Contents

Chapter 1

plotting methods

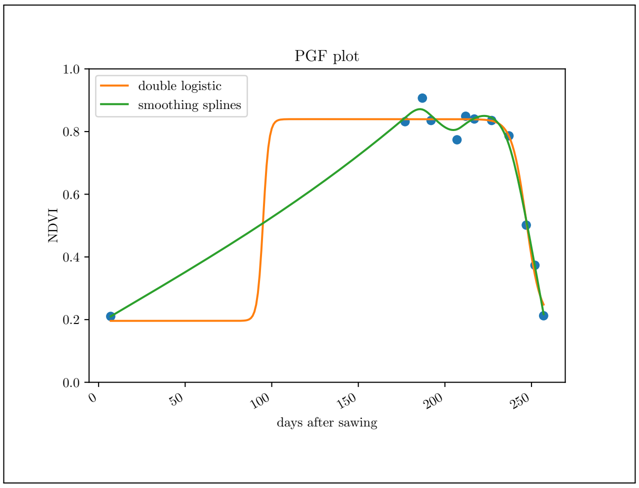


Figure 1.1: A PDF plot from `matplotlib`.

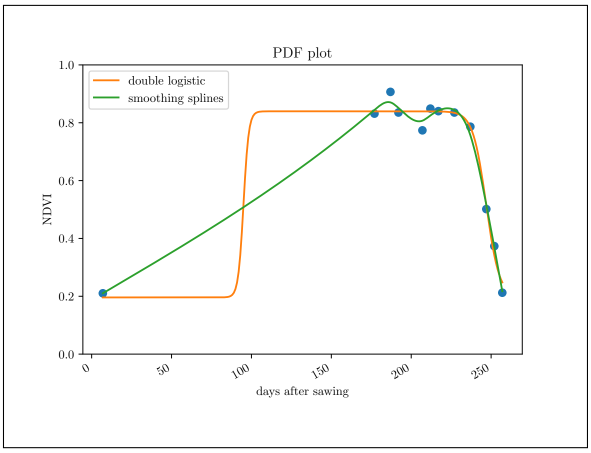


Figure 1.2: A PGF plot from `matplotlib`.

Chapter 2

Interpolation Methods

2.1 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 models:

2.2 Methods - Description

2.2.1 Kernel Regression

As described previously, we would like to estimate

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f_X(x)}, \qquad (2.2.1.1)$$

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 2.2.1.1 we arrive at the *Nadaraya-Watson* kernel estimator:

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

Pros: - can be assigned degrees of freedom (trace of the hat-matrix) - estimation of the noise variance $\hat{\sigma}_{\varepsilon}^2$ (XXX c.f. CompStat 3.2.2)

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

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

2.2.2 loess

2.2.3 Savitzky-Golay 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:

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

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

For equidistant points this can efficiently be calculated by

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

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

Pros - popular technique in signal processing - efficient calculation for equidistant points

Cons - no natural way of how to estimate points which are not in the data. XXX

Interpolation and Smoothing

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 but this time weight the observations by $w_i = \min\left(1, 1 - \frac{X_i^1 - X_i^0}{\max_i \|X_i^1 - X_i^0\|}\right)$. This reduces the impact of outliers[2] and by repeating this step the authors obtain an "upper NDVI envelope"

Pros - Upper envelope matches intuition for the NDVI. Therefore, it is robust against outliers with small values.

Cons - Not generalizable to other spectral indices. - 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.

2.2.4 Double Logistic

The Double Logistic smoothing as introduced 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).

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

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

Where the five free parameters: Y_{max}, d_0, d_1, t_0, t_1 are initially estimated by least squares. Similar as for the Savitzky-Golay Filter (c.f. section 2.2.3) we reestimate (only once) the parameters by giving less weight to the overestimated observations and more weight to the underestimated observations[4].

2.2.5 Fourier Approximation

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

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

Beck, Atzberger, Høgda, Johansen, and Skidmore (Beck et al.) shows in their lag-plots a heavy autocorrelation of resiudals

2.2.6 Cubic Smoothing Splines

We interpolate with a function in C^2 (space of three time continuous differentiable functions) which is defined piecewise by cubic polynomials. **Pros** Regression splines (B-splines) Wood, Pya, and Säfken (Wood et al.) use a basis of the spline space (e.g. B-splines or j-th cardinal basis) and fit the splines of degree k to approximate the data.

[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

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

2.2.7 B-splines

from Lyche and Morken (Lyche and Morken)

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

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

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

Pros - can be assigned degrees of freedom - extendable to ”smooth” version - performs also well if points are not equidistant

Cons - smoothing process does not translate well to a interpretation (unlike smoothing splines) - choice of smoothing parameter s

2.2.8 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.

Pros: - 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)

Cons: - choose λ

2.2.9 Penalized Regression Splines

Intuition: similar as Natural Smoothing Splines, but we choose knots

2.2.10 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 ($\forall x \notin [x_1, x_n] : \hat{m}''(x) = 0$)

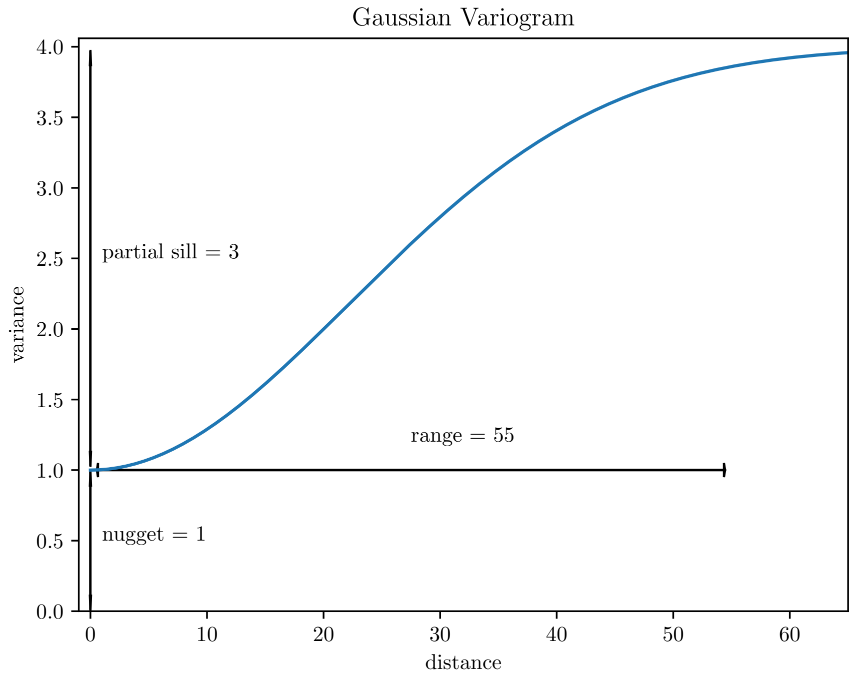


Figure 2.1: Gaussian Variogram with nugget=1, partial sill=3, range=55

Definitions and Assumptions

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

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

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

And decide to use a gaussian Variogram defined by

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

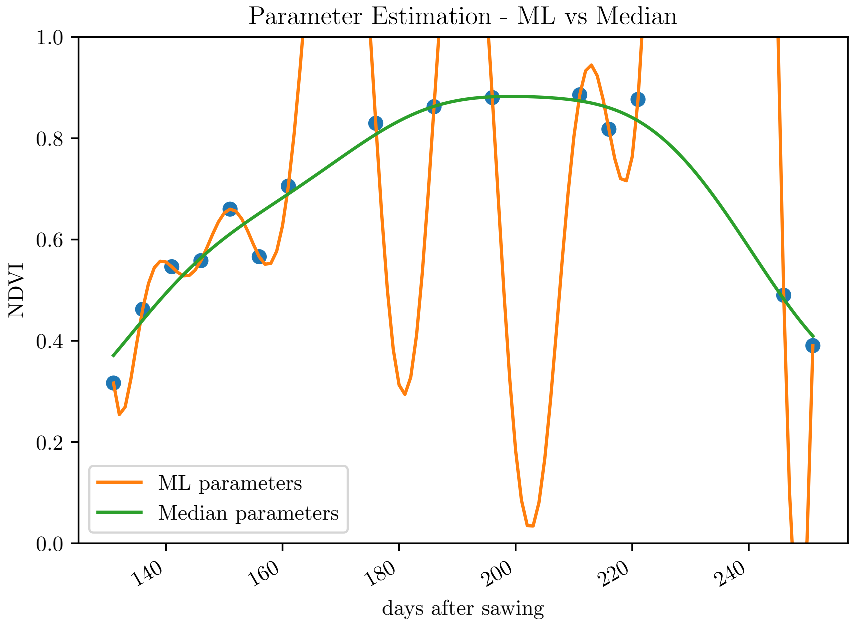
where h is the distance, n is the nugget, r is the range and p is the partial sill visuilized in figure 2.1.[9]

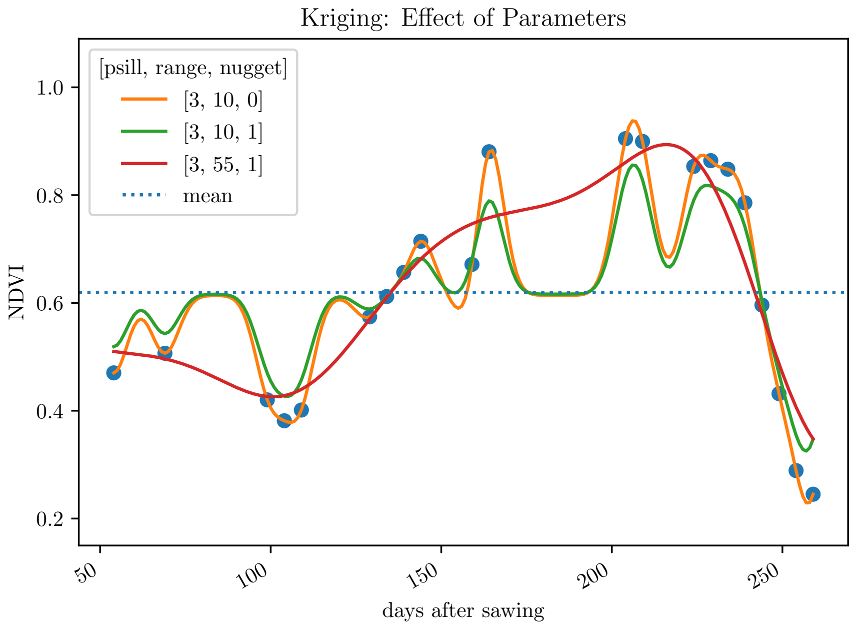
2.2.11 Other Methods to study:

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)

[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.


Parameter Estimation - ML vs Median
NDVI
days after sawing
ML parameters
Median parameters
0.0
0.2
0.4
0.6
0.8
1.0
140
160
180
200
220
240


Kriging: Effect of Parameters
NDVI
days after sawing
[psill, range, nugget]
[3, 10, 0]
[3, 10, 1]
[3, 55, 1]
mean
0.2
0.4
0.6
0.8
1.0
50
100
150
200
250

(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).

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.

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.

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.

Wood, S. N., N. Pya, and B. Säfken. Smoothing Parameter and Model Selection for General Smooth Models. *111*(516), 1548–1563.

Appendix A

Hi Mom