



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 Instructions and the relevant code needed to reproduce this thesis can be found in the
18 GitHub repository and to use our results we recommend the provided R-package.
19 More information is given in the appendix A.

20 Acknowledgements

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

24 This endeavor would not have been possible without Gregor Perich. His high personal com-
25 mitment, reliability as well as the weekly instructive supervision meetings were essential
26 for this work.

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

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

35 Abstract

36 xxx abbreviations

37 Multispectral satellite imagery is utilized to estimate Time Series (TS) of spectral indices
38 at ground-level. As such, the TS of the Normalized Difference Vegetation Index (NDVI) is
39 used to model vegetation development. Satellite measurements may not match the ground
40 signal due to contamination by atmospheric effects (e.g., clouds or shadows). Therefore,
41 traditional approaches try to filter out contaminated observations before extracting and
42 subsequently interpolating the NDVI. After filtering, remaining contaminated observations
43 and resulting data gaps are the two challenges for interpolation that we address in this
44 thesis. For this purpose, we use cereal crop yield maps from 2017-2021 of a farm in Switzer-
45 land and corresponding Sentinel 2 satellite image TS published by the European Space
46 Agency. Contaminated observations were filtered with the provided Scene Classification
47 Layer (SCL). We give a benchmark-supported review of different interpolation methods.
48 Based on it, we opt for Smoothing Splines as a flexible non-parametric method and Double
49 Logistic approximation as a parametric method with implicit shape assumptions. In addition,
50 we generalize an iterative technique which robustifies interpolation methods against
51 outliers by reducing their weight. In most cases, this robustification successfully decreased
52 the 50% and 75% quantiles of the absolute out-of-bag residuals. Moreover, we present a
53 general interpolation procedure that utilizes additional information to correct the target
54 variable with an uncertainty estimate and then performs a weighted interpolation. In our
55 setting, the target variable is the NDVI and as additional information we use the SCL,
56 the observed NDVI and the spectral bands. Consequently, we no longer filter using the
57 SCL but weight observations according to their reliability. Applying this procedure, the
58 variance in crop yield explained by the resulting NDVI TS decreases by 5.4%.

59 Contents

60	Notation	v
61	1 Introduction	1
62	2 Data and Methods	3
63	2.1 Sentinel 2 Data	3
64	2.2 Crop Yield Data	3
65	2.3 Normalized Difference Vegetation Index	4
66	2.4 Transformation of Timescale	5
67	2.5 The Concept of a ‘Pixel’	6
68	2.6 Illustration of S2 Images	6
69	2.7 Estimation Evaluation Criteria	6
70	2.7.1 Score Functions	8
71	2.7.2 Out-Of-Bag (OOB) and Leave-One-Out-Cross-Validation (LOOCV)	8
72	3 Interpolation Methods (IMs)	9
73	3.1 Interpolation Setup	9
74	3.2 Parametric Regression	9
75	3.2.1 Double Logistic (DL)	11
76	3.2.2 Fourier Series (FS)	11
77	3.2.3 Optimization Issues	12
78	3.3 Non-Parametric Regression	12
79	3.3.1 Kernel Regression: Nadaraya-Watson (NW)	12
80	3.3.2 Universal Kriging (UK)	13
81	3.3.3 Savitzky-Golay Filter (SG)	15
82	3.3.4 Locally Weighted Regression (LOESS)	16
83	3.3.5 B-Splines (BS)	17
84	3.3.6 Smoothing Splines (SS)	17
85	3.4 Tuning Parameter Estimation	18
86	3.5 Robustification	19
87	3.5.1 Our Adjustment:	20
88	3.5.2 Examples and Conclusions	20
89	3.5.3 Upper Envelope Approach — Penalty for Negative Residuals	20
90	3.6 Performance Assessment	21
91	4 NDVI Correction	22
92	4.1 Considering other SCL Classes	22
93	4.2 Correction Models	23
94	4.2.1 Ordinary Least Squares (OLS)	23
95	4.2.2 Least Absolute Shrinkage and Selection Operator (LASSO)	24
96	4.2.3 General Additive Model (GAM)	25
97	4.2.4 Random Forest (RF)	25
98	4.2.5 Multivariate Adaptive Regression Splines (MARS)	26
99	4.3 Weighted Interpolation	27
100	4.4 Resulting Interpolation Strategies (ISs)	27
101	4.5 Evaluation via (relative) Yield Prediction Error (relative YPE)	28

102	5 Results	31
103	5.1 Goodness of Fit for Selected IMs	31
104	5.2 YPE for Tested ISs	31
105	6 Discussion	33
106	6.1 IMs	33
107	6.1.1 Data Gaps in Time Series	33
108	6.1.2 Preselection	34
109	6.1.3 Candidate Selection	34
110	6.2 NDVI Correction	34
111	6.2.1 Choose IS	34
112	6.2.2 Investigation of Error Sources in Yield Estimation	35
113	6.2.3 NDVI Correction as Unsupervised Learning	35
114	6.2.4 Using Additional Covariates	36
115	7 Conclusion	37
116	7.1 Future Work	38
117	7.1.1 Time Series Correction-Interpolation as a General Method	38
118	7.1.2 Minor Improvements	39
119	Bibliography	40
120	A Reproducibility	43
121	A.1 Reproduce Results	43
122	A.2 R-Package	43
123	B Further Material	45
124	B.1 Data and Methods	45
125	B.1.1 GDD	45
126	B.2 Interpolation	46
127	B.3 NDVI correction	47
128	B.3.1 OLS ^{SCL} Model Outputs	47

129 Notations

- 130 Since this thesis, despite its applied nature, is located at the Mathematics Department,
131 we adhere to the convention of speaking in the first-person plural “we”.
132 Furthermore, only equations that are referenced elsewhere are equipped with a number.

133 Variables

c	a (vector of) constant(s)
$\lambda \in \mathbb{R}$	a scalar
$n \in \mathbb{N}$	sample size
i, j	indices in $\{1, \dots, n\}$
$n \in \mathbb{R}^n$	time, usually in GDD
$w \in \mathbb{R}^n$	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 $y - \hat{y}$
$X \in \mathbb{R}^{n \times p}$	the design matrix. Each row corresponds to one observation and each column to one covariate.
$X_{[:,j]}$	the j^{th} column of X
$X_{[i,:]}$	the i^{th} row of X

134 Abbreviations and Objects

NDVI	Normalized Difference Vegetation Index (Rouse, 1974).
TS	Time Series.
IM	Interpolation Method. That is a simple ¹ method that interpolates data $(t_i, y_i)_{i=1, \dots, n}$ and yields a function $f(t) = y$, approximating the data.
IS	Interpolation Strategy. This is the category of functions that map $(t_i, y_i)_{i=1, \dots, n}$ to a function $f(t) = y$, approximating the data. So a IS describes a strategy of how to arrive at an interpolation starting from the data $(t_i, y_i)_{i=1, \dots, n}$. For this, initial data may be corrected (cf. chapter 4), (possibly different) IMs (iteratively) used, weightings applied (cf. robustification in section 3.5). Note, that strictly speaking, every IM also is an IS. But usually, we expect an IS to involve a more ‘complex’ procedure.
S2	Sentinel 2 satellites. Two multi-spectral image satellites deployed by the European Space Agency.

¹I.e., no combination of various methods.

SCL	Scene Classification Layer provided by the European Space Agency 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, see table 2.2
Pixel	A pixel originates of an image pixel and describes a square of 10 x 10 meters in the field that 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	the observed data (weather and spectral bands) at time t and the location of one pixel.
P	a pixel. We see it as a collection of all the observations at the specified location within one season. More formally, $P := \{P_t t \text{ is a valid sample time within a defined season}\}$
P^{SCL45}	is similar to P but we only consider observations that 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.
DAS	Days After Sowing
GDD	Growing Degree Days – cumulative sum of “max(0, temperature – threshold)”
YPE	(Relative) Yield Prediction Error. See Definition 4.5.0.1
OOB	Out Of the Box. Describes the procedure of estimating the value for a point by a model that has not seen this point before (see section 2.7.2).
LOOCV	Leave One Out Cross Validation. Describes the procedure of estimating the value for a point by a model that has seen all the points except the current one (see section 2.7.2).

135 Statistical Models

DL	Double Logistic (see section 3.2.1)
FS	Fourier Series (see section 3.2.2)
NW	Nadaraya-Watson (see section 3.3.1)
UK	Universal Kriging (see section 3.3.2)
SG	Savitzky-Golay Filter (see section 3.3.3)
LOESS	Locally Weighted Regression (see section 3.3.4)
BS	B-splines (see section 3.3.5)
SS	Smoothing Splines (see section 3.3.6)
OLS	Ordinary Least Squares (see section 4.2.1)
OLS^{SCL}	OLS using only the observed NDVI and SCL classes (as factor variables)
OLS^{all}	OLS using the covariates OLS^{SCL} uses and the spectral bands
LASSO	Least Absolute Shrinkage and Selection Operator (see section 4.2.2)
GAM	General Additive Model (see section 4.2.3)
RF	Random Forest (see section 4.2.4)
MARS	Multivariate Adaptive Regression Splines (see section 4.2.5)

136 **Chapter 1**

137 **Introduction**

138 Remote sensing aims to measure target variables efficiently from a distance. Large scale
139 monitoring of forest and agricultural vegetation dynamics is of great interest to author-
140 ties, insurance companies and research. Examples include crop classification for subsi-
141 dizing farmers ([Henits et al., 2022](#)) and the creation of crop models for estimating crop
142 yields or nitrogen concentrations ([Courault et al., 2021; Perich et al., 2021](#)). For this,
143 freely distributed multi-spectral satellite imagery from the Sentinel-2 (S2) satellites are
144 used ([ESA, 2022b](#)). In order to transform the high dimensional satellite images into easily
145 interpretable metrics, spectral indices such as the Normalized Difference Vegetation Index
146 (NDVI) are used ([Rouse, 1974](#)). The NDVI serves as a proxy for vegetation density (or
147 chlorophyll content to be more precise), and thus the corresponding NDVI Time Series
148 (TS) reflects the vegetation development. The quality of a satellite image, however, de-
149 pends on atmospheric conditions and thus in case of a dense cloud cover, the information
150 content derived from the NDVI is impaired. Therefore, [ESA \(2022a\)](#) also provides a Scene
151 Classification Layer (SCL), which provides additional metadata about what is observed
152 (e.g., shadows, clouds, vegetation, etc.) . So when extracting the NDVI TS from the
153 Sentinel 2 satellite imagery TS, we can filter out the contaminated observations using the
154 SCL classification. However, due to this filtration it may occur that we have no observa-
155 tions for several weeks, especially in winter. It is also possible that some observations are
156 wrongly classified by the SCL (e.g., as vegetation) and thus result in an erroneous NDVI
157 causing an outlier in the TS. Consequently, the main challenge is to interpolate an NDVI
158 TS, which can contain both large data gaps and outliers.

159 Currently, there are several approaches to address these issues. One is to look at the
160 observed evolution of vegetation density and assume its bell shape for the NDVI TS given
161 the strong correlation between NDVI and photosynthetic activity ([Gamon et al., 1995](#)).
162 Approaches to model this include a 2nd order Fourier approximation ([Stöckli and Vidale,
163 2004](#)) or a Double Logistic function ([Beck et al., 2006](#)). On the other hand, assumptions
164 are made about more abstract properties of the curve, such as smoothness. We divide
165 these into local and global approaches. Nadaraya-Watson ([Strbac et al., 2017](#)), Savitzky-
166 Golay Filter ([Chen et al., 2004](#)) and Locally Reweighted Regression ([Omori et al., 2021](#)) use
167 a sliding window to interpolate the TS stepwise. Global methods like B-Splines ([Gurung
168 et al., 2009](#)) and Smoothing Splines ([Cai et al., 2017](#)) reduce the squares of all residuals
169 simultaneously, and Universal Kriging fits a Gaussian process to the data ([Chandola and
170 Vatsavai, 2010](#)).

171 The research questions pursued in this thesis are:

- 172 i.) Which IMs are used in the context of NDVI, and what are their advantages and
173 disadvantages?
- 174 ii.) How may contaminated data be dealt with?
- 175 iii.) How do data gaps affect interpolation?
- 176 iv.) How to deal with data gaps?
- 177 v.) How can we recognize a good interpolation of the NDVI?

178 In this thesis, we will discuss the strengths and weaknesses of Interpolation Methods
179 (IMs) and evaluate them with respect to NDVI interpolation. For this purpose, we use
180 the Sentinel 2 satellite image TS and crop yield maps of different fields of different cereal
181 species on a farm in Witzwil, Switzerland over the years 2017-2021. After presenting the
182 available data, illustrating challenges and defining different concepts in chapter 2, we turn
183 to the two main blocks of this thesis. One covers the study of IMs and the other presents
184 a general procedure of correcting (NDVI) TS with uncertainty estimation by utilizing
185 additional information. On the first block, in chapter 3 we examine parametric and non-
186 parametric IMs and discuss their strengths and weaknesses (question i.). We generalize
187 and test an iterative technique that makes IMs more robust to outliers by weighting them
188 less (question ii.). To evaluate IMs, we present an approach that uses out-of-bag residuals
189 (question v.). In section 6.1.1, we discuss how different IMs respond to data gaps (question
190 iii.), and in section 6.1.2 we preselect IMs. We evaluate this preselection in the results
191 section 5.1 and select two candidates from different IMs in section 6.1.3. For the second
192 block, we correct possibly contaminated data with statistical models in chapter 4 (question
193 ii.) and utilize previously ignored observations, which we hope will further reduce data
194 gaps (question iv.). Thus, we no longer filter the observations a priori via the SCL, but
195 instead correct the observed NDVI and weight the observations via estimated uncertainties.
196 By combining different statistical models and IMs, we get 28 ISs. We compare those with
197 a vegetation-oriented quality measure (question v.) and describe the results in section 5.2.
198 Based on these results, in section 6.2 we argue what the best IS is. In addition, we justify
199 why our NDVI correction can be understood as unsupervised learning and why we relied
200 only on satellite imagery and not on meteorological data for the NDVI correction. Our
201 conclusions of this thesis, recommendations, as well as an outlook on future work is given
202 in chapter 7.

203 **Chapter 2**

204 **Data and Methods**

205 This section describes the available data and the challenges associated with it. Our study
206 region is a farm of over 800ha, which is located in western Switzerland. From Perich et al.
207 (2022) we acquired Sentinel-2 (S2) satellite image data (section 2.1), yield maps of several
208 cereals from 2017 to 2021 (section 2.2), and meteorological data (section 2.5). Methods to
209 evaluate an estimator or model are given in section 2.7.

210 For IMs we refer to sections 3.2 and 3.3, for a robust IS to section 3.5. In section 3.4
211 we describe a method to objectively determine the quality of an interpolation, and in
212 chapter 4 we present the NDVI correction together with an adapted IS.

213 xxx define IMs and ISs?

214 **2.1 Sentinel 2 Data**

215 The European Space Agency (ESA, 2022b) freely distributes images of the S2 satellites.
216 Together, both satellites have a revisit time of 5 days in our study region.

217 The S2 images contain 12 spectral bands with spatial resolutions of up to 10 meters (see
218 table 2.1). Bands with a lower resolution (20 and 60 meters) were upscaled to 10 meter
219 resolution using cubic interpolation (Perich et al., 2022). In order to decrease the effect of
220 atmospheric conditions like reflections and scattering, bottom-of-atmosphere, radiometric
221 corrected Level-2A data was used. ? supplies the Scene Classification Layer (SCL). It is
222 a model output that for each location assigns the observed pixel to one of 11 SCL-classes
223 (cf. table 2.2). In this thesis, we will use this classification to filter out data points that
224 we believe to be less informative. These are all observations in which the SCL-class does
225 not correspond to vegetation or bare soils (classes 4 and 5). We define the set SCL45 as
226 the observations that are labelled as SCL-class 4 or 5.

227 **2.2 Crop Yield Data**

228 The crop yield data were collected from a combine harvester. Equipped with a satellite-
229 based navigation system, the harvester drives over the fields and continuously estimates the
230 dry crop yield density in t/ha (see fig. 2.1a). We use the data set presented in Perich et al.
231 (2022), where error-prone measurement points (such as during a tight curve of the combine

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

232 harvester) were removed and then the yield map was rasterized using linear interpolation
 233 (cf. fig. 2.1b). We summarize the rasterized dry yield values by the following statistics:

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

235 Comparing the average per-field crop yield reported by the farmer with the yield estimated
 236 by the combine harvester shows that the latter overestimates crop yield by ca. 10% (Perich
 237 et al., 2022). Since the relative estimation error is approximately constant and we do not
 238 aim for an absolute yield prediction, we will not consider this deviation.

239 2.3 Normalized Difference Vegetation Index

240 The well-known NDVI introduced by Rouse (1974) is used to approximate vegetation in
 241 remote sensing due to a healthy, photosynthesizing plant exhibiting a large increase in

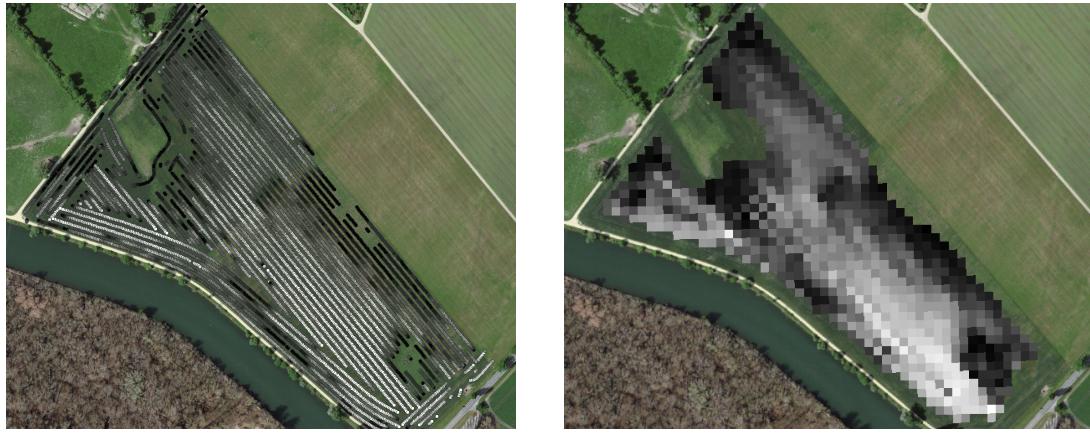


Figure 2.1: Crop yield density map of a field. Ranges from 0.1 t/ha (black) to 5.35 t/ha (white)

reflectance between the red and infrared region of the light spectrum. It is calculated using the S2 bands $B4$ (red) and $B8$ (infrared) (table 2.1) by:

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

Since we measure the NDVI via the S2 satellites from space, we cannot expect to obtain the same NDVI as measured with a ground-based spectroradiometer. This is especially true if the ground signal is obfuscated by either the clouds directly or by cloud shadows. Even if we only use SCL45 observations flagged as cloud-free, we still encounter measurement errors, as described in section 2.6. Therefore, we call the calculated values merely the observed NDVI. In the following chapters, we will study the resulting NDVI TS extensively. Such a TS is shown in figure 2.2.

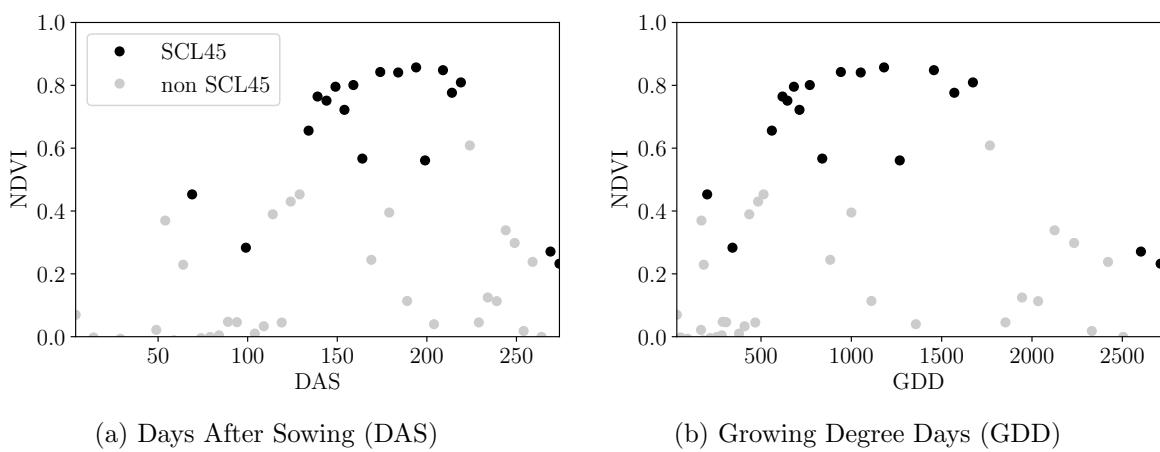


Figure 2.2: NDVI TS plotted against DAS and GDD. GDD are introduced in section 2.4.

250

2.4 Transformation of Timescale

Two drawbacks become apparent when using Days After Sowing (DAS) as the timescale (see figure 2.2a): First, this scale makes it difficult to compare two NDVI TS because

wheat is not always sown on the same day of the year and in some years plants begin to emerge earlier. Second, because there are only few SCL45 observations in autumn and winter, we face significant data gaps during this period. To fix both problem [McMaster and Wilhelm \(1997\)](#) proposes a transform the timescale into a meaningful temperature based one. The resulting Growing Degree Days (GDD) are defined as the cumulative sum of temperature above a given base temperature T_{base} since sowing. For cereals, we use $T_{base} = 0$ ([Holzkämper et al., 2015](#)). Thus, the GGD for n days after sowing will be equal to:

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

Important plant growth stages and their corresponding GDD values are tabulated in [B.1.1](#). In figure [2.2](#) we see an example for comparison of the DAS and GDD timescale. Here we see that the first 120 DAS are compressed to just 500 GDD and hence the gap in observations was successfully compressed. Due to the reasons mentioned above, from now on we will only consider GDD.

2.5 The Concept of a ‘Pixel’

Now we create a new data structure that we call Pixel. This originates from the pixels of the S2 satellite images. It will contain all the information needed to answer the research questions in the following chapters.

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

2.6 Illustration of S2 Images

Using an example pixel, we illustrate the challenges in working with S2 image data. Figure [2.3](#) shows a selection of 6 satellite images of a field, one selected Pixel and the NDVI TS of this pixel. In February (image a), we see no vegetation but bare soil and thus also a low NDVI. At the beginning of May (b), we observe a cloudless dark green field with a high NDVI. In (c) heavy cloud cover (SCL class 9) leads to a complete loss of plant information in this S2 observation. Figure (d) shows that the SCL classification is not reliable, since we evidently observe clouds which is also reflected in a sudden NDVI drop. Even though SCL indicates that (e) are thin cirrus clouds, we see a pale green, and we also note a NDVI. Therefore, we remark that some SCL45 observations are not accurate and even though a few non-SCL45 observations contain useful information, most of them are too unreliable (e.g., all SCL 9 observations). Thus, we aim to substitute the unreliable ones with interpolated versions and correct corrupt ones.

2.7 Estimation Evaluation Criteria

In this section, we define score functions for estimation accuracy and techniques to adequately apply it.

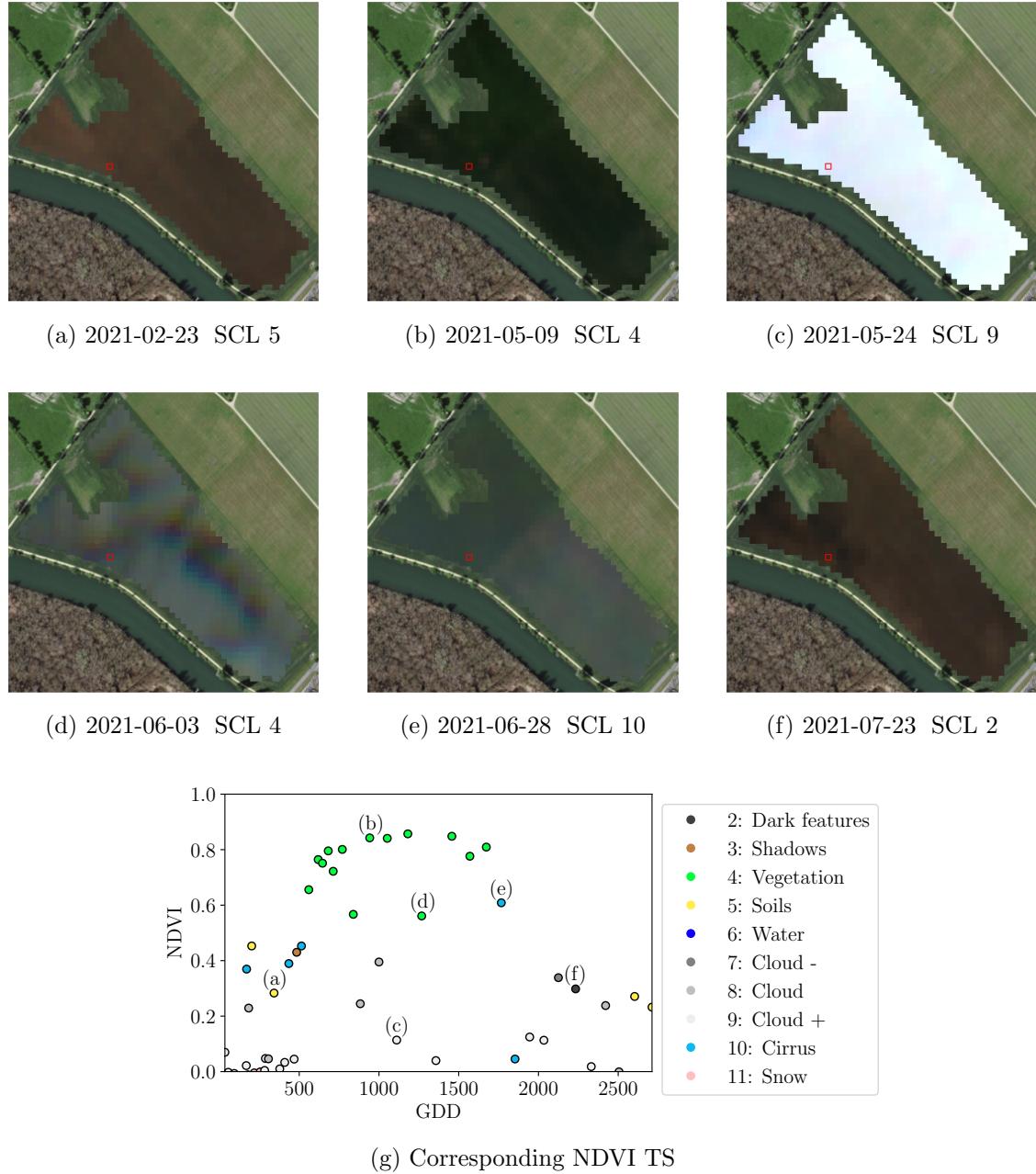


Figure 2.3: Satellite images of a field at selected times with a static background for orientation. Moreover, the NDVI TS of the red-highlighted pixel is shown in (g) colored by the SCL labels.

293 **2.7.1 Score Functions**

294 Now, we define the Root Mean Square Error (RMSE), a prominent and outlier-sensitive
 295 score function.

296 **Definition 2.7.1.1.** *XXX (RMSE) Given a vector $y \in \mathbb{R}^n$ and its estimator \hat{y} , we define
 297 the RMSE as:*

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

298 Thus, the RMSE measures how close the estimated \hat{y} are to the original y by considering
 299 the squares of errors. The lower the RMSE, the more similar are the fitted values to the
 300 original values. Note that one strong outlier may corrupt this score function. Therefore,
 301 we will also define the $x^{\text{th}}\%$ Quantile of Absolute Residuals (QAR x) that can handle some
 302 fraction of outliers:

303 **Definition 2.7.1.2.** *(QAR x) Given a percentage $x \in \{1, 2, \dots, 100\}$, a vector $y \in \mathbb{R}^n$ and
 304 its estimator \hat{y} , we assume that $|y_1 - \hat{y}_1| \leq |y_2 - \hat{y}_2| \leq \dots \leq |y_n - \hat{y}_n|$. Then, we define
 305 the QAR x as the biggest residual under the $x\%$ smallest ones. Formally:*

$$\text{QAR}^{x\%} := \max \left\{ |y_i - \hat{y}_i| : i \leq \frac{x}{100} n \right\}$$

306 Note that QAR 50 coincides with the median of absolute residuals and QAR 100 with the
 307 maximum of absolute residuals. Hence, the higher, x the fewer outliers QAR x can handle.
 308 Consequently, if we expect the data to have 5% outliers, we should choose x smaller than
 309 95. Furthermore, if an estimator attains lower QAR x values than a competing estimator
 310 for $x = 50, 75, 85$, we conclude¹ that it also produces smaller residuals in most cases.

311 **2.7.2 Out-Of-Bag (OOB) and Leave-One-Out-Cross-Validation (LOOCV)**

The rationale for OOB and LOOCV is that we intend to evaluate a model M with unseen data. That is, if D describes the entire dataset, and we train a model on a subset of D , we can use the remaining data to evaluate the model. To formally introduce this, let:

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

312 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
 313 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
 314 obtain $\hat{y} := (\hat{y}_1, \dots, \hat{y}_n)$ the OOB estimator for $y \in \mathbb{R}^n$.

315 In the case that $M^{(-i)}$ was fitted on the set $D \setminus \{(X_i, y_i)\}$ (i.e., not a true subset), we
 316 call the corresponding \hat{y}_i also the LOOCV estimator. If we optimize some parameter via
 317 OOB (or LOOCV) this means that we search for the parameter that minimizes some loss
 318 function which takes the OOB (or LOOCV) residuals. In the bootstrap (e.g., random
 319 forest) framework, we define \hat{y}_i to be the average of all computed and admissible $M^{(-i)}$.

¹Strictly speaking, this conclusion assumes that the distribution of the residuals is unimodal (i.e., has only one ‘peak’).

320 **Chapter 3**

321 **Interpolation Methods (IMs)**

322 In section 2.6 we have established the need for interpolating the NDVI TS. In this chapter,
323 we first specify a setting for the interpolation and divide the IMs into those that
324 make fundamental shape assumptions (parametric) and those that are more flexible (non-
325 parametric). We give an introduction for each method with a compact definition, highlight
326 adjustments or give remarks where appropriate, and then point out strengths and weak-
327 nesses of each method. Additionally, a brief overview of the considered IMs is provided
328 in table 3.1. Afterwards, we extract a robustification strategy from the one IM and gen-
329 eralize it, so we can use it for all methods that allow for a priori weighted observations.
330 Finally, using LOOCV, we tune the parameters (where necessary) and get a first idea of
331 the performance of each method.

332 **3.1 Interpolation Setup**

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

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

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

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

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

336 **3.2 Parametric Regression**

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

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

	Assumptions	Advantages	Disadvantages	w	b
Double- Logistic	<ul style="list-style-type: none"> - Function first increases then decreases - NDVI has a minimal value 	<ul style="list-style-type: none"> - Good for evergreen plants (if snow masks NDVI) - Upper envelope 	<ul style="list-style-type: none"> - Parameter estimation can be very difficult - Strange behavior for long data-gaps 	Yes	(Yes)
Fourier Series	<ul style="list-style-type: none"> - NDVI can be approximated by a 2cd order Fourier series. 	<ul style="list-style-type: none"> - Incorporates periodical growth-cycles 	<ul style="list-style-type: none"> - Parameter estimation can be very difficult - Curve easily exceeds the bounds of the NDVI 	Yes	No
Nadaraya- Watson (Kernel Smooth- ing)	<ul style="list-style-type: none"> - Close points are related to each other via a kernel function 	<ul style="list-style-type: none"> - Simple - Computationally very fast 	<ul style="list-style-type: none"> - Biased, especially at ‘peaks’ and ‘valleys’ - Bandwidth: fails if there are big data-gaps 	Yes	Yes
Universal Kriging	<ul style="list-style-type: none"> - Function is a realization of a stationary Gaussian process 	<ul style="list-style-type: none"> - Informative parameters - Flexible 	<ul style="list-style-type: none"> - Regression to the mean - Assumptions clearly not met 	Yes	(Yes)
SG	<ul style="list-style-type: none"> - High frequencies are noise (Low-Pass-Filter) - Equidistant points - Local polynomials 	<ul style="list-style-type: none"> - Computationally very fast 	<ul style="list-style-type: none"> - Cannot deal natively with missing data (need some interpolation) 	No	(Yes)
SG + NDVI	<ul style="list-style-type: none"> - Upper envelope - Vegetation cannot grow faster than some slope 	<ul style="list-style-type: none"> - Biological knowledge 	<ul style="list-style-type: none"> - Bad “upper envelope” since weights are not used for the estimation itself 	(No)	(Yes)
LOESS	<ul style="list-style-type: none"> - Local polynomial with points closer to the estimated point are more important 	<ul style="list-style-type: none"> - Flexible - Generalization of SG - Weighting function makes intuitive sense 	<ul style="list-style-type: none"> - Computationally expensive 	Yes	(Yes)
B-Splines (Smoothed)	<ul style="list-style-type: none"> - Function can be approximated by a linear combination of B-splines basis functions 	<ul style="list-style-type: none"> - General assumption - Flexible shape 	<ul style="list-style-type: none"> - Unbounded - No intuitive meaning for smoothing 	Yes	No
Smoothing splines	<ul style="list-style-type: none"> - 2cd derivative of function is integrable 	<ul style="list-style-type: none"> - Intuitive meaning of penalty - General assumptions - Flexible shape 	<ul style="list-style-type: none"> - Choice of smoothing parameter 	Yes	(Yes)

340 **3.2.1 Double Logistic (DL)**

The Double Logistic smoothing as described in Beck et al. (2006) heavily relies on shape assumptions of the fitted curve (i.e., the NDVI TS). First, we assume that there is a minimum NDVI level y_{\min} in the winter (e.g., due to evergreen plants), which might be masked by snow. This can be estimated beforehand, taking several years into account. Second, we assume that the growth cycle can be divided into an increase and a decrease period, where the TS follows a logistic function. The maximum increase (or decrease) is observed at t_0 (or t_1) with a slope of d_0 (or d_1). The equation of the double-logistic fit is given by:

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

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

343 **Robustification**

- 344 Similar as for the SG (cf. section 3.3.3) one can reestimate (only once) the parameters by
 345 giving less weight to the overestimated observations and more weight to the underestimated
 346 observations. For the details on the choice of the weights, we refer to Beck et al. (2006).
 347 We will not apply this reestimation, but rather the robustification introduced later in
 348 section 3.5.

Advantages	Disadvantages
<ul style="list-style-type: none"> — Incorporates subject specific knowledge in the case of evergreen plants covered in snow. — Optimized parameters have an intuitive meaning. 	<ul style="list-style-type: none"> — Strong shape assumptions on the NDVI curve. — Parameter optimization might go wrong. This can be mitigated to some extent by providing bounds for the parameters. — Strange behavior in regions with little observations (cf. figure 3.1).

349 **3.2.2 Fourier Series (FS)**

Stöckli and Vidale (2004) approximates the NDVI curve using a second order FS:

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

- 350 where $\Phi = 2\pi \times (t - 1)/n$. If we set the period to match one year, this would coincide
 351 with the notion that plants grow every year. Analogous to section 3.2.1 we fit it to the
 352 data by least squares. Example fits can be seen in figure 3.1

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

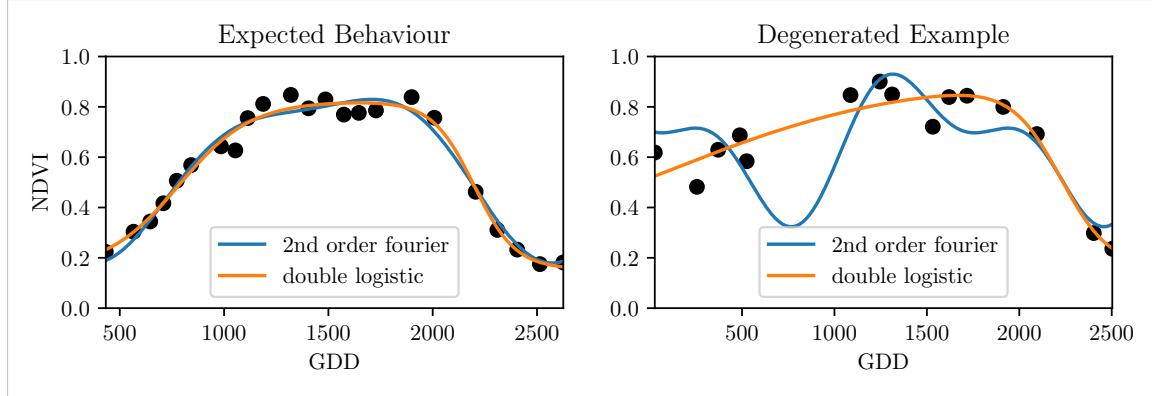


Figure 3.1: Here we observe the possibilities of a precise fit for the two parametric methods but notice also some misbehavior.

353 3.2.3 Optimization Issues

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

359 3.3 Non-Parametric Regression

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

363 3.3.1 Kernel Regression: Nadaraya-Watson (NW)

364 As described in section 3.1, we aim to estimate

$$\mathbb{E}[Y \mid T = t] = \int_{\mathbb{R}} y f_{Y|T}(y \mid t) dy = \frac{\int_{\mathbb{R}} y f_{T,Y}(t,y) dy}{f_T(t)}, \quad (3.3.1.1)$$

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

$$\hat{f}_T(t) = \frac{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right)}{nh}, \quad \hat{f}_{T,Y}(t,y) = \frac{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right) K\left(\frac{y-Y_i}{h}\right)}{nh^2}, \quad (3.3.1.2)$$

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

where h , the bandwidth, symbolizes the window size of to consider. By using the above function in equation (3.3.1.1) we arrive at the NW kernel estimator:

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

367 Common choices for the kernel are the normal function or a uniform function (also called
368 ‘bot’ function).

369 Choose Bandwidth

370 Note that we still need to choose the bandwidth of the function. This can be done with
371 the help of LOOCV while optimizing the RMSE. For non-equidistant data we refer to
372 Brockmann et al. (1993) where a local adaptive bandwidth selection is presented.

Advantages	Disadvantages
— Flexible due to different possible kernels.	— If the $t \mapsto K(t)$ is not continuous, \hat{m} isn’t either.
— Can be assigned degrees of freedom (trace of the hat-matrix).	— Choice of bandwidth, especially if t_i are not equidistant.
— Estimation of the noise variance $\hat{\sigma}_\varepsilon^2$. ²	

373 3.3.2 Universal Kriging (UK)

374 UK as described in dig (2007) was developed in geostatistics to deal with autocorrelation
375 of the response variable at locations that are spatially close. By applying the notion that
376 two spectral indices that are timewise close should also take similar values, we justify the
377 application of UK. In the end, we would like to fit a smooth Gaussian process to the data.

378 A Gaussian Process $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \dots, S(t_k))$ has a multi-
379 variate Gaussian distribution for every collection of times t_1, \dots, t_k . S can be fully charac-
380 terized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') := \text{Cov}(S(t), S(t'))$.
381 Furthermore, we will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be
382 constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following
383 only $\gamma(h)$.³

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

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

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

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

384 Here h denotes the distance, n is the nugget, r is the range and p is the partial sill. The
385 influence of the parameters is visualized in figure 3.2.⁴

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.

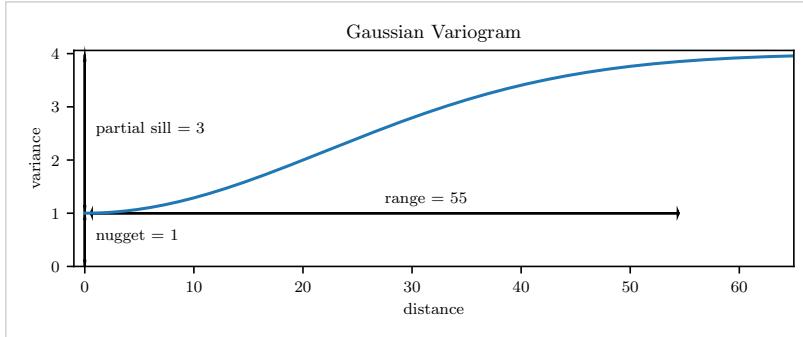


Figure 3.2: Gaussian Variogram with nugget=1, partial sill=3, range=55

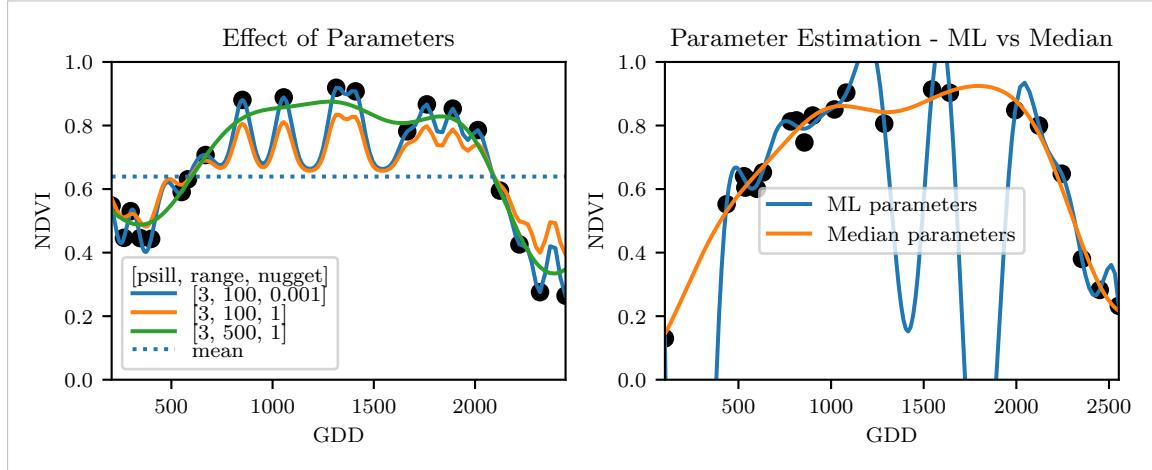


Figure 3.3: On the left, we see how the interpolation changes if we increase the nugget and the range parameter. On the right, we compare two UK interpolations, where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such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

Finally, we consider a one-dimensional Gaussian process G_γ with Variogram γ and tune the Variogram parameters using maximum likelihood⁵. Let z be a vector with the new values to extrapolate, then we can determine the values $m(z) = \mathbb{E}[G_\gamma(z)|(t, y)]$ using Bayes rule⁶. For an example fit, we refer to figure 3.3.

390 Violated Assumption

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

⁵ As illustrated in figure 3.3 maximum likelihood estimation can lead to overfitting. Thus, we will in practice sample several such optimized parameters and use their median in the end.

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

Advantages	Disadvantages
<ul style="list-style-type: none"> — It is a well-studied method. — Variogram parameters have an intuitive meaning. — Flexible covariance structure. 	<ul style="list-style-type: none"> — Regression to the mean. — Violated assumption of constant mean and constant variance. Thus, the NDVI is not a stationary process. — Pure maximum likelihood can result in overfitting.

395 **3.3.3 Savitzky-Golay Filter (SG)**

396 The SG, introduced in [Savitzky and Golay \(1964\)](#) is a technique in signal processing and
 397 can be used to filter out high frequencies (low-pass filter) ([Schafer, 2011](#)). Furthermore,
 398 it can also be used for smoothing by filtering high frequency noise while keeping the low
 399 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:

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

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

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

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

401 [Chen et al. \(2004\)](#) developed a ‘robust’ IM for the NDVI based on the SG. The method
 402 is based on the assumption that due to atmospheric effects the observed NDVI tends to
 403 be underestimated and that it cannot increase too quickly. The latter is argued by the
 404 biological impossibility of such fast vegetation changes. Their proposed algorithm is:

- 405 i.) Remove non-SCL45 points.
- 406 ii.) Remove points that would indicate an increase greater than 0.4 within 20 days.
- 407 iii.) Linearly interpolate to obtain an equidistant TS X^0 .
- 408 iv.) Apply the SG to obtain a new TS X^1 .
- 409 v.) Update X^1 by applying again a SG. Repeat this until $w^T |X^1 - X^0|$ stops decreasing,
 410 where w is a weight vector with $w_i = \min \left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|} \right)$. This reduces the
 411 penalty introduced by outliers⁷ and by repeating this step we approach the “upper
 412 NDVI envelope”.

413 **Extension: Spatial-Temporal SG**

414 One notable adaptation of the SG is the presented by [Cao et al. \(2018\)](#). The key difference
 415 is the additional assumption of the cloud cover being discontinuous and that we can

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

416 improve by looking at adjacent pixels⁸. Because we are working with rather high-resolution
 417 satellite data, and we need the variance in the predictors, we will waive this extension.

Advantages	Disadvantages
<ul style="list-style-type: none"> — Popular technique in signal processing. — Efficient calculation for equidistant points. — The upper envelope matches intuition for the NDVI. Therefore, it is robust against outliers with small values. 	<ul style="list-style-type: none"> — No natural way of how to estimate points that are not in the data. — Not generalizable to other spectral indices. — Linear interpolation to account for missing data might not be appropriate. — No smooth interpolation between two measurements.

418 **3.3.4 Locally Weighted Regression (LOESS)**

419 The LOESS introduced by [Cleveland \(1979\)](#) can be understood as a generalization of the
 420 SG (cf. sec. 3.3.3).

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

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

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

423 **Differences between the Robust LOESS and the SG**

424 The LOESS smoother takes a fraction of points instead of a fixed number, and therefore
 425 automatically adapts to the size of the data we wish to interpolate. However, we run
 426 into the danger of considering too little observations, since the estimation breaks down if
 427 $\lceil \alpha n \rceil < d + 1$.⁹ Furthermore, LOESS gives less weight to points further away. This yields a
 428 “smoother” estimate, since when we slide the window (e.g., for estimating the next value)
 429 an influential point at the border does not suddenly get zero weight from being weighted
 430 equally before. Finally, the LOESS can also be used for non-equidistant data and allows
 431 for arbitrary interpolation.

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

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

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

432 **3.3.5 B-Splines (BS)**

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

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

433 where B are basis functions and recursively defined by:

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

Assuming that all t_i are distinct, this yields an interpolation that fits the data perfectly. To reduce the amount of overfitting and increase the smoothness, we relax the constraint that we must 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. — Also performs well if points are not equidistant. | <ul style="list-style-type: none"> — Smoothing process does not translate well to an interpretation (unlike SS). — Choice of smoothing parameter s. |
|---|--|
-

435 **3.3.6 Smoothing Splines (SS)**436 Let \mathcal{F} be the Sobolev space (the space of functions of which the second derivative is
437 integrable). Then the unique¹¹ minimizer

$$\hat{m} := \arg \min_{f \in \mathcal{F}} \sum_{i=1}^n w_i (y_i - f(t_i))^2 + \lambda \int f''(t)^2 dt \quad (3.3.6.1)$$

438 is a cubic spline (i.e., a piecewise cubic polynomial function). The objective function
439 ensures that we decrease the curvature while keeping the RMSE low.440 **Whittaker — Discrete Version with Higher Order Derivatives**

The Whittaker smoother introduced in [Eilers \(2003\)](#) is closely reminiscent of the SS and is also used for the NDVI TS ([Atzberger and Eilers, 2011](#)). Similar to SS, we minimize the following expression over $z \in \mathbb{R}^n$:

$$(y - z)^T W (y - z) + \lambda z^T D^T D z,$$

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.

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

441 where W is a diagonal weight-matrix, λ our parameter and D a matrix that serves the
 442 purpose of approximating a differentiation of k -th order. In essence, this minimization
 443 function is the same as equation 3.3.6.1. The only differences are, that we substitute the
 444 integral by a sum and that we are more flexible with the order of the derivatives we are
 445 using. The main drawback is that we do not get a smooth function that interpolates, and
 446 that the sum behaves worse than the integral for non-equidistant data points. Thus, we
 447 will not consider the Whittaker further but consider the more general SS.

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). — Bounded within the data range if λ is chosen a priori. 	<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).

448 3.4 Tuning Parameter Estimation

449 Many of the IMs introduced in section 3.2 and 3.3 include a free parameter. To determine
 450 this parameter for a specific IM, we will estimate the absolute residuals using OOB esti-
 451 mation and then optimize the parameter using a score function. We clarify the procedure
 452 step by step:

- 453 i.) Construct a set Λ of candidate parameters that generously covers the parameter
 454 space.
- 455 ii.) Consider \mathcal{P} , a set of Pixels.
- 456 iii.) For each parameter $\lambda \in \Lambda$ consider the individual pixels and compute the LOOCV¹²
 457 for the absolute residuals of the specific NDVI IM for all Pixels in \mathcal{P} and store them
 458 in the set R_λ .
- 459 iv.) Determine $\lambda_{optimal} = \arg \min_{\lambda \in \Lambda} q_{90}(R_\lambda)$, where we describe the 90% quantile with
 460 q_{90} .

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

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

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

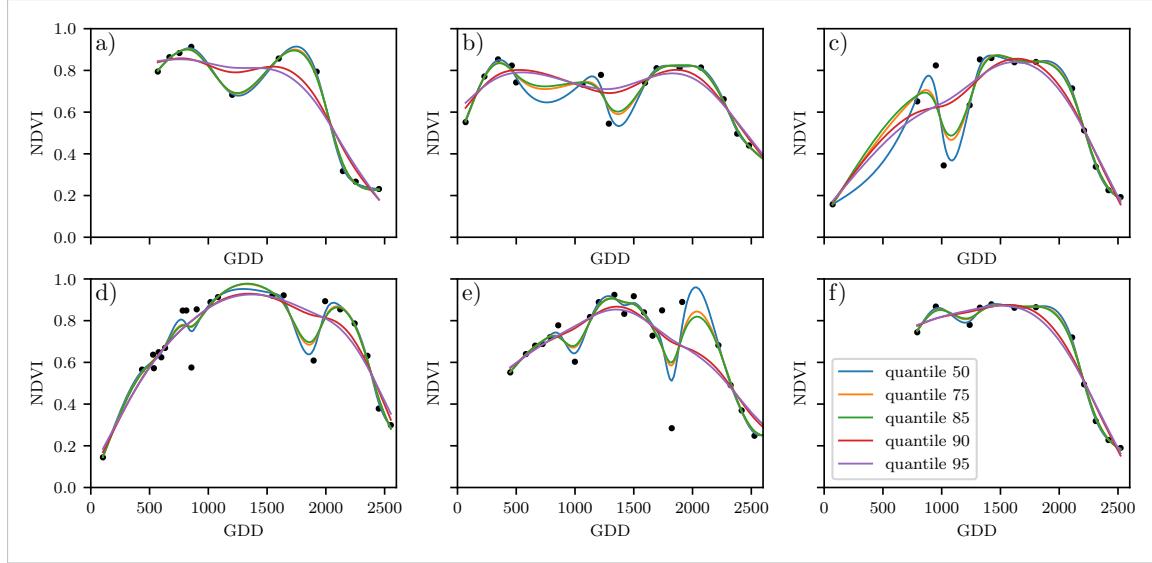


Figure 3.4: Smoothing splines fit with smoothing parameter optimized by minimizing the given quantile of the absolute leave-one-out residuals. Note that the larger the considered quantile is, the smoother the resulting curve becomes.

465 3.5 Robustification

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

469 Even though the procedure is taken from the robust version of the LOESS smoother (cf.
 470 section 3.3.4 and Cleveland (1979)), we can apply it to every IM that allows for prior
 471 weighting of observations.

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

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

474 Next, we compute new weights by

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

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

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

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

¹⁴The breakdown point relates only to outliers in the y values. Note that we do not require the IMs 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).

479 **3.5.1 Our Adjustment:**

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

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

480 for $r, w \in \mathbb{R}^n$.481 **3.5.2 Examples and Conclusions**

482 Examples of the first four iterative fits using SS are shown in figure 3.5 for six pixels.
 483 For the analogous figures of the other IMs cf. figures B.1, B.2, B.3 and B.1. Indeed, we
 484 observe how the interpolated TS is less affected by outliers after each iteration. We notice
 485 the biggest difference in the first iteration. Furthermore, in the plot at the bottom left we
 486 see how the interpolation ‘escapes’ from the right endpoint with each successive iteration,
 487 even though our intuition does not necessarily identify this point as an outlier. Therefore,
 488 in the following, we will always stop after one iteration.

consider naming the subplots

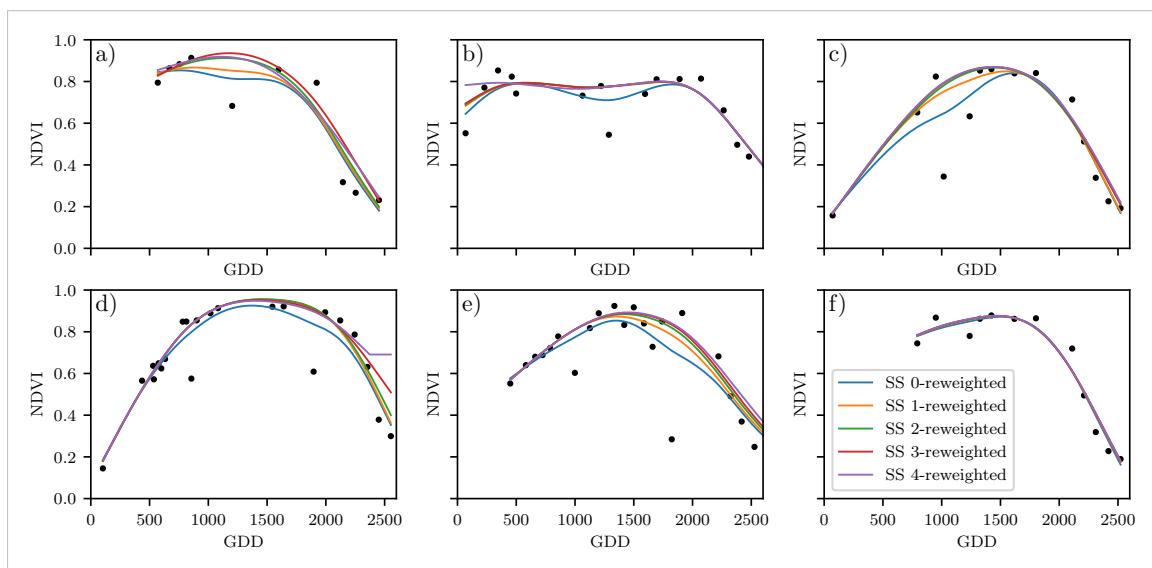


Figure 3.5: Smoothing splines fitted to different (SCL45) NDVI TS. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed.

489 **3.5.3 Upper Envelope Approach — Penalty for Negative Residuals**

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

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

497 **3.6 Performance Assessment**

498 Next, we will benchmark the in section 6.1.2 preselected IMs with and without robustifi-
499 cation. For this, we will use the same technique as we did for the parameter determination
500 in section 3.4. On B_λ we apply the RMSE and different quantiles.

501 The results are presented in section 5.1 and are discussed in section 6.1. The double logistic
502 turns out to be the best convincing parametric method, and from the non-parametric
503 methods we choose the SS.

504 **Chapter 4**

505 **NDVI Correction**

506 Let's remind ourselves that the data from the S2 satellites is distributed with an SCL and
507 we therefore have some evidence about what is observed at each pixel for each sampled
508 time (cf. table 2.2). So far, we have only considered points, labeled as cloud- and shadow-
509 free (SCL45). However, we remind ourselves of the satellite images in figure 2.3d, where
510 we had cloudy images despite the 'vegetation' label and see vegetation in figure 2.3e even
511 though we are supposed to observe 'cirrus clouds'.

512 In this chapter, we will try to improve our NDVI interpolation by not relying only on the
513 observed NDVI, but by training our own model to correct the NDVI using all S2 bands.
514 For this, we introduce several statistical modelling approaches and discuss the strengths
515 and weaknesses for each of them. After correcting the observed NDVI, we will assess the
516 uncertainties of our corrections and translate them into weights. These will be used for
517 the subsequent interpolation. This step-by-step procedure is illustrated by the figure B.4
518 in the appendix. Finally, we will evaluate which combination of IMs and correction model
519 performs the best.

520 **4.1 Considering other SCL Classes**

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

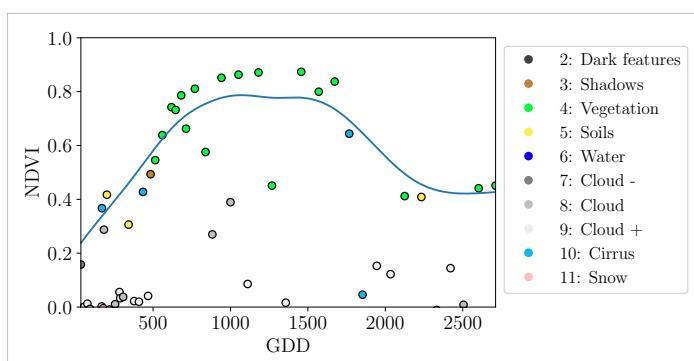


Figure 4.1: A smoothing splines fit considering green and yellow points (SCL45).

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

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

527 **Assumption 4.1.0.1.** The “true” NDVI value at time t can be successfully estimated by
 528 robustified LOOCV interpolation using high-quality observations. That is, the interpolated
 529 value (using a robustified IM from chapter 3) considering the points $P^{SCL45} \setminus P_t$. In the
 530 following, we will call this estimate the “true”-NDVI.

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

538 It might be tempting to just include some of the mentioned SCL classes for interpolation.
 539 But on the one hand, the choice would not be objective and on the other hand, the
 540 correlation seems to be weaker than for SCL45. Therefore, in the following section, we
 541 will correct the observed NDVI and estimate the uncertainty of each correction.

542 4.2 Correction Models

543 For training an NDVI correction model, we require ground-truth data which we will aim
 544 to model using informative covariates. Since ground-truth NDVI data is not available,
 545 we will again use the assumption 4.1.0.1 and use the “true” NDVI instead. There is no
 546 canonical answer to the question of which covariates we should use. It is a tradeoff between
 547 simplicity, generalizability, and performance (with the danger of overfitting). Our desire
 548 with the NDVI correction is to develop a product that is simple to use and understand.
 549 Therefore, in the subsequent, we will only take the spectral data of the satellite (i.e., all
 550 the bands) and the observed NDVI derived from it as covariates. We organize the chosen
 551 covariates in the design matrix X^1 , where each row corresponds to a P_t (i.e., a pixel at a
 552 time t) and each column to one covariate.

553 In the following, we will introduce different approaches, to model the relationship between
 554 the response $y := \text{NDVI}_{\text{true}} \in \mathbb{R}^n$ and the design matrix $X \in \mathbb{R}^{n \times p}$. First, we will study
 555 the basic OLS. Second, we look at the LASSO, a penalized adaptation of the OLS which
 556 is known to successfully deal with highly correlated covariates. Afterwards, GAMs are
 557 introduced, which model the response similar to OLS but allow for non-linear relations.
 558 Last but not least, we discuss RF and MARS, which are both flexible modelling approaches.

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

561 4.2.1 Ordinary Least Squares (OLS)

562 The OLS is a linear model that aims to minimize the sum of the squared residuals. We
 563 assume a linear relationship between y and X and allow for Gaussian noise. That is:

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

¹Strictly speaking, we include also the intercept and introduce one dummy variable for each SCL-class.

564 Assuming that $(X^T X)$ is regular, we can estimate the regression coefficients β by

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

565 We will train two models, one using all covariates discussed above and one using only the
566 SCL-classes and the observed NDVI.

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

567 **4.2.2 Least Absolute Shrinkage and Selection Operator (LASSO)**

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

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

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

573 Tibshirani (2011) shows that the LASSO solution tends to be sparse. That is $\beta_i = 0$ for
574 most $i = 1, \dots, p$. The larger λ , the more $\beta_i = 0$ and hence the simpler the resulting
575 model.

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

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

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

³The last two terms are equivalent by lagrangian optimization.

⁴The cross validated Root Mean Square Error is the mean of the RMSE's obtained for each fold using the model trained on the remaining folds.

⁵This is if our covariates are $\{1, a, b\}$, then we will now use $\{1, a, b, ab, a^2, b^2\}$.

583 **4.2.3 General Additive Model (GAM)**

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

$$g_{add}(x) = \mu + \sum_{i=1}^p g_j(x_j).^6$$

588 To estimate the non-parametric functions, we can use SS (ref sec. [3.3.6](#)). For this let \mathcal{S}_j
 589 be the function that takes some $z \in \mathbb{R}^n$ and returns the SS fitted to $(X_{:,j}, z)$ where the
 590 smoothing parameter is optimized by LOOCV⁷. Since we cannot fit all g_j simultaneously,
 591 we will use a strategy named Backfitting. We basically cycle through the indices $1, \dots, p$
 592 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

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

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

594 **4.2.4 Random Forest (RF)**

595 To define a random Forest introduced by [Breiman \(2001\)](#) we will first define what a Tree
 596 is. A (decision) Tree is a graph (V, E) without circles, a distinct root node, every node
 597 has at most two children and every leaf has a value assigned to it. At each node there
 598 is a boolean condition testing if one variable is greater than some value and a pointer to
 599 one child depending on the boolean value. To evaluate a tree we start at the root node,
 600 test the boolean expression and go to the node indicated by the resulting pointer. This
 601 we repeat until we end up at a leaf-node, where we return the value assigned to it.

602 To build such a Tree, we will recursively partition the covariate space using greedy splits⁸
 603 decreasing the RMSE⁹ each time. If the set we want to split contains less than a certain
 604 amount of training points, we stop.

⁶Where g_j is a real-valued function. For identifiability we also demand $\mathbb{E}[g_j(X_{:,j})] = 0$ for $j = 1, \dots, p$.

⁷For efficiency a proxy of the LOOCV is used called generalized cross validation.

⁸For computational reasons, we will only use splits along one covariate. So we ‘cut’ our covariate space into rectangles.

⁹To 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).

605 To build a Random Forest we will bootstrap-aggregate¹⁰ many such Trees¹¹. The prediction
 606 of the Random Forest for a new point x is then the mean of the predictions from all
 607 the 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.

608 4.2.5 Multivariate Adaptive Regression Splines (MARS)

609 A MARS model as introduced in [Friedman \(1991\)](#) can be described by

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

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

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

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

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

$$\mathcal{C} := \{(h(\cdot)b_1(\cdot), h(\cdot)b_2(\cdot)) \mid h \in \mathcal{M}, (b_1, b_2) \in \mathcal{B}\} \quad (4.2.5.1)$$

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

620 Finally, to avoid overfitting, we prune the set \mathcal{M} by optimizing a LOOCV score.¹²

621 To reduce computational complexity, we follow the recommendation from [Stephen \(2021\)](#)
 622 and restrict h in equation (4.2.5.1) to be of degree one (so it is also in a pair of \mathcal{B}).
 623 Consequently, \mathcal{C} contains functions with a degree of at most 2.

¹⁰That is we will sample (with replacement) several times n observations from our original data and fit a Tree to each such sample.

¹¹Building 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.

¹²This 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 that – when excluding from \mathcal{M} – leads to the best LOOCV score.

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).

624 4.3 Weighted Interpolation

625 Once we corrected the NDVI using the models described in the previous section, we are left
 626 with the problem that not every correction is equally reliable.¹³. Hence, we are interested
 627 in a measure of how uncertain an estimate is. We achieve this analogously as we corrected
 628 the NDVI, by replacing the response NDVI-“true” with the absolute residuals $v := |y - \hat{y}|$
 629 and modeling their relationship with the covariates defined by X . In this way, we obtain
 630 a model for the absolute residuals v and the estimator \hat{v} .

631 In the following, we will convert our uncertainty estimate into weights that can be used
 632 for interpolation. For this, consider a pixel P , $\hat{y}^{(P)}$ its corrected NDVI values and $\hat{v}^{(P)}$
 633 the estimated uncertainties of $\hat{y}^{(P)}$. In order to interpolate $\hat{y}^{(P)}$, we will give less weight
 634 to unreliable observations. Thus, we define the link function connecting \hat{v} with weights:

$$w_{\tau}^{(P)} := \frac{1}{R} \frac{1}{\hat{v}_{\tau}^{(P)}}, \quad \text{for } \tau = 1, \dots, n_P \quad (4.3.0.1)$$

635 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.
 636 The normalization is needed since for some IMs, inflating the sum of weights would decrease
 637 the effect of the smoothing.

638 4.4 Resulting Interpolation Strategies (ISs)

639 We have developed the following procedure to obtain a new interpolation (keyword-wise):
 640 i.) LOOCV Interpolation (+ robustify?) to get “true” NDVI
 641 ii.) Correction
 642 iii.) Uncertainty estimation
 643 iv.) Interpolation (+ robustify?)

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

- 645 — Interpolation: SS / DL
- 646 — Robustify: Yes / No
- 647 — Correction & uncertainty estimation: RF / OLS – considering only SCL-classes /
 648 OLS – considering all selected covariates / MARS / GAM / LASSO / no correction.

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

¹³One correction is illustrated in the figure B.4f. In this figure, the outer points (labeled as clouds) have a large scatter.

- 651 — We use the same IM each time.
 652 — Either we robustify both times, or we do not robustify at all.
 653 — We use the same underlying method for correction and uncertainty estimation.
 654 In this fashion, we obtain 28 distinct ISs, which we will benchmark in the next section.

655 4.5 Evaluation via (relative) Yield Prediction Error (relative YPE)

656 In this section, we introduce the relative YPE and utilize it to evaluate the 28 ISs from
 657 section 4.4. The fundamental assumption is that the closer the interpolated NDVI TS is
 658 to the true one, the better it can be used to determine crop yield. Implicitly, we believe
 659 that an NDVI TS that better models yield will incorporate more true information about
 660 the underlying vegetation. Therefore, we want to determine a comparable YPE for each
 661 IS and choose it as a benchmark criterion. This is an objective measure since we have not
 662 considered crop yield in any of our previous steps. Moreover, this criterion is justified by
 663 the fact that yield estimation has been a motivation for the interpolation.

664 **Definition 4.5.0.1.** (*Relative YPE*) Let $y \in \mathbb{R}^n$ be the yield, M be a model for estimating y ,
 665 and $\hat{y} = M(X)$ where X describes the data¹⁴. We define the relative YPE as the relative
 666 RMSE in yield estimation. Formally expressed:

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

667 where \bar{y} denotes the sample mean. For the (non-relative) YPE do not divide by \bar{y} .

668 We would like to estimate the yield from the NDVI TS produced by all the ISs for all
 669 pixels. However, given the high dimensionality and different lengths of the interpolation
 670 (not every TS has the same start and end point), we must first map each NDVI TS into a
 671 low-dimensional vector space of covariates. For this, we will use the following statistics:

- Maximum slope
- Minimum slope
- Integral¹⁵ over all
- Peak (i.e., maximal NDVI)
- GDD for the Peak
- Integral¹⁵ up to the peak
- Integral¹⁵ after peak
- Integral¹⁵ from 0-685 GDD
- Integral¹⁵ from 685-1075 GDD

672 For the choice we were inspired by (cf. table 2 in Kamir et al. (2020)). However, we
 673 deliberately omit any statistic that involves the minimum (e.g., the NDVI-range), since
 674 we regard the minimum as a very error-prone measure due to the large influence of clouds
 675 in the TS.

676 As a result, for each IS, a matrix is obtained in which each row corresponds to a pixel
 677 and both the yield and the covariates (computed by applying the above statistics) are
 678 contained. Using this matrix, we train a random forest for yield estimation, and compute

¹⁴We will use the matrixes derived in section 4.5.

¹⁵We will only consider the integral of the function $\max(0, NDVI - 0.3)$, where 0.3 is assumed to be a minimal NDVI value (cf. satellite images 2.3a and 2.3f with their NDVI in plot 2.3g).

679 the integrated OOB¹⁶ estimates \hat{y} . Note that the choice of the modeling approach does not
680 matter much, as long as it is general enough (i.e., able to approximate any function) and
681 we use the same one for each IS. Finally, for each IS, we calculate the YPE and describe
682 the results in section 5.2.

¹⁶By 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 $\frac{1}{e}$).

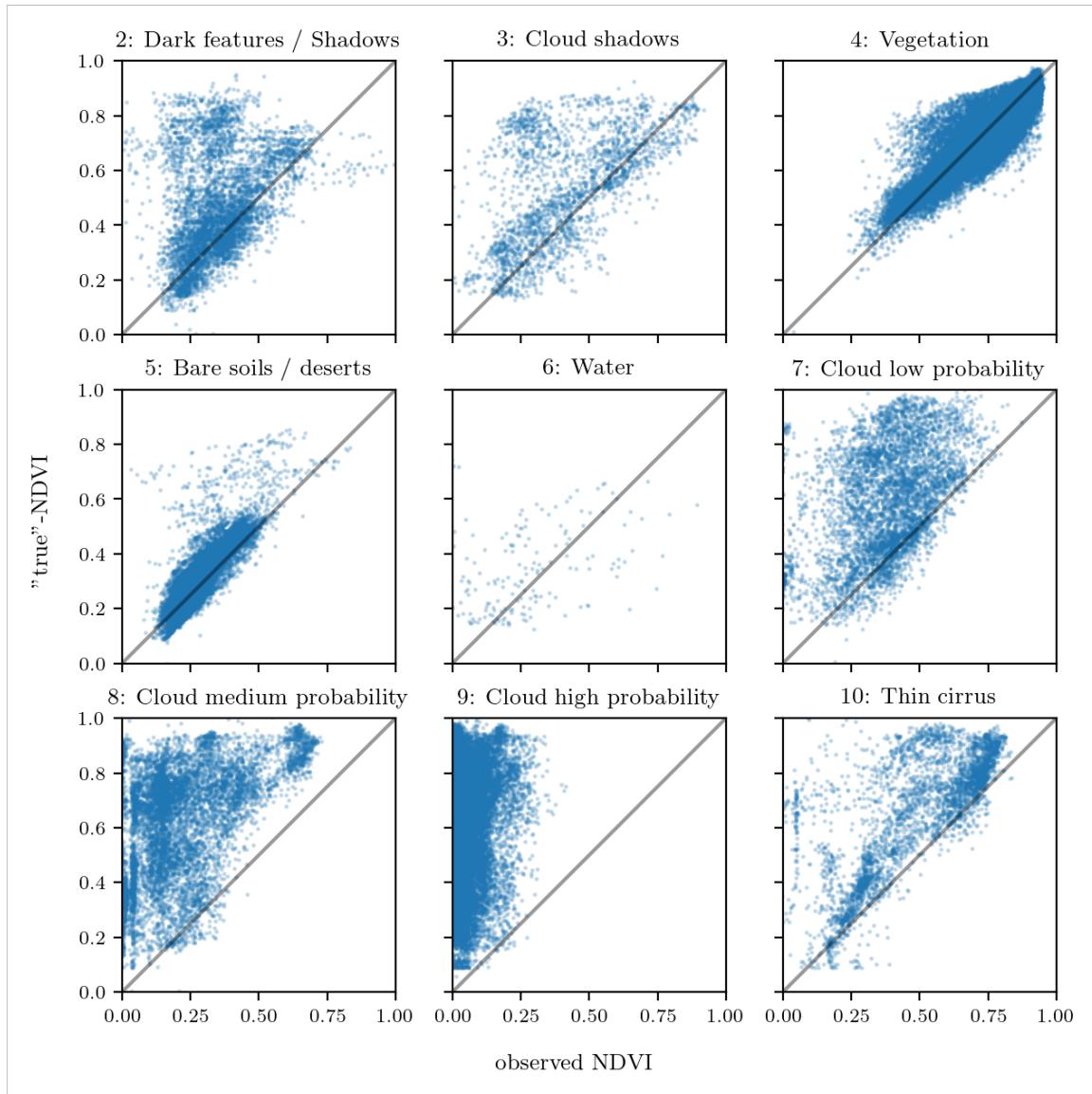


Figure 4.2: For each SCL class, we compare the true NDVI with the observed NDVI. (The true NDVI was estimated with LOOCV smoothing splines, and we used all observations of 10% of the total pixels.)

683 **Chapter 5**

684 **Results**

685 **5.1 Goodness of Fit for Selected IMs**

686 Table 5.1 shows the benchmarks of the selected¹ IMs (on P^{SCL45}) with respect to various
 687 score functions. The score functions summarize the absolute values of the LOOCV residuals
 688 (the smaller, the better). For each of the 5 selected IMs, we consider the basic and
 689 the robustified (see section 3.5) version.

Table 5.1: Comparing the goodness of fit for selected IMs (on P^{SCL45}) measured with the score functions (that take the LOOCV residuals as input) listed in the left column. q_X denotes here the $X\%$ quantile. Colored row-wise.

	SS	LOESS	DL	BS	FS	SS^{rob}	$LOESS^{rob}$	DL^{rob}	BS^{rob}	FS^{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

690 DL performs the best among both robustified and non-robustified with respect to most of
 691 the score functions used (all except q95) and is especially superior to the other parametric
 692 approach, which is FS. Especially the robust FS performs poorly. The LOESS is superior
 693 on every score function than all other non-parametric methods, but is closely followed by
 694 the SS. The BSPL exhibits the worst performance out of all non-parametric method tested
 695 here.

696 **5.2 YPE for Tested ISs**

697 The YPE for the 28 chosen ISs (cf. section 4.4) is given in table 5.2. We note that
 698 robustification does not improve the quality of the fit (measured via the YPE) in most
 699 cases. In addition, SS tend to be better than DL (with and without robustification) in
 700 terms of YPE, especially if no correction is made. The IS that leads to the lowest YPE is
 701 the OLS^{SCL} with SS. Given that the OLS^{SCL} models have very good interpretability, we

¹ For the discussion which IMs have been selected cf. section 6.1.2.

Table 5.2: Relative YPE for various ISs. For the non-relative YPE and the coefficient of determination (R^2) cf. table B.1 and B.2.

	RF	OLS ^{SCL}	OLS ^{all}	MARS	GAM	LASSO	no corrections
SS	0.155	0.140	0.143	0.142	0.142	0.142	0.149
SS ^{rob}	0.155	0.143	0.147	0.149	0.146	0.145	0.148
DL	0.156	0.151	0.152	0.152	0.149	0.149	0.158
DL ^{rob}	0.157	0.153	0.152	0.145	0.148	0.150	0.157

702 also present the regression equations below. The corrected NDVI is calculated using

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

703 where $\mathbb{1}_{SCL=2}$ is equal to one if the current observation corresponds to SCL class 2 and
 704 zero otherwise². Whereas, we obtain the estimated absolute residuals by:

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

705 Thus, if we observe a pixel with SCL class 4 ('vegetation') but a NDVI of only 0.4, the
 706 corrected NDVI would be $0.711 \cdot 0.4 + 0.21 = 0.494$ with an estimated absolute residual
 707 of $-0.133 \cdot 0.4 + 0.146 = 0.93$. In the equation 5.2.0.1, we notice the strongest upwards
 708 correction for SCL classes 8, 9 and 11 (correspond to 'medium probability clouds', 'high
 709 probability clouds' and 'thin cirrus clouds'). The estimated absolute residuals, however,
 710 are the smallest for SCL classes 4 and 5 ('vegetation' and 'bare soil'). Furthermore, the
 711 higher the observed NDVI the lower are the estimated absolute residuals.

712 For the R-output of the `summary` function of the two models, we refer to the appendix B.3.1.

² $\mathbb{1}$ is also called an indicator function or characteristic function in mathematics.

713 **Chapter 6**

714 **Discussion**

715 In the first part of the discussion, we examine IMs for compatibility with data gaps and
716 argue choices for selected IMs. In the second part, we identify the best IS and discuss
717 issues that have arisen in the context of the NDVI correction.

718 **6.1 IMs**

719 **6.1.1 Data Gaps in Time Series**

720 NW estimates the value for t by relating to the points near t . To determine what “near”
721 means, a bandwidth h is used (cf. equation 3.3.1.2). This gets problematic as soon as the
722 data gaps become larger than h , since no points are left that are close to t .

723 Regarding the GK, we expect that due to the stationarity assumption, the interpolation
724 will always tend to the mean if data gaps are present (cf. figure 3.3).

725 Since the SG requires equidistant points, data gaps will break it. REFXXX proposes a
726 linear interpolation to restore missing data points. However, due to the timescale trans-
727 formation to GDD in section 2.4, the requirement of equidistant remains an unresolved
728 issue.

729 We do not trust the FS interpolation if there are noticeable data gaps. As could be
730 observed in figure 3.1 the curve can escape there. Additionally, the poor goodness-of-fit
731 values for the robustified variant in table 5.1 illustrate the unreliability of this IM method.
732 These are meaningful in describing the ability to cope with data gaps, since more data
733 points are ignored during the robustification and thus data gaps are simulated.

734 Similarly, for SS, LOESS, DL, and BS we compare the values in table 5.1 between the
735 robustified and non-robust variant. We find that the robust variant does not differ strongly
736 from the non-robust variant (unlike as for FS). Thus, we conclude that these methods do
737 not have systematic failures.

738 Regarding the LOESS, in case of data gaps, the weights can attain non-intuitive values.
739 The result can be a strongly fluctuating behavior, as observed in figure B.1 in plot (c).
740 There, a strange peak between the first and second observation is visible. This peak
741 originates from local weighting. In this case, the first data point in the plot, although
742 adjacent to the peak, is given a low weight compared to the points to the right of the peak
743 (for estimating the value at this peak).

744 In our experience, the DL handles data gaps well, but it may happen that the model
 745 describes the NDVI increase as abrupt. This, however, was fixed, by bounding the first
 746 derivative (cf. section 3.2.3).

747 6.1.2 Preselection

748 Here we justify our preselection of the IMs tested in section 3.6. We decided against NW
 749 Because of its systematic errors at peaks and valleys. Moreover, this method handles
 750 data gaps poorly (cf. 6.1.1). Moreover, UK will not be considered since the underlying
 751 stationarity assumption is not met and therefore a systematic bias is introduced. On top
 752 of that, maximum likelihood parameter estimation occasionally might lead to overfitting
 753 (cf. 3.3). Also, we do not include the SG in the next selection, since we see it as a special
 754 case of LOESS. The remaining IMs are thus SS LOESS DL BS and FS.

755 6.1.3 Candidate Selection

756 Given that DL convinces regarding most of the selected score functions in table 5.1 we
 757 will apply this method also in chapter 4. Moreover, we see that the robustification mostly
 758 improved the score regarding the 50, 75, 85, and 90 % Quantiles. Only for the outlier-
 759 sensitive score functions (RMSE and q95)¹ we notice significant worsening (we consider the
 760 robust FS separately in section 6.1.1). Consequently, we will also use the robustification
 761 in section 4. In order to not only rely on the form assumptions of the DL, we further
 762 choose a non-parametric method for further consideration. Despite the LOESS slightly
 763 dominating the SS in table 5.1, we choose the SS. This is due to the strange behavior of
 764 the LOESS in case of data gaps (see section 6.1.1) and the good interpretability of the SS
 765 using the minimization function 3.3.6.1.

766 6.2 NDVI Correction

767 6.2.1 Choose IS

768 The evaluation of various ISs via the YPE (cf. section 5.2) shows that SS are better suited
 769 than DL for yield estimation. Moreover, it seems surprising that robustification tends
 770 to worsen the results, despite reducing LOOCV residuals in most cases (cf. section 5.1).
 771 We conjecture that the correction models handle outliers by themselves (by correcting or
 772 down-weighting them) and thus do not benefit from an external robustification. Indeed,
 773 for OLS^{SCL} we see in equation 5.2.0.2 that the smaller the observed NDVI of a point,
 774 the larger the estimated residual — yielding a lower weight. This is consistent with our
 775 experience that outliers usually underestimate the NDVI. Our conjecture is consistent with
 776 the fact that if we do not correct, robustification produces a marginal improvement.

777 Using the best IS with correction (SS+OLS^{SCL}), instead of the best IS without correction
 778 (SS^{rob}), we can additionally explain $(0.148 - 0.140)/0.148 = 5.4\%$ of the variance in yield
 779 prediction (cf. table 5.2. To give a context, 100% would allow us to model the yield
 780 perfectly.

781 Note that the results discussed here depend strongly on the link function used (cf. equa-
 782 tion 4.3.0.1). Once we change it, we should also repeat this analysis.

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

6.2.2 Investigation of Error Sources in Yield Estimation

Although the YPE was not our primary goal but was only used as a means to select the best IS, we compare our values with the corresponding ones by Perich et al. (2022). There, a YPE 1.00 [t/ha] was obtained using weather data in addition to NDVI TS. Since our error is only about 3.3% larger (cf. table B.1), we consider our results to be competitive. Especially as we did not use meteorological data aside from the timescale transformation (cf. section 2.4) and in contrary did not scale the yield down by 10% (cf. section 2.2). In the following, we ask ourselves how much modelling performance we can actually expect.

This will be limited by multiple sources of uncertainty in the data:

- i.) Uncertainty in yield data collected by the combine harvester.
- ii.) Uncertainty in yield data through rasterization.
- iii.) Uncertainty in satellite images through “measurement errors” introduced via clouds and other atmospheric effects.
- iv.) Uncertainty introduced by assumed homogeneity of vegetation on pixel resolution.
- v.) Uncertainty introduced by interpolating NDVI TS (especially when long data-gaps are present).

find citations for the above

Furthermore, even if we would have a perfect NDVI curve, it contains only a fraction of the information about the underlying vegetation. Nonetheless, Perich et al. (2022) manages to explain up to 86% of the variance in crop yield with only the NDVI TS and weather data (Table 5). Although the authors divided the data into training and test data, this subdivision was done randomly at pixel level (without subdividing into fields or years). Thus, there are pixels in the training data that are neighboring pixels from the test data and consequently exhibit high correlations (in yield and NDVI). We suspect that overfitting via high-correlation pixels is responsible for these high values. On the other hand, the authors observe poor results for cross-year-validation² (table 6) and account them to uneven (extreme) weather. If this is not rather caused by the suspected overfitting, could be investigated by performing a cross-field-validation³. Nevertheless, we claim, that our results are affected by spatial correlation of neighboring pixels. This is because we expect all tested ISs to benefit equally from this correlation in terms of YPE, and we are only interested in the relative differences. our result is not a ‘good’ YPE, but the selected IS. So

6.2.3 NDVI Correction as Unsupervised Learning

The question arises if we can build the correction model on the same year as we want to apply it on. Usually, a similar approach might carry the danger of overfitting. However, we have not used any ground truth at any point (until the evaluation). Instead, we estimated the “true” NDVI with the assumption 4.1.0.1 via OOB. In other words, we have not used any ground truth but rather developed an unsupervised learner of the NDVI. Consequently, we reason that we can apply our method to a new (comparable) dataset.

²By cross-year-validation we understand a cross validation with respect to the RMSE, where each year represents a single fold.

³By cross-field-validation we understand the same as with cross-year-validation but with splitting each fold (i.e., a year) further into the respective fields. Since we have multiple fields per year, during evaluation each model trained will have seen the weather of all years but no adjacent pixels.

822 **6.2.4 Using Additional Covariates**

823 In section 4.2 we have only used covariates derived from spectral data. We decided against
824 using meteorological data, since we consider five years of data not to be sufficient to
825 model pattern of how vegetation reacts to various weather events. Moreover, we expect
826 the weather in our study region to be rather homogeneous, which is suggested by the fact
827 that the weather data published by Meteoswiss are for a grid with a resolution of 1 km.
828 On the other hand, we want the underlying model not to learn improper relationships.
829 For example, the model might automatically predict a high NDVI for a day in summer
830 (detected by high GDD or many sunshine hours) just because it is “used” to observing a lot
831 of vegetation in summer. Including temporally (e.g., P_{t-1} and P_{t+1}) and geographically
832 adjacent pixels would likely improve performance. However, for simplicity, we omit it
833 here⁴.

⁴This 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). Another complication is a border-pixel with some adjacent pixels outside the field.

834 **Chapter 7**

835 **Conclusion**

836 In this thesis, we investigated how to model vegetation dynamics through NDVI TS derived
837 from satellite images. The major challenges faced, were how to deal with contaminated
838 observations (due to clouds or shadows) and how to interpolate the observed NDVI values.
839 A summary of the IMs considered can be found in the table 3.

840 Filtering the observations contaminated by clouds and shadows via SCL introduces data
841 gaps, especially in winter. Therefore, we aim for IMs that handle such data gaps well.
842 The Nadaraya-Watson kernel estimator struggles when there are no or too few points
843 in the window of interest; Universal Kriging is biased towards the mean, particularly in
844 environments with no data (cf. figure 3.3); 2nd order Fourier series can deviate strongly
845 within data gaps (cf. figure 3.1) and the Savitzky-Golay filter depends on equidistant
846 observations (cf. section 6.1.1). Occasionally, a generalization of the Savitzky-Golay filter
847 — the Locally Weighted Regression — has also shown surprising behavior in data gaps
848 (cf. figure B.1).

849 In contrast, the latter performed well in Leave-One-Out-Cross-Validation (LOOCV) (cf.
850 table 5.1). Nevertheless, we prefer the Smoothing Splines (SS) as they perform only slightly
851 worse there, but produce a much smoother curve (cf. figure 3.5 and B.1). SS flexibly
852 approximate the data while keeping curvature low (cf. equation 3.3.6.1). B-splines, on
853 the other hand, were worse than SS with respect to every score function tested, and their
854 smoothing mechanism is also less interpretable. However, the best performing method
855 here is the approximation by a Double logistic (DL), which makes strong assumptions
856 about the shape of the NDVI curve. Problems for the parameter estimation of the DL
857 (and the Fourier series) have been resolved by restricting the parameter space by generous
858 but realistic values. Problems with overfitting in universal kriging were overcome by
859 determining the variogram parameters for a subsample of NDVI TS and finally using the
860 median of each parameter. In the end, we choose DL and SS as our preferred IMs.

861 The traditional answer to the question of how to deal with contaminated observations is
862 that we only consider observations that are labeled as vegetation or bare soil by the SCL
863 (SCL45). The unreliability of this labeling, however, is illustrated in figure 2.3. Moreover,
864 filtered observations (non-SCL45) might still contain valuable information (see section 4.1).
865 Therefore, we do not adhere to traditional (SCL) filtration, but instead consider all ob-
866 servations and correct the observed NDVI with uncertainty estimation. For this, we use
867 statistical models that take additional information such as the remaining spectral bands,
868 the current SCL label and the observed NDVI into account. But before we interpolate

869 the corrected NDVI values, we assign a weight to each observation, corresponding to its
870 uncertainty. The uncertainty is estimated analogously as the NDVI has been corrected.
871 That is, taking the same covariates but replacing the old response ($\text{NDVI}^{\text{true}}$) with a new
872 one ($\text{abs}(\text{NDVI}^{\text{corrected}} - \text{NDVI}^{\text{true}})$). By combining different IMs with various statistical
873 models, we obtain 28 different Interpolation Strategies (ISs) (see section 4.4). To assess
874 which of these ISs is best, we assume that the better the IS, the better it allows interpo-
875 lated NDVI TS to predict yield. Surprisingly, the best strategy is the one with SS and the
876 simplest static model considered, which uses only the observed NDVI and SCL classifica-
877 tion. Let us recapitulate the best IS: First, we estimate the “true” NDVI (c.f. assumption
878 4.1.0.1) using SS via LOOCV. Then obtain the corrected NDVI using the OLS^{SCL} model
879 (cf. equation 5.2.0.1). Subsequently, we estimate the absolute error with the OLS^{SCL}
880 model (cf. equation 5.2.0.1) and thereby obtain weights which are supposed to reflect the
881 reliability of the corrected NDVI (cf. equation 4.3.0.1). Finally, we perform a weighted
882 interpolation with SS.

883 To make the IMs more robust to contaminated observations (outliers) that remained af-
884 ter SCL filtration, we generalized an iterative technique. After an initial fit, in each
885 iteration we give less weight to observations with comparatively large residuals and then
886 perform a weighted interpolation (see section 3.5). However, after too many iterations,
887 non-contaminated points might get ignored (i.e., given a zero weight). The greatest im-
888 provements, on the other hand, were perceived after the first iteration (see figure 3.5). For
889 evaluating the generalized robustification technique, we used raw LOOCV performance
890 on the one hand, and the ability to model the NDVI TS for crop yield estimation on the
891 other hand. On the one hand, robustification (narrowly) misses the target of being part
892 of the best IS. On the other hand, we see in table 5.1 that robustification leads to smaller
893 LOOCV residuals in most cases. That is (except for the Fourier approximation) the 50%
894 and 75% quantiles of the absolute residuals are smaller for the robustified ones. Hence,
895 when we expect contaminated observations, we advise to robustify the interpolation.

896 As to the question of which IM we recommend, we consider two cases. If one only intends
897 to fit a curve to the NDVI TS as precisely as possible, we recommend the robustified DL,
898 since it minimizes the LOOCV residuals in most cases (cf. table 5.1). In the event that one
899 requires an interpolation that contains as much information about the plant as possible,
900 we recommend the SS. This recommendation is especially valid if we traditionally consider
901 only SCL45 observations without correcting the proposed NDVI. However, we recommend
902 the abovementioned IS with NDVI correction, because it reduces the variance of the NDVI-
903 based yield prediction by 5.4% (cf. section 6.2.1). Considering all the error sources (cf.
904 section 6.2.2) and the fact that we only consider the NDVI TS, we consider the 5.4% to
905 be a solid improvement.

906 7.1 Future Work

907 7.1.1 Time Series Correction-Interpolation as a General Method

908 Throughout this thesis, we developed a correction and IM for the NDVI. However, we
909 never relied on any properties of the NDVI. Only the parameter estimated via cross-
910 validation in chapter 3.4 depends on the scale of the TS. For simplicity, we could thus
911 determine the parameter using Generalized Cross Validation (Ripley and Maechler, 2022).
912 Therefore, our approach of interpolation and correction of TS can be applied to arbitrary
913 TS if additional information is available. This includes TS outside of satellite imagery

914 or remote sensing. However, further research is required, to demonstrate the general
915 usefulness of this approach.

916 As an example, we could develop cloud-correction with uncertainty estimation and inter-
917 polation. In the same manner as we corrected the NDVI TS for a pixel in chapter 4, one
918 could look at each spectral band separately and correct it with an uncertainty estimate.
919 Subsequently, one reassembles the corrected bands and translates the multiple estimated
920 uncertainties into one. Optionally, the TS can also be interpolated before merging, as in
921 chapter 4.3. The resulting question would be how well this approach performs.

922 **7.1.2 Minor Improvements**

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

- 925 — **Data:** The method how the combine harvester point cloud has been extrapolated to
926 the 10m grid of S2 could possibly be improved.
- 927 — **Data:** We have not included the spectral bands that have a resolution of 60 m. But
928 precisely these seem to be promising for cloud correction, since they are a proxy of
929 the water (content and form) in the atmosphere.
- 930 — **Data:** Raiyani et al. (2021) presents a machine learning approach that supposedly
931 improves the SCL and thus could improve our results that are based on the SCL.
- 932 — **NDVI Correction:** Explore the effect of different link and normalizing functions in
933 section 4.3. Currently, we run into the danger of some outer points getting nearly
934 ignored just because one estimated absolute residual for some interior point is close
935 to zero.
- 936 — **NDVI Correction:** Yield is not the only target variable of interest. Other variables
937 like protein content could also be used in section 4.5 for the method evaluation.

938 Bibliography

- 939 (2007). Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.),
940 *Model-Based Geostatistics*, pp. 46–78. New York, NY: Springer.
- 941 Atzberger, C. and P. H. Eilers (2011, September). A time series for monitoring vegetation
942 activity and phenology at 10-daily time steps covering large parts of South America.
943 *International Journal of Digital Earth* 4(5), 365–386.
- 944 Bailey, S. J. (2018, July). Using Growing Degree Days to Predict Plant Stages. pp. 8.
- 945 Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore (2006,
946 February). Improved monitoring of vegetation dynamics at very high latitudes: A new
947 method using MODIS NDVI. *Remote Sensing of Environment* 100(3), 321–334.
- 948 Breiman, L. (2001, October). Random Forests. *Machine Learning* 45(1), 5–32.
- 949 Brockmann, M., T. Gasser, and E. Herrmann (1993, December). Locally Adaptive Band-
950 width Choice for Kernel Regression Estimators. *Journal of the American Statistical
951 Association* 88(424), 1302–1309.
- 952 Bühlman, P. and M. Maechler (2020, October). Computational Statistics.
- 953 Cai, Z., P. Jönsson, H. Jin, and L. Eklundh (2017, December). Performance of Smoothing
954 Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology
955 from MODIS Data. *Remote Sensing* 9(12), 1271.
- 956 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
957 spatiotemporal information with the Savitzky-Golay filter. *Remote Sensing of Environment* 217, 244–257.
- 958 Chandola, V. and R. R. Vatsavai (2010). Scalable time series change detection for biomass
959 monitoring using Gaussian Processes. *Conference on Intelligent Data Understanding*,
14.
- 960 Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004, June). A simple method for reconstructing a high-quality NDVI time-series data set based on the
961 Savitzky–Golay filter. *Remote Sensing of Environment* 91(3), 332–344.
- 962 Cleveland, W. S. (1979, December). Robust Locally Weighted Regression and Smoothing
963 Scatterplots. *Journal of the American Statistical Association* 74(368), 829–836.
- 964 Courault, D., L. Hossard, V. Demarez, H. Dechatre, K. Irfan, N. Baghdadi, F. Flamain,
965 and F. Ruget (2021, July). STICS crop model and Sentinel-2 images for monitoring rice
966 growth and yield in the Camargue region. *Agronomy for Sustainable Development* 41(4),
967 49.
- 968

- 972 Eilers, P. H. C. (2003, July). A Perfect Smoother. *Analytical Chemistry* 75(14), 3631–3636.
- 973 ESA (2022a, August). European Space Agency: Level-2A Algorithm Overview.
974 <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithms>.
- 976 ESA (2022b, August). Sentinel-2. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>.
- 979 Friedman, J. H. (1991, March). Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19(1), 1–67.
- 980 Gamon, J. A., C. B. Field, M. L. Goulden, K. L. Griffin, A. E. Hartley, G. Joel, J. Penue-
981 las, and R. Valentini (1995). Relationships Between NDVI, Canopy Structure, and
982 Photosynthesis in Three Californian Vegetation Types. *Ecological Applications* 5(1),
983 28–41.
- 984 Gurung, R. B., F. J. Breidt, A. Dutin, and S. M. Ogle (2009, October). Predicting
985 Enhanced Vegetation Index (EVI) curves for ecosystem modeling applications. *Remote
986 Sensing of Environment* 113(10), 2186–2193.
- 987 Hastie, T. and R. Tibshirani (1987, June). Generalized Additive Models: Some Applica-
988 tions. *Journal of the American Statistical Association* 82(398), 371–386.
- 989 Henits, L., Á. Szerletics, D. Szokol, G. Szlovák, E. Gojdár, and A. Zlinszky (2022, Jan-
990 uary). Sentinel-2 Enables Nationwide Monitoring of Single Area Payment Scheme and
991 Greening Agricultural Subsidies in Hungary. *Remote Sensing* 14(16), 3917.
- 992 Holzkämper, A., D. Fossati, J. Hiltbrunner, and J. Fuhrer (2015, January). Spatial and
993 temporal trends in agro-climatic limitations to production potentials for grain maize
994 and winter wheat in Switzerland. *Regional Environmental Change* 15(1), 109–122.
- 995 Jaramaz, D., V. Perović, S. Belanovic Simic, E. Saljnikov, D. Cakmak, V. Mrvić, and
996 L. Zivotic (2013, May). The ESA Sentinel-2 mission Vegetation variables for Remote
997 sensing of Plant monitoring.
- 998 Kamir, E., F. Waldner, and Z. Hochman (2020, February). Estimating wheat yields
999 in Australia using climate records, satellite image time series and machine learning
1000 methods. *ISPRS Journal of Photogrammetry and Remote Sensing* 160, 124–135.
- 1001 Lyche, T. and K. Morken (2005, January). Spline Methods.
- 1002 McMaster, G. S. and W. W. Wilhelm (1997, December). Growing degree-days: One
1003 equation, two interpretations. *Agricultural and Forest Meteorology* 87(4), 291–300.
- 1004 Omori, K., T. Sakai, J. Miyamoto, A. Itou, A. N. Oo, and A. Hirano (2021, April).
1005 Assessment of paddy fields' damage caused by Cyclone Nargis using MODIS time-series
1006 images (2004–2013). *Paddy and Water Environment* 19(2), 271–281.
- 1007 Perich, G., H. Aasen, J. Verrelst, F. Argento, A. Walter, and F. Liebisch (2021, Jan-
1008 uary). Crop Nitrogen Retrieval Methods for Simulated Sentinel-2 Data Using In-Field
1009 Spectrometer Data. *Remote Sensing* 13(12), 2404.
- 1010 Perich, G., M. O. Turkoglu, L. V. Graf, J. D. Wegner, H. Aasen, A. Walter, and F. Liebisch
1011 (2022, July). Pixel-based crop yield mapping and prediction using spectral indices and
1012 neural networks on Sentinel-2 time series data.

- 1013 Raiyani, K., T. Gonçalves, L. Rato, P. Salgueiro, and J. R. Marques da Silva (2021,
1014 January). Sentinel-2 Image Scene Classification: A Comparison between Sen2Cor and
1015 a Machine Learning Approach. *Remote Sensing* 13(2), 300.
- 1016 Ripley, B. D. and M. Maechler (2022, August). R: Fit a Smoothing Spline.
1017 <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/smooth.spline.html>.
- 1018 Rouse, J. W. (1974, May). Monitoring the vernal advancement and retrogradation (green
1019 wave effect) of natural vegetation. Technical Report NASA-CR-139243.
- 1020 Savitzky, A. and M. J. E. Golay (1964, July). Smoothing and Differentiation of Data by
1021 Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), 1627–1639.
- 1022 Schafer, R. W. (2011, July). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE*
1023 *Signal Processing Magazine* 28(4), 111–117.
- 1024 Stephen, M. (2021, July). Earth: Multivariate Adaptive Regression Splines.
- 1025 Stöckli, R. and P. L. Vidale (2004, September). European plant phenology and climate
1026 as seen in a 20-year AVHRR land-surface parameter dataset. *International Journal of*
1027 *Remote Sensing* 25(17), 3303–3330.
- 1028 Strbac, O., M. Milanovic, and V. Ogrizovic (2017, July). Estimation the evapotrasnpiration
1029 of urban parks with field based and remotely sensed datasets. pp. 13.
- 1030 Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective.
1031 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(3), 273–
1032 282.

1033 **Appendix A**

1034 **Reproducibility**

1035 **A.1 Reproduce Results**

1036 For reproducibility of the whole computations, we refer to our codebase at:

1037 <https://github.com/LGraz/MasterThesis-Code>

1038 In order to reproduce our computations and results, set up the directory as described in the
1039 README. The ‘Yield Mapping’ Data used, is published alongside Perich et al. (2022). Ex-
1040 ecute the computations via the script `./shell_scripts/reproduce.sh` and do not execute
1041 the python and R files by hand (unless you follow the order in `./shell_scripts/reproduce.sh`).

1042 **A.2 R-Package**

1043 We also provide an R package for a general time series correction and interpolation if
1044 additional data is available at:

1045 <https://github.com/LGraz/CorrectTimeSeries>

1046 In our case, we consider the NDVI time series and the additional data consists of the
1047 unused spectral bands.

1048 We recommend installing it via the `devtools` package by:

1049 `devtools::install_github("LGraz/CorrectTimeSeries")`

1050 In the following, we shall give a stand-alone example of how the R package can be used:

```
1051 1 library(CorrectTimeSeries)
1052 2
1053 3 # load a list of dataframes, each one describes one pixel with the covariates and
1054 4 # the response
1055 5 data(timeseries_list)
1056 6 str(timeseries_list[[1]])
1057 7
1058 8 # Train/Load RF
1059 9 train_model_myself <- TRUE
1060 10 if (train_model_myself){
1061 11     # Add "true" NDVI (or generally the response), by Out-Of-Bag estimation
1062 12     timeseries_list <- lapply(timeseries_list, function(df) {
1063 13         df$oob_ndvi <- OOB_est(df$gdd, df$ndvi_observed) # gdd is the time-axis
1064 14         df
1065 15     })
1066 16     # Train correction model
1067 17     formula <- "oob_ndvi ~ B02+B03+B04+B05+B06+B07+B08+B8A+B11+B12+scl_class"
1068 18     RF <- train_RF_with_fromula(formula, timeseries_list, robustify=TRUE)
1069 19 } else {
```

```
1071 19  data(RF_for_NDVI)
1072 20  RF <- RF_for_NDVI
1073 21 }
1074 22
1075 23 # ADD CORRECTION
1076 24 timeseries_list <- lapply(timeseries_list, function(df) {
1077 25   df$corrected_ndvi <- randomForest:::predict.randomForest(RF, df)
1078 26   df
1079 27 })
1080 28
1081 29 # Get interpolation for each timeseries
1082 30 newx <- 1:1000
1083 31 lapply(timeseries_list, function(df){
1084 32   ss <- smoothing_spline(df$gdd, df$corrected_ndvi)
1085 33   predict(ss, newx)$y
1086 34 })
```

Example of how to use the `CorrectTimeSeries` package

1088 **Appendix B**

1089 **Further Material**

1090 **B.1 Data and Methods**

1091 **B.1.1 GDD**

1092 Bailey (2018) tabulates the corresponding GDD for each stage of wheat.

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

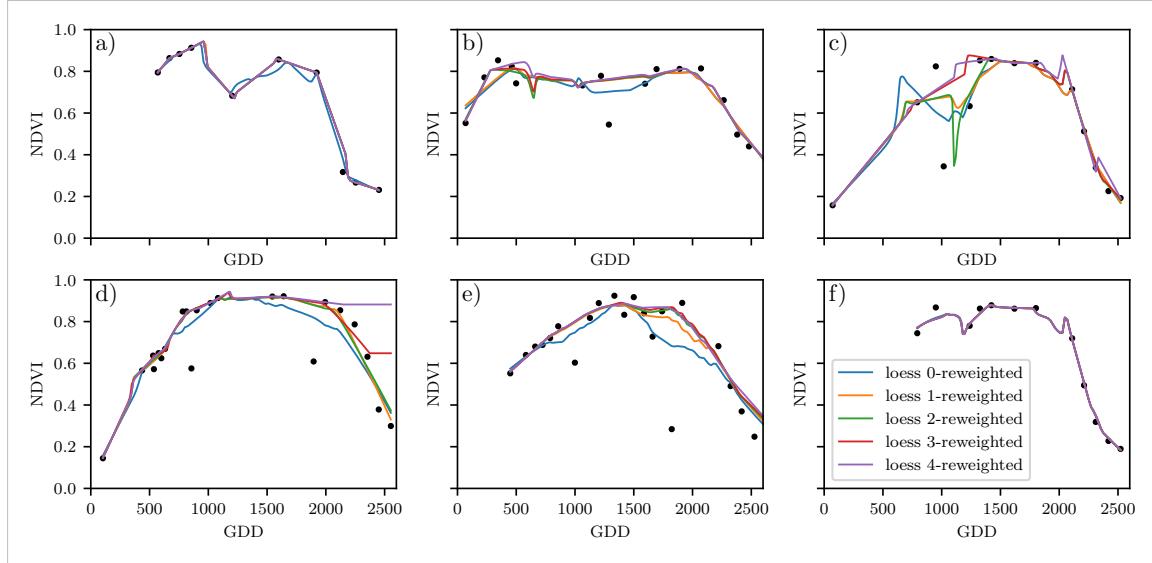
1095 **B.2 Interpolation**

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

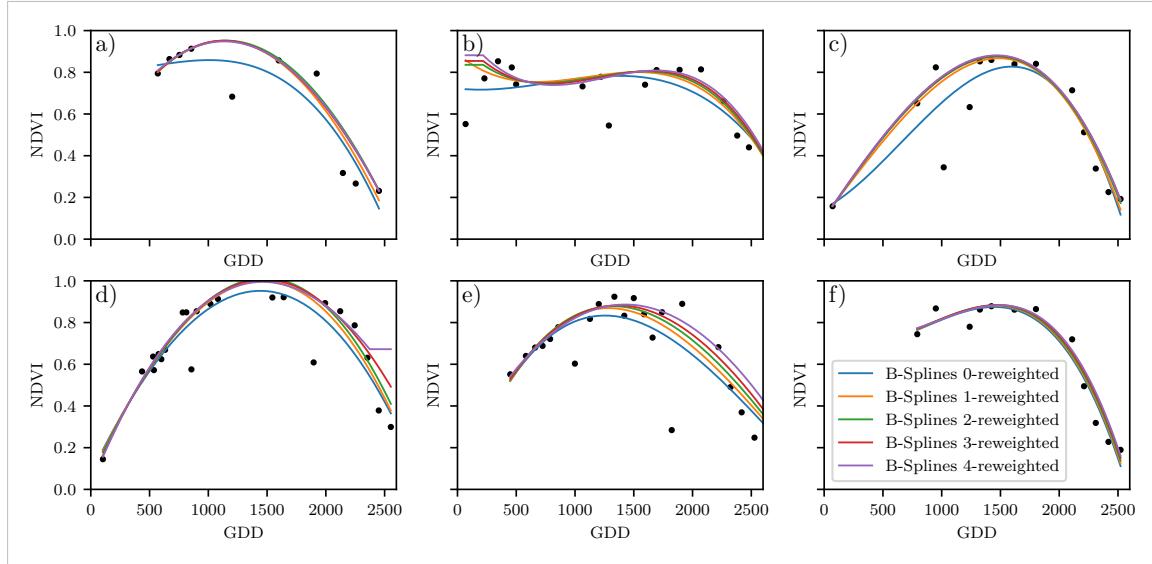


Figure B.2: B-splines fitted to different (SCL45) NDVI TS. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed.

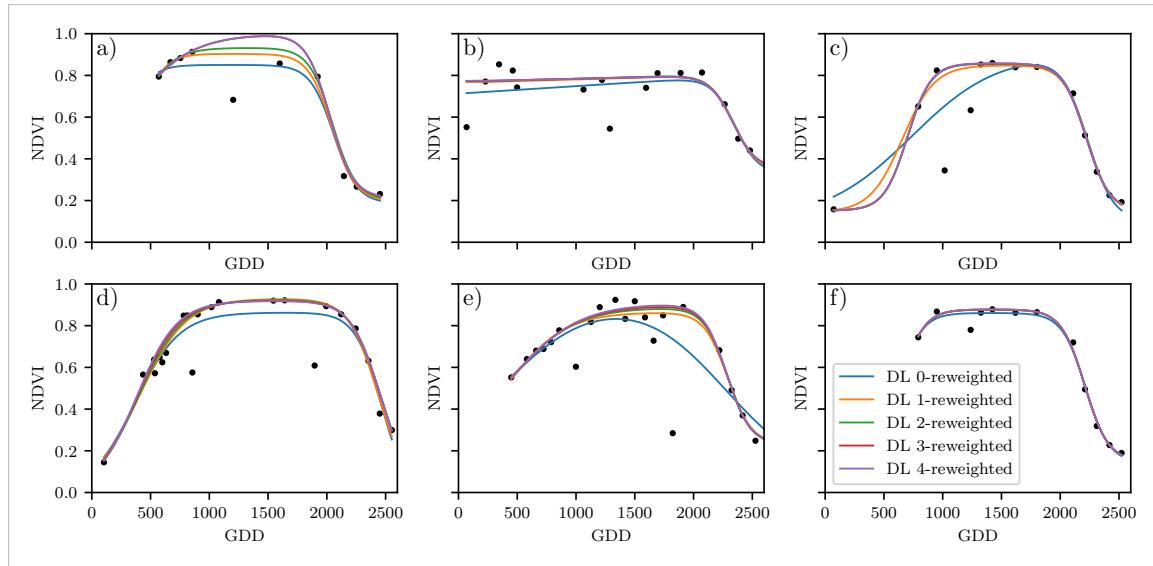


Figure B.3: A Double Logistic curve fitted to different (SCL45) NDVI TS. Iterations of a robustifying refit (as indicated in section 3.5) are also displayed.

1096 B.3 NDVI correction

Table B.1: Non-relative RMSE for yield prediction in [t/ha] (cf. table 5.2)

	RF	OLS ^{SCL}	OLS ^{all}	MARS	GAM	LASSO	no corrections
SS	1.144	1.033	1.051	1.042	1.046	1.042	1.095
SS ^{rob}	1.144	1.054	1.084	1.094	1.072	1.071	1.091
DL	1.150	1.115	1.116	1.116	1.097	1.098	1.159
DL ^{rob}	1.159	1.128	1.117	1.064	1.093	1.105	1.156

Table B.2: Coefficient of determination (R^2) of yield prediction (cf. table 5.2)

	RF	OLS ^{SCL}	OLS ^{all}	MARS	GAM	LASSO	no corrections
SS	0.431	0.486	0.477	0.481	0.479	0.481	0.455
SS ^{rob}	0.431	0.475	0.461	0.456	0.467	0.467	0.457
DL	0.427	0.445	0.444	0.444	0.454	0.453	0.423
DL ^{rob}	0.423	0.439	0.444	0.470	0.456	0.450	0.424

1097 B.3.1 OLS^{SCL} Model Outputs

```

1098
1099 1 Call:
1100 2 lm(formula = (paste(response, " ~ ", "ndvi_observed + scl_class"))),
1101 3     data = ndvi_df)
1102 4
1103 5 Residuals:
1104 6     Min      1Q  Median      3Q      Max
1105 7 -0.7997 -0.0717  0.0039  0.0695  0.6632
1106 8
1107 9 Coefficients:
1108 10             Estimate Std. Error t value Pr(>|t|)
1109 11 (Intercept) 0.21465    0.00230   93.46 < 2e-16 ***

```

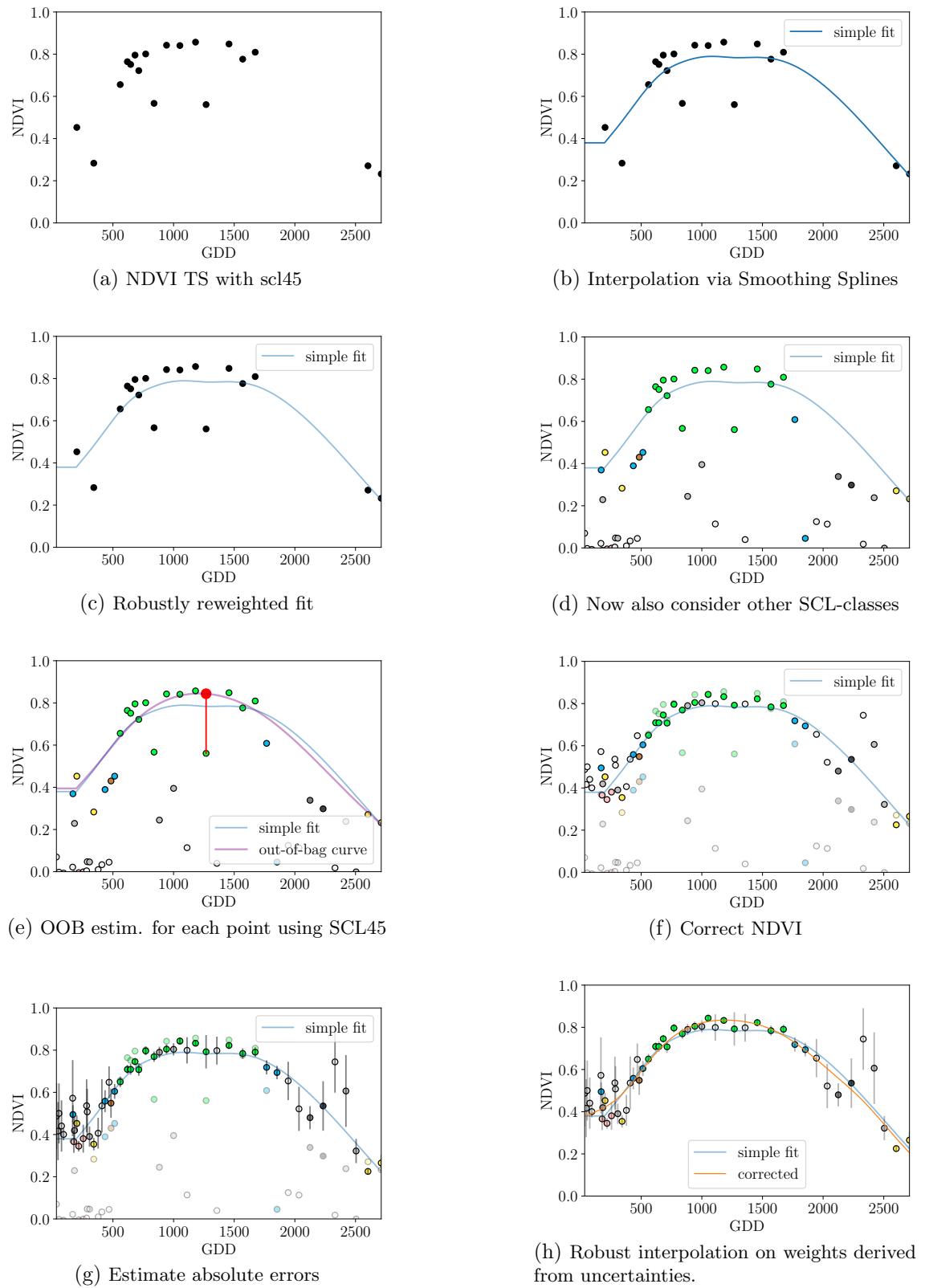


Figure B.4: Stepwise illustration of robust NDVI-Correction. For the color encoding of the SCL classes we refer to table 2.2.

1110	12	ndvi_observed	0.71116	0.00346	205.65	< 2e-16 ***
1111	13	scl_class3	0.02205	0.00356	6.20	5.8e-10 ***
1112	14	scl_class4	-0.00431	0.00251	-1.72	0.085 .

```

1113 15  scl_class5    -0.09875    0.00234   -42.15   < 2e-16 ***  

1114 16  scl_class6    -0.05301    0.01104   -4.80    1.6e-06 ***  

1115 17  scl_class7    0.11245    0.00274   41.09    < 2e-16 ***  

1116 18  scl_class8    0.25963    0.00253   102.57   < 2e-16 ***  

1117 19  scl_class9    0.35994    0.00236   152.47   < 2e-16 ***  

1118 20  scl_class10   0.09091    0.00308   29.54    < 2e-16 ***  

1119 21  scl_class11   0.29784    0.00392   76.06    < 2e-16 ***  

1120 ---  

1121 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

1122  

1123 Residual standard error: 0.146 on 124978 degrees of freedom  

1124 Multiple R-squared:  0.532,      Adjusted R-squared:  0.532  

1125 F-statistic: 1.42e+04 on 10 and 124978 DF,  p-value: <2e-16

```

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

```

1127  

1128 1 Call:  

1129 2 lm(formula = (paste(get_res(), " ~ ", "ndvi_observed + scl_class")),  

1130 3   data = ndvi_df)  

1131  

1132 5 Residuals:  

1133 6   Min     1Q   Median     3Q     Max  

1134 7 -0.2051 -0.0427 -0.0074  0.0329  0.6589  

1135  

1136 9 Coefficients:  

1137          Estimate Std. Error t value Pr(>|t|)  

1138 (Intercept) 0.18647  0.00126 147.74   < 2e-16 ***  

1139 ndvi_observed -0.13265  0.00190 -69.80   < 2e-16 ***  

1140 scl_class3 -0.00180  0.00196 -0.92    0.3587  

1141 scl_class4 -0.04069  0.00138 -29.55   < 2e-16 ***  

1142 scl_class5 -0.09698  0.00129 -75.32   < 2e-16 ***  

1143 scl_class6 -0.01906  0.00606 -3.14    0.0017 **  

1144 scl_class7  0.01641  0.00150 10.91    < 2e-16 ***  

1145 scl_class8 -0.00560  0.00139 -4.02    5.7e-05 ***  

1146 scl_class9 -0.01384  0.00130 -10.67   < 2e-16 ***  

1147 scl_class10 -0.00690  0.00169 -4.08    4.5e-05 ***  

1148 scl_class11 -0.01446  0.00215 -6.72    1.8e-11 ***  

1149 ---  

1150 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

1151  

1152 Residual standard error: 0.08 on 124978 degrees of freedom  

1153 Multiple R-squared:  0.352,      Adjusted R-squared:  0.352  

1154 F-statistic: 6.8e+03 on 10 and 124978 DF,  p-value: <2e-16

```

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

1156 check quantile und LOOCV definitions

1157 figure spacing (caption zu nah dran — manuell vspace einfügen wo nötig)