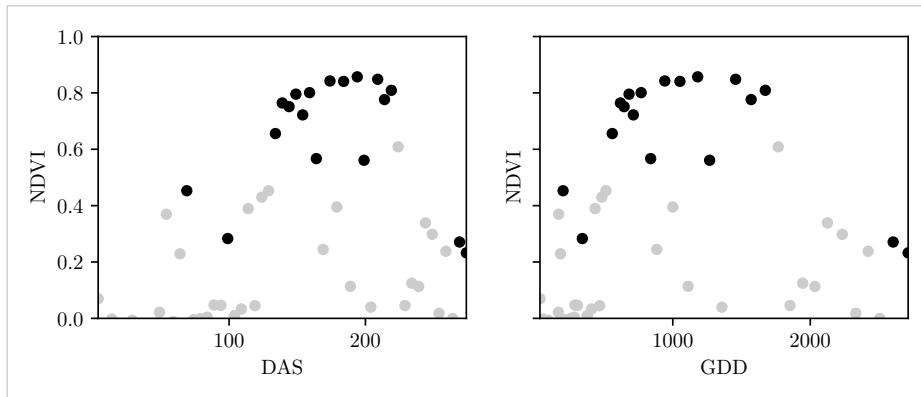


Figure 3.1: The same NDVI time-series, on the left with DAS as the timescale, on the right GDD is the timescale. SCL45 are colored black. Non-SCL45 (clouds and shadows) are colored in gray.

336

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

337

Ähnliche struktur sich überlegen

338

3.3 Parametric Regression

339

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

342

3.3.1 Double Logistic

343

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

346

Assumptions:

347

Die Aufzählung ist hier m.M.n. nicht so passend für einen "Fliesstext".

348

- 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.

351

- 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:

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

354

Where the five free parameters: y_{\max} , d_0 , d_1 , t_0 , t_1 are initially estimated by least squares.

355

Such fit can be seen in figure [3.2](#).

356 Similar as for the Savitzky-Golay Filter (cf. section 3.4.3) we reestimate (only once) the
 357 parameters by giving less weight to the overestimated observations and more weight to
 358 the underestimated observations¹.

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. (cf. figure 3.2)

359 **3.3.2 Fourier Approximation**

Similar as in section 3.3.1 we fit a parametric curve to the data by least squares. Here we take the second order Fourier series:

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

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

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 (cf. figure 3.2) — Hard to interpret estimated parameters — Parameter estimation can go wrong. Introducing bounds can help.

361 **3.3.3 Optimization Issues**

362 We shall mention some optimization issues we countered during implementation. Since we
 363 aim to minimize the residual sum of squares over 5 (or 6) parameters, we try to solve a
 364 non-convex optimization problem. Thus, the algorithm² either struggles to find the global
 365 minimum or fails to converge. This was fixed by providing for each parameter reasonable
 366 initial values and generous bounds (which match our experience).

367 **3.4 Non-Parametric Regression**

369 In non-parametric curve estimation, the curve does no longer have to be fully determined
 370 by parameters, but we allow it to also depend on the data. Note, that we do not exclude
 371 the use of tuning-parameters.

TODO:
include
Weighted
versions

¹For the details on the weights we refer to Beck, Atzberger, Høgda, Johansen, and Skidmore (Beck et al.)

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

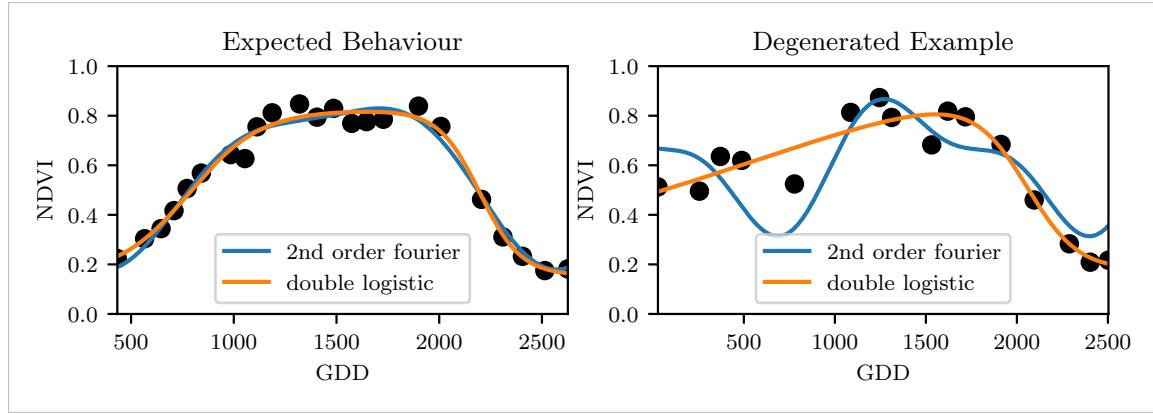


Figure 3.2: Here we observe the possibilities of a precise fit for the two parametric methods but notice also some misbehavior

372 3.4.1 Kernel Regression

373 As described previously (XXX REF Setup section), we would like to estimate

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f_X(x)}, \quad (3.4.1.1)$$

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

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

By using the above function in equation (3.4.1.1) we arrive at the *Nadaraya-Watson* kernel estimator:

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

374 Common choices for the kernel are the normal function or a uniform function (also called
375 'box' function.). Note that we still need to choose the bandwidth of the function (in
376 the case of the normal function, this is σ the standard deviation). For local adaptive
377 bandwidth selection we refer to [Brockmann, Gasser, and Herrmann \(Brockmann et al.\)](#).

Die
Aufzählung
ist hier
m.M.n.
nicht so
passend
für
einen
"Fliess-
text".

Advantages	Disadvantages
<ul style="list-style-type: none"> — flexible due to different possible kernels — can be assigned degrees of freedom (trace of the hat-matrix) — estimation of the noise variance $\hat{\sigma}_\varepsilon^2$ (REF cf. CompStat 3.2.2) 	<ul style="list-style-type: none"> — if the $x \mapsto K(x)$ is not continuous, \hat{m} isn't either — choice of bandwidth, especially if x_i are not equidistant.

378 3.4.2 Kriging

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

ohne
un-
terkapi-
tel
struk-
tur

384 Definitions and Assumptions

386 **Definition 3.4.2.1.** (*Gaussian Process*) A Gaussian Process $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \dots, S(t_k))$ has a multivariate Gaussian distribution for every collection of times t_1, \dots, t_k . S can be fully characterized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') = \text{Cov}(S(t), S(t'))$

390 **Assumption 1.** We will assume the Gaussian process to be stationary. That is for $\mu(t)$
 391 to be constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the
 392 following only $\gamma(h)$.³

Definition 3.4.2.2. (*Variogram*) We also define the variogram of a Gaussian process as

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

And decide to use a Gaussian Variogram defined by

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

393 where h is the distance, n is the nugget, r is the range and p is the partial sill visualized
 394 in figure 3.3.⁴

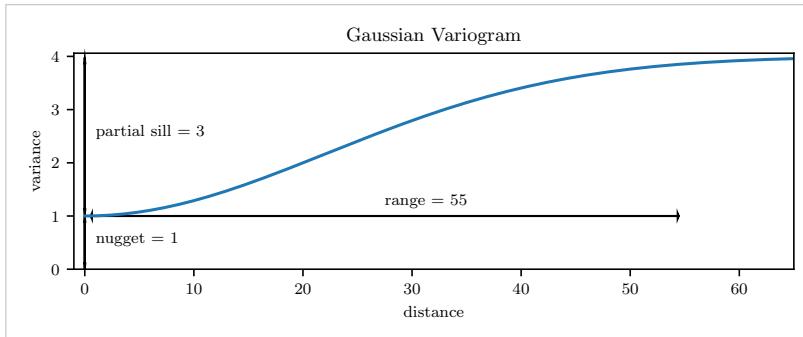


Figure 3.3: Gaussian Variogram with nugget=1, partial sill=3, range=55

395 Next, we consider a one-dimensional Gaussian process G_γ with variogram γ . We tune the
 396 variogram parameters using maximum likelihood⁵. Let z be a vector with the new values
 397 to extrapolate, then we can determine the values $m(z) = \mathbb{E}[G_\gamma(z)|(x, y)]$ using Bayes
 398 rule⁶. For an example fit, we refer to figure 3.4.

399 Since we observe a clear pattern of a growth period in spring and harvest in the end of
 400 summer, we have to admit that assumption 1 with the constant mean is clearly violated.
 401 This is also the reason why we observe (for every variogram parameter) a tendency to the
 402 mean, as indicated in figure 3.4.

³Note 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.

⁴Strictly speaking we use a scaled version of the variogram. Thus, only the ratio of p/n matters.

⁵As illustrated in figure 3.4 maximum likelihood estimation can lead to overfitting. Thus, we will in practice sample several such optimized parameters and use their median in the end.

⁶Bayes rule generally claims, that for two random variables A and B we have that $P(A|B) = P(B|A)/P(B)$

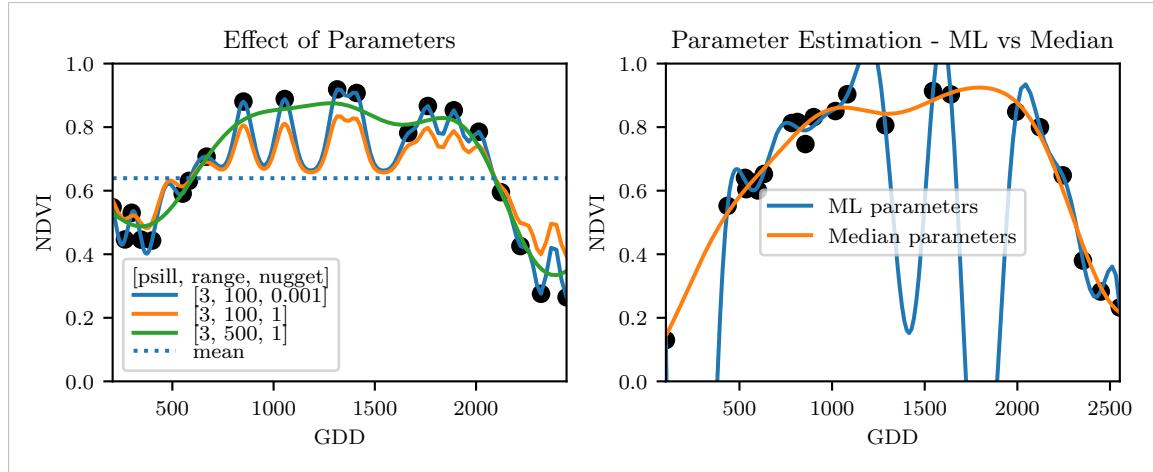


Figure 3.4: On the left, we see how the interpolation change if we increase the nugget and the range parameter. On the right, we compare two kriging interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such numerical optimizations.

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. — Skewness of errors is not taken into account.

403 3.4.3 Savitzky-Golay Filter (SG Filter)

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

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

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

For equidistant points this can efficiently be calculated by

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

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

406 Adaptation to the NDVI

407 [Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh \(Chen et al.\)](#) developed a ‘robust’ 408 interpolation method for the NDVI based on the SG Filter. The method is based on the

figure /
tabelle
/ pseudocode
anstatt
aufzählung

- assumption that due to atmospheric effects the observed NDVI tends to be underestimated and that it cannot increase too quickly. The latter is argued by the biological impossibility of such fast vegetation changes. Their proposed algorithm is:
- i.) Remove points which are labeled as cloudy.
 - ii.) Remove points which would indicate an increase greater than 0.4 within 20 days.
 - iii.) Linearly interpolate to obtain an equidistant time series X^0 .
 - iv.) Apply the SG Filter to obtain a new time series X^1 .
 - v.) Update X^1 by applying again a SG Filter. Repeat this until $w^T |X^1 - X^0|$ stops decreasing, where w is a weight vector with $w_i = \min\left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|}\right)$. This reduces the penalty introduced by outliers⁷ and by repeating this step we approach the “upper NDVI envelope”.

Advantages

- Popular technique in signal processing.
- Efficient calculation for equidistant points.
- Upper envelope matches intuition for the NDVI. Therefore, it is robust against outliers with small values.

Disadvantages

- No natural way of how to estimate points which are not in the data.
- Not generalizable to other spectral indices.
- Linear interpolation to account for missing data might be not appropriate.
- No smooth interpolation between two measurements.

Extension: Spatial-Temporal-Savitzky-Golay Filter

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

3.4.4 Locally Weighted Regression (LOESS)

The Locally Weighted Regression (LOESS) introduced by Cleveland (Cleveland) can be understood as a generalization of the SG Filter (cf. sec. 3.4.3).

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

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

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

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

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

⁹If too many weights are set to zero, we might end up considering not enough observations and thus

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

432 The LOESS smoother takes a fraction of points instead of a fixed number and therefore
 433 automatically adapts to the size of the data we wish to interpolate. However, we run
 434 into the danger of considering too little observations, since the estimation breaks down if
 435 $\lceil \alpha n \rceil < d + 1$.⁹ Furthermore, LOESS gives less weight to points further away. This yields
 436 a "smoother" estimate, since when we slide the window (e.g. for estimating the next value)
 437 an influential point at the border does not suddenly get zero weight from being weighted
 438 equally before. Finally, the LOESS also can be used for non-equidistant data and allows
 439 for arbitrary interpolation.

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)

440 **3.4.5 B-splines**

B-splines as discussed in [Lyche and Mørken](#) ([Lyche and Mørken](#)) 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:

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

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 functions¹⁰ 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

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.

¹⁰So 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

441 **3.4.6 Natural Smoothing Splines**

442 Let \mathcal{F} be the Sobolev space (the space of functions of which the second derivative is
 443 integrable). Then the unique¹¹ minimizer

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

444 is a natural¹² cubic spline (i.e. a piecewise cubic polynomial function). The objective
 445 function ensures that we decrease the curvature while keeping the RMSE low.

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).

446 **3.5 Tuning Parameter Estimation**

447 Many of the interpolation methods introduced in section 3.3 and 3.4 include a free parameter.
 448 To determine this parameter for a specific interpolation method, we will estimate the
 449 absolute residuals using OOB estimation and then optimize the parameter using a score
 450 function. We clarify the procedure step by step:

- 451 i.) Construct a set Λ of candidate parameters that generously covers the parameter
 452 space.
- 453 ii.) Consider \mathcal{P} , a set of Pixels.
- 454 iii.) For each parameter $\lambda \in \Lambda$ consider the individual pixels and compute the LOOCV¹³
 455 for the absolute residuals of the specific NDVI-interpolation method for all Pixels in
 456 \mathcal{P} and store them in the set R_λ .
- 457 iv.) Determine $\lambda_{optimal} = \arg \min_{\lambda \in \Lambda} q_{90}(R_\lambda)$, where we describe the 90% quantile with
 458 q_{90} .

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

461 Figure 3.5 exemplifies the effect of the optimization function (different quantiles). To
 462 summarize, we may say that the higher the quantile, the stronger the smoothing.

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

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

¹³For a definition of the leave-one-out-cross-validation we refer to section 2.2.2

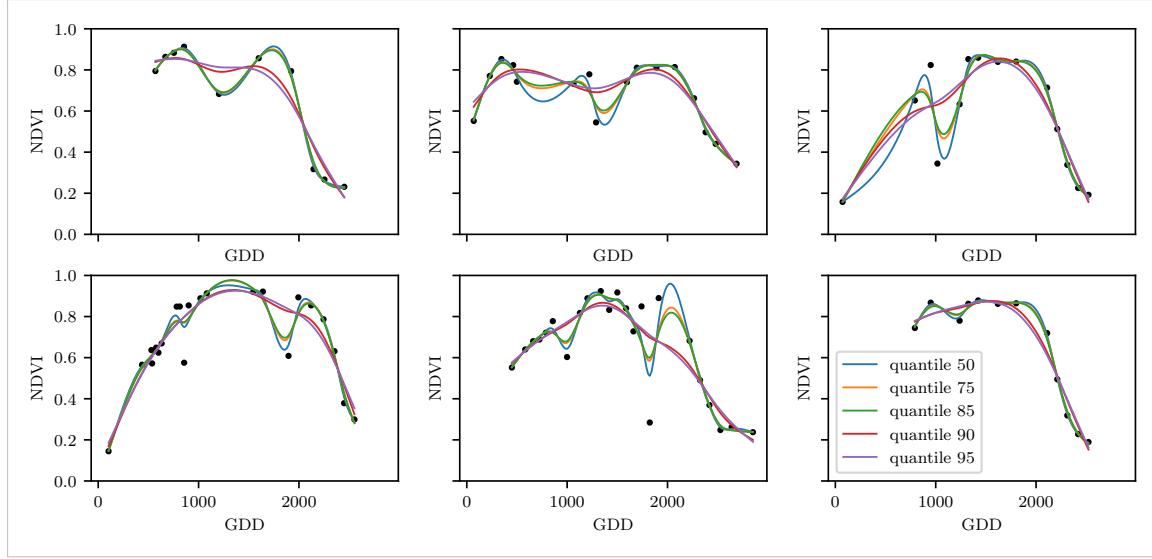


Figure 3.5: Smoothing splines fit with smoothing parameter optimized by minimizing the given quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

3.6 Robustification

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

Even though the procedure is taken from the robust version of the LOESS smoother (cf. section 3.4.4 and [Cleveland \(Cleveland\)](#)), we can apply it to every interpolation method that allows for prior weighting of observations.

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

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

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

$$w_i^{\text{new}} := w_i^{\text{old}} B\left(\frac{|r_i|}{6 \text{ med}(|r_1|, \dots, |r_n|)}\right); \quad r_i := y_i - \hat{y}_i \quad (3.6.0.1)$$

and interpolate again using the new weights. We can iterate this reweighting and stop after several steps or when the change of the values is smaller than some tolerance.

Note that this procedure is indeed robust since we use the median for the normalization which has a breakdown point¹⁴ of 50%.¹⁵

¹⁴Intuitively, 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%.

¹⁵The breakdown point relates only to outliers in the y values. Note that we do not require the interpo-

476 **3.6.1 Our Adjustment:**

In the case that we would like to apply prior weights, we want to prevent low-weighted observations to corrupt our estimation of scale (the median) and thus we use the weighted median. This can be defined as

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

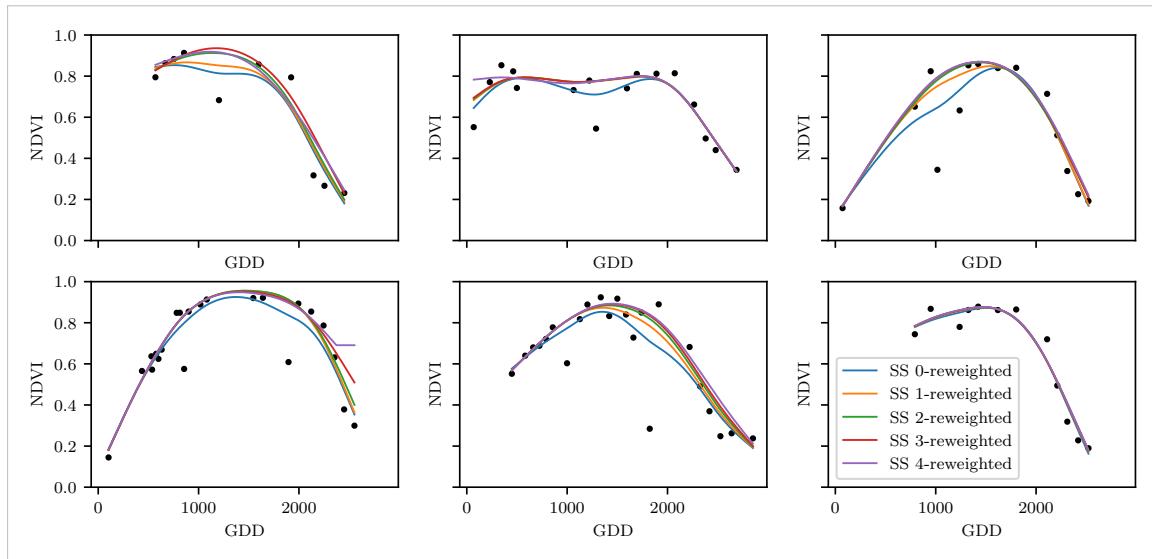
477 for $r, w \in \mathbb{R}^n$. ¹⁶478 **3.6.2 Examples and Conclusions**

Figure 3.6: Smoothing Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

479 In figure 3.6 we observe for six pixels how the NDVI time series interpolated with smoothing
480 splines looks after 0, 1, 2, 3, 4 iterations (See appendix figures B.1, B.2, B.3 and B.1 for the
481 analogous figures of the other interpolation methods).

482 Indeed, we observe how the interpolated time series is less affected by outliers after each
483 iteration. We notice the biggest difference in the first iteration. Furthermore, in the plot
484 at the bottom left we see how the interpolation ‘escapes’ from the right endpoint with
485 each successive iteration, even though our intuition does not necessarily identify this point
486 as an outlier. Therefore, in the following, we will always perform only one iteration and
487 then stop.

consider
naming
the sub-
plots

488 **3.6.3 Upper Envelope Approach - Penalty for Negative Residuals**

489 If we artificially increase the negative residuals in 3.6.0.1 by multiplying (e.g. factor 2),
490 the corresponding points will get less weight in the next iteration. This allows us to create

lation 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).

¹⁶This adjustment is also necessary to keep the scale estimation meaningful during the iterations.

Table 3.2: Comparing the goodness of fit for different interpolation methods measured with the statistics listed in the left column. Considering only SCL45 points, we get the out-of-bag estimates using the given interpolation method (TODO XXX: link to table explaining the methods?). Consequently, we compute the absolute (value of the) residuals and apply the given statistic to it.

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

491 an interpolation that resembles an upper envelope. Intuitively, this upper envelope can be
 492 thought of as a sheet that is laid on top of the points.

493 This approach is based on the premise that we tend to underestimate the NDVI (as in
 494 REF-savitzky-golay). Since we want to develop a general method that is in principle not
 495 related to the NDVI, we will not pursue this approach further.

496 3.7 Performance Assessment

497 Next, we will benchmark the different interpolation methods with and without robustification.
 498 For this, we will use the same technique as we did for the parameter determination
 499 in section 3.5. On B_λ we apply the RMSE and different quantiles and present the results
 500 in table 3.2.

501 3.8 XXX Evaluation

- 503 – ss dominate (i.e. have better benchmark values w.r.t. all considered statistics) b-splines
 504 (robustified and non-robustified)
- 505 – dl dominate Fourier (robustified and non-robustified)
- 506 – loess slightly dominates ss, but we prefer ss because of the smoothness guarantees (com-
 507 pare the figures B.1 and 3.6).
- 508 – use dl and ss in the following (keeping robustified and non-robustified variants)

write
out key-
words,
after
final re-
sults

509 **Chapter 4**

510 **NDVI Correction**

511 Let's remind ourselves that the data from the S2 is distributed with an SCL and we
512 therefore have some information about what is observed at each pixel for each sampled
513 time (cf. table 2.2). So far, we have only considered cloud-free points (i.e., SCL-classes 4
514 and 5). In this chapter, we would like to improve the NDVI interpolation by inspecting
515 also other SCL-classes and by using more information than just the two bands used to
516 calculate the NDVI (B4 and B8).

517 **4.1 Considering other SCL Classes**

518 In figure 4.1 we notice that some blue points which correspond to the SCL-class 10 (thin
519 cirrus clouds) follow the interpolated line closely and that they might be useful in improving
520 an interpolation fit.

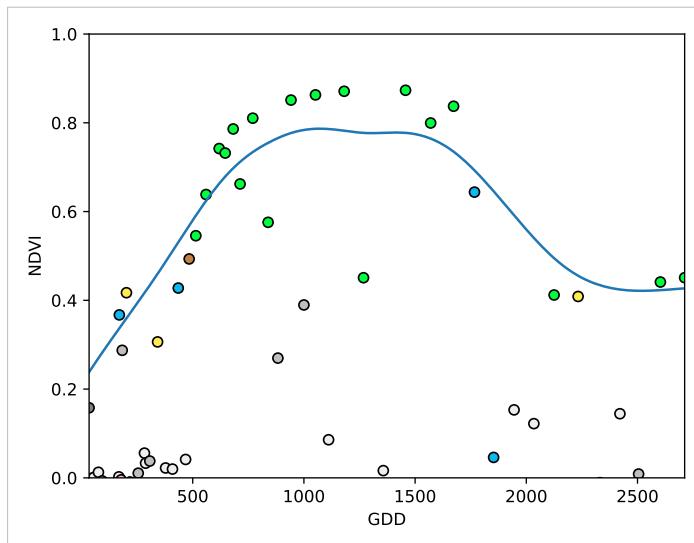


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45)

521 To get an impression of whether there is some useful information contained in the remaining
522 SCL-classes (all except 4 and 5) we would like to compare the observed NDVI with the
523 true NDVI. But since we do not have any ground truth data, we will make the following
524 assumption:

525 **Assumption 1.** The “true” NDVI value at time t can be successfully estimated by out-of
 526 bag (OOB) interpolation using high-quality observations. That is, the interpolated value
 527 — using an interpolation method from chapter 3 — considering the points $P^{SCL45} \setminus P_t$.
 528 In the following, we will call this estimate the “true”-NDVI.

529 We would like to get an idea if there is any information we can recover from SCL-classes
 530 other than 4 and 5. For that, we will check for the other SCL-classes if there is a relation
 531 between the “true” NDVI (derived with Smoothing Splines) and the observed NDVI. Thus,
 532 we pair each “true” NDVI with its observed one, collect all pairs, and create a scatter plot
 533 for each SCL-class in fig 4.2. As expected, the “true” and the observed NDVI seem to be
 534 highly correlated for SCL45. But we can also detect some patterns of correlation in the
 535 SCL-classes 2, 3, 7, 8 and 10.

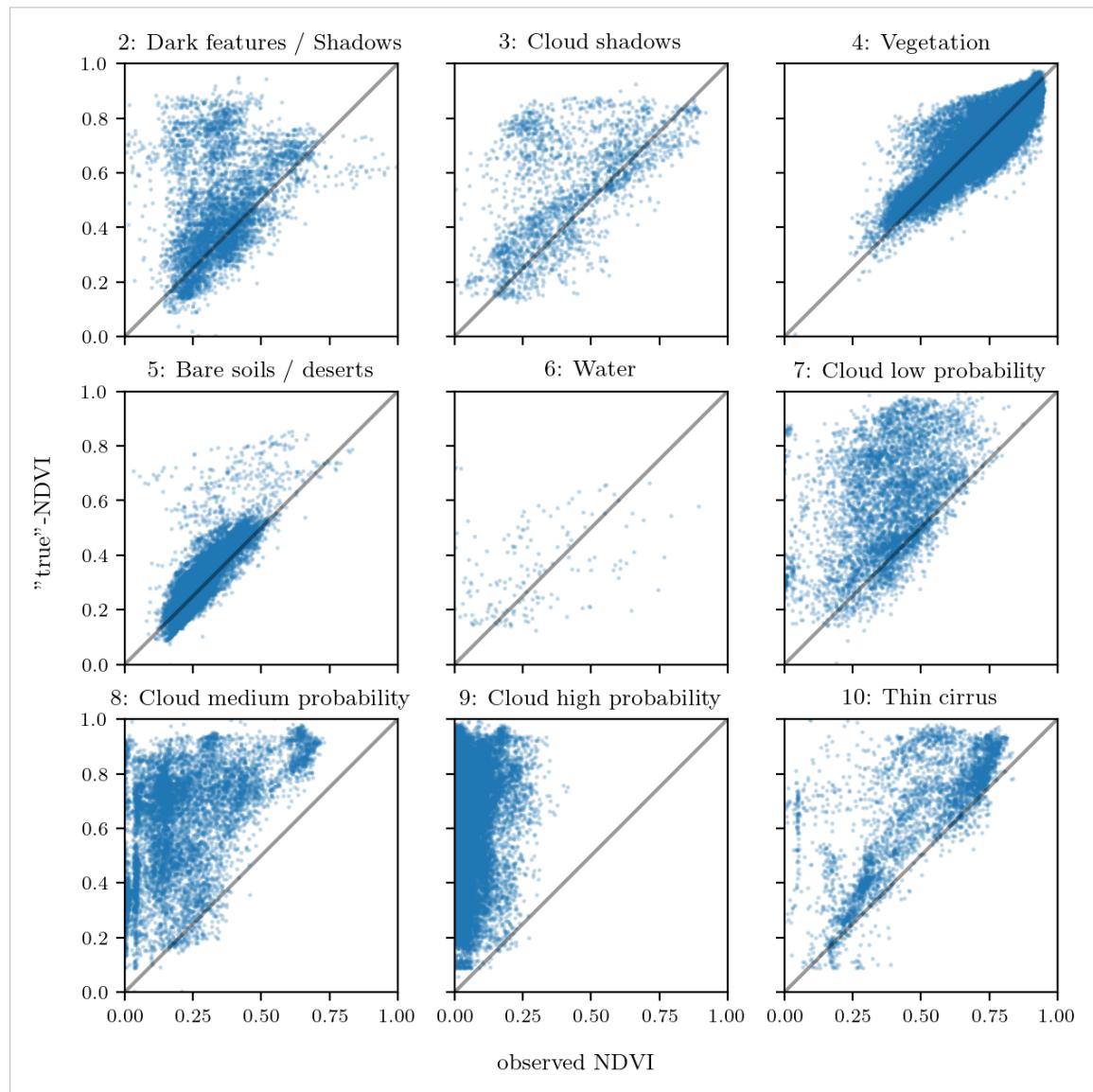


Figure 4.2: For each SCL class, we compare the true NDVI with the observed NDVI. (The true NDVI was estimated with OOB smoothing splines, and we used all observations of 10% of the total training pixels.)

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

540 4.2 Correction

541 We recall the satellite images in figure 2.1d, where we had cloudy images despite SCL4
 542 labeled and see fragments in figure 2.1e even though we are supposed to see clouds (SCL
 543 10 - Cirrus clouds). The SCL classification is based only on a mixed model trained using
 544 the s2 bands.

545 We will improve our NDVI interpolation by not relying on the existing SCL classifica-
 546 tion, but by training our own model to estimate/correct NDVI using all S2 bands (see
 547 sections 4.2.1 and 4.2.2). After we have corrected the observed NDVI, we will assess the
 548 uncertainty of our corrections and translate it into weights (in section 4.2.3). These will
 549 be used for the subsequent interpolation. This step-by-step procedure is illustrated by the
 550 REF graph in the appendix.

551 Finally, in section 4.4 we will evaluate this correction procedure, considering different
 552 interpolation methods and correction models.

hier
könn-
te man
auch
wieder
bezug
nehmen
auf die
originale
Sektion,
wo man
SCL
einführt.
dort
fehlt
mo-
mentan
auch
die Erk-
lärung,
dass
SCL ein
model
output
ist

553 4.2.1 Response and Covariates

554 For training an NDVI correction model, we need ground-truth (response) and informative
 555 covariates. We organize those in a table, where each row corresponds to a P_t (i.e., a pixel at
 556 a time t). Since ground-truth NDVI data is not available, we will again use the assumption
 557 1 and use the “true” NDVI instead. There is no canonical answer to the question of
 558 which covariates we should use. It is a tradeoff between simplicity/generalizability and
 559 performance (with the danger of overfitting). Our desire with the NDVI correction is
 560 to develop a product that is simple for others to understand and use. Therefore, in the
 561 subsequent, we will only take the spectral data of the satellite (i.e. all the bands) and the
 562 observed NDVI derived from it as covariates.

563 4.2.2 Correction Methods

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

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

569 Ordinary Least Squares (OLS)

570 The OLS is a linear model which aims to minimize the sum of the squared residuals. Let
 571 $y \in \mathbb{R}^n$ be the vector of responses and $X \in \mathbb{R}^{n \times p}$ be the design matrix, where each row
 572 corresponds to one pixel and each column consist of one covariate¹. We assume a linear

¹Strictly speaking, since SCL-classes are dummy variables

573 relationship between y and X and allow for Gaussian noise. That is:

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

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

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

575 We will train two models, one using only the SCL-classes as covariates and the other one
576 using all covariates (which are discussed in section 4.2.1).

Advantages	Disadvantages
<ul style="list-style-type: none"> — Simple method with good interpretability of coefficients. — Computationally cheap. 	<ul style="list-style-type: none"> — Catches only linear relationships. — No integrated variable selection.²

577 LASSO

578 The Lasso can be similarly expressed than the OLS but adds a penalty to the minimization
579 problem:

$$\hat{\beta}_\lambda = \arg \min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 = \arg \min_{\beta \in \mathbb{R}^p \text{ and } \|\beta\|_1 < \lambda} \|y - X\beta\|_2^2. \quad (4.2.2.1)$$

580 Even though we do not have a closed form solution for equation (4.2.2.1) we can solve
581 it easily via optimization, since the function $\beta \in \{\beta \in \mathbb{R}^p \mid \|\beta\|_1 < \lambda\} \mapsto \|y - X\beta\|_2^2$ is
582 continuous and convex.

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

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

590 We will apply the Lasso using the selected covariates in section 4.2.1 and their second
591 degree of interactions.⁵

Advantages	Disadvantages
<ul style="list-style-type: none"> — Usually yields a sparse solution. This tends to give better generalizability (prediction performance on unseen data). — Successfully deals with correlation in covariates. — Interpretable results. 	<ul style="list-style-type: none"> — Estimate is biased. — Computationally expensive.

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 $\{a, b, ab, a^2, b^2\}$.

592 **Random Forest (*RF*)**

593 To define a random Forest introduced by [Breiman \(Breiman\)](#) we will first define what a
 594 Tree is. A (*decision*) Tree is a graph (V, E) without circles, a distinct root node, every
 595 node has at most two children and every leaf has a value assigned to it. At each node there
 596 is a boolean condition testing if one variable is greater than some value and a pointer to
 597 one child depending on the boolean value. To evaluate a tree we start at the root node,
 598 test the boolean expression and go to the node indicated by the resulting pointer. This
 599 we repeat until we end up at a leaf-node, where we return the value assigned to it.

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

603 To build a *Random Forest* we will bootstrap-aggregate⁸ many such Trees⁹. The prediction
 604 of the Random Forest for a new point x is then the mean of the predictions from all the
 605 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.

606 **Multivariate Adaptive Regression Splines (*MARS*)**

607 A MARS model as introduced in [Friedman \(Friedman\)](#) can be described by

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

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

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

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

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

$$\mathcal{C} := \{(h(\cdot)b_1(\cdot), h(\cdot)b_2(\cdot)) \mid h \in \mathcal{M}, (b_1, b_2) \in \mathcal{B}\} \quad (4.2.2.2)$$

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.

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

Finally, to avoid overfitting, we prune the set \mathcal{M} by optimizing a generalized cross validation score (GCV).¹⁰

To reduce computational complexity, we follow the recommendation from REF Stephen (Stephen) and restrict h in equation (4.2.2.2) to be of degree one (so it is also in a pair of \mathcal{B}). Consequently, \mathcal{C} contains functions with a degree of at most 2.

Advantages	Disadvantages
— Catches non-linear relationships.	— Computationally expensive (can be reduced by restricting the degree of interactions).
— Interpretability via functions in \mathcal{M} and their coefficients.	
— Allows for interactions with variable selection.	

623 General Additive Model (*GAM*)

624 GAMs as described in [Hastie and Tibshirani](#) ([Hastie and Tibshirani](#)) are a special case of
 625 Projection Pursuit Regression, where only the p directions parallel to the coordinate axes
 626 are considered. The result is different to a linear model since the coordinate functions are
 627 not restricted to be linear but are assumed to be non-parametric functions. The model
 628 can be written as:

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

629 To estimate the non-parametric functions, we can use smoothing splines (ref sec. 3.4.6).
 630 For this let \mathcal{S}_j be the function which takes some $z \in \mathbb{R}^n$ and returns the smoothing splines
 631 fitted to $(X_{:,j}, z)$ where the smoothing parameter is optimized by GCV. Since we cannot
 632 fit all g_j simultaneously, we will use a strategy named Backfitting. We basically cycle
 633 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

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

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}

$\{h\}$. We discard the function which – when excluding from \mathcal{M} – leads to the best GCV score.

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

Advantages	Disadvantages
— Captures non-linearity.	— No automatic variable selection.
— Good interpretability.	— Computationally expensive.

635

636 4.2.3 Uncertainty Estimation

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

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

643 4.2.4 Interpolation

644 Consider now a pixel P , $\hat{y}^{(P)}$ its corrected NDVI and $\hat{v}^{(P)}$ the estimated uncertainties of
 645 $\hat{y}^{(P)}$. In order to interpolate $\hat{y}^{(P)}$, we will give less weight to unreliable observations. Thus,
 646 we define the weight function:

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

647 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.
 648 The normalization is needed since for some interpolation methods, inflating the sum of
 649 weights would decrease the effect of the smoothing.

650 4.3 Resulting Interpolation Strategies

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

- 652 i.) OOB Interpolation (+ robustify?)
- 653 ii.) Correction
- 654 iii.) Uncertainty estimation
- 655 iv.) Interpolation (+ robustify?)

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

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

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

- 663 — We use the same interpolation method each time.

Ich
finde die
sections,
in denen
Du die
Modelle
erklärest,
gut.
Aller-
dings
fehlt
mir die
Über-
leitung/Ein-
leitung,
warum
die
Modelle
ge-
braucht
werden

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

- 664 — Either we robustify both times, or we do not robustify at all.
 665 — We use the same underlying method for correction and uncertainty estimation.
 666 In this fashion, we obtain 28 distinct interpolation strategies, which we will benchmark in
 667 the next section.

668 4.4 Evaluation Method

669 In this section, we introduce the relative yield-estimation-accuracy (*RYEA*) and utilize it
 670 to evaluate the interpolation strategies from section 4.3.

671 **Definition 4.4.0.1.** (*RYEA*) Let $y \in \mathbb{R}^n$ be the yield, M be a model for estimating y , and
 672 $\hat{y} = M(X)$ where X describes the data¹³. We define the *RYEA* as the relative RMSE in
 673 yield estimation. Formally expressed:

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

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

675 4.4.1 Idea

676 The fundamental assumption is that the closer the interpolated NDVI time series is to
 677 the true one, the better it can be used to determine crop yield. Implicitly, we believe that
 678 an NDVI time series which better models yield will incorporate more true information
 679 about the underlying vegetation. Therefore, we want to determine a comparable RYEA
 680 for each interpolation strategy and choose it as a benchmark criterion. This is an objective
 681 measure, since we have not considered crop yield in any of our previous steps. Moreover,
 682 this criterion is justified by the fact that yield estimation has been a motivation for the
 683 interpolation.

684 4.4.2 Yield Estimation

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

- 690 — Maximum slope
- 691 — Minimum slope
- 692 — Integral¹⁴ over all
- 693 — Peak (i.e. maximal NDVI)
- 694 — Peak GDD (i.e. value at which the peak is attained)
- 695 — Integral¹⁴ up to the peak
- 696 — Integral¹⁴ after peak

tabelle
wäre
sauberer

¹³We will use the matrixes derived in section 4.4.2

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

- 697 — Integral¹⁴ from 0-685 GDD
698 — Integral¹⁴ from 685-1075 GDD

699 For the choice we were inspired by REF-kamir. However, we deliberately omit any statistic
700 that involves the minimum (e.g. the NDVI-range), since we regard the minimum as a very
701 error-prone measure due to the large influence of clouds in the time series.

check
reference

702 As a result, for each interpolation strategy, a matrix is obtained in which each row corre-
703 sponds to a pixel and both the yield and the characterizing statistics are contained. Using
704 this matrix, we train a random forest for yield estimation, and compute the integrated
705 OOB estimates¹⁵ \hat{y} . Note that the choice of the modeling approach does not matter much,
706 as long as it is general enough (i.e. able to approximate any function) and we use the same
707 one for each interpolation strategy. Finally, for each interpolation strategy, we calculate
708 the RYEA. The results are shown in table 6.1.

welche
charac-
terizing
statis-
tics
genau?

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 $\frac{1}{e}$).

709 **Chapter 5**

710 **Results**

711 **5.1 XXX small recap from “Interpolation Methods”**

712 shoud w write 1:1 the sam es in the end of section 3

713 **5.2 Robustification and NDVI-Correction**

714 var(yield) 4.034559

715 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.1066 6.1855 7.5595 7.3592 8.7564 13.3508

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

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

Table 5.1: XXX RMSE of yield prediction

	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

Table 5.2: 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

Table 5.3: XXX RMSE of yield prediction

	RF	OLS-SCL	OLS-all	MARS	GAM	lasso	no-correction
ss	nan	nan	nan	nan	nan	nan	nan
dl	nan	nan	nan	nan	nan	nan	nan
ss-rob	nan	nan	nan	nan	nan	nan	nan
dl-rob	nan	nan	nan	nan	nan	nan	nan

716 **Chapter 6**

717 **Discussion**

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

719 **6.1 Interpolation Methods**

720 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?

721 XXX discuss results from table

722 **6.2 NDVI Correction**

723 **6.2.1 Bootstrap**

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

730 **6.2.2 Using Additional Covariates**

732 In section 4.2.1 we have only used the spectral data (and the observational NDVI calculated
733 from them) as covariates. Since we have the weather data available (cf. REF-SEC), it
734 would be a small effort to incorporate it, together with statistics collected from it (i.e.
735 GDD or 'rainfall in the last 30 days').

736 We decided against using this data, because on the one hand we have the problem that
737 we have practically too few observations (we observe only 5 years) and we expect the
738 weather in our study region to be rather homogeneous which is suggested by the fact
739 that the weather data published by Meteoswiss are for a grid with a resolution of 1 km.
740 On the other hand, we want the underlying model not to learn improper relationships.
741 For example, the model might automatically predict a high NDVI for a day in summer

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

742 (detected by high GDD / many sunshine hours / high temperature) just because it is
743 “used” to observing a lot of vegetation in summer. Including temporally (e.g., P_{t-1} and
744 P_{t+1}) and geographically adjacent pixels would likely improve performance. However, for
745 simplicity, we omit it here¹.

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

749 6.2.3 High RMSE in Yield Prediction

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

- 751 i.) Uncertainty in Yield data collected by the combine harvester
- 752 ii.) Uncertainty in Yield data through rasterization
- 753 iii.) Uncertainty in satellite images through “measurement errors” introduced via clouds
754 and other atmospheric effects
- 755 iv.) Uncertainty introduced by interpolating (especially when long data-gaps are present)

¹This 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).

756 **Chapter 7**

757 **Conclusion**

758
759 - itpl methods,
760 parametric dl
761 non-param
762 discarded
763 kernel methods because of strong bias
764 kriging because assumptions and highdim parameters
765 savitzky-golay filter since we will investigate the LOESS which can be thought a
766 loess slightly best performing itpl method but we notice non-smooth behaviour if
767 loess > ss > bspl
768 choose ss because of its meaningful definition (minimizing the integral of the second
769 - robustifying useful?
770

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

772 **7.1 Future Work**

773 **7.1.1 Time Series Correction-Interpolation as a General Method**

774 Throughout this thesis, we developed a correction and interpolation method for the NDVI.
775 However, we never used features of the NDVI. Only the parameter estimated via cross-
776 validation in chapter 3.5 depends on the scale of the time series. For simplicity, we could
777 thus determine the parameter using Generalized Cross Validation (as Ripley and Maechler
778 suggests). Therefore, our approach of interpolation and correction of time series can be
779 applied to arbitrary time series as long as additional information is available. However,
780 further research is required, to demonstrate the usefulness of this approach in general.

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

782 This generalization can be used in particular for cloud correction. In the same manner as
783 we corrected the NDVI time series in chapter 4, we can correct each spectral band and
784 reunite the corrected bands with the uncertainties. If desired, the time series can also be
785 interpolated before merging as in chapter 4.2.4. The resulting question would be how well
786 this approach performs.

787 **7.1.2 Minor Improvements**

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

- 790 — **Data:** Method how data has been extrapolated to the grid could possibly be improved
- 791 — **Data:** For computational reasons, we mostly considered all years and split the data
792 (on the pixel level) randomly into a train/test set. A leave one year out cross
793 validation might yield more accurate results.
- 794 — **Data:** We have not included the spectral bands which have a resolution of 60m. But
795 precisely these seem to be promising for cloud correction, since they are a proxy of
796 the water (content and form) in the atmosphere.
- 797 — **NDVI Correction:** Explore the effect of different link functions between the esti-
798 mated absolute residuals and the weights in section [4.2.4](#).
- 799 — **NDVI Correction:** Yield is not the only target variable of interest. Other variables
800 like protein content could also be used in section [4.4](#) for the method evaluation.

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

801 Bibliography

- 802 Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.), *Model-
803 Based Geostatistics*, pp. 46–78. Springer.
- 804 Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. Improved
805 monitoring of vegetation dynamics at very high latitudes: A new method using MODIS
806 NDVI. *100*(3), 321–334.
- 807 Breiman, L. Random Forests. *45*(1), 5–32.
- 808 Brockmann, M., T. Gasser, and E. Herrmann. Locally Adaptive Bandwidth Choice for
809 Kernel Regression Estimators. *88*(424), 1302–1309.
- 810 Cao, R., Y. Chen, M. Shen, J. Chen, J. Zhou, C. Wang, and W. Yang. A simple method to
811 improve the quality of NDVI time-series data by integrating spatiotemporal information
812 with the Savitzky–Golay filter. *217*, 244–257.
- 813 Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh. A simple method
814 for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay
815 filter. *91*(3), 332–344.
- 816 Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots.
817 *74*(368), 829–836.
- 818 Friedman, J. H. Multivariate Adaptive Regression Splines. *19*(1), 1–67.
- 819 Hastie, T. and R. Tibshirani. Generalized Additive Models: Some Applications. *82*(398),
820 371–386.
- 821 Jaramaz, D., V. Perović, S. Belanovic Simic, E. Saljnikov, D. Cakmak, V. Mrvić, and
822 L. Zivotic. The ESA Sentinel-2 mission Vegetation variables for Remote sensing of
823 Plant monitoring.
- 824 Lyche, T. and K. Mørken. Spline Methods.
- 825 McMaster, G. S. and W. W. Wilhelm. Growing degree-days: One equation, two interpre-
826 tations. *87*(4), 291–300.
- 827 Ripley, B. D. and M. Maechler. R: Fit a Smoothing Spline.
- 828 Rouse, J. W. Monitoring the vernal advancement and retrogradation (green wave effect)
829 of natural vegetation.
- 830 Savitzky, A. and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified
831 Least Squares Procedures. *36*(8), 1627–1639.
- 832 Schafer, R. W. What Is a Savitzky–Golay Filter? [Lecture Notes]. *28*(4), 111–117.

- 833 Skakun, S., E. Vermote, B. Franch, J.-C. Roger, N. Kussul, J. Ju, and J. Masek. Winter
834 Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface
835 Reflectance, Through Phenological Fitting, into Regression Yield Models. *11*(15), 1768.
- 836 Stephen, M. Earth: Multivariate Adaptive Regression Splines.
- 837 Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *73*(3),
838 273–282.

839 **Appendix A**

840 **Reproducibility**

841 **A.1 Reproduce Results**

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

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

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

848 **A.2 R-Package**

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

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

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

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

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

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

```
857 1 library(CorrectTimeSeries)
858 2
859 3 # load a list of dataframes, each one describes one pixel with the covariates and
860 4 # the response
861 5 data(timeseries_list)
862 6 str(timeseries_list[[1]])
863 7
864 8 # Train/Load RF
865 9 train_model_myself <- TRUE
866 10 if (train_model_myself){
867 11   # Add "true" NDVI (or generally the response), by Out-Of-Bag estimation
868 12   timeseries_list <- lapply(timeseries_list, function(df) {
869 13     df$oob_ndvi <- OOB_est(df$gdd, df$ndvi_observed) # gdd is the time-axis
870 14     df
871 15   })
872 16   # Train correction model
873 17   formula <- "oob_ndvi ~ B02+B03+B04+B05+B06+B07+B08+B8A+B11+B12+scl_class"
874 18   RF <- train_RF_with_fromula(formula, timeseries_list, robustify=TRUE)
875 19 } else {
```

```
877 19  data(RF_for_NDVI)
878 20  RF <- RF_for_NDVI
879 21 }
880 22
881 23 # ADD CORRECTION
882 24 timeseries_list <- lapply(timeseries_list, function(df) {
883 25   df$corrected_ndvi <- randomForest:::predict.randomForest(RF, df)
884 26   df
885 27 })
886 28
887 29 # Get interpolation for each timeseries
888 30 newx <- 1:1000
889 31 lapply(timeseries_list, function(df){
890 32   ss <- smoothing_spline(df$gdd, df$corrected_ndvi)
891 33   predict(ss, newx)$y
892 34 })
```

Example of how to use the `CorrectTimeSeries` package

894 **Appendix B**

895 **Further Material**

896 **B.1 Interpolation**

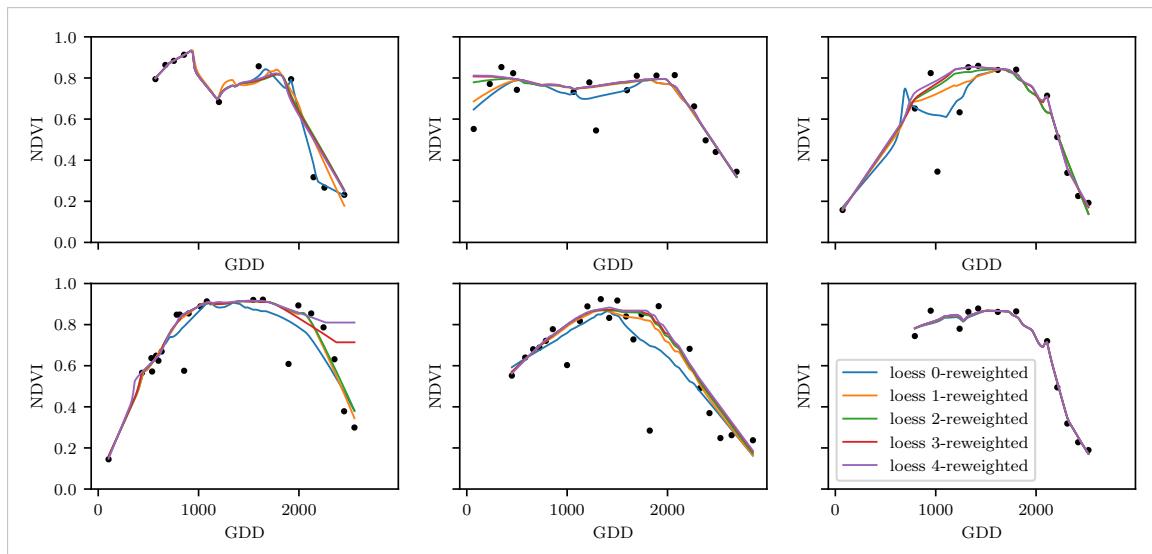


Figure B.1: The LOESS smoother fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

897 **B.2 NDVI correction**

page breaks

```
898
899
900 1 Call:
901 2 lm(formula = (paste(response, " ~ ", "ndvi_observed + scl_class")),
902 3     data = ndvi_df)
903
904 5 Residuals:
905 6   Min     1Q   Median     3Q    Max
906 7 -0.7997 -0.0717  0.0039  0.0695  0.6632
907
908 9 Coefficients:
909 10             Estimate Std. Error t value Pr(>|t|)
910 11 (Intercept) 0.21465   0.00230   93.46 < 2e-16 ***
911 12 ndvi_observed 0.71116   0.00346  205.65 < 2e-16 ***
```

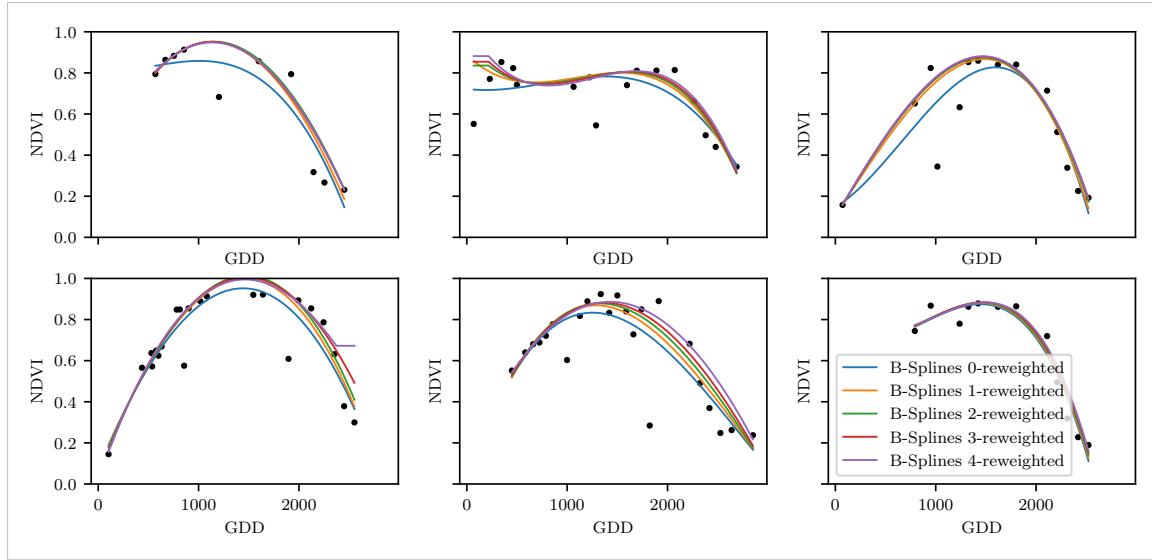


Figure B.2: B-Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

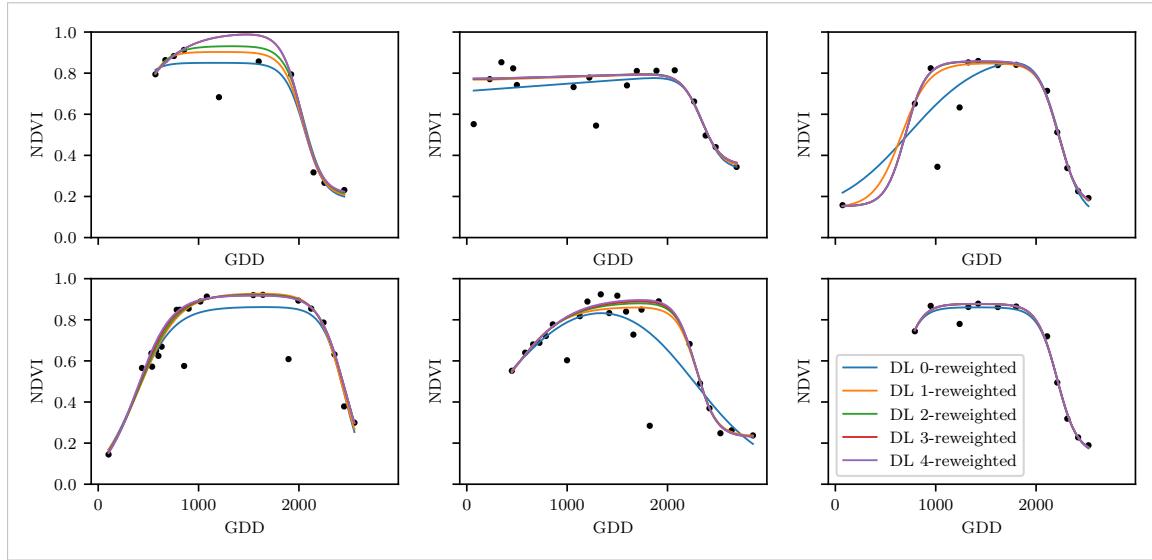


Figure B.3: A Double Logistic curve fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

```

912 13 scl_class3      0.02205    0.00356    6.20   5.8e-10 ***  
913 14 scl_class4     -0.00431    0.00251    -1.72   0.085 .  
914 15 scl_class5     -0.09875    0.00234    -42.15  < 2e-16 ***  
915 16 scl_class6     -0.05301    0.01104    -4.80   1.6e-06 ***  
916 17 scl_class7     0.11245    0.00274    41.09  < 2e-16 ***  
917 18 scl_class8     0.25963    0.00253    102.57 < 2e-16 ***  
918 19 scl_class9     0.35994    0.00236    152.47 < 2e-16 ***  
919 20 scl_class10    0.09091    0.00308    29.54  < 2e-16 ***  
920 21 scl_class11    0.29784    0.00392    76.06  < 2e-16 ***  
921 ---  
922 23 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
923 24  
924 25 Residual standard error: 0.146 on 124978 degrees of freedom  
925 26 Multiple R-squared:  0.532,          Adjusted R-squared:  0.532

```

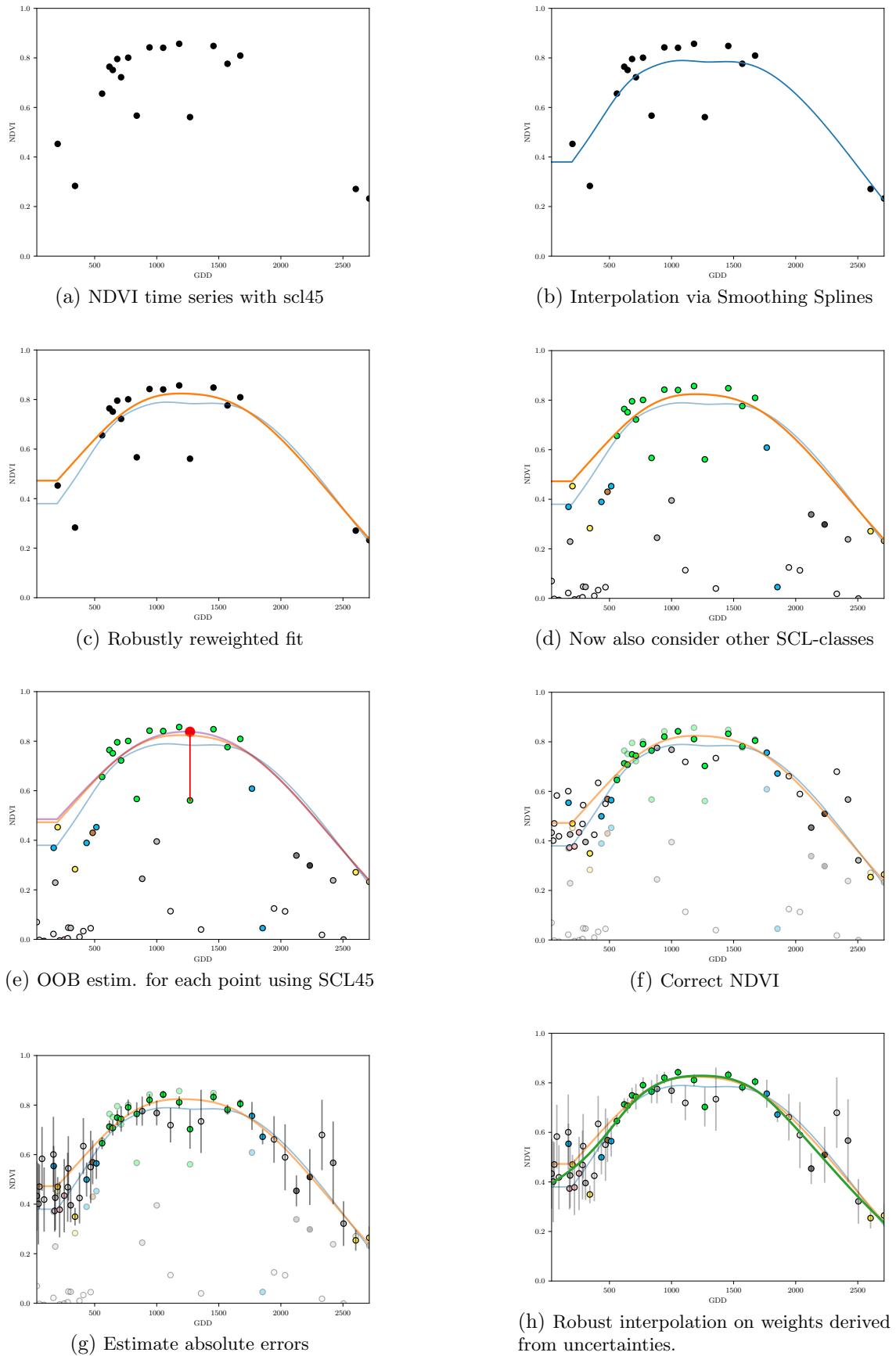


Figure B.4: Stepwise illustration of robust NDVI-Correction. For the color encoding of the SCL classes we refer to table 2.2.

```
926 27 F-statistic: 1.42e+04 on 10 and 124978 DF, p-value: <2e-16
```

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

```

928
929 1 Call:
930 2 lm(formula = (paste(get_res(), " ~ ", "ndvi_observed + scl_class")),
931 3     data = ndvi_df)
932 4
933 5 Residuals:
934 6     Min      1Q   Median      3Q      Max
935 7 -0.2051 -0.0427 -0.0074  0.0329  0.6589
936 8
937 9 Coefficients:
938 10            Estimate Std. Error t value Pr(>|t|)
939 11 (Intercept) 0.18647  0.00126 147.74 < 2e-16 ***
940 12 ndvi_observed -0.13265  0.00190 -69.80 < 2e-16 ***
941 13 scl_class3 -0.00180  0.00196 -0.92  0.3587
942 14 scl_class4 -0.04069  0.00138 -29.55 < 2e-16 ***
943 15 scl_class5 -0.09698  0.00129 -75.32 < 2e-16 ***
944 16 scl_class6 -0.01906  0.00606 -3.14  0.0017 **
945 17 scl_class7  0.01641  0.00150 10.91 < 2e-16 ***
946 18 scl_class8 -0.00560  0.00139 -4.02 5.7e-05 ***
947 19 scl_class9 -0.01384  0.00130 -10.67 < 2e-16 ***
948 20 scl_class10 -0.00690  0.00169 -4.08 4.5e-05 ***
949 21 scl_class11 -0.01446  0.00215 -6.72 1.8e-11 ***
950 22 ---
951 23 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
952 24
953 25 Residual standard error: 0.08 on 124978 degrees of freedom
954 26 Multiple R-squared:  0.352,    Adjusted R-squared:  0.352
955 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)