Contents

Chapter 1

Interpolation Methods

1.1 Setting

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

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

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

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

Different assumptions on m will lead to the following methods:

1.2 Overview

here will be my nice summary table :)

1.3 Methods - Description

1.3.1 Kernel Regression

As described previously, we would like to estimate

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f_X(x)}, \qquad (1.3.1.1)$$

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

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

By plugging the above into equation 1.3.1.1 we arrive at the *Nadaraya-Watson* kernel estimator:

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

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

XXX

	assumtpions	pros	cons	weights	bounded
Savitzky-Golay filter	– high frequencies are noise (low.pass filter) – equidistant points		– cannot deal natively with missing data (need some interpolation)	no	(yes)
+ NDVI	– upper envelope – vegetation cannot grow faster than some slope	– biological knowledge	– bad "upper envelope" since weights are not used for the estimation itselfe	(no)	(yes)
Loess					
Smoothing Splines	– 2cd derivative of function is integrable	– intuitive meaning of penalty – general assumptions – flexible shape	– unbounded	yes	no
B-Splines + Smooth	– function can be approximated by a linear combination of B-splines basis functions	– general assumtion – flexible shape	– unbounded – no intuitive meaning for smoothing		no
(Gaussian) Kernel Smoothing		– simple – general assumptions	– bandwidh: failes if there are big data-gaps	yes	yes
Fourier					
Double-Logistic	– function first increases then decreases – ndvi has a minimal value	– good for evergreen plants (if snow masks ndvi) –upper envelope	– parameterestimation can go seriously wrong – strange behaviour for long data-gaps	yes	yes
Universal Kriging	– function is a realization of a stationary gaussian process	– informative parameters – flexible	– regression to the mean – assumptions clearly not met	yes	yes

1.3.2   Savitzky-Golay Filter (SG Filter)

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

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

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

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

For equidistant points this can efficiently be calculated by

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

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

Adaptation to the NDVI

In a rather famous paper Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh (Chen et al.) a "robust" method based on the Savitzky-Golay has been used. The method is based on the assumption that due to atmospheric effects the observed NDVI tends to be underestimated and that it cannot increase too quickly[1].

Algorithm:

   i.) Remove points which are labeled as cloudy

  ii.) Remove points which would indicate an increase greater than 0.4 within 20 days

 iii.) Linearly interpolate to obtain an equidistant time series X^0

 iv.) Apply the Savitzky-Golay Filter to obtain a new time series X^1

 v.) Update X^1 by applying again a Savitzky-Golay Filter. Repeat this until $w^T|X^1 - X^0|$ stops decreasing, where w is a weight vector with $w_i = \min\left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|}\right)$. This reduces the penalty introduced by outliers[2] and by repeating this step we approach the "upper NDVI envelope".

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

---

[1]The latter is argued by the biological impossibility of such fast vegetation changes
[2]Here we call a point i an outlier if $X_i^0 < X_i^1$.

Extension: Spatial-Temporal-Savitzky-Golay Filter

One notable adaptation of the Savitzky-Golay is the presented by Cao, Chen, Shen, Chen, Zhou, Wang, and Yang (Cao et al.). The key difference is the additional assumption of the cloud cover being discontinuous and that we can improve by looking at adjacent pixels[3]. Because we are working with rather high resolution satellite data, and we need the variance in the predictors we will waive this extension.

1.3.3   Robust Locally Weighted Regression (Robust LOESS)

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

The Locally Weighted Regression (LOESS) can be understood as a generlalization of the Savitzky-Golay Filter (c.f. sec. 1.3.2).

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

$$w_i(x_j) = \begin{cases} \left(1 - \left(\frac{x_j}{h_i}\right)^2\right)^2, & \text{for } |x_j| < h_i \\ 0, & \text{for } |x_j| \geqslant h_i \end{cases},$$

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

How does the Robust LOESS differ from the SG Filter?

Pros	Cons
• 1	• 1
• 2	• 2

1.3.4   Double Logistic

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

Assumptions:

- 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 into several years into account.

- The growth cycle can be divided into an increase and a decrease period where the time series follows a logistic function. The maximum increase (or decrease) is observed at t_0 (or t_1) with a slope of d_0 (or d_1).

---

[3]Here, 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

The equation of the double-logistic fit is given by:

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

Where the five free parameters: $Y_{\max}$, d_0, d_1, t_0, t_1 are initially estimated by least squares. Such fit can be seen in figure 1.1.

Similar as for the Savitzky-Golay Filter (c.f. section 1.3.2) we reestimate (only once) the parameters by giving less weight to the overestimated observations and more weight to the underestimated observations[4].

Pros	**Cons**
• Incorperates subject specific knowledge in the case of evergreen plants covered by snow. • Optimized parameters have an intuitive meaning.	• Strong shape assumptions on the NDVI curve. • Parameter optimization might go wrong. This can be mitigated to some extend to provide bounds for the paramters • Strange behaviour in regrions with little observations. (cf. figure 1.1)

1.3.5   Fourier Approximation

Similar as in section 1.3.4 we fit a parametric curve to the data by least squares. Here we take the second order fourier series:

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

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

Pros	**Cons**
• Assumtion of periodicity can be helpful if we are modelling multiyear grow cycles • Flexible curve shape	• Bad behaviour in regions with little data (cf. figure 1.1) • Hard to interpret estimated parameters • Parameter estimation can go wrong. Introducing bounds can help.

1.3.6   B-splines

from Lyche and Morken (Lyche and Morken)

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

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

---

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

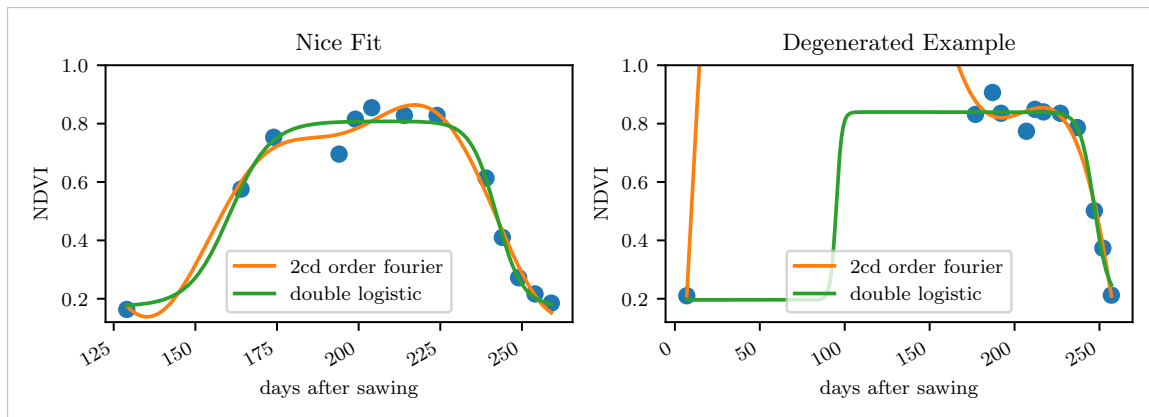Figure 1.1: Here we observe the nice fitting possibilitys of the two parametric methods but notice also some misbehaviour

Smoothing: We can relax the constrain that we have to perfectly interpolate. Thus we use the minimum number of knots[5] such that: $\sum_{i=1}^{n}(w(y_i - \hat{y}_i))^2 \leq s$

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

1.3.7 Natural Smoothing Splines

Let $\mathcal{F}$ be the Sobolev space (the space of functions of which the second derivative is integrable). Then the unique[6] minimizer

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

is a natural[7] cubic spline (i.e. a piecewise cubic polynomial function). The objective function has an intuitive meaning, as to avoid lateral acceleration it is desirable to move the steering wheel as little as possible, when driving a car.

1.3.8 Whittaker Smoother

1.3.9 Kriging

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

---

[5]SciPy uses FITPACK and DFITPACK, the documentation suggests that smoothness is achieved by reducing the number knots used

[6]Strictly speaking it is only unique for $\lambda > 0$

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

Pros	Cons
• 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) • performs also well if points are not equidistant • fixes the Runge's phenomenon (fluctuation of high degree polynomial interpolation)	• choose λ

Pros	Cons
• 1 • 2	• 1 • 2

Definitions and Assumptions

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

Assumption: We will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following only $\gamma(h)$.[8]

We also define the variogram of a Gaussian process as

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

And decide to use a gaussian Variogram defined by

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

where h is the distance, n is the nugget, r is the range and p is the partial sill visuilized in figure 1.2.[9]

1.3.10   Other Methods to study:

From inroduction of Chen, Jönsson, Tamura, Gu, Matsushita, and Eklundh (Chen et al.): (1) threshold- based methods, such as the best index slope extraction algorithm (BISE) (Viovy et al., 1992); (2) Fourier-based fitting methods (Cihlar, 1996; Roerink et al., 2000; Sellers et al., 1994); and (3) asymmetric function fitting methods such as the asymmetric Gaussian function fitting approach (Jonsson Eklundh, 2002) and the weighted least-squares linear regression approach (Swets et al., 1999).

---

[8]Note 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.

[9]Strictrly speaking we use a scaled version of the variogram. Thus only the ratio of p/n matters.
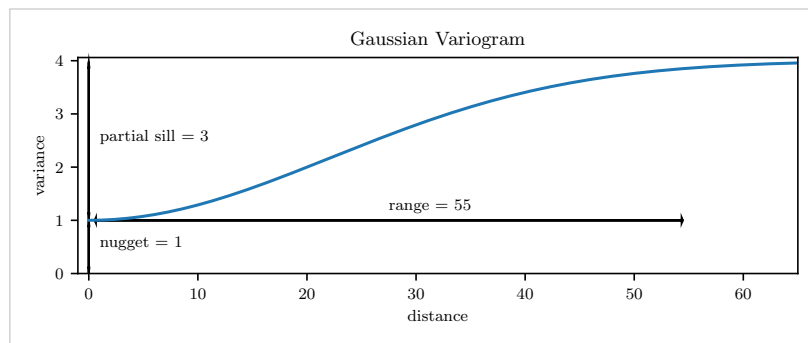
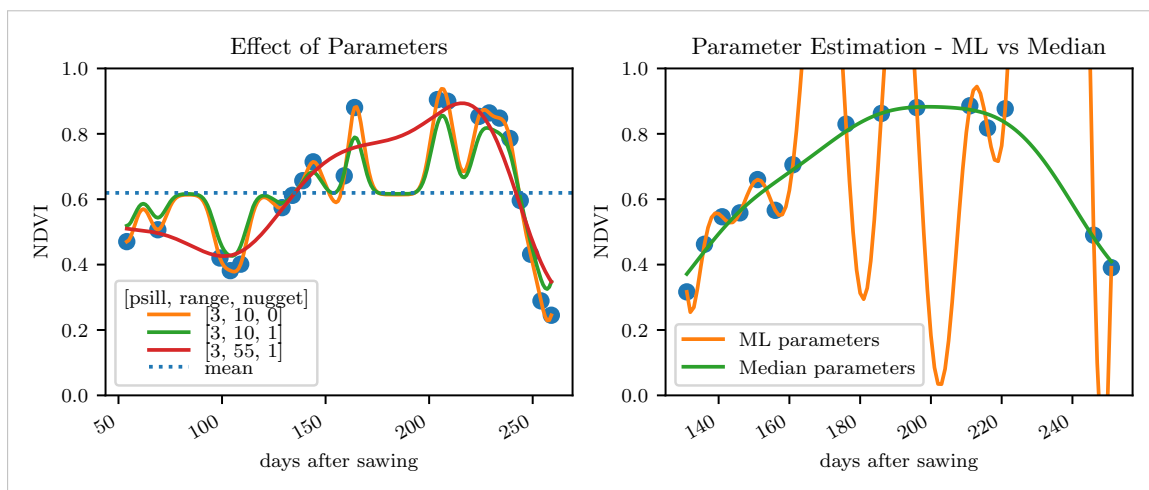Figure 1.2: Gaussian Variogram with nugget=1, partial sill=3, range=55

Figure 1.3: On the left we see how the interpolation change if we increase the nugget and the range parameter. On the right we compare two kriging interpolations where one takes parameters by numerically maximizing the (which results in a very small nugget) and the other takes the median of many such numerical optimizations.

Pros	Cons
• It is a well-studied method	• Regression to the mean
• Parameters have an intuitive meaning	• Assumption of constant mean and constant variance is clearly violated. Thus, the NVI is not a stationary process.
• Flexible covariance structure	• Skewness of errors is not taken into account.

Bibliography

Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.), *Model-Based Geostatistics*, pp. 46–78. Springer.

Beck, P. S. A., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *100*(3), 321–334.

Cao, R., Y. Chen, M. Shen, J. Chen, J. Zhou, C. Wang, and W. Yang. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *217*, 244–257.

Cappellari, M., R. M. McDermid, K. Alatalo, L. Blitz, M. Bois, F. Bournaud, M. Bureau, A. F. Crocker, R. L. Davies, T. A. Davis, P. T. de Zeeuw, P.-A. Duc, E. Emsellem, S. Khochfar, D. Krajnović, H. Kuntschner, R. Morganti, T. Naab, T. Oosterloo, M. Sarzi, N. Scott, P. Serra, A.-M. Weijmans, and L. M. Young. The ATLAS3D project - XX. Mass-size and mass-sigma distributions of early-type galaxies: Bulge fraction drives kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function. *432*, 1862–1893.

Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *91*, 332–344.

Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *74*(368), 829–836.

Lyche, T. and K. Morken. Spline Methods.

Savitzky, A. and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *36*(8), 1627–1639.

Schafer, R. W. What Is a Savitzky-Golay Filter? [Lecture Notes]. *28*(4), 111–117.

Appendix A

Hi Mom