



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 Sentinel 2 Data	3
50	2.2 Crop Yield Data	3
51	2.3 The Concept of a ‘Pixel’	5
52	2.3.1 Normalized Difference Vegetation Index (NDVI)	5
53	2.3.2 Challenges in S2 Data	5
54	2.4 General Methods	7
55	2.4.1 Root Mean Square Error (RMSE)	7
56	2.4.2 Out-Of-Bag (<i>OOB</i>) and Leave-One-Out-Cross-Validation (<i>LOOCV</i>)	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	12
64	3.4 Non-Parametric Regression	12
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	31
93	6.1 Interpolation Methods	31
94	6.2 NDVI Correction	31
95	6.2.1 Bootstrap	31
96	6.2.2 Using Additional Covariates	31
97	6.2.3 High RMSE in Yield Prediction	32
98	7 Conclusion	33
99	7.1 Future Work	33
100	7.1.1 Time Series Correction-Interpolation as a General Method	33
101	7.1.2 Minor Improvements	34
102	Bibliography	35
103	A Reproducibility	37
104	A.1 Reproduce Results	37
105	A.2 R-Package	37
106	B Further Material	39
107	B.1 Interpolation	39
108	B.2 NDVI correction	39

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 Why do we do interpolation in NDVI (and other indices) time series? What are 113 possible shortcomings thereof?	1
114 Hier bitte noch eine kleine Beschreibung der Crop-yields reinnehmen. Also was sind 115 die Ertragswerte der Daten.	3
116 Please clarify this in more detail. We used pixels flagged with SCL 4 & 5, but as 117 can be seen in Fig. 2.1 d), this can yield erroneous NDVI values, etc.	5
118 Für den Leser wäre es interessant, wenn Du noch kurz die wichtigsten GDD Werte 119 aus der Literatur beschreiben würdest (D.h. z.B. Sowing, Emergence of Plants, 120 Anthesis, Senescence, Harvest)	5
121 necessary info for what? To answer the research questions asked in section XXX . . 122 which?	5
123 Hier noch eine NDVI Zeitreihe parallel dazu zeigen. Ansonsten wird nicht klar, 124 warum wir die Interpolation überhaupt machen.	5
125 which challenges? were they introduced earlier? E.g. in the introduction?	5
126 verdeutliche dem lesrer, dass ein auftrag das findne von interpolationmethoden war .	8
127 put section in methods / data	8
128 Findet man hier noch Literatur, in welcher ähnliches diskutiert wurde, die man 129 zitieren kann?	8
130 add fourier and fix order to match chapters	9
131 Paper zitieren wo eingeführt oder wo benutzt (falls einführung fast schon trivial) .	9
132 Ähnliche struktur sich überlegen	10
133 Die Aufzählung ist hier m.M.n. nicht so passend für einen "Fliesstext".	10
134 TODO: include Weighted versions	12
135 Die Aufzählung ist hier m.M.n. nicht so passend für einen "Fliesstext".	12
136 ohne unterkapitel struktur	13
137 figure / tabelle / pseudocode anstatt aufzählung	15
138 consider naming the sub-plots	19
139 write out keywords, after final results	20
140 hier könnte man auch wieder bezug nehmen auf die originale Sektion, wo man SCL 141 einführt. dort fehlt momentan auch die Erklärung, dass SCL ein model output 142 ist	23
143 Ich finde die sections, in denen Du die Modelle erklärt, gut. Allerdings fehlt mir 144 die Überleitung/Einleitung, warum die Modelle gebraucht werden	26
145 tabelle wäre sauberer	28
146 check reference	28
147 welche characterizing statistics genau?	29

148	shoud w write 1:1 the sam es in the end of section 3	30
149	Here in the discussion, you should take up the points you mentioned in the introduction	31
150	You already capture the "main" structure of your thesis with the interpolation and	
151	the NDVi correction sections. Can you combine them both in a "synthesis"	
152	subsection at the end of the discussion?	31
153	where does this section belong to? Chapter 'NDVI Correction' or 'Further Work'? .	31
154	which data? I assume the combine harvester point data?	34
155	page breaks	39
156	replace space before ref by tilda	42

¹⁵⁷ Notations

¹⁵⁸ Variables

c	a (vector of) constant(s)
$\lambda \in \mathbb{R}$	a scalar
$n \in \mathbb{N}$	sample size
i, j	are indices in $\{1, \dots, n\}$
$x \in \mathbb{R}^n$	covariable in 1-dim interpolation setting
¹⁵⁹ $w \in \mathbb{R}^n$	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 $y - \hat{y}$

¹⁶⁰ 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 4.4.0.1
OOB	Out Of the Box. Describes the procedure of estimating the value for a point but not consider the point itself (c.f. section 2.4.2)

¹⁶¹ XXX ML models and their shortnames

¹⁶² European Space Agency (ESA)

¹⁶³ **MATLAB Matrix Notation**

¹⁶⁴ We will use the MATLAB ‘:’ notation to indicate rows and columns of a matrix. That is
¹⁶⁵ 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
¹⁶⁶ X .

167 **Chapter 1**

168 **Introduction**

169 **1.1 XXX motivation - why is it important**

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

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

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

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

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

179 **1.2 XXX problembaum / fragestellungen**

180 problem schilderung anhand referenzen und evtl. eines bileds:

181 **1.3 XXX State-of-the-art**

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

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

- 184 — Doublelogistic (winter-ndvi)
- 185 — parametric / non-parametric approaches
- 186 — spatio-temporal approaches

187 **1.4 Research Questions**

188 XXX

189 1.5 Roadmap – anderer name XXX

190 This thesis is structured as follows: XXX

191 **Chapter 2**

192 **Data and Methods**

193 We will start by describing the available data and the challenges associated with it. Our
194 study region is a farm of over 800ha, which is located in western Switzerland. From
195 [Perich, Turkoglu, Graf, Wegner, Aasen, Walter, and Liebisch \(2022\)](#) we acquire satellite
196 image data (section 2.1), yield maps of several cereals from 2017 to 2021 (section 2.2),
197 and meteorological data (section 2.3). Afterwards, we will introduce general methods in
198 section 2.4, which will be used in the remaining chapters.

199 **2.1 Sentinel 2 Data**

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

204 The S2 images contain 12 spectral bands with spatial resolutions up to 10 meters (see
205 [2.1](#)). Bands with a lower resolution (20 and 60 meters) were upscaled to 10 meter reso-
206 lution using cubic interpolation ([Perich et al. \(2022\)](#)). In order to decrease the effect of
207 atmospheric conditions like reflections and scattering, bottom-of-atmosphere, radiometric
208 corrected Level-2A data was used². The ESA also supplies an algorithm³ produces Scene
209 Classification Layer (*SCL*) where for each location the observed subject is assigned to one
210 of 11 *SCL*-classes (c.f. [table 2.2](#)). In this thesis, we will use this classification to filter out
211 data points, which we believe to be less informative. That are all observations which *SCL*-
212 class does not correspond to vegetation or bare soils (classes 4 and 5). For convenience,
213 we define the set *SCL45* as the observations which belong to *SCL*-class 4 or 5.

214 **2.2 Crop Yield Data**

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

¹REF: <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”

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

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

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)

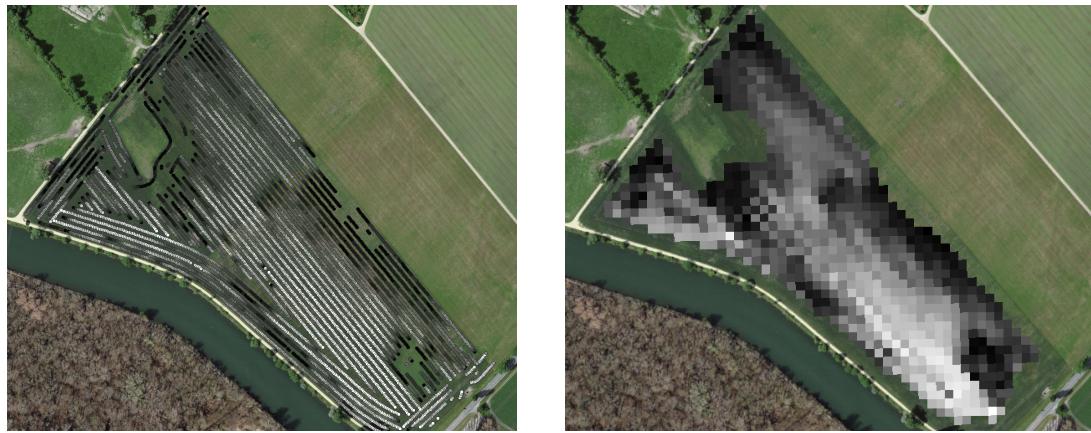
Color	No.	Class	Color	No.	Class
	0:	Missing Data		6:	Water
	1:	Saturated or defective pixel		7:	Cloud low probability
	2:	Dark features / Shadows		8:	Cloud medium probability
	3:	Cloud shadows		9:	Cloud high probability
	4:	Vegetation		10:	Thin cirrus cloud
	5:	Bare soils		11:	Snow or ice

218 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](#)).

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

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

223 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)

2.3 The Concept of a ‘Pixel’

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

2.3.1 Normalized Difference Vegetation Index (NDVI)

The well-known (*NDVI*) introduced in [Rouse \(1974\)](#) can be calculated using the bands *B4* and *B8* (table 2.1) by:

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

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

To define a timescale, we consider Days After Sowing (*DAS*) and a transformed timescale, Growing Degree Days (*GDD*) ([McMaster and Wilhelm \(1997\)](#)REF). The latter are defined as the cumulative sum (since sowing) of temperature above a given base temperature T_{base} . For cereals, we use $T_{base} = 0$ ([Perich et al. \(2022\)](#)). 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).

2.3.2 Challenges in S2 Data

The figure 2.2 shows a selection of 6 satellite images of a field, which display our challenges. In February (image a), we see no vegetation but bare soil. At the beginning of May, we

Please clarify this in more detail. We used pixels flagged with SCL 4 & 5, but as can be seen in Fig. 2.1 d), this can yield erroneous NDVI values, etc.

Für den Leser wäre es interessant,

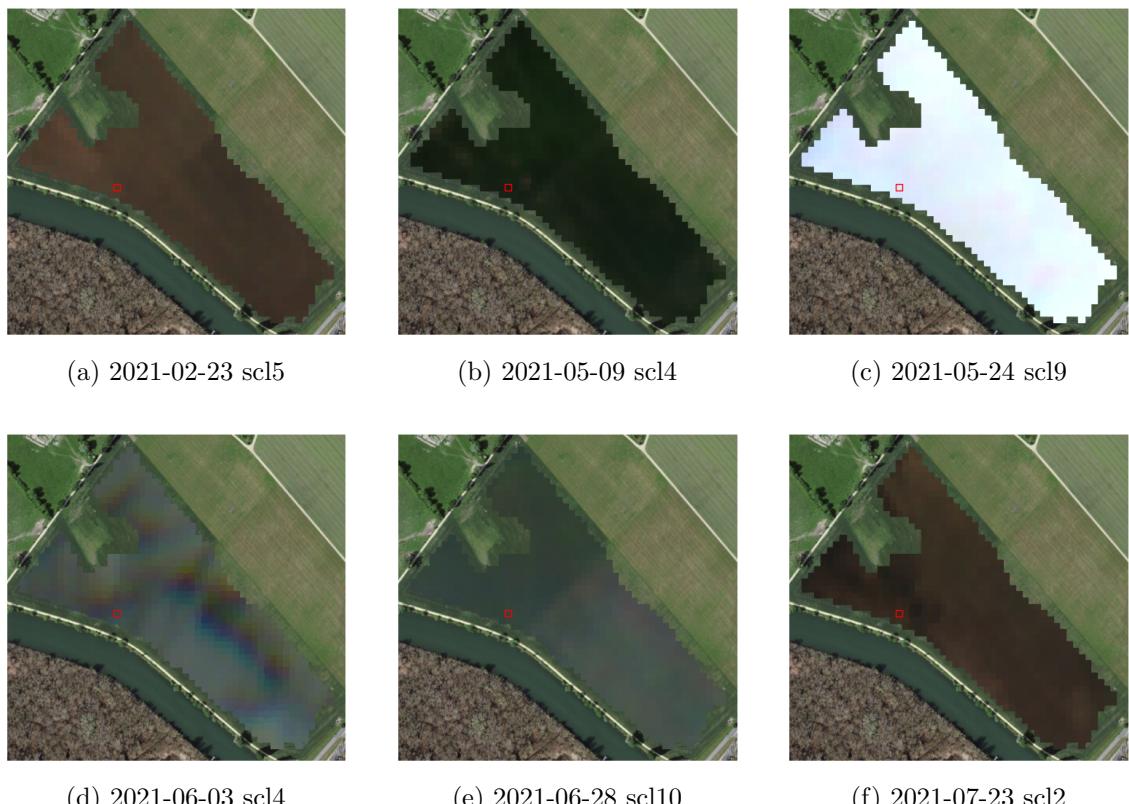


Figure 2.2: Satellite images of a field at selected times with a static background for orientation. The SCL-class of the highlighted pixel is provided in the respective subtitle. (???xxx include scl legend?)

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

255 2.4 General Methods

256 Here we will only introduce Methods which will accure in several places. For interpolation
 257 methods we refer to sections 3.3 and 3.4, for a robustification strategy to section 3.6. In
 258 section 3.5 we describe a method to objectively determine the quality of an interpolation,
 259 and in section 4.2 we present the NDVI correction together with an adapted interpolation
 260 strategy.

261 2.4.1 Root Mean Square Error (RMSE)

262 In this section we describe different criteria to evaluate models. Hence, given a vector
 263 $y \in \mathbb{R}^n$ and its estimator \hat{y} (estimated using the model), we define the RMSE as:

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

264 2.4.2 Out-Of-Bag (*OOB*) and Leave-One-Out-Cross-Validation (*LOOCV*)

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

To formally introduce this, let:

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

268 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
 269 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
 270 obtain $\hat{y} := (\hat{y}_1, \dots, \hat{y}_n)$ the OOB estimator for $y \in \mathbb{R}^n$.

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

273 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
 274 the corresponding \hat{y}_i also the LOOCV estimator.

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

278 **Chapter 3**

279 **Interpolation Methods**

280

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

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

287 Afterwards, we introduce and discuss each method.

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

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

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

292 **3.1 DAS vs. GDD**

294 Prior to interpolating the NDVI time series, we should decide on a timescale. We can
295 choose between DAS and GDD (c.f. section 2.3 and equation 2.3.1). In figure 3.1 we see
296 an example for comparison of the two. Here we see that the first 120 DAS are compressed
297 to just 500 GDD. This has several advantages. First, it makes the scales comparable (in
298 terms of plant growth) because the plants are not concerned with the month of the year
299 but the current temperature. Second, in winter we tend to have higher cloud cover and
300 thus fewer SCL45 observations. Hence, this gap in observations is compressed. Therefore,
301 we will only use GDD in the subsequent.

put sec-
tion in
methods
/ data

302 **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)

³⁰³ We will introduce parametric and non-parametric approaches to estimate m in section 3.3 and 3.4.

³⁰⁵ 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.

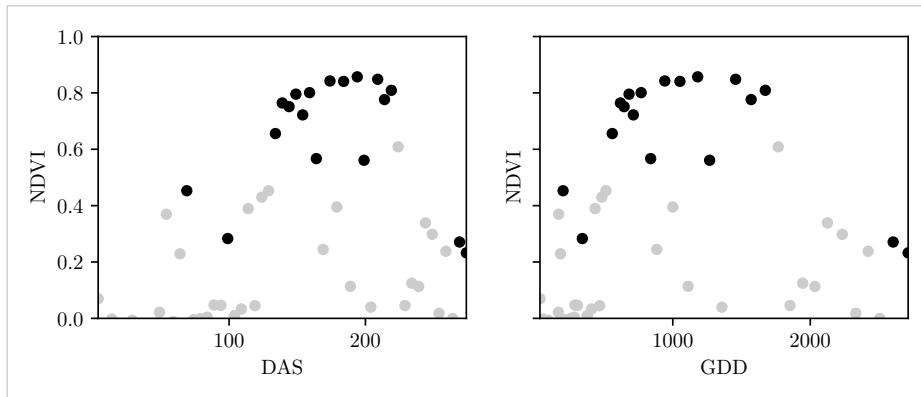


Figure 3.1: The same NDVI time-series, on the left with DAS as the timescale, on the right GDD is the timescale. SCL45 are colored black. Non-SCL45 (clouds and shadows) are colored in gray.

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

308 Ähnliche struktur sich überlegen

309 3.3 Parametric Regression

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

313 3.3.1 Double Logistic

314 The Double Logistic smoothing as described in Beck, Atzberger, Høgda, Johansen, and
 315 Skidmore (2006)REF heavily relies on shape assumptions of the fitted curve (i.e. the NDVI
 316 time series).

317 Assumptions:

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

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

322 — The growth cycle can be divided into an increase and a decrease period, where
 323 the time series follows a logistic function. The maximum increase (or decrease) is
 324 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:

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

325 Where the five free parameters: y_{\max} , d_0 , d_1 , t_0 , t_1 are initially estimated by least squares.
 326 Such fit can be seen in figure 3.2.

³²⁷ Similar as for the Savitzky-Golay Filter (c.f. section 3.4.3) we reestimate (only once) the parameters by giving less weight to the overestimated observations and more weight to ³²⁸ the underestimated observations¹. ³²⁹

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 3.2)

³³⁰ 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:

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

³³¹ 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 (c.f. figure 3.2) — Hard to interpret estimated parameters — Parameter estimation can go wrong. Introducing bounds can help.

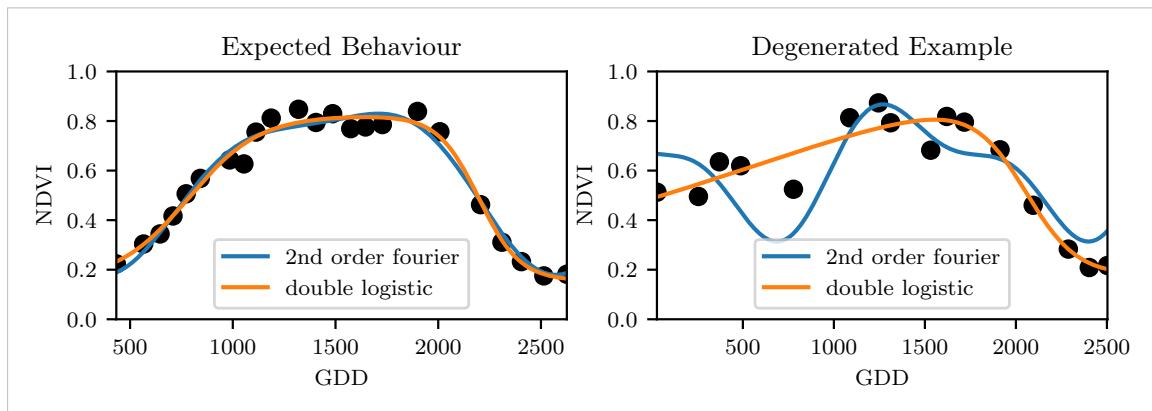


Figure 3.2: Here we observe the possibilities of a precise fit for the two parametric methods but notice also some misbehavior

¹For the details on the weights we refer to Beck et al. (2006)

332 **3.3.3 Optimization Issues**

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

338 **3.4 Non-Parametric Regression**

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

TODO:
include
Weighted
versions

343 **3.4.1 Kernel Regression**

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

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

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

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

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

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

345 Common choices for the kernel are the normal function or a uniform function (also called
 346 ‘box’ function.). Note that we still need to choose the bandwidth of the function (in
 347 the case of the normal function, this is σ the standard deviation). For local adaptive
 348 bandwidth selection we refer to [Brockmann, Gasser, and Herrmann \(1993\)](#).

Die
Aufzäh-
lung
ist hier
m.M.n.
nicht so
passend
für
einen
”Fliess-
text”.

Advantages

- 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 c.f. CompStat 3.2.2)

Disadvantages

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

349 **3.4.2 Kriging**

350 Kriging was developed in geostatistics to deal with autocorrelation of the response variable
 351 at locations which are spatially close. By applying the notion that two spectral indices
 352 which are (timewise) close should also take similar values, we justify the application of
 353 Kriging. In the end, we would like to fit a smooth Gaussian process to the data. For this
 354 subsection, we will follow [Diggle and Ribeiro \(2007\)](#).

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

ohne
un-
terkapi-
tel
struk-
tur

356 Definitions and Assumptions

357 **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 $\gamma(t, t') = \text{Cov}(S(t), S(t'))$

361 **Assumption 1.** We will assume the Gaussian process to be stationary. That is for $\mu(t)$
 362 to be constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the
 363 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

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

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

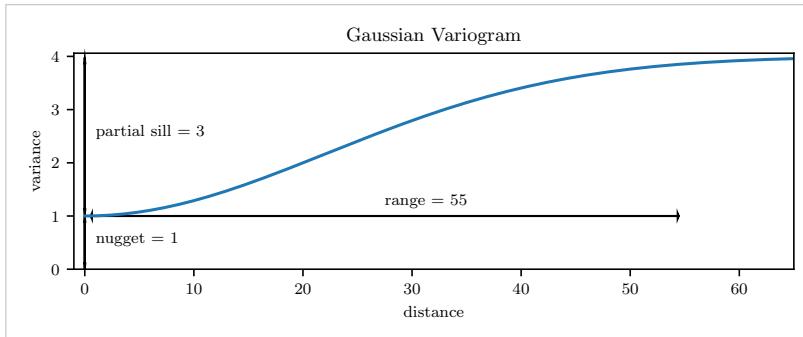


Figure 3.3: Gaussian Variogram with nugget=1, partial sill=3, range=55

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

370 Since we observe a clear pattern of a growth period in spring and harvest in the end of
 371 summer, we have to admit that assumption 1 with the constant mean is clearly violated.
 372 This is also the reason why we observe (for every variogram parameter) a tendency to the
 373 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 $P(A|B) = P(B|A)/P(B)$

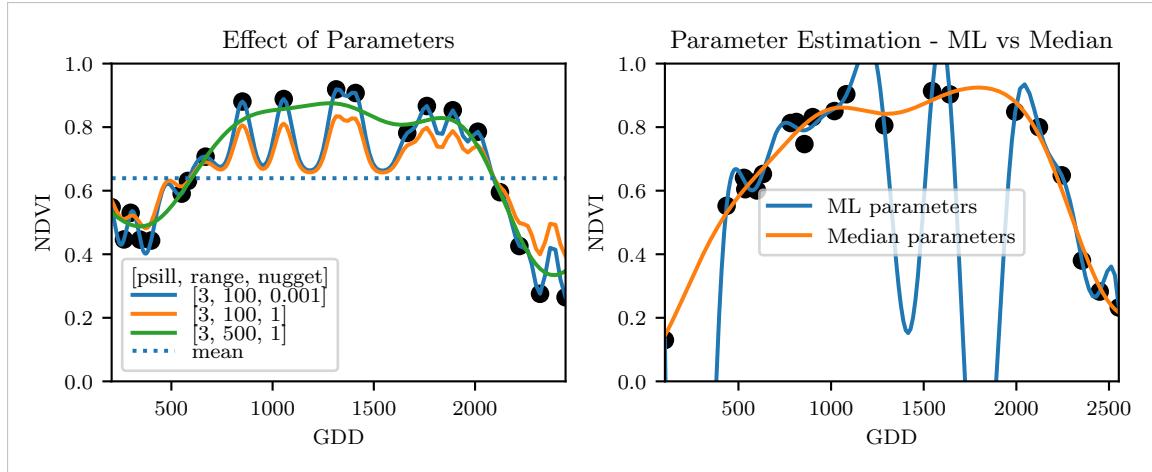


Figure 3.4: On the left, we see how the interpolation change if we increase the nugget and the range parameter. On the right, we compare two kriging interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such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

374 3.4.3 Savitzky-Golay Filter (SG Filter)

The *Savitzky-Golay Filter*, introduced in [Savitzky and Golay \(1964\)](#) is a technique in signal processing and can be used to filter out high frequencies (low-pass filter) ([Schafer, 2011](#)). 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:

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

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

For equidistant points this can efficiently be calculated by

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

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

377 Adaptation to the NDVI

378 [Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh \(2004\)](#) developed a ‘robust’ 379 interpolation method for the NDVI based on the SG Filter. The method is based on the

figure /
tabelle
/ pseudocode
anstatt
aufzählung

- 380 assumption that due to atmospheric effects the observed NDVI tends to be underestimated
 381 and that it cannot increase too quickly. The latter is argued by the biological impossibility
 382 of such fast vegetation changes. Their proposed algorithm is:
- 383 i.) Remove points which are labeled as cloudy.
 - 384 ii.) Remove points which would indicate an increase greater than 0.4 within 20 days.
 - 385 iii.) Linearly interpolate to obtain an equidistant time series X^0 .
 - 386 iv.) Apply the SG Filter to obtain a new time series X^1 .
 - 387 v.) Update X^1 by applying again a SG Filter. Repeat this until $w^T |X^1 - X^0|$ stops
 388 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
 389 reduces the penalty introduced by outliers⁷ and by repeating this step we approach
 390 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.

391 Extension: Spatial-Temporal-Savitzky-Golay Filter

392 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
 393 being discontinuous and that we can improve by looking at adjacent pixels⁸. Because we
 394 are working with rather high resolution satellite data, and we need the variance in the
 395 predictors, we will waive this extension.

397 3.4.4 Locally Weighted Regression (LOESS)

398 The Locally Weighted Regression (LOESS) introduced by [Cleveland \(1979\)](#) can be under-
 399 stood as a generalization of the SG Filter (c.f. 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

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

400 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
 401 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

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

403 The LOESS smoother takes a fraction of points instead of a fixed number and therefore
 404 automatically adapts to the size of the data we wish to interpolate. However, we run
 405 into the danger of considering too little observations, since the estimation breaks down if
 406 $\lceil \alpha n \rceil < d + 1$.⁹ Furthermore, LOESS gives less weight to points further away. This yields
 407 a "smoother" estimate, since when we slide the window (e.g. for estimating the next value)
 408 an influential point at the border does not suddenly get zero weight from being weighted
 409 equally before. Finally, the LOESS also can be used for non-equidistant data and allows
 410 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)

411 **3.4.5 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:

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

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

412 **3.4.6 Natural Smoothing Splines**

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

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

415 is a natural¹² cubic spline (i.e. a piecewise cubic polynomial function). The objective
 416 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).

417 **3.5 Tuning Parameter Estimation**

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

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

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

432 Figure 3.5 exemplifies the effect of the optimization function (different quantiles). To
 433 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 2.4.2

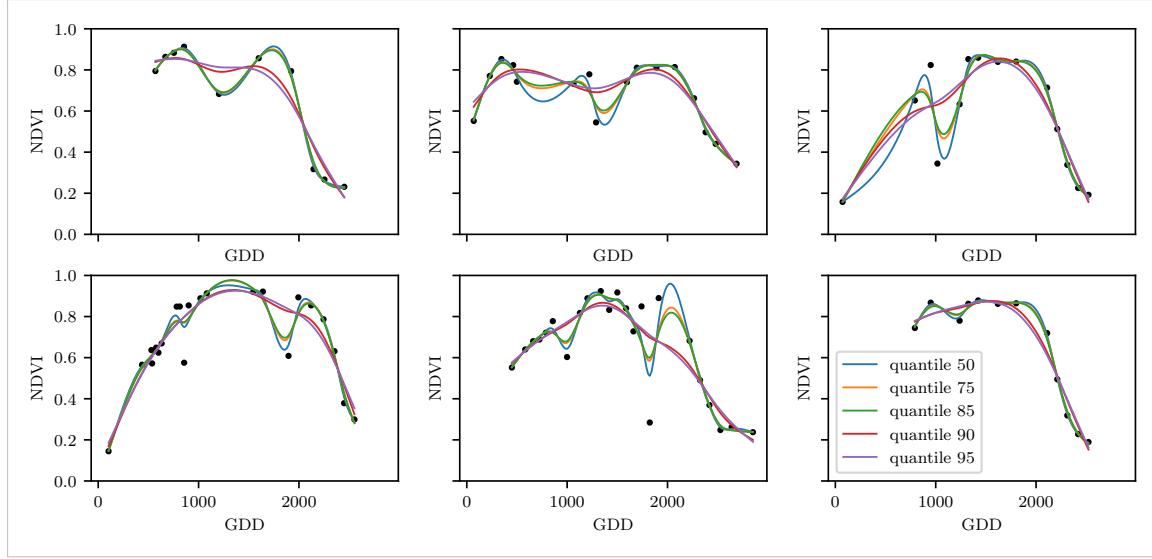


Figure 3.5: Smoothing splines fit with smoothing parameter optimized by minimizing the given quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

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 (c.f. section 3.4.4 and Cleveland (1979)), 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-

447 **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

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

448 for $r, w \in \mathbb{R}^n$. ¹⁶

449 **3.6.2 Examples and Conclusions**

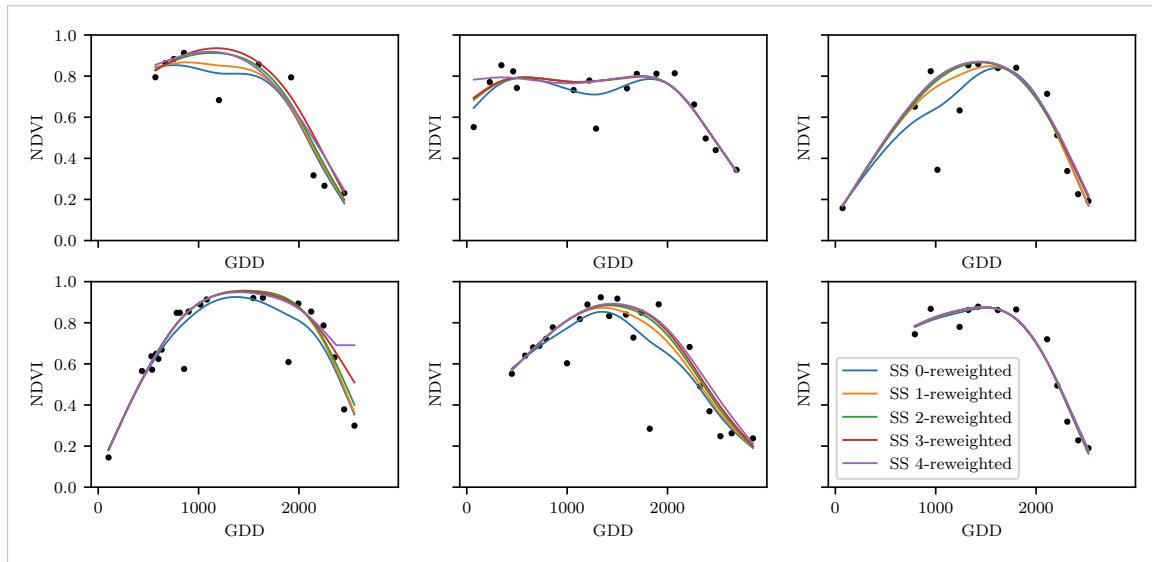


Figure 3.6: Smoothing Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

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

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

consider naming the subplots

459 **3.6.3 Upper Envelope Approach - Penalty for Negative Residuals**

460 If we artificially increase the negative residuals in 3.6.0.1 by multiplying (e.g. factor 2),
 461 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

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

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

467 3.7 Performance Assessment

468 Next, we will benchmark the different interpolation methods with and without robustifi-
 469 cation. For this, we will use the same technique as we did for the parameter determina-
 470 tion in section 3.5. On B_λ we apply the RMSE and different quantiles and present the results
 471 in table 3.2.

472 3.8 XXX Evaluation

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

write
out key-
words,
after
final re-
sults

480 **Chapter 4**

481 **NDVI Correction**

482 Let's remind ourselves that the data from the S2 is distributed with an SCL and we
483 therefore have some information about what is observed at each pixel for each sampled
484 time (c.f. table 2.2). So far, we have only considered cloud-free points (i.e., SCL-classes
485 4 and 5). In this chapter, we would like to improve the NDVI interpolation by inspecting
486 also other SCL-classes and by using more information than just the two bands used to
487 calculate the NDVI (B4 and B8).

488 **4.1 Considering other SCL Classes**

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

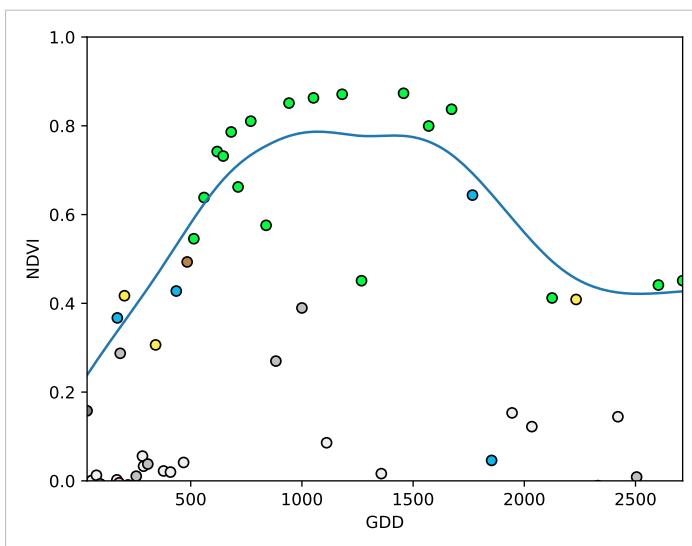


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45)

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

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

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

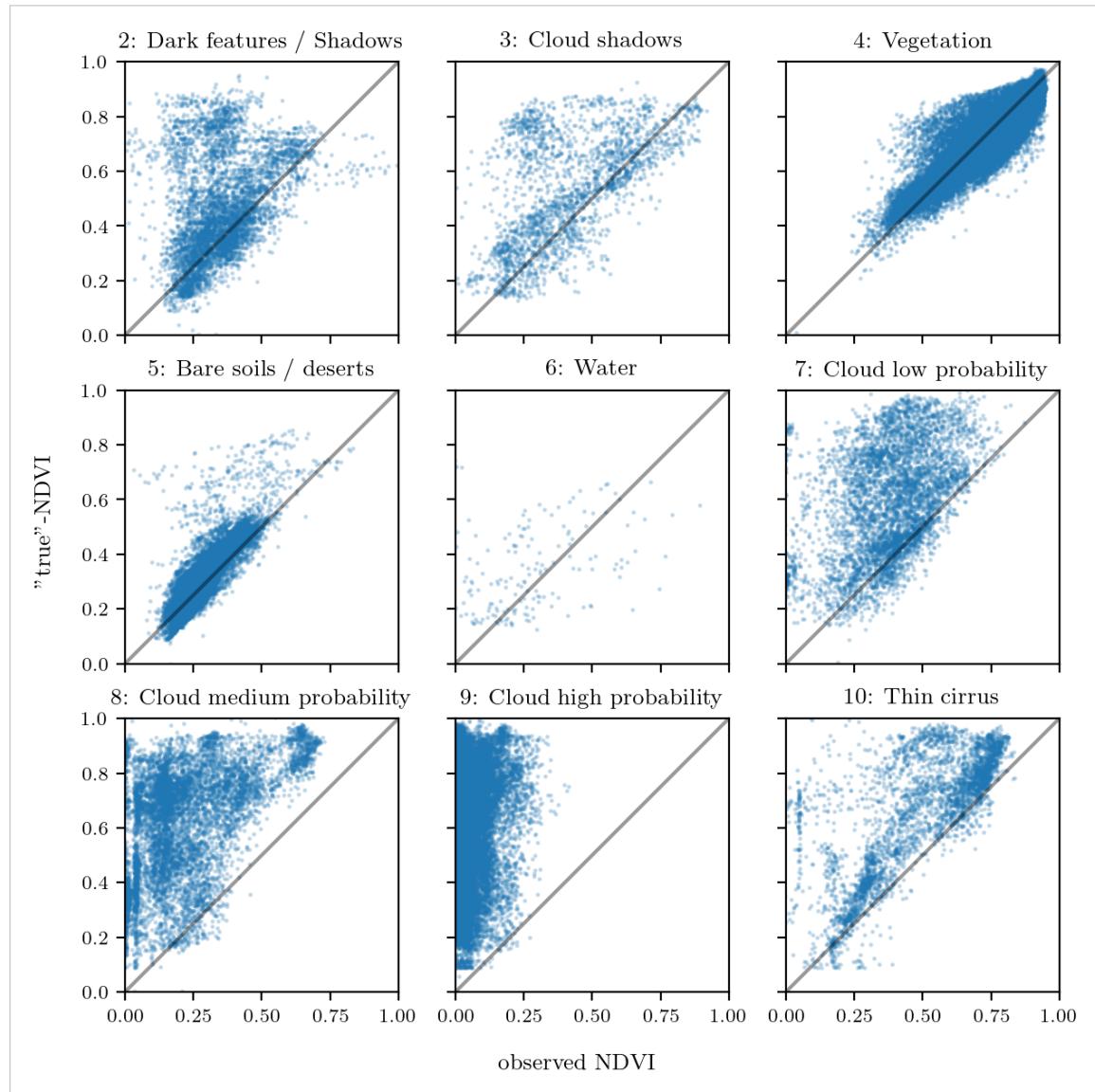


Figure 4.2: For each SCL class, we compare the true NDVI with the observed NDVI. (The true NDVI was estimated with OOB smoothing splines, and we used all observations of 10% of the total training pixels.)

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

511 4.2 Correction

512 We recall the satellite images in figure 2.2d, where we had cloudy images despite SCL4
 513 labeled and see fragments in figure 2.2e even though we are supposed to see clouds (SCL
 514 10 - Cirrus clouds). The SCL classification is based only on a mixed model trained using
 515 the S2 bands.

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

522 Finally, in section 4.4 we will evaluate this correction procedure, considering different
 523 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

524 4.2.1 Response and Covariates

525 For training an NDVI correction model, we need ground-truth (response) and informative
 526 covariates. We organize those in a table, where each row corresponds to a P_t (i.e., a pixel at
 527 a time t). Since ground-truth NDVI data is not available, we will again use the assumption
 528 1 and use the “true” NDVI instead. There is no canonical answer to the question of
 529 which covariates we should use. It is a tradeoff between simplicity/generalizability and
 530 performance (with the danger of overfitting). Our desire with the NDVI correction is
 531 to develop a product that is simple for others to understand and use. Therefore, in the
 532 subsequent, we will only take the spectral data of the satellite (i.e. all the bands) and the
 533 observed NDVI derived from it as covariates.

534 4.2.2 Correction Methods

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

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

540 Ordinary Least Squares (OLS)

541 The OLS is a linear model which aims to minimize the sum of the squared residuals. Let
 542 $y \in \mathbb{R}^n$ be the vector of responses and $X \in \mathbb{R}^{n \times p}$ be the design matrix, where each row
 543 corresponds to one pixel and each column consist of one covariate¹. We assume a linear

¹Strictly speaking, since SCL-classes are dummy variables

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

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

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

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

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

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

548 LASSO

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

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

551 Even though we do not have a closed form solution for equation (4.2.2.1) we can solve
552 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
553 continuous and convex.

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

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

561 We will apply the Lasso using the selected covariates in section 4.2.1 and their second
562 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 $\{a, b, ab, a^2, b^2\}$.

563 **Random Forest (*RF*)**

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

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

574 To build a *Random Forest* we will bootstrap-aggregate⁸ many such Trees⁹. The prediction
 575 of the Random Forest for a new point x is then the mean of the predictions from all the
 576 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.

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

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

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

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

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

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

584 and the set $\mathcal{M} = \{1\}$ of all functions currently in the model. Now, consider \mathcal{C} the set of
 585 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.

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

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

590 To reduce computational complexity, we follow the recommendation from REFStephen
 591 (2021) and restrict h in equation (4.2.2.2) to be of degree one (so it is also in a pair of \mathcal{B}).
 592 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).

593 General Additive Model (*GAM*)

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

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

598 To estimate the non-parametric functions, we can use smoothing splines (ref sec. 3.4.6).
 599 For this let \mathcal{S}_j be the function which takes some $z \in \mathbb{R}^n$ and returns the smoothing splines
 600 fitted to $(X_{:,j}, z)$ where the smoothing parameter is optimized by GCV. Since we cannot
 601 fit all g_j simultaneously, we will use a strategy named Backfitting. We basically cycle
 602 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

603 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.

604

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 $j = 1, \dots, p$.

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

605 **4.2.3 Uncertainty Estimation**

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

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

612 **4.2.4 Interpolation**

613 Consider now a pixel P , $\hat{y}^{(P)}$ its corrected NDVI and $\hat{v}^{(P)}$ the estimated uncertainties of
 614 $\hat{y}^{(P)}$. In order to interpolate $\hat{y}^{(P)}$, we will give less weight to unreliable observations. Thus,
 615 we define the weight function:

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

616 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.
 617 The normalization is needed since for some interpolation methods, inflating the sum of
 618 weights would decrease the effect of the smoothing.

619 **4.3 Resulting Interpolation Strategies**

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

- 621 i.) OOB Interpolation (+ robustify?)
- 622 ii.) Correction
- 623 iii.) Uncertainty estimation
- 624 iv.) Interpolation (+ robustify?)

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

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

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

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

635 In this fashion, we obtain 28 distinct interpolation strategies, which we will benchmark in
 636 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.

637 4.4 Evaluation Method

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

640 **Definition 4.4.0.1.** (*RYEA*) Let $y \in \mathbb{R}^n$ be the yield, M be a model for estimating y , and
 641 $\hat{y} = M(X)$ where X describes the data¹³. We define the *RYEA* as the relative RMSE in
 642 yield estimation. Formally expressed:

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

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

644 4.4.1 Idea

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

653 4.4.2 Yield Estimation

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

- 659 — Maximum slope
- 660 — Minimum slope
- 661 — Integral¹⁴ over all
- 662 — Peak (i.e. maximal NDVI)
- 663 — Peak GDD (i.e. value at which the peak is attained)
- 664 — Integral¹⁴ up to the peak
- 665 — Integral¹⁴ after peak
- 666 — Integral¹⁴ from 0-685 GDD
- 667 — Integral¹⁴ from 685-1075 GDD

tabelle
wäre
sauberer

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

check
refer-
ence

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

671 As a result, for each interpolation strategy, a matrix is obtained in which each row corre-
672 sponds to a pixel and both the yield and the characterizing statistics are contained. Using
673 this matrix, we train a random forest for yield estimation, and compute the integrated
674 OOB estimates¹⁵ \hat{y} . Note that the choice of the modeling approach does not matter much,
675 as long as it is general enough (i.e. able to approximate any function) and we use the same
676 one for each interpolation strategy. Finally, for each interpolation strategy, we calculate
677 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 $\frac{1}{e}$).

678 **Chapter 5**

679 **Results**

680 **5.1 XXX small recap from “Interpolation Methods”**

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

682 **5.2 Robustification and NDVI-Correction**

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

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

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

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

Table 5.1: 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

687 **Chapter 6**

688 **Discussion**

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

690 SCL is prone to errors as can be seen in figure 2.2. A machine learning approach like the
691 one developed in [Raiyani, Gonçalves, Rato, Salgueiro, and Marques da Silva \(2021\)](#) could
692 be used instead.

693 **6.1 Interpolation Methods**

694 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?

695 XXX discuss results from table

696 **6.2 NDVI Correction**

697 **6.2.1 Bootstrap**

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

704 **6.2.2 Using Additional Covariates**

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

709 We decided against using this data, because on the one hand we have the problem that
710 we have practically too few observations (we observe only 5 years) and we expect the

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

712 weather in our study region to be rather homogeneous which is suggested by the fact
713 that the weather data published by Meteoswiss are for a grid with a resolution of 1 km.
714 On the other hand, we want the underlying model not to learn improper relationships.
715 For example, the model might automatically predict a high NDVI for a day in summer
716 (detected by high GDD / many sunshine hours / high temperature) just because it is
717 “used” to observing a lot of vegetation in summer. Including temporally (e.g., P_{t-1} and
718 P_{t+1}) and geographically adjacent pixels would likely improve performance. However, for
719 simplicity, we omit it here¹.

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

723 6.2.3 High RMSE in Yield Prediction

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

- 725 i.) Uncertainty in Yield data collected by the combine harvester
- 726 ii.) Uncertainty in Yield data through rasterization
- 727 iii.) Uncertainty in satellite images through “measurement errors” introduced via clouds
728 and other atmospheric effects
- 729 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).

730 **Chapter 7**

731 **Conclusion**

732

733

```
- 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 > bspl  
  choose ss because of its meaningful definition (minimizing the integral of the second  
  - robustifying useful?
```

744

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

746 **7.1 Future Work**

747 **7.1.1 Time Series Correction-Interpolation as a General Method**

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

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

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

761 **7.1.2 Minor Improvements**

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

- 764 — **Data:** Method how data has been extrapolated to the grid could possibly be improved
- 765 — **Data:** For computational reasons, we mostly considered all years and split the data
766 (on the pixel level) randomly into a train/test set. A leave one year out cross
767 validation might yield more accurate results.
- 768 — **Data:** We have not included the spectral bands which have a resolution of 60m. But
769 precisely these seem to be promising for cloud correction, since they are a proxy of
770 the water (content and form) in the atmosphere.
- 771 — **NDVI Correction:** Explore the effect of different link functions between the esti-
772 mated absolute residuals and the weights in section [4.2.4](#).
- 773 — **NDVI Correction:** Yield is not the only target variable of interest. Other variables
774 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?

775 Bibliography

- 776 (2007). Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.),
777 *Model-Based Geostatistics*, pp. 46–78. New York, NY: Springer.
- 778 Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore (2006,
779 February). Improved monitoring of vegetation dynamics at very high latitudes: A new
780 method using MODIS NDVI. *Remote Sensing of Environment* 100(3), 321–334.
- 781 Breiman, L. (2001, October). Random Forests. *Machine Learning* 45(1), 5–32.
- 782 Brockmann, M., T. Gasser, and E. Herrmann (1993, December). Locally Adaptive Band-
783 width Choice for Kernel Regression Estimators. *Journal of the American Statistical
784 Association* 88(424), 1302–1309.
- 785 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
786 spatiotemporal information with the Savitzky-Golay filter. *Remote Sensing of Environ-
787 ment* 217, 244–257.
- 789 Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004, June). A
790 simple method for reconstructing a high-quality NDVI time-series data set based on the
791 Savitzky–Golay filter. *Remote Sensing of Environment* 91(3), 332–344.
- 792 Cleveland, W. S. (1979, December). Robust Locally Weighted Regression and Smoothing
793 Scatterplots. *Journal of the American Statistical Association* 74(368), 829–836.
- 794 Friedman, J. H. (1991, March). Multivariate Adaptive Regression Splines. *The Annals of
795 Statistics* 19(1), 1–67.
- 796 Hastie, T. and R. Tibshirani (1987, June). Generalized Additive Models: Some Applica-
797 tions. *Journal of the American Statistical Association* 82(398), 371–386.
- 798 Jaramaz, D., V. Perović, S. Belanovic Simic, E. Saljnikov, D. Cakmak, V. Mrvić, and
799 L. Zivotic (2013, May). The ESA Sentinel-2 mission Vegetation variables for Remote
800 sensing of Plant monitoring.
- 801 Lyche, T. and K. Mørken (2005, January). Spline Methods.
- 802 McMaster, G. S. and W. W. Wilhelm (1997, December). Growing degree-days: One
803 equation, two interpretations. *Agricultural and Forest Meteorology* 87(4), 291–300.
- 804 Perich, G., M. O. Turkoglu, L. V. Graf, J. D. Wegner, H. Aasen, A. Walter, and F. Liebisch
805 (2022, July). Pixel-based yield mapping and prediction from Sentinel-2 using spectral
806 indices and neural networks.

- 807 Raiyani, K., T. Gonçalves, L. Rato, P. Salgueiro, and J. R. Marques da Silva (2021,
808 January). Sentinel-2 Image Scene Classification: A Comparison between Sen2Cor and
809 a Machine Learning Approach. *Remote Sensing* 13(2), 300.
- 810 Ripley, B. D. and M. Maechler. R: Fit a Smoothing Spline. [https://stat.ethz.ch/R-
811 manual/R-patched/library/stats/html/smooth.spline.html](https://stat.ethz.ch/R-manual/R-patched/library/stats/html/smooth.spline.html).
- 812 Rouse, J. W. (1974, May). Monitoring the vernal advancement and retrogradation (green
813 wave effect) of natural vegetation. Technical Report NASA-CR-139243.
- 814 Savitzky, A. and M. J. E. Golay (1964, July). Smoothing and Differentiation of Data by
815 Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), 1627–1639.
- 816 Schafer, R. W. (2011, July). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE
817 Signal Processing Magazine* 28(4), 111–117.
- 818 Skakun, S., E. Vermote, B. Franch, J.-C. Roger, N. Kussul, J. Ju, and J. Masek (2019,
819 July). Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating
820 Surface Reflectance, Through Phenological Fitting, into Regression Yield Models.
821 *Remote Sensing* 11(15), 1768.
- 822 Stephen, M. (2021, July). Earth: Multivariate Adaptive Regression Splines.
- 823 Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective.
824 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(3), 273–
825 282.

826 Appendix A

827 Reproducibility

828 A.1 Reproduce Results

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

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

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

835 A.2 R-Package

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

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

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

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

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

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

```
844
845 1 library(CorrectTimeSeries)
846 2
847 3 # load a list of dataframes, each one describes one pixel with the covariates and
848 4 # the response
849 5 data(timeseries_list)
850 6 str(timeseries_list[[1]])
851 7
852 8 # Train/Load RF
853 9 train_model_myself <- TRUE
854 10 if (train_model_myself){
855 11     # Add "true" NDVI (or generally the response), by Out-Of-Bag estimation
856 12     timeseries_list <- lapply(timeseries_list, function(df) {
857 13         df$oob_ndvi <- OOB_est(df$gdd, df$ndvi_observed) # gdd is the time-axis
858 14         df
859 15     })
860 16     # Train correction model
861 17     formula <- "oob_ndvi ~ B02+B03+B04+B05+B06+B07+B08+B8A+B11+B12+scl_class"
862 18     RF <- train_RF_with_fromula(formula, timeseries_list, robustify=TRUE)
863 } else {
```

```
864 19  data(RF_for_NDVI)
865 20  RF <- RF_for_NDVI
866 21 }
867 22
868 23 # ADD CORRECTION
869 24 timeseries_list <- lapply(timeseries_list, function(df) {
870 25   df$corrected_ndvi <- randomForest:::predict.randomForest(RF, df)
871 26   df
872 27 })
873 28
874 29 # Get interpolation for each timeseries
875 30 newx <- 1:1000
876 31 lapply(timeseries_list, function(df){
877 32   ss <- smoothing_spline(df$gdd, df$corrected_ndvi)
878 33   predict(ss, newx)$y
879 34 })
```

Example of how to use the `CorrectTimeSeries` package

881 **Appendix B**

882 **Further Material**

883 **B.1 Interpolation**

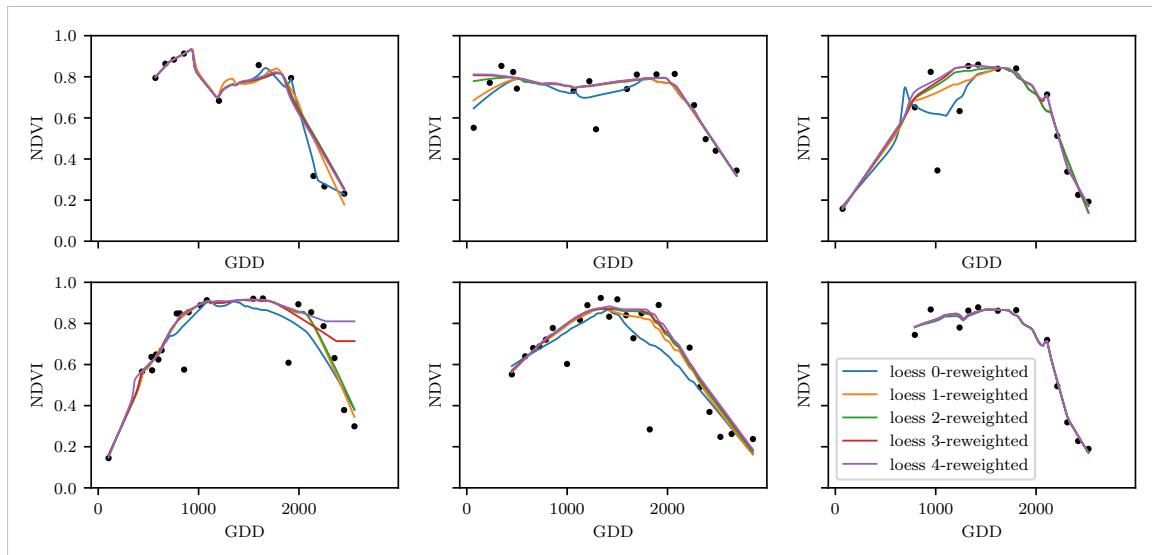


Figure B.1: The LOESS smoother fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

884 **B.2 NDVI correction**

885 page breaks

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

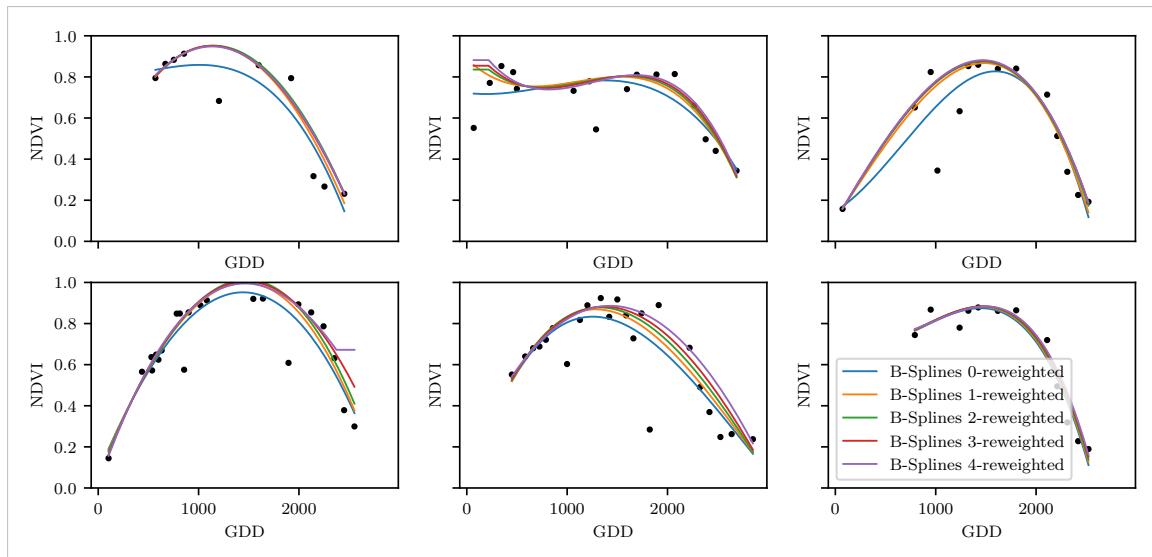


Figure B.2: B-Splines fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

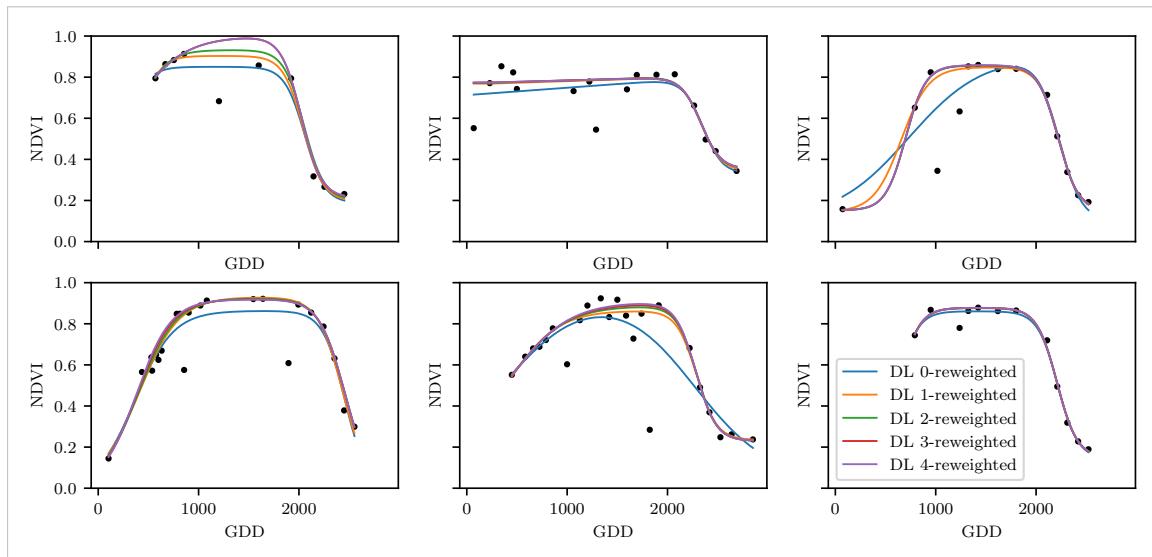


Figure B.3: A Double Logistic curve fitted to different (SCL45) NDVI time series. Iterations of a robustifying refit (as indicated in section 3.6) are also displayed

```

886
887 1 Call:
888 2 lm(formula = (paste(response, " ~ ", "ndvi_observed + scl_class")),
889 3     data = ndvi_df)
890
891 5 Residuals:
892 6   Min      1Q  Median      3Q      Max
893 7 -0.7997 -0.0717  0.0039  0.0695  0.6632
894
895 9 Coefficients:
896 10             Estimate Std. Error t value Pr(>|t|)
897 11 (Intercept)  0.21465   0.00230  93.46 < 2e-16 ***
898 12 ndvi_observed 0.71116   0.00346 205.65 < 2e-16 ***
899 13 scl_class3    0.02205   0.00356   6.20  5.8e-10 ***

```

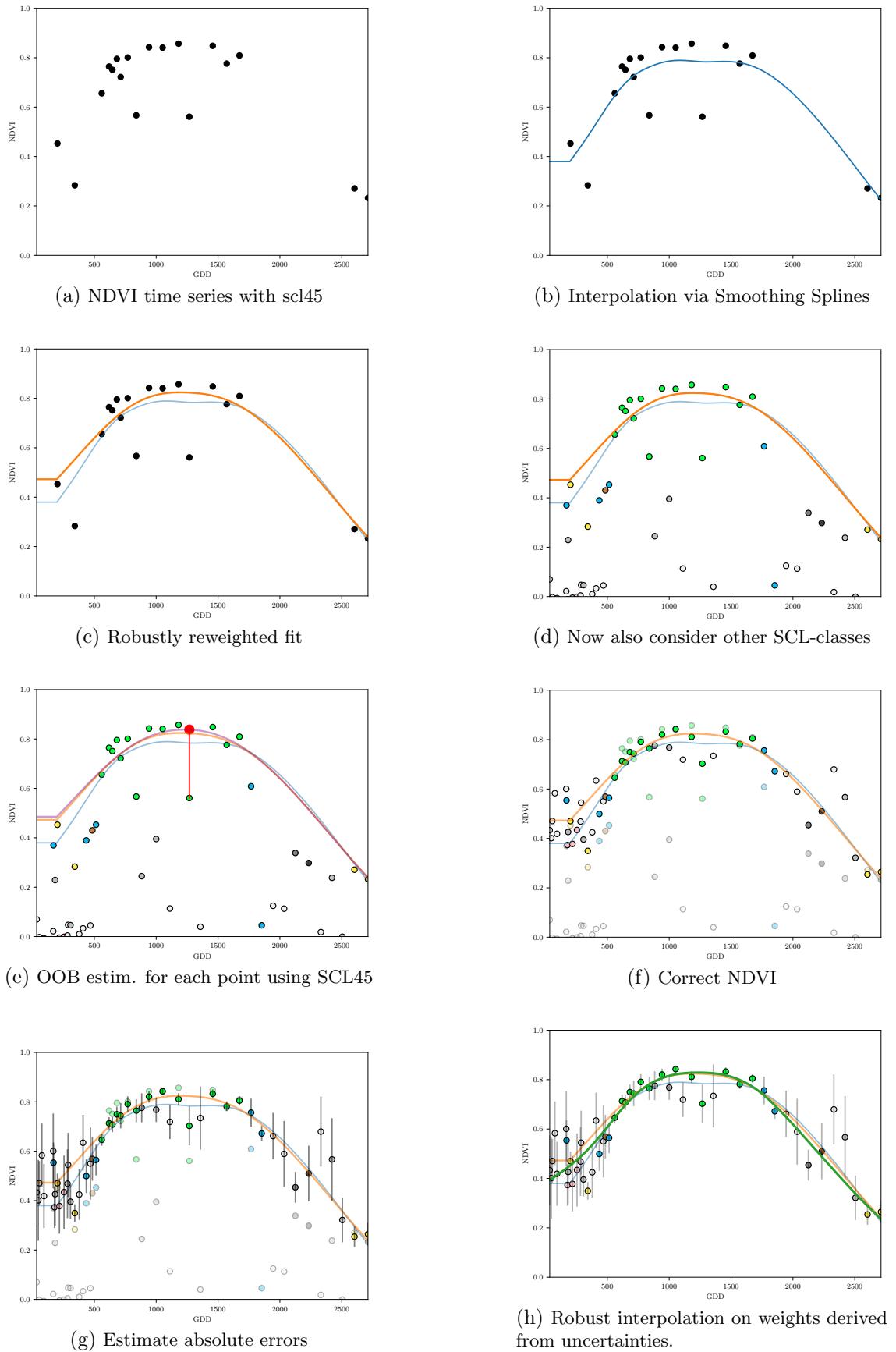


Figure B.4: Stepwise illustration of robust NDVI-Correction. For the color encoding of the SCL classes we refer to table 2.2.

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

```

900 14 | scl_class4    -0.00431   0.00251   -1.72    0.085 .
901 15 | scl_class5   -0.09875   0.00234   -42.15   < 2e-16 ***
902 16 | scl_class6   -0.05301   0.01104   -4.80    1.6e-06 ***
903 17 | scl_class7    0.11245   0.00274   41.09    < 2e-16 ***
904 18 | scl_class8    0.25963   0.00253   102.57   < 2e-16 ***
905 19 | scl_class9    0.35994   0.00236   152.47   < 2e-16 ***
906 20 | scl_class10   0.09091   0.00308   29.54    < 2e-16 ***
907 21 | scl_class11   0.29784   0.00392   76.06    < 2e-16 ***
908 22 |
909 23 | Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
910 24
911 25 | Residual standard error: 0.146 on 124978 degrees of freedom
912 26 | Multiple R-squared:  0.532,      Adjusted R-squared:  0.532
913 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)

```

915
916 1 | Call:
917 2 | lm(formula = (paste(get_res(), " ~ ", "ndvi_observed + scl_class")),
918 3 |   data = ndvi_df)
919
920 5 | Residuals:
921 6 |   Min     1Q   Median     3Q   Max
922 7 | -0.2051 -0.0427 -0.0074  0.0329  0.6589
923
924 9 | Coefficients:
925 10 |            Estimate Std. Error t value Pr(>|t|)
926 11 | (Intercept) 0.18647   0.00126 147.74   < 2e-16 ***
927 12 | ndvi_observed -0.13265  0.00190 -69.80   < 2e-16 ***
928 13 | scl_class3   -0.00180  0.00196 -0.92    0.3587
929 14 | scl_class4   -0.04069  0.00138 -29.55   < 2e-16 ***
930 15 | scl_class5   -0.09698  0.00129 -75.32   < 2e-16 ***
931 16 | scl_class6   -0.01906  0.00606 -3.14    0.0017 **
932 17 | scl_class7   0.01641   0.00150 10.91    < 2e-16 ***
933 18 | scl_class8   -0.00560  0.00139 -4.02    5.7e-05 ***
934 19 | scl_class9   -0.01384  0.00130 -10.67   < 2e-16 ***
935 20 | scl_class10  -0.00690  0.00169 -4.08    4.5e-05 ***
936 21 | scl_class11 -0.01446  0.00215 -6.72    1.8e-11 ***
937 22 |
938 23 | Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
939 24
940 25 | Residual standard error: 0.08 on 124978 degrees of freedom
941 26 | Multiple R-squared:  0.352,      Adjusted R-squared:  0.352
942 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)

replace space before ref by tilda