Research Article

Corinna Harmening*, Hans Neuner

Choosing the Optimal Number of B-spline Control Points (Part 1: Methodology and Approximation of Curves)

DOI 10.1515/jag-2016-0003
received March 15, 2016; accepted July 13, 2016.

Abstract: Due to the establishment of terrestrial laser scanner, the analysis strategies in engineering geodesy change from pointwise approaches to areal ones. These areal analysis strategies are commonly built on the modelling of the acquired point clouds.

Freeform curves and surfaces like B-spline curves/surfaces are one possible approach to obtain space continuous information. A variety of parameters determines the B-spline's appearance; the B-spline's complexity is mostly determined by the number of control points. Usually, this number of control points is chosen quite arbitrarily by intuitive trial-and-error-procedures. In this paper, the Akaike Information Criterion and the Bayesian Information Criterion are investigated with regard to a justified and reproducible choice of the optimal number of control points of B-spline curves. Additionally, we develop a method which is based on the structural risk minimization of the statistical learning theory. Unlike the Akaike and the Bayesian Information Criteria this method doesn't use the number of parameters as complexity measure of the approximating functions but their Vapnik-Chervonenkis-dimension. Furthermore, it is also valid for non-linear models. Thus, the three methods differ in their target function to be minimized and consequently in their definition of optimality.

The present paper will be continued by a second paper dealing with the choice of the optimal number of control points of B-spline surfaces.

Keywords: AIC, BIC, B-spline Curves, Structural Risk Minimization, VC-dimension

***Corresponding author: Corinna Harmening**, Department of Geodesy and Geoinformation, TU Wien, Gusshausstr. 25–29 / E120, 1040 Vienna, Austria, e-mail: corinna.harmening@tuwien.ac.at
Hans Neuner, Department of Geodesy and Geoinformation, TU Wien, Vienna, e-mail: hans.neuner@geo.tuwien.ac.at

1 Introduction

The development of areal measurement techniques like laser scanning has shifted the focus in engineering geodesy from point-wise approaches to areal ones. Although the areal measurement techniques yield many benefits like the contactless and fast data acquisition as well as a resulting high point density, they also hold new challenges, especially concerning the respective analysis strategies [20].

Deformation analysis as a classical task in engineering geodesy is particularly affected by these new challenges as the acquired points are not longer reproducible so that the assignment of homologous points of different measuring epochs becomes impossible [29]. In order to face this challenge, a surface orientated description of deformations is demanded [18].

The present study is part of a project that deals with a surface orientated description of deformations by means of a least-squares collocation (LSC) which is continuous in space and time so that it is possible to describe deformations at arbitrary times and at arbitrary places of the object. LSC is a well-established method in geodesy, which differs from other methods in that the observed phenomena are not only modelled by means of a deterministic trend, but also by means of a statistical signal [18]. Based on this spatio-temporal collocation, a non-linear sensitivity analysis will be performed in order to determine optimal acquisition configurations (see [22] for initial studies concerning this task).

This article focuses on the description of the collocation's geometric part. The trend-modelling is realized according to the geometry-based approach, which is one of five possible ways to handle point clouds with regard to deformation analysis (for further details see [29]). This approach involves a geometric modelling of the point clouds of each measuring epoch. The resulting models are the basis for the subsequent deformation analysis (see for example [3], [34]). In the presented project B-splines are used to model the point clouds of the individual measuring epochs.

Freeform surfaces like B-splines are particularly suitable to model point clouds as they are a flexible tool to model even arbitrary geometries [30]. While being a standard tool in computational geometry [16], they also find their way into geodesy: They are used to model the vertical total electron content of the Earth's atmosphere [31], to approximate height changes [19], to determine a bridge's deflection under traffic load [28] and as a basis for plant phenotyping [15]. As they can easily be extended to the time-dimension (see for example [24]), and as they have the property of local support [30], meaning that a change of some parameters affects the surface only locally, they are just right for being used with regard to deformation analysis.

The B-spline's complexity is to a large extent determined by means of the number of control points. Usually, this parameter group is chosen quite arbitrarily by intuitive trial-and-error-procedures, trying to find a balance between parsimony of the curve/surface and the approximation quality. Model selection is a scientific field dealing with the problem to find the optimal model; the optimality is defined by a certain target function. Model selection criteria like the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used in geodesy in order to find the best mathematical function to model ice mass variations [4], whereas methods of the field of statistical learning theory (SLT) like structural risk minimization (SRM) are used to determine the structure of artificial neural networks [27]. The support vector machines, which are a concrete realization of the SRM-principle and a standard tool to solve classification problems, are most known in geodesy [12].

In this contribution methods of the field of model selection are investigated with regard to their applicability to choose the number of curve control points in a justified way. Furthermore, methods of the field of statistical learning theory are refined in order to establish an alternative to the model selection criteria whose use is limited to linear models.

The paper is structured as follows: Section 2 gives a short overview of the estimation of B-spline curves and introduces the basic principles for choosing the optimal number of control points of B-spline curves. Section 3 provides the mathematical basis for model selection in general as well as for the two model selection criteria AIC and BIC in particular. In section 4 the basic principles of SLT are introduced, focussing on the Vapnik-Chervonenkis (VC)-dimension of function classes and SRM. In section 5 a classifier which allows the determination of the VC-dimension of B-splines is developed. This is a fundamental precondition to apply SRM to the determination of the optimal number of control points. AIC, BIC and SRM are used to determine the optimal number of curve control points in section 6. The results are analyzed, evaluated and compared, and a final conclusion is drawn in section 7. Section 8 gives a short outlook for future investigations.

2 Theoretical Basics

2.1 Estimation of B-spline Curves

The shape of a B-spline curve $\mathbf{C}(u)$ of degree p is predefined by its $n+1$ control points $\mathbf{P}_i$ [30]:

$$\mathbf{C}(u) = \sum_{i=0}^{n} N_{i,p}(u) \cdot \mathbf{P}_i, \quad u = [0,\ldots,1]. \tag{1}$$

Therefore, a curve point, whose position on the curve is specified by the curve parameter u, results as the weighted average of the control points $\mathbf{P}_i$. The corresponding weights are defined by the B-spline basis functions $N_{i,p}(u)$, which can be computed recursively by means of the Cox-de-Boor-algorithm (see [14]; [5]). The definition of the knot vector $U = [u_0, \ldots, u_r]$, which splits the B-spline's domain into knot spans, is a prerequisite for the computation of the basis functions.

The calculation of the knot vector requires the allocation of parameters to the observations as shown in [30], [32] and [8]. A suitable method was introduced in [17]. The parameterization of the observations is also a prerequisite for the linear estimation of the B-spline curve according to [30].

In order to obtain a linear relationship between the l observations $\mathbf{C}(u_k)$, $k = 1,\ldots,l$, and the unknowns $\mathbf{P}_i$, the B-spline's degree is usually specified a priori: The use of cubic B-splines with $p = 3$ is a generally accepted choice [30].

2.2 Problem Definition

The remaining parameter type to be specified prior to the estimation of the control points is the number of control points. It has an immense influence on the B-spline's appearance as it determines its complexity: Figure 1 shows a simulated B-spline curve with $n + 1 = 6$ control points which is superimposed by white noise (blue points). Based on this point cloud, three B-spline

curves with different numbers of control points are estimated:

The more complex the estimated model becomes by increasing n, the better it approximates the data [26]. This behaviour is reflected by the sum of squared residuals which is largest for the green curve and smallest for the black one.

The green curve has four control points and is obviously too inflexible to model the complete phenomenon. Models leading to curves like the green one have a large bias and **underfit** the data. They are not able to reflect all information [9].

Although it leads to the smallest sum of squared residuals, the black curve with 30 control points is much too complex and is consequently not able to separate the noise from the actual phenomenon; this model **overfits** the data, as it encodes noise in the estimated control points. Therefore, one would intuitively choose the red curve rather than the black one to be the correct one.

Generally speaking, the model's bias becomes smaller when the number of the model's free parameters increases, whereas the model's **variance** grows: Different samples of the same phenomenon will lead to very different curves, as the noise is modelled too [9].

Usually, intuitive trial-and-error-procedures like these are used to choose the optimal number of control points. However, these intuitive approaches show a lack of scientific validity as they are based on a subjective decision rather than on an optimality criterion, which has to be fulfilled by the best model. Furthermore, the subjective decision on the best model is by no means as distinct as in figure 1 if the task of model selection includes neighboured models with $n+1=5$ and $n+1=7$.

For this reason, different possibilities to choose the optimal number of control points in a justified way are presented, compared and evaluated in the following.

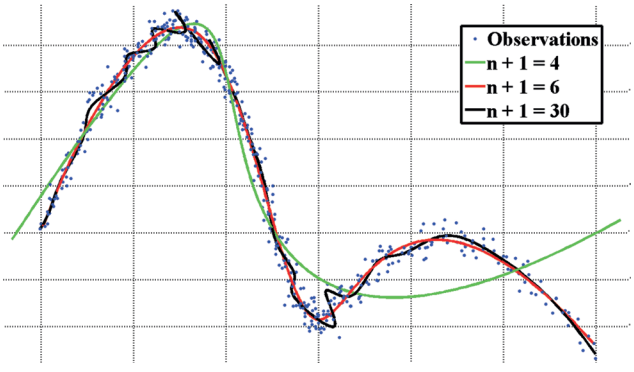

Figure 1: B-spline curves with different number of control points $n+1$.

The problem outlined above goes beyond the classical problem of parameter estimation, where the model's structural form $g(\text{data}, \boldsymbol{\theta})$ is assumed to be known and only the optimal model parameters $\hat{\boldsymbol{\theta}}$ have to be estimated by means of e. g. least squares (LS) methods or by means of the maximum likelihood (ML) theory [23]. Now, a set of candidate models $g(\text{data}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with $\boldsymbol{\Theta}$ being a set of parameters, is available. That function which approximates the underlying phenomenon in an optimal manner has to be selected [12].

A selected model should approximate the underlying phenomenon rather than the available data set, which is only a finite and noisy realization of the phenomenon: The aim is to model the information which is provided by the data rather than the data itself [10]. Only if the model separates the noise from the information, a reliable prediction of future data sets is possible [26].

For this reason, the optimal model has to find a balance between simplicity (small variance, large bias) and complexity (large variance, small bias), which is known as the **bias-variance trade off**. Techniques which implement this trade off between underfitting and overfitting can be found in the field of model selection as well as in the field of SLT. All of these techniques implement the **principle of parsimony**, which states that a good model should approximate the phenomenon as well as possible while being as simple as possible [9].

3 Model Selection

Model selection deals with the problem of choosing an optimal model from a set of possible candidate models. Information criteria are the tools which are used in model selection to compare different models. They evaluate each model according to a certain optimality criterion and choose that model to be optimal which achieves the best score [9].

3.1 Classification of Model Selection Criteria

There exists a variety of model selection criteria, which can be classified to be either asymptotic efficient or asymptotic consistent [26].

– Asymptotic efficient criteria operate according to the quote of George Box that "all models are wrong, but some are useful" [6]: The underlying phenomenon is assumed to have infinite dimension and is too

complex to be captured by means of a finite sample [9]. As a consequence, the "true" functional dependency cannot be described by a model and is not contained in the set of candidate models. Asymptotic efficient criteria aim to identify that candidate model with finite dimension which approximates the true dependency in an optimal manner [26]. The respective optimality criterion is based on a distance measure which defines the closeness between the model and the truth. For infinite sample sizes those criteria identify that model which minimizes the prediction error [1]. Asymptotic efficient criteria choose more complex models to be optimal when the sample size increases, because more details of the infinite complex phenomenon can be unveiled with growing sample size [9].

– Asymptotic consistent criteria are based on the assumption that the true dependency is of finite dimension and is contained in the set of candidate models. This class of criteria aims to identify the true dependency; it works successfully with probability of 1 when the sample size is infinite [26]. In contrast to the asymptotic efficient criteria, which choose more complex models when the sample size increases, asymptotic consistent criteria settle at a certain model when the sample size increases [1].

In the following the two most frequently used information criteria, AIC and BIC, are presented, before they are investigated in terms of their applicability to choose the correct number of B-spline control points.

Both criteria are based on ML theory, which chooses those model parameters $\hat{\boldsymbol{\theta}}$ to be optimal that maximize the conditional probability $\mathcal{L}(\boldsymbol{\theta}|\text{data})$ in consideration of the given data set. It has to be noted that there are different ways to derive the criteria. The original derivations can be found in [2] (AIC) and [33] (BIC), whereas this paper follows derivations according to [9] and [13] respectively, which are more intuitively accessible and less restrictive.

3.2 Akaike Information Criterion (AIC)

The AIC belongs to the asymptotic efficient information criteria. It uses the Kullback-Leibler(KL)-distance $I(f, g)$ between the true distribution $f(x)$ and the candidate distribution $g(x \mid \theta)$:

$$I(f,g) = \int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx. \tag{2}$$

The KL-distance can either be interpreted as a directed distance between two distributions or as the loss of information which is sustained when the candidate model $g(x \mid \theta)$ is used instead of the truth $f(x)$. In order to find the model which is optimal in the sense of the AIC, the KL-distance has to be minimized [9].

As the truth $f(x)$ is unknown, a transition from absolute distances to relative ones is made:

$$I(f,g) = \int f(x) \log(f(x))\, dx - \int f(x) \log(g(x|\theta))\, dx \tag{3}$$

$$= E_f[\log(f(x))] - E_f[\log(g(x|\theta))] \tag{4}$$

$$= C - E_f[\log(g(x|\theta))]. \tag{5}$$

$E_f[.]$ is the expectation operator with respect to the distribution f.

The value of C is constant across all models so that the second term – the relative distance between the two distributions – becomes the important term to be determined. However, the actual parameters $\boldsymbol{\theta}$ of the candidate model are also unknown and can only be estimated by means of the available data set. For this reason, the objective function T changes to be the expected relative KL-distance, which has to be minimized in order to find the optimal model [9]:

$$T = E_{data}\, E_f[\log(g(x|\hat{\theta}(\text{data})))]. \tag{6}$$

In [2] it is shown that the maximized log-likelihood is a biased estimator of this objective function and that the bias is asymptotically equal to the number of estimated parameters K, provided that the candidate models are similar to the true phenomenon [10]. This result leads to the information criterion

$$\text{AIC} = -2 \log(\mathcal{L}(\hat{\theta}|\text{data})) + 2K. \tag{7}$$

In case of linear models as well as independent and normal distributed residuals ϵ with constant variance, the criterion in equation (7) can be adapted to the LS-case [9]:

$$\text{AIC} = l \log(\hat{\sigma}^2) + 2K \tag{8}$$

$$\hat{\sigma}^2 = \frac{\sum_i \hat{\epsilon}_i^2}{l}. \tag{9}$$

In the equations above the sample size is denoted by l and the ML-estimator of the variance factor by $\hat{\sigma}^2$. The

absolute number K of parameters includes the variance factor $\hat{\sigma}^2$.

The estimator of the relative KL-distance in equation (7) has been proven to be a poor estimator in case the number of parameters is large in relation to the sample size. These observations lead to an improved criterion for small sample sizes [21]:

$$\mathrm{AIC}_C = \mathrm{AIC} + \frac{2K(K+1)}{l-K-1}. \tag{10}$$

In [9] it is recommended to use the AIC_C if

$$\frac{l}{K} < 40. \tag{11}$$

When selecting a model according to AIC, the corresponding score is computed for each candidate model. The candidate model having the smallest score is chosen to be optimal. However, the size of the scores is not meaningful: Due to the consideration of relative distances instead of absolute ones, it is impossible to evaluate how good a model is in an absolute sense, but only how good it is compared to the other candidate models.

3.3 Bayesian Information Criterion (BIC)

The BIC as a representative of the asymptotic consistent criteria was originally developed by Schwarz in [33] and is therefore also called the Schwarz Information Criterion (SIC) in some literature. The original derivation is based on severe restrictions on the probability distributions, which have to belong to the regular exponential family. However, in subsequent publications these restrictions were eased (see for example [11] or [13]). The restrictions on the data sets, which have to be independent and identically distributed (iid), and on the linearity of the models are identical to the prerequisites of AIC.

The BIC aims to select that model from a set of available models which is a posteriori the most likely [11]. The derivation of this criterion is based on Bayes' theorem, which computes the posteriori probability of the j-th model M_j given the data [13]:

$$P(M_j|\mathrm{data}) = \frac{P(M_j)}{f(\mathrm{data})} \underbrace{\int f(\mathrm{data}|M_j,\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j|M_j)\,d\boldsymbol{\theta}_j}_{\lambda_j(\mathrm{data})}. \tag{12}$$

According to equation (12), the posterior $P(M_j|\mathrm{data})$ is computed by means of three quantities:

- The prior probability of the j-th candidate model $P(M_j)$. Typically, the priors of the models are unknown and, therefore, are assumed to be equal [40]. As a consequence, this quantity has no influence when the BIC-scores of different models are compared.
- The prior probability of the data $f(\mathrm{data})$. This measure is constant over the set of models and does neither influence the computation of the posteriori probabilities.
- The marginal likelihood $\lambda_j(\mathrm{data})$ of the j-th candidate model. The marginal likelihood is computed by means of the likelihood $f(\mathrm{data}|M_j,\boldsymbol{\theta}_j)$ and the prior density of the respective parameter $\boldsymbol{\theta}_j$, given the j-th model $\pi(\boldsymbol{\theta}_j|M_j)$. This is the important quantity to be evaluated:

$$\lambda_j(\mathrm{data}) = \int f(\mathrm{data}|M_j,\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j|M_j)\,d\boldsymbol{\theta}_j \tag{13}$$

$$= \int \mathcal{L}_j(\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j|M_j)\,d\boldsymbol{\theta}_j. \tag{14}$$

These considerations lead to the exact definition of the BIC [13]:

$$\mathrm{BIC} = 2\log(\lambda_{l,j}(\mathrm{data})). \tag{15}$$

As a closed evaluation of the marginal likelihood is typically impossible, this expression cannot be evaluated exactly. Alternatively, the expression is approximated by means of Laplace's method, resulting in the common information criterion

$$\mathrm{BIC} = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\mathrm{data})) + \log(l)\,K. \tag{16}$$

3.4 AIC vs. BIC

The two model selection criteria AIC and BIC have a very similar structure: Both formulas are penalized log-likelihood criteria with BIC imposing a stronger penalty on the number of estimated parameters than AIC for $l \geq 8$ [13]. Although the criteria's derivations are based on objective functions which do not implement the principle of parsimony per se, the resulting criteria do, as they find a balance between approximation quality and complexity: The more parameters are included into the adjustment, the better the model fits the data and the smaller the first term in equations (7) and (16) becomes. However, as every additional parameter increases the result's uncertainty

with respect to prediction, the number of parameters is penalized by means of the second term in equations (7) and (16) [9].

Despite their similar structure, the criteria are based on two completely different philosophies [10]: AIC is an asymptotic efficient criterion and aims finding that model which is closest to the true phenomenon by approximating an information-theoretical measure. On the contrary, BIC is an asymptotic consistent criterion which aims finding the truth in a set of candidate models by approximating the posteriori probabilities of the models.

It can be summarized that AIC and BIC differ in the definition of the models' quality, as they approximate two completely different objective functions. As a consequence, the criteria do not necessarily choose the same model to be optimal although both criteria select a model which is optimal in a certain sense [25].

The question, which model selection criterion is the better, cannot be answered in general, but depends on the respective model selection problem as well as on the respective data sample. Furthermore, the properties described above refer to the asymptotic case: These properties are not necessarily valid for finite sample sizes [9].

4 Statistical Learning Theory (SLT)

In model selection a function's complexity is expressed by means of the number of free parameters. However, for some functions the number of parameters is not an appropriate complexity measure: Figure 2 shows a family of curves whose complexity is infinite, although its appearance is determined by only one parameter [12]. Functions like these require a complexity measure, which goes beyond the number of parameters. Such a complexity measure is the Vapnik-Chervonenkis(VC)-dimension, which plays an important role in SLT.

SLT was established by Vladimir Vapnik and provides a theoretical framework for the estimation of functional dependencies on a given set of data [38]. Based on this framework, learning problems like regression

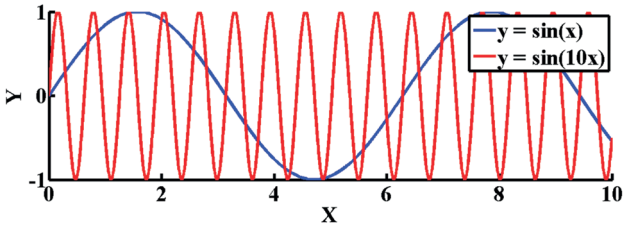

Figure 2: Complexity of $y = \sin(ax)$.

estimation, pattern recognition or density estimation can be solved. In contrast to model selection, whose assertions refer to the asymptotic case, SLT describes the estimation of functional dependencies in finite samples and explicitly takes the sample size into consideration. Furthermore, SLT can also be applied to non-linear models [12]. Unlike BIC, which requires the truth to be of finite dimension, SLT does not have any requirements for the finiteness of the model's dimension. As a consequence, the truth has not to be included into the set of candidate models.

4.1 Definition of the Learning Problem

The general learning problem is based on an input vector $\mathbf{x}$, which is independently generated according to an unknown probability distribution $P(\mathbf{x})$. Based on this input vector, a system generates an output vector $\mathbf{y}$ which follows the fixed, but unknown probability distribution $P(\mathbf{y}|\mathbf{x})$. The pair of input and output is denoted as $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ in the following.

The learning problem aims to select that function from a set of functions $g(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, which approximates the dependency between input and output in an optimal manner [12]. The quality of this approximation is measured by a loss function

$$L(\mathbf{y}, g(\mathbf{x}, \boldsymbol{\theta})) = Q(\mathrm{z}, \boldsymbol{\theta}), \tag{17}$$

which expresses the discrepancy between the actual output $\mathbf{y}$ and the estimated output $g(\mathbf{x}, \boldsymbol{\theta})$. In case of least-squares regression, this loss function is given by

$$L(\mathbf{y}, g(\mathbf{x}, \boldsymbol{\theta})) = (\mathbf{y} - g(\mathbf{x}, \boldsymbol{\theta}))^2. \tag{18}$$

The expected value of the loss, the risk functional

$$R(\boldsymbol{\theta}) = \int \tilde{Q}(\mathbf{z}, \boldsymbol{\theta})\, dP(\mathbf{z}), \tag{19}$$

is minimized by the optimal function $\tilde{g}(\mathbf{x}, \boldsymbol{\theta}_0)$.

However, as $dP(\mathbf{z})$ is unknown, the risk functional (19) cannot be computed, and the minimization of the true risk is replaced by the minimization of the empirical risk

$$R_{emp}(\boldsymbol{\theta}) = \frac{1}{l} \sum_{i=1}^{l} Q(\mathbf{z}_i, \boldsymbol{\theta}), \tag{20}$$

which is based on the iid training data $\mathbf{z}$ as a finite realization of $dP(\mathbf{z})$. The approximation of the function $g(\mathbf{x}, \boldsymbol{\theta}_0)$, which minimizes the true risk, by means of the

function $g(\mathbf{x}, \boldsymbol{\theta}_e)$, which minimizes the empirical risk, is called empirical risk minimization (ERM). Inserting equation (18) in (20) results in the classical least-squares regression [38].

To improve the estimation with increasing sample size, the ERM has to be consistent: The expected value for the empirical risk $R(\boldsymbol{\theta})$ as well as the empirical risk $R_{emp}(\boldsymbol{\theta})$ itself have to converge to the minimal possible value of the risk $R(\boldsymbol{\theta}_0)$ [12]:

$$R(\boldsymbol{\theta}) \to R(\boldsymbol{\theta}_0) \tag{21}$$

$$R_{emp}(\boldsymbol{\theta}) \to R(\boldsymbol{\theta}_0). \tag{22}$$

In this context, Vapnik defines several necessary and sufficient conditions for consistency of the ERM (see [37] for further details). The most important one is the finiteness of the VC-dimension of the function class, as it is independent of the unknown probability density $p(\mathbf{z})$ and as it refers to a general property of the set of approximating functions [12].

4.2 VC-dimension of a Set of Functions

Originally, SLT was developed for pattern recognition problems, before the respective concepts were extended to the general learning problem [12]. The definition of the VC-dimension is based on indicator functions. However, the basic idea can be extended to real-valued functions.

A set of indicator functions **shatters** a sample of l points, if it can separate them in all 2^l possible ways [38]. According to this definition, a 2D-line can shatter a set of three points, as can be seen in figure 3.

The **VC-dimension** h of a set of indicator functions is defined by the maximum number of samples which can be shattered by the set of functions [12]. In case of the 2D-line, the VC-dimension equals three: While it can shatter a set consisting of three points, it cannot separate a set of four points in all possible ways (see figure 4 for a classification example which cannot be achieved by a 2D-line).

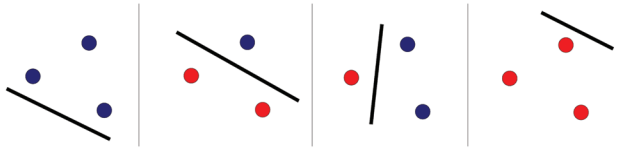

Figure 3: A 2D-line shatters a set of three points, as it can separate the points in $2^l = 2^3 = 8$ possible ways (only four ways are pictured, the other four ways can be achieved by switching the labels). The colours indicate the class affiliation.

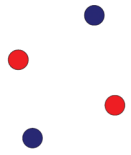

Figure 4: A 2D-line cannot shatter a set of four points.

The VC-dimension is an essential tool in SLT as it describes the complexity of a set of functions [35]. As stated above, the finiteness of VC-dimension is a necessary and sufficient condition for the consistency of ERM. Furthermore, the VC-dimension can be used in order to establish constructive distribution-independent bounds on the discrepancy between the empirical and the true risk. Those bounds form the basis for the structural risk minimization, which can be used as an alternative to model selection in order to choose the optimal model $g(\mathbf{x}, \boldsymbol{\theta})$ [12].

4.3 Structural Risk Minimization (SRM)

The considerations of consistency of the ERM refer to the asymptotic case and lead to the question how much the empirical risk and the true risk differ in case of finite sample sizes. The answer to this question is given by several bounds on the generalization (see [37] for further details). As the majority of these bounds is distribution-dependent and the distribution of laser scanner data is unknown, the focus of this paper lies on the distribution-independent bounds, which are based on the VC-dimension [12].

In case of regression, the true risk is bounded by the empirical risk with probability $1 - \eta$ as follows:

$$R(\boldsymbol{\theta}) \leq \frac{R_{emp}(\boldsymbol{\theta})}{(1 - c\sqrt{e})_+} \tag{23}$$

with

$$e = a_1 \frac{h\left(\ln\frac{a_2 l}{h} + 1\right) - \ln(\eta/4)}{l}. \tag{24}$$

In [12] the constant values are set to $c = 1$, $a_1 = 1$ and $a_2 = 1$, and the confidence level is chosen according to

$$\eta = \min\left(\frac{4}{\sqrt{l}}, 1\right) \tag{25}$$

so that the bound is a function of the sample size l and the VC-dimension h.

On the one hand, these bounds justify the usage of ERM if the ratio l/h is large. In that case, $e \approx 0$ holds and the empirical risk is close to the true risk [12]. On the other hand, they emphasize the need of a new principle if the ratio l/h is small. In that case it is not sufficient to minimize the empirical risk; rather a simultaneous minimization of the numerator and a maximization of the denominator in equation (23) are necessary. The former is solely determined by the empirical risk, whereas the latter is a function of the VC-dimension. Again an implementation of the principle of parsimony is necessary: The empirical risk has to be minimized and the complexity of the function class has to be controlled simultaneously [38]. The SRM is a formal mechanism for solving this problem. The starting point is the structuring of the set S of loss functions into nested subsets $S_k = Q(\mathbf{z}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}_k$, which can be ordered according to their finite VC-dimension h:

$$S_1 \subset S_2 \subset \ldots \subset S_k \subset \ldots \tag{26}$$

and

$$h_1 \leq h_2 \leq \ldots \leq h_k \leq \ldots. \tag{27}$$

SRM starts with the minimization of the empirical risk of each element S_k of the structure. Based on the empirical risks, the upper bound on the true risk is computed by equation (23). The model g_k leading to the smallest guaranteed risk is chosen to be optimal [36].

4.4 Estimating the VC-dimension

For the estimation of the upper bound (23) of the true risk the VC-dimension of the respective set of functions has to be determined. There exist only few classes of functions whose VC-dimension is known; the class of linear indicator functions, which was used in section 4.2 to demonstrate the basic idea, is one of the few exceptions. As an analytical computation of the VC-dimension is not possible either, a method to estimate the VC-dimension empirically is proposed in [35].

The estimation of the VC-dimension uses an analytical formula for the maximum deviation between the frequencies of errors, which are caused by the respective classifier on two randomly labelled data sets [12]. According to this formula, the maximum deviation $\xi(l)$ is bounded by a function Φ, which depends solely on the ratio $\tau = l/h$:

$$\xi(l) \leq \Phi(\tau), \tag{28}$$

with

$$\Phi(\tau) = \begin{cases} 1, & \text{if } \tau < 0.5 \\ a\dfrac{\ln(2\tau)+1}{\tau - k}\left(\sqrt{1+\dfrac{b(\tau-k)}{\ln(2\tau)+1}}+1\right) & \text{otherwise.} \end{cases} \tag{29}$$

The constants $a = 0.16$ and $b = 1.2$ have been empirically estimated by [35], and k is set to $k = 0.14928$ in order to achieve $\Phi(0.5) = 1$ [12].

A detailed description of the empirical determination of $\xi(l)$ is given in [12]: A randomly labelled set $\mathbf{Z}_{2l}$ of size $2l$ is split into two sets $\mathbf{Z}_l^1$ and $\mathbf{Z}_l^2$ of size l. The labels of the latter are flipped, and the fusion of $\mathbf{Z}_l^1$ and $\mathbf{Z}_{l,flipped}^2$ is used to estimate the parameters of a classifier belonging to the set of functions whose VC-dimension has to be determined. By means of this classifier, the training data is classified, and the number of misclassified points is used to determine the error rates on the subset $\mathbf{Z}_l^1$ (Error($\mathbf{Z}_l^1$)) and on the subset $\mathbf{Z}_l^2$ (Error($\mathbf{Z}_l^2$)) respectively, which directly lead to the wanted quantity:

$$\xi(l) = \max(|\text{Error}(\mathbf{Z}_l^1) - \text{Error}(\mathbf{Z}_l^2)|). \tag{30}$$

A schematic overview of the procedure can be seen in figure 5.

The procedure described above is performed for varying sample sizes l_1, l_2, ... l_k. In order to reduce the influence of the random sampling, the maximum deviation $\xi(l_i)$ is averaged over several repeated experiments for each sample size. As a result of this procedure, tuples of the kind $\{l_i, \xi(_i)\}$ are available for $i = 1, \ldots, k$. The VC-dimension h results as the integer value which produces the best fit between the empirically determined tuples and the formula (29).

5 Estimating the VC-dimension of B-spline Curves

The difficulty in estimating the VC-dimension of B-spline curves with the help of the procedure mentioned above is to use them as a classifier. The support vector machine (SVM) is a classifier, which is closely linked to the SLT, as it is an implementation of the SRM [12]. This connection encourages the investigation if the concept of the SVMs can be used to perform a B-spline classification in order to estimate the VC-dimension of B-splines.

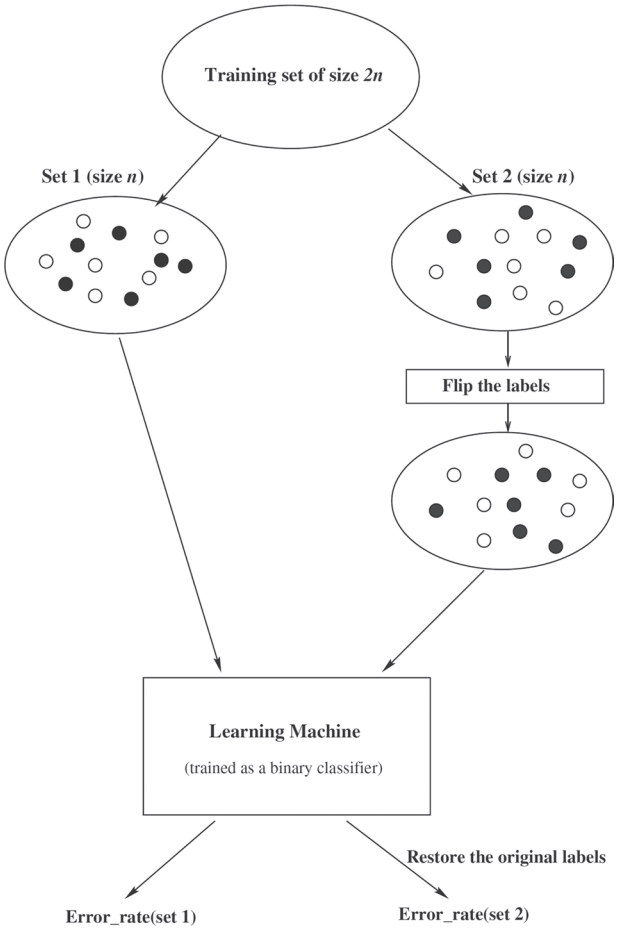

Figure 5: Measuring the maximum deviation between the error rates observed on two independent data sets [12].

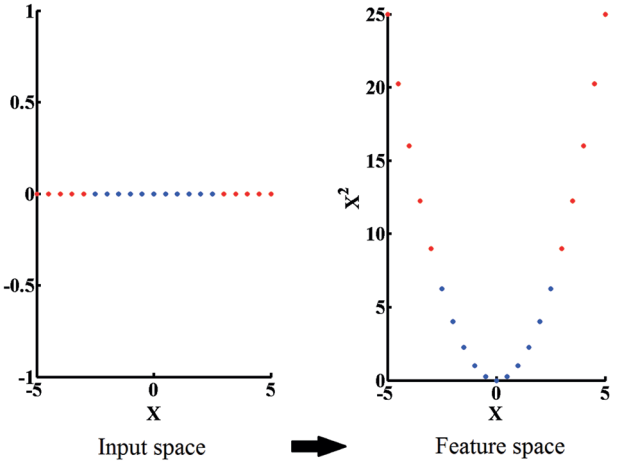

Figure 6: Mapping into a high-dimensional feature space: The input data (left) is not linearly separable. A transformation maps the input data into the feature space $\mathbf{x}_{FS} = (\mathbf{x}, \mathbf{x}^2)$, in which the data is linearly separable (right).

5.1 Basic Ideas of Support Vector Machines

Basically, SVMs aim to find **linear decision boundaries** that separate two classes according to their predefined label. However, the input data consisting of coordinates and labels is usually not linearly separable in a satisfactory manner. For this reason, the input data is **mapped into a high-dimensional feature space** by means of non-linear functions, and the classification is performed in this feature space. A linear decision boundary in the feature space is non-linear in the input space [12]. The principle of such a mapping can be seen in figure 6: The original data in the one dimensional input space is not linearly separable. A mapping into the two dimensional feature space $\mathbf{x}_{FS} = (\mathbf{x}, \mathbf{x}^2)$ solves the problem, because in this feature space it is possible to find a linear decision boundary which separates the data without errors.

Although this idea benefits from the advantages of linear optimization, it has a severe drawback as the dimensionality and consequently the computational complexity of the optimization problem can significantly increase with increasing number of features [38]. To circumvent this problem, the SVMs use the **kernel trick**. The use of the kernel trick requires that the input vector $\mathbf{x}$ appears solely in form of the inner product $(\mathbf{x} \cdot \mathbf{x}^T)$ in the optimization problem. This condition is satisfied by SVMs as shown in [37]. Instead of computing the inner product in the feature space directly, the inner product is replaced by a kernel function, which transforms the result of the inner product in the input space to the feature space:

$$K(\mathbf{x}, \mathbf{x}^T) = (\mathbf{x}_{FS} \cdot \mathbf{x}_{FS}^T). \tag{31}$$

As can be seen, this kernel function computes the inner product in the feature space indirectly via the input vector $\mathbf{x}$ so that the dimensionality of the optimization problem corresponds to the dimensionality of the input space rather than to the dimensionality of the feature space [12].

Basically, there exists a variety of kernel functions of the kind given in (31), which realize mappings into different feature spaces. All those functions have to be symmetric and have to fulfil Mercer's condition

$$\iint K(\mathbf{x}, \mathbf{x}^T)\, \varphi(\mathbf{x})\, \varphi(\mathbf{x}^T) d\mathbf{x} d\mathbf{x}^T > 0 \tag{32}$$

for all

$$\varphi \neq 0, \int \varphi^2(\mathbf{x})\, d\mathbf{x} < \infty. \tag{33}$$

Based on these principles, the basic idea of the SVM-classification can be seen in figure 7: Given are two classes, which have to be linearly separated with minimal

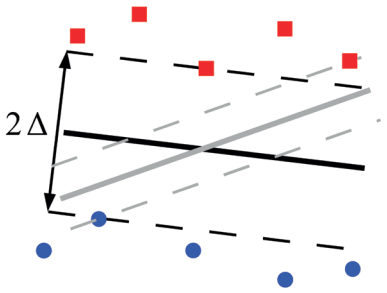

Figure 7: The basic principle of SVMs. The black line and the grey line separate both the two classes without error. The black one is chosen to be optimal as it maximizes the margin Δ.

classification error. For the sake of simplicity, figure 7 depicts the case of two linearly separable classes so that a classification without errors is possible. However, there are several linear decision boundaries, which lead to the same empirical risk $R_{emp} = 0$. According to the SRM, the decision boundary with minimal complexity is chosen to be the optimal. This complexity control is implemented by imposing a constraint on the optimization problem: The optimal decision boundary should maximize the distance to its closest point, the margin Δ. This constraint is justified because the margin directly influences the functions' complexity: The larger the margin, the smaller the subset of functions with this margin and the smaller the complexity of this subset [12].

However, the minimizing of the VC-dimension has a serious drawback with regard to the estimation of the VC-dimension: As only the subset of linear indicator functions with maximum margin is used to determine empirical values for $\xi(l)$, only the VC-dimension of this subset can be determined rather than that of the whole set of linear indicator functions of a given type.

For this reason, the classification by means of SVMs is not suitable to estimate the VC-dimension of a set of functions. However, the basic principles of the SVMs (linear optimization, mapping into a high-dimensional feature space, kernel trick) are used in the following to build a classifier which is suitable.

5.2 Requirements for an Appropriate Classifier

The previous remarks on the SVM classification point out the concepts for an appropriate classifier which is based on B-splines curves:

- The optimization problem to be solved should be linear.
- In contrast to the SVM classification, there is no need to restrict the optimization problem and, as a consequence, the VC-dimension by means of constraints.
- To determine non-linear decision boundaries, the optimization problem has to be transferred into a form which allows the application of the kernel trick.
- An appropriate kernel, which leads to decision boundaries of B-spline form, has to be established.

The developed classifier is based on the principle of the ridge regression, which minimizes the loss function:

$$L = \sum_{i=1}^{l} \epsilon_i^2 + \gamma \, || \boldsymbol{\theta}_{FS} ||^2 \tag{34}$$

$$= (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_{FS})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_{FS}) + \gamma \, || \boldsymbol{\theta}_{FS} ||^2 . \tag{35}$$

The classical least-squares problem (first term) is stabilized by the second term, whose strength of evidence is determined by the regularisation parameter γ. This form is also known as Tikhonov regularisation and is usually used to solve ill-posed problems [37]. In this context, the regularisation property is only indirectly relevant, as it allows the kernelization of the form (34).

Compared to the classical least-squares regression, the ridge regression has a severe drawback as it is a biased estimation. However, in this context the estimation's quality does not play a role as it is only used as a tool to determine the VC-dimension. For this reason, the ridge regression is by all means a suitable expedient.

The kernelization is based on the duality principle of convex optimization problems as in general only the dual form of an optimization problem can be kernelized rather than the primal form. The duality of convex optimization is a very wide topic. The following section gives only a short overview of how to transform a primal problem to its identical dual problem. For further details, please refer to [7], which was also the basis for the following section.

5.3 Duality in Convex Optimization

The general optimization problem is defined as the minimization of a loss function $L(\boldsymbol{\theta})$, subject to some inequality constraints $f_i^{in}(\boldsymbol{\theta}) \leq 0$, $i = 1, \ldots, n_i$ as well as some equality constraints $f_j^{eq}(\boldsymbol{\theta}) = 0$, $j = 1, \ldots, n_e$.

An optimization problem is convex if the inequality constraints as well as the loss function are convex, whereas the equality constraint is linear. As stated in section 5.2, in this paper only linear optimization in a high dimensional feature space as a special case of convex optimization is used.

In case of convex optimization, the original problem – the primal optimization problem – can be transformed

into an alternative form, the dual form. Starting point is the constraint loss function $L^*(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\nu})$, which is a function of the primal optimization variable $\boldsymbol{\theta}$ and of the Lagrange multipliers $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, which are denoted as dual variables:

$$L^*(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\nu}) = L(\boldsymbol{\theta}) + \sum_{i=1}^{n_i} \mu_i f_i^{in}(\boldsymbol{\theta}) + \sum_{j=1}^{n_e} \nu_j f_j^{eq}(\boldsymbol{\theta}). \quad (36)$$

The dual form of the constraint loss function (36) is a function of solely the dual variables and is defined as the infimum of the primal form over $\boldsymbol{\theta}$:

$$L_d^*(\boldsymbol{\mu}, \boldsymbol{\nu}) = \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L^*(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\nu}). \quad (37)$$

Under certain conditions **strong duality** holds; that means that the solution for the dual problem $\hat{d}$ is identical to that of the primal problem $\hat{p}$:

$$\hat{d} = \hat{p}. \quad (38)$$

A precondition for strong duality of optimization problems is the fulfilment of the Slater condition (see [37] for further information).

In case of a convex optimization problem, for which strong duality holds, there exist sufficient conditions which have to be fulfilled by the optimal primal and dual parameters $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\nu}}$.

$$f_i^{in}(\hat{\boldsymbol{\theta}}) \le 0 \qquad \forall i = 1,\ldots,n_i \quad (39)$$

$$f_j^{eq}(\hat{\boldsymbol{\theta}}) = 0 \qquad \forall j = 1,\ldots,n_e \quad (40)$$

$$\hat{\mu}_i \ge 0 \qquad \forall i = 1,\ldots,n_i \quad (41)$$

$$\hat{\mu}_i f_i^{in}(\hat{\boldsymbol{\theta}}) = 0 \qquad \forall i = 1,\ldots,n_i \quad (42)$$

$$0 = \nabla L^*(\hat{\boldsymbol{\theta}}) + \sum_{i=1}^{n_i} \hat{\mu}_i \nabla f_i^{in}(\boldsymbol{\theta}) + \sum_{j=1}^{n_e} \hat{\nu}_j \nabla f_j^{eq}(\boldsymbol{\theta}) \quad (43)$$

with ∇ being the vector differential operator.

These conditions are known as the Karush-Kuhn-Tucker(KKT)-conditions and are an extension of the Lagrange multipliers for optimization problems with equality constraints.

5.4 Ridge Classification with B-spline Kernel

The primal solution of the ridge regression

$$\theta_p = (\mathbf{X}'\mathbf{X} + \gamma\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \quad (44)$$

is obtained by determining the derivations of (34) with respect to the unknown parameters $\boldsymbol{\theta}$ and by equating them to zero. It is obvious that in the primal solution the input vector $\mathbf{x}$ is not contained solely in form of the inner product, but also separated in the factor $\mathbf{X}^T\mathbf{y}$. As a consequence, this form cannot be kernelized.

The following derivation of the kernelized ridge regression is carried out analogously to the derivation of the kernelized SVMs in [37]. At first the optimization problem (34) is reformulated: Minimize

$$\epsilon^T\epsilon \quad (45)$$

subject to the equality constraint

$$\epsilon = \mathrm{y} - \mathbf{X}\boldsymbol{\theta} \quad (46)$$

and the inequality constraint

$$\boldsymbol{\theta}^T\boldsymbol{\theta} \le C, \quad (47)$$

which results from the regularization term in equation (34). The primal Lagrangian of this optimization problem in matrix notation is

$$L_p = \epsilon^T\epsilon + \boldsymbol{\nu}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta} - \epsilon) + \mu(\boldsymbol{\theta}^T\boldsymbol{\theta} - C). \quad (48)$$

Taking the derivatives of the primal loss function with respect to the primal variables ϵ and $\boldsymbol{\theta}$ and equating them to 0 leads to

$$\epsilon = \frac{1}{2}\boldsymbol{\nu} \quad (49)$$

$$\boldsymbol{\theta} = \frac{1}{2\lambda}\mathrm{X}^T\boldsymbol{\nu}. \quad (50)$$

Inserting these relations into equation (48) and substituting $\alpha = \frac{1}{2\lambda}\boldsymbol{\nu}$ leads to the dual loss function

$$L_d = -\lambda^2\alpha^T\alpha + 2\lambda\alpha^T\mathbf{y} - \lambda\alpha^T\mathbf{X}\mathbf{X}^T\alpha - \lambda C. \quad (51)$$

Computing the derivations with respect to the dual variables and equating them to zero leads to the dual solution of the ridge regression

$$\alpha = [(\mathbf{X}\mathbf{X}^T) + \lambda\mathbf{I}]^{-1}\mathrm{y}. \quad (52)$$

When comparing equations (44) and (52), it becomes evident that in the dual problem – unlike the primal one – the input vector $\mathbf{x}$ appears solely in form of the inner

product. As a consequence, a kernelization by means of equation (31) is possible:

$$\alpha = [K(\mathbf{x}, \mathbf{x}^T) + \lambda \mathbf{I}]^{-1} \mathbf{y}. \tag{53}$$

For a linear ridge classification the output $\mathbf{y}$ of the system is given by the class labels $\mathbf{L}_c = \{-1, 1\}$, whereas the design matrix $\mathbf{X}$ contains the d-dimensional Cartesian coordinates of the l points to be classified:

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^d & 1 \\ x_2^1 & x_2^2 & \dots & x_2^d & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_l^1 & x_l^2 & \dots & x_l^d & 1 \end{bmatrix}. \tag{54}$$

Thus, the points are classified by the sign of their distance to the linear decision boundary.

The last step to establish an appropriate B-spline-curve-classifier is to define a kernel function. The procedure is the same as shown in [39] for the Spline-Kernel. Starting point is the functional form of a B-spline curve given in equation (1). This function is linear in the $n+1$-dimensional feature space spanned by the B-spline basis functions

$$N_{0,p}(u), N_{1,p}(u), \dots, N_{n,p}(u).$$

The inner product in this feature space and as a consequence the kernel function for B-spline curves is given by

$$K(u_s, u_t) = \sum_{i=0}^{n} N_{i,p}(u_s) \cdot N_{i,p}(u_t). \tag{55}$$

This kernel function completes the B-spline classifier. The main steps to obtain the classifier are summarized in figure 8.

The kernel's structure leads to a coordinate-wise classification rather than to a point-wise one. This coordinate-wise classification is implied by the coordinate-wise description of B-splines (see equation 1), which is commonly used (see for example [30]). As a consequence, the resulting VC-dimension also refers to a single coordinate of a B-spline curve rather than to a B-spline curve as a whole. Despite this coordinate-wise classification, the resulting labels of a point's coordinates are always identical. This is caused by the B-spline kernel, which only depends on the curve parameter u rather than on the Cartesian coordinates. Because of the coordinate-wise definition a classification in 3D and, as a consequence, the determination of

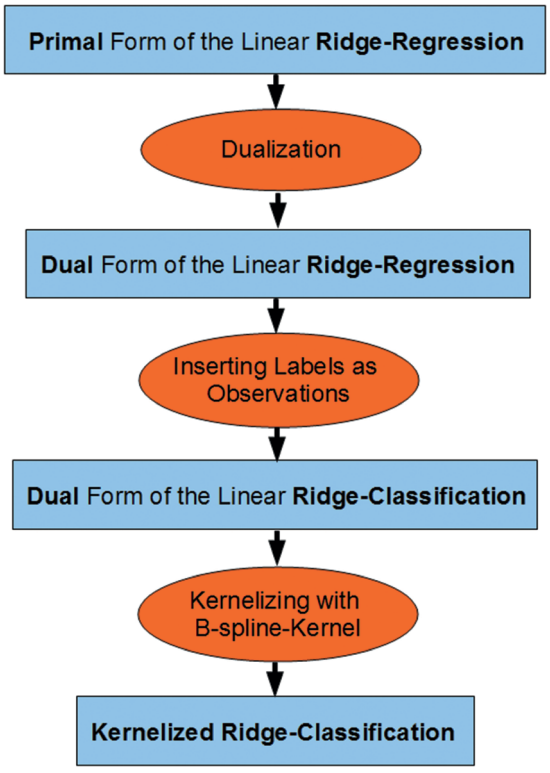

Figure 8: Main steps for obtaining a B-spline ridge classifier.

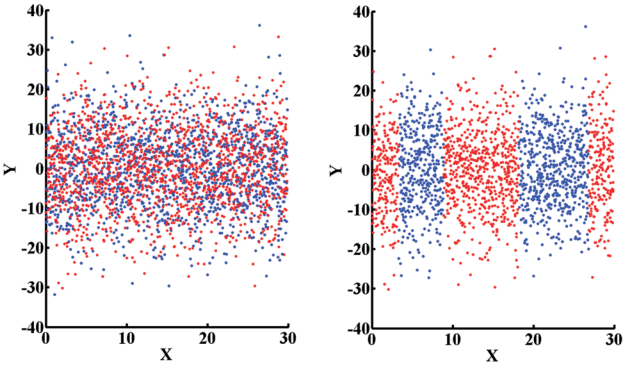

Figure 9: Ridge classification with B-spline kernel. A randomly labelled data set with $l = 2000$ (left) is classified by means of a B-spline curve with six control points (right).

the VC-dimension of a 3D-B-spline curves is also possible without defining an "above" or "below" a 3D-curve.

The practical use of the established classifier can be seen in figure 9. Input of the classification is a randomly labelled data set (left side). The result of the classification by means of a B-spline curve with six control points can be seen in figure 9 (right side). The points with a positive distance to the B-spline curve are allocated to the red class, whereas the distances of the points belonging to the blue class have a negative sign.

5.5 VC-dimension of B-splines

After having provided the basics for a classification based on B-spline curves, their VC-dimension can be determined according to the procedure discussed in section 4.4.

The general procedure is exemplarily demonstrated by means of a B-spline curve with six control points. Starting point is a point cloud which is labelled randomly (see figure 9 (left)). Twenty subsamples of sizes $l_1 = 100$, $l_2 = 200$, ..., $l_{20} = 2000$ are randomly drawn out of this point cloud, and each of the subsample is classified by means of the B-spline classifier presented above. In figure 9 (right side) such a classification is exemplarily shown for the sample size l_{20}. The classification results are used to determine values for $\xi(l_1)$, ..., $\xi(l_{20})$ according to equation (30). To reduce the influence of the random drawing, this procedure is repeated twenty times, and the results are averaged to $\bar{\xi}(l_1)$, ..., $\bar{\xi}(l_{20})$.

The empirical values are graphically presented in figure 10 (blue dots). The analytical function (29) is fitted through these points, and the integer value h, which leads to the best fit between the empirical values and the analytical function, is chosen to be the VC-dimension. The VC-dimension of the B-spline curve with six control points is estimated at $\hat{h} = 6$.

The regularization parameter γ, which has to be chosen a priori, determines the influence of the regularization term in the ridge regression. For an appropriate classifier, the constraints in the optimization problem should be as weak as possible (see also section 5.2). Ideally, γ should be equal 0 resulting in a classical least squares classification with no constraints. However, the dual classification problem becomes singular in this case. As a compromise, γ should be chosen to be as small as possible to achieve a regular classification problem. This approach is validated by estimating the VC-dimension of a two-dimensional line. The estimated value $\hat{h} = 3$ corresponds with the theoretical known value $h = 3$.

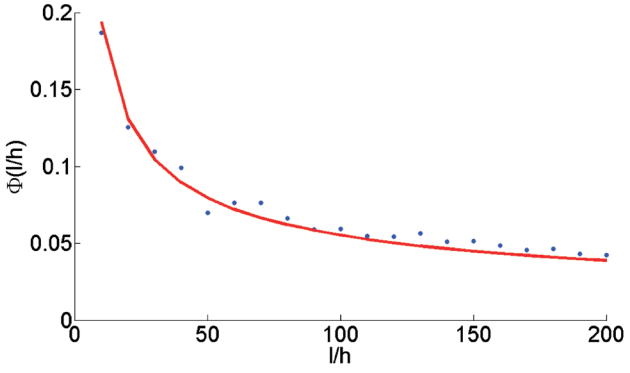

Figure 10: Estimating the VC-dimension of a B-spline curve with six control points. $h = 6$ produces the best fit between the analytical function $\Phi(l / h)$ (red curve, equation (29)) and the empirical values $\xi(l_i)$ (blue dots).

Table 1: VC-dimension of a B-spline curve with $n + 1$ control points of different degree p.

n + 1	4	5	6	7	8	9	10
p = 3	4	4	6	7	7	8	9
p = 4		5	6	7	7	8	9
p = 5			6	6	7	8	9
p = 6				6	7	8	9
p = 7					7	8	9
p = 8						8	9

In table 1 the estimated VC-dimensions of B-spline curves with different numbers of control points and varying degree are listed.

As might be expected, the curve's complexity increases if the number of control points increases, too. Interestingly, the degree of the B-spline curve does not seem to influence the VC-dimension. The few variations of the VC-dimension, which occur if the degree increases, can be explained by the estimation's uncertainty.

Generally, the estimated VC-dimension is always close to the number of estimated control points. As the estimation of the control points is treated as a linear estimation problem in this paper, this behaviour could be expected and can be seen as another validation.

6 Choosing the Optimal Number of Curve Control Points

The sections above provide the theoretical aspects for the selection of an optimal model. In the following the three introduced possibilities AIC, BIC and SRM are applied to determine the optimal number of B-spline curve control points. The applicability of the approaches to this concrete problem is analyzed, the results are compared, and the methods are evaluated.

The theoretical basics of the three procedures suggest to investigate their behaviour under various aspects:

- The **correctness** of the methods: How much does the determined number of control points differ from the actual one? These investigations require the knowledge about the true functional form of the phenomenon.
- The **repeatability** of a method [9]: If the computations are based on different realizations of the same

phenomenon, stable results are characteristic of a good procedure. How often does the method provide the correct/the most frequent result?
- The methods' **precision** is closely linked to the repeatability. It reflects how much the results vary.
- AIC and BIC are based on asymptotic assumptions, whereas SRM explicitly takes the sample size into account. For this reason, the **influence of the sample size** on the methods' performances will be investigated.

Figure 11: Simulated Bézier curves which are used to examine the applicability of AIC, BIC and SRM to the determination of the optimal number of control points.

6.1 Bézier Curves

Although the final objective of the presented project is the determination of the optimal number of surface control points, this paper focusses on B-spline curves, in order to achieve an initial impression of the methods' performances.

The mathematical basics in section 2 reveal that the curve estimation requires the a priori determination of appropriate curve parameters and of an appropriate knot vector. As both determinations are based on the noisy data set, they cause an additional degree of uncertainty which influences the estimation result and, as a consequence, the determination of the optimal number of control points. In order to keep this additional uncertainty as low as possible, the initial investigations are based on Bézier curves, which are a special case of B-spline curves with $n = p$. The knot vector of a Bézier curve reduces to $U = [0, ..., 0, 1, ... 1]$ and does not have to be determined a priori any longer [30].

As the actual number of control points has to be known in order to evaluate the criteria's correctness, the following computations are based on simulated data sets: The basis for the investigations are two Bézier curves with different number of control points ($n_1 + 1 = 6$ and $n_2 + 1 = 12$) – the phenomena to be investigated. Each phenomenon is realized by means of concrete data points. The respective sample size varies between $l_1 = 100$ and $l_6 = 5000$. Each of these curves is superimposed by white noise with a standard deviation of $\sigma = 2$ mm. In order to investigate the criteria's repeatability, the noise generation is repeated five times for each sample size, resulting in 30 data sets for each curve forming the basis for the following investigations. One realization of each curve is exemplarily shown in figure 11.

The criteria presented above can be directly applied to these data sets: The best fitting Bézier curve is estimated for varying numbers of control points, and the resulting residuals are used to compute the empirical risk or the log-likelihood respectively. The penalization results in a function of the VC-dimension in case of SRM and in a function of the number of estimated parameters in case of AIC and BIC. In order to keep the analysis consistent and comparable, all computations are implemented coordinate-wise, caused by the coordinate-wise definition of the VC-dimension (see section 5). The number of estimated parameters, which is penalized by AIC and BIC, is, therefore, given by

$$K = (n + 1) + 1. \tag{56}$$

This procedure results in an optimal number of control points in x-direction $\hat{n}_x + 1$ as well as in an optimal number of control points in y-direction $\hat{n}_y + 1$ in case of the simulated 2D Bézier curves. The overall optimal number of control points $\hat{n} + 1$ is chosen to be the maximum of these two numbers:

$$\hat{n} = \max(\hat{n}_x, \hat{n}_y). \tag{57}$$

The complete results of these computations can be found in table 5 in the appendix 9. These results are summarized by some statistical quantities:
- The correctness of the computations is evaluated by means of the arithmetic mean. In order to enable a better comparison of the three methods, the values are not rounded to the next integer value. As the influence of outliers on the arithmetic mean which is based on 30 samples is quite large, the median is computed, too. However, a mean or a median close to the correct value does not necessarily indicate a good performance as values far above and below the correct value can add up to the correct mean. For this reason, the correctness has always been regarded in combination with the following two quantities.
- The frequency of the mode f_m as well as the frequency of the correct value f_c can be seen as a measure for the repeatability.
- The methods' precision is evaluated by means of the standard deviation as well as by the range of the results, given in terms of the minimum and the maximum value.

The statistical quantities are listed in table 2. In case of the curve C_1 with $n_1+1=6$ control points, SRM evidently outperforms the model selection criteria AIC and BIC.

In none data set AIC chooses the actual number of control points to be optimal, rather AIC shows a distinct tendency to overestimate the number of control points. This bias is indicated by the arithmetic mean, the median as well as the mode. A coherence between the tendency to overestimate and the sample size is not recognizable (see table 5). The bad performance of AIC is emphasized by the criterion's precision. The chosen number of control points varies between $n_1+1=7$ and $n_1+1=12$ even over data sets having the same sample size, resulting in a comparatively large standard deviation of 1, 8. The bad precision of AIC is also reflected by the stability in terms of the mode's frequency: Only in nine cases out of a total of 30 the mode is chosen to be optimal.

The correctness of BIC is significantly better than the AIC's one: Although the correct number of control points is detected only in three cases, the optimal number of control points is close to the actual one in most cases. The median as well as the mode miss the actual number of control points just by one. As this missing of the actual value is quite stable (17 cases out of the total of 30), the repeatability is also improved compared to AIC. Although there is a certain variability even over data sets with the same sample size, the range of this variation is smaller than that of the AIC: With the exception of four outliers with $n_1+1\leq 10$, BIC detects 6–9 control points to be optimal. However, these outliers influence the computation of the mean as well as the computation of the standard deviation, which is only slightly better than that of AIC.

Table 5 in the appendix shows that for the smaller sample sizes the results of SRM and BIC are almost identical. However, with growing sample size SRM performs significantly better than BIC does as it settles at the correct number of control points while showing minimal variability over the different data sets. This behaviour is also reflected by the values in table 2. Median and mode are identical to the actual number of control points. The mean is only slightly too large as the correct number of control points is chosen in 16 cases and as in the remaining 14 cases the estimated number of control points is close to the actual one. As a consequence, the standard deviation is significantly smaller than the ones of AIC and BIC.

Table 2: Statistical quantities of the experiments which have been performed in order to determine the optimal number of control points of two different Bézier curves (left column: $n_1+1=6$, right column: $n_2+1=12$). The statistical quantities are based on 30 data sets.

	AIC	BIC	SRM	AIC	BIC	SRM
mean	10,2	7,7	6,6	13,7	12,7	12,1
median	11	7	6	13	12	12
mode	12	7	6	12	12	11
f_m	9	17	16	8	12	14
f_c	0	3	16	8	12	9
min	7	6	6	11	11	11
max	12	12	9	16	16	16
std	1,8	1,5	0,8	1,8	1,4	1,5

In case of the curve C_2 with $n_2+1=12$ control points, the results are slightly different and by far not as distinct. Although the correctness of AIC is significantly better than in case of C_1, AIC still performs worse than BIC and SRM, in particular with regard to the precision and the repeatability. The results of BIC and SRM are quite balanced with BIC having a slightly better precision, while SRM performs better concerning the repeatability. With regard to the correctness, BIC scores better than SRM. However, viewed over all samples, SRM performs better as it results in a mean which is almost identical to the actual number of control points.

Comparing the results of C_1 and C_2, the deterioration of the SRM's precision is particularly apparent. This instability can be explained by the curve parameters' u uncertainty, which was already mentioned: In case of the second curve, the a priori determined curve parameters are comparatively unstable. The parameter's instability directly affects the repeatability of SRM.

6.2 B-spline Curves

The observations above suggest that the results will become worse if B-splines rather than Bézier curves are used and the knot vector has also be determined a priori. For this reason, the investigations above are repeated with two cubic B-spline curves, both having the same control points, but two different knot vectors U_1 and U_2. The two curves are exemplarily shown in figure 12.

The results for these two curves differ completely as can be seen in the tables 3 and 6 in the appendix. In case of the curve with knot vector U_1, the comparison of the three methods results in similar conclusions as it does for the Bézier curve: AIC shows a distinct tendency to overestimate the number of control points and chooses the correct number of control points only in few cases. Although the value of f_m prompts a good repeatability, it has to be noted that this value is not meaningful as the mode is identical to the maximal number of control points which has been investigated. It can be expected that the

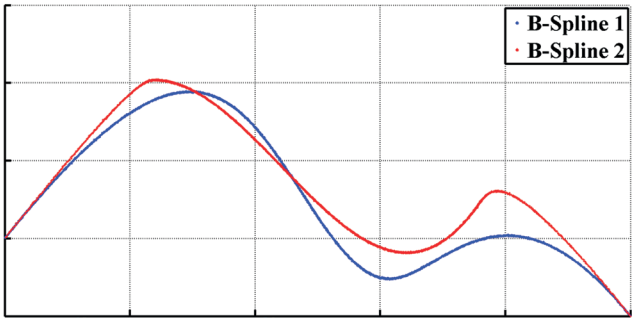

Figure 12: Simulated B-spline curves which are used to examine the applicability of AIC, BIC and SRM to the determination of the optimal number of control points.

Table 3: Statistical quantities of the experiments which have been performed in order to determine the optimal number of control points of two different B-spline curves (left column: $n_1+1=6$, knot vector U_1, right column: $n_2+1=6$, knot vector U_2). The statistical quantities are based on 30 data sets.

	AIC	BIC	SRM	AIC	BIC	SRM
mean	11,1	7,8	6	11,6	11,6	11,6
median	12	6	6	12	12	12
mode	12	6	6	12	12	12
f_m	24	20	30	19	19	18
f_c	3	20	30	0	0	0
min	6	6	6	11	11	11
max	12	12	6	12	12	12
std	2,0	2,7	0	0,5	0,5	0,5

value of f_m is going to decrease if the maximal number of control points increases.

BIC performs better than AIC and chooses the correct number of control points in 60 % of the data sets, resulting in a good correctness and repeatability. However, the huge standard deviation results in a bad precision. It is caused by the fact that BIC chooses either the correct number of control points or the maximal one and seems to provide an "all-or-nothing"-solution.

SRM achieves the best possible result: It chooses the correct number of control points for all 30 data sets.

Unexpectedly, for all three methods the entirety of the results is better than that one that is achieved for the Bézier curve: It seems that the additional a priori determination of the knot vector balances the uncertainty introduced by means of the parameters rather than compounding it.

This observation does not hold for the second curve with knot vector U_2: For this curve all three methods perform equally bad, choosing either 11 or 12 control points to be optimal.

These different results can be explained by means of the quality of the a priori determined knot vector $\hat{U}$: For the first curve the determined knot vector is similar to the actual one

$$\hat{U}_1 \approx U_1, \tag{58}$$

whereas the determined knot vector differs substantially from the actual one in case of the second curve:

$$\hat{U}_2 \neq U_2. \tag{59}$$

This discrepancy of the knot vectors is compensated by additional control points, resulting in the overestimated number of control points. The bad performance of the three methods in case of the second curve is thus not caused by the methods themselves, but by the poor determination of the knot vector. This statement can be confirmed by concrete values of $\hat{\sigma}$ in equation (9) which can be seen as a measure of the approximation quality. According to this quality measure, the approximation of the first data set by means of a B-spline curve with $n_1+1=6$ control points is very satisfying as $\sigma_{1,6}=0.0014$ is smaller than the simulated noise of $\sigma_{sim}=0.002$. If a B-spline curve with $n_2+1=6$ control points is used to model the second data set, this value increases significantly to $\sigma_{2,6}=0.0093$. Only the addition of control points leads to an improved approximation as indicated by $\sigma_{2,12}=0.0017$.

Obviously, the knot determination according to [30] is not sufficient in case of the second curve.

6.3 Point-wise vs. Coordinate-wise Considerations

For reasons declared above, all investigations in this paper are based on coordinate-wise computations. However, AIC and BIC can easily be extended to a point-wise view, which leads to significantly improved results with regard to correctness, repeatability and precision (see table 4). A similar improvement can be expected if SRM is extended to a point-wise computation. However, this extension is not straightforward as the B-spline kernel used in this paper implements the common coordinate-wise description of B-splines. In order to extend SRM to a point-wise computation, a point-wise description of B-splines has to be implemented. Based on this description, a kernel, which computes the VC-dimension of a B-spline as a whole rather than the VC-dimension of single coordinate directions, has to be built.

Table 4: Optimal number of control points in case of coordinate-wise (c) and point-wise (p) computation of AIC and BIC for the Bézier curve with $n+1=6$ control points.

l	AIC (c)	AIC (p)	BIC (c)	BIC (p)
100	12	6	8	6
100	11	6	7	6
100	10	6	7	6
100	8	7	8	6
100	8	6	8	6
5000	11	7	11	6
5000	7	6	7	6
5000	8	6	7	6
5000	11	10	11	6
5000	7	6	7	6

6.4 Final Evaluation

All in all it can be stated that

- AIC has been proven to be unsuitable to choose the optimal number of curve control points as it shows a distinct tendency to overestimate and as the respective results are much too unstable.
- BIC and SRM both deliver satisfactory results with SRM performing slightly better than BIC.
- However, it has to be taken into account that the estimation of the VC-dimension and, as a consequence, the SRM demands much more computational efforts than the computation of the BIC-values does.
- On the contrary, BIC is per definition only applicable to linear problems. The sole estimation of B-spline control points is a linear estimation problem, whereas it becomes nonlinear if the determination of the knot vector is included into the adjustment.
- We recommend SRM as it is more general than BIC and as it is also suitable for nonlinear cases. For linear problems, BIC is an appropriate alternative, offering the advantage of computational parsimony.
- In case of both methods, it is essential that the curve parameters and the knot vector are determined in an appropriate way.
- As a point-wise computation is much more stable than a coordinate-wise one, the former should be preferred to the latter if possible.

7 Conclusion

The extension of point-based analysis approaches to areal ones is a current research topic in engineering geodesy. The presented project aims to develop a spatio-temporal collocation whereby deformations can be described at arbitrary times and at arbitrary places of an object.

In order to model the deterministic trend of the deformation process, B-spline surfaces have emerged to be a powerful tool. One of the main challenges regarding the estimation of such freeform surfaces is to determine the optimal number of control points in a scientifically justified way, balancing the surface's complexity and the quality of the approximation.

In this paper the applicability of the model selection criteria AIC and BIC to the selection of the optimal number of B-spline curve control points has been investigated. Both criteria maintain a balance in the quality of the approximation and the number of estimated parameters. As these criteria are limited to linear models and as the number of parameters is not always a suitable complexity measure, SRM has been investigated as an alternative. In order to implement the latter, a ridge classification with B-spline kernel was developed to determine the VC-dimension of B-splines, which is the classical complexity measure of the statistical learning theory and which forms the basis for SRM.

AIC, BIC and SRM have been used to determine the optimal number of several Bézier and B-spline curves. The results have been evaluated with regard to their correctness, their repeatability and their precision. For all data sets, AIC performs unsatisfactorily, whereas BIC and SRM have been proven to be suitable in order to determine the optimal number of B-spline curve control points.

8 Outlook

For reasons of simplicity and intelligibility, this paper deals with the choice of the optimal number of B-spline curve control points. As the presented project deals with B-spline surfaces rather than curves, further studies concerning the B-spline surfaces are inevitable. However, these studies would be beyond the scope of the paper so that they are going to be presented separately. In the follow-up paper, the methods will also be evaluated by means of real laser scanner data rather than by simulated data sets.

Acknowledgments: The presented paper shows results developed during the research project "Integrierte raumzeitliche Modellierung unter Nutzung korrelierter Messgrößen zur Ableitung von Aufnahmekonfiguationen und Beschreibung von Deformationsvorgängen" (IMKAD) (1706-N29), which is funded by the Austrian Science Fund (FWF).

References

[1] Ken Aho, DeWayne Derryberry and Teri Peterson, Model selection for ecologists: the worldviews of AIC and BIC, *Ecology* **95** (2014), pp. 631–636.

[2] Hirotogu Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle*, Selected Papers of Hirotugu Akaike (Emanuel Parzen, Kunio Tanabe and Genshiro Kitagawa, eds.), Springer Series in Statistics, Springer New York, 1998, pp. 199–213.

[3] Mario Alba, Luigi Fregonese, Federico Prandi, Marco Scaioni and Paolo Valgoi, Structural Monitoring of a Large Dam by Terrestrial Laserscanning, *The ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden, Deutschland* (2006).

[4] Oliver Baur, Michael Kuhn and Will E. Featherstone, *GRACE-Derived Linear and Non-linear Secular Mass Variations Over Greenland*, VII Hotine-Marussi Symposium on Mathematical Geodesy (Nico Sneeuw, Pavel Novák, Mattia Crespi and Fernando Sansò, eds.), International Association of Geodesy Symposia 137, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 381–386.

[5] Carl de Boor, On calculating with B-splines, *Journal of Approximation Theory* **6** (1972), pp. 50–62.

[6] George E. P. Box, Science and Statistics, *Journal of the American Statistical Association* **71** (1976), pp. 791–799.

[7] Stephen P. Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, UK and New York, 2004.

[8] Johannes Bureick, Hamza Alkhatib and Ingo Neumann, Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis, *Journal of Applied Geodesy* **10** (2016), pp. 27–35.

[9] Kenneth P. Burnham and David R. Anderson, *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed, Springer, New York, 2002.

[10] Kenneth P. Burnham and David R. Anderson, Multimodel Inference: Understanding AIC and BIC in Model Selection, *Sociological Methods & Research* **33** (2004), pp. 261–304.

[11] Joseph Cavanaugh and Andrew Neath, Generalizing the derivation of the Schwarz information criterion, *Communications in Statistics - Theory and Methods* **28** (1999), pp. 49–66.

[12] Vladimir S. Cherkassky and Filip Mulier, *Learning from data: Concepts, theory, and methods*, 2nd ed, IEEE Press and Wiley-Interscience, Hoboken, N. J., 2007.

[13] Gerda Claeskens and Nils Lid Hjort, *Model selection and model averaging*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, Cambridge and New York, 2008.

[14] Maurice G. Cox, The Numerical Evaluation of B-Splines, *IMA Journal of Applied Mathematics* **10** (1972), pp. 134–149.

[15] Jan Dupuis, Christoph Holst and Heiner Kuhlmann, Laser Scanning Based Growth Analysis of Plants as a New Challenge for Deformation Monitoring, *Journal of Applied Geodesy* **10** (2016), pp. 37–44.

[16] Gerald E. Farin, *Curves and surfaces for CAGD: A practical guide*, 5th ed, The Morgan Kaufmann series in computer graphics and geometric modeling, Morgan Kaufmann and Academic Press, San Francisco, CA and London, 2002.

[17] Corinna Harmening and Hans Neuner, A constraint-based parameterization technique for B-spline surfaces, *Journal of Applied Geodesy* **9** (2015), pp. 143–161.

[18] Otto Heunecke, Heiner Kuhlmann, Walter Welsch, Andreas Eichhorn and Hans Neuner, *Handbuch Ingenieurgeodäsie: Auswertung geodätischer Überwachungsmessungen*, 2nd ed, Wichmann, H, Heidelberg, Neckar, 2008.

[19] Christoph Holst and Heiner Kuhlmann, Mathematische Modelle zur flächenhaften Approximation punktweise gemessener Bodensenkungen auf Basis von Präzisionsnivellements, in: *Geomonitoring* (Wolfgang Busch, Wolfgang Niemeier and Ingo Neumann, eds.), pp. 189–206, 2015.

[20] Christoph Holst and Heiner Kuhlmann, Challenges and Present Fields of Action at Laser Scanner Based Deformation Analyses, *Journal of Applied Geodesy* **10** (2016), pp. 17–25.

[21] Clifford M. Hurvich and Chih-Ling Tsai, Bias of the corrected AIC criterion for underfitted regression and time series models, *Biometrika* **78** (1991), pp. 499–509.

[22] Stephanie Kauker and Volker Schwieger, Approach for a Synthetic Covariance Matrix for Terrestrial Laser Scanner, *Proceedings of the 2nd International Workshop: Integration of point- and area-wise geodetic monitoring for structures and natural objects, March 23–24, 2015, Stuttgart* (2015).

[23] Karl-Rudolf Koch, *Parameterschätzung und Hypothesentests in linearen Modellen*, 3rd ed, Dümmlerbuch 7892, Dümmler, Bonn, 1997.

[24] Karl-Rudolf Koch, NURBS surface with changing shape, *Allgemeine Vermessungsnachrichten* (2010), pp. 83–89.

[25] Jouni Kuha, AIC and BIC: Comparisons of Assumptions and Performance, *Sociological Methods & Research* **33** (2004), pp. 188–229.

[26] Allan D. R. McQuarrie and Chih-Ling Tsai, *Regression and time series model selection*, World Scientific, Singapore and River Edge, N. J., 1998.

[27] Hans Neuner, Model selection for system identification by means of artificial neural networks, *Journal of Applied Geodesy* **6** (2012), pp. 117–124.

[28] Hans Neuner, Claudius Schmitt and Ingo Neumann, Zur Bestimmung der verkehrsseitig verursachten Dehnung an einem Brückentragwerk mittels terrestrischem Laserscanning, *Ingenieurvermessung 14: Beiträge zum 17. Internationalen Ingenieursvermessungskurs Zürich, 2014* (2013).

[29] Johannes Ohlmann-Lauber and Thomas Schäfer, Ansätze zur Ableitung von Deformationen aus TLS-Daten, *Terrestrisches Laserscanning – TLS 2011 mit TLS-Challenge* (2011), pp. 147–158.

[30] Les A. Piegl and Wayne Tiller, *The NURBS book*, 2nd ed, Monographs in visual communications, Springer, Berlin and New York, 1997.

[31] Michael Schmidt, Denise Dettmering and Florian Seitz, *Using B-Spline Expansions for Ionosphere Modeling*, Handbook of Geomathematics (Willi Freeden, M. Zuhair Nashed and Thomas Sonar, eds.), Springer, Berlin, Heidelberg, 2014, pp. 1–40.

[32] Claudius Schmitt and Hans Neuner, Knot estimation on B-Spline curves, *Österreichische Zeitschrift für Vermessung und Geoinformation (VGI)* **103** (2015), pp. 188–197.

[33] Gideon Schwarz, Estimating the Dimension of a Model, *The Annals of Statistics* **6** (1978), pp. 461–464.

[34] Rinske van Gosliga, Roderik Lindenbergh and Norbert Pfeifer, Deformation Analysis of a bored tunnel by means of Terrestrial Laserscaning, *Proceedings on ISPRS Commission V Symposium, Dresden* (2006).

[35] Vladimir Vapnik, Esther Levin and Yann Le Cun, Measuring the VC-Dimension of a Learning Machine, *Neural Computation* **6** (1994), pp. 851–876.
[36] Vladimir N. Vapnik, Principles of Risk Minimization for Learning Theory, *Advances in Neural Information Processing Systems* **4** (1992).
[37] Vladimir N. Vapnik, *Statistical learning theory*, Adaptive and learning systems for signal processing, communications, and control, Wiley, New York, 1998.
[38] Vladimir. N. Vapnik, An overview of statistical learning theory, *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* **10** (1999), pp. 988–999.
[39] Vladimir N. Vapnik, Steven E. Golowich and Alex Smola, Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, *Advances in Neural Information Processing Systems* **9** (1996), pp. 281–287.
[40] Larry Wasserman, Bayesian Model Selection and Model Averaging, *Journal of mathematical psychology* **44** (2000), pp. 92–107.

Appendix

Complete Results of the Determination of the Optimal Number of Control Points

In the tables 5 and 6 the complete results of the estimation of the optimal number of control points for the Bézier and the B-spline curve are listed.

Table 5: Optimal number of control points of two simulated Bézier curves with $n_1+1=6$ and $n_2+1=12$ according to AIC, BIC and SRM. Correct choices are highlighted in bold characters.

l	AIC	BIC	SRM	AIC	BIC	SRM
100	12	8	**6**	**12**	**12**	**12**
100	11	7	7	13	11	11
100	10	7	**6**	11	11	11
100	8	8	8	11	11	11
100	8	8	8	11	11	11
250	10	7	7	**12**	**12**	**12**
250	10	7	7	13	13	**12**
250	7	7	7	16	15	15
250	12	7	7	15	**12**	15
250	9	9	7	15	13	15
500	11	7	**6**	16	13	11
500	12	**6**	**6**	16	11	11
500	7	7	**6**	13	**12**	11
500	12	9	9	14	14	11
500	11	7	7	**12**	**12**	**12**
1000	11	7	7	**12**	**12**	**12**
1000	9	7	**6**	16	16	16
1000	12	10	**6**	**12**	**12**	11
1000	11	8	**6**	13	13	13
1000	12	7	**6**	**12**	**12**	**12**
2000	12	12	**6**	13	13	13
2000	12	7	7	16	**12**	**12**
2000	12	**6**	**6**	15	14	11
2000	11	7	**6**	16	**12**	11
2000	10	**6**	**6**	16	16	11
5000	11	11	**6**	16	14	11
5000	7	7	**6**	**12**	**12**	**12**
5000	8	7	**6**	15	15	15
5000	11	11	7	**12**	**12**	**12**
5000	7	7	7	14	14	11

Table 6: Optimal number of control points of two simulated B-spline curves with $n+1=6$ and two different knot vectors U_1 and U_2 according to AIC, BIC and SRM. Correct choices are highlighted in bold characters.

l	AIC	BIC	SRM	AIC	BIC	SRM
100	12	**6**	**6**	11	11	11
100	9	**6**	**6**	11	11	11
100	9	9	**6**	11	11	11
100	12	**6**	**6**	11	11	11
100	12	**6**	**6**	11	11	11
250	12	**6**	**6**	11	11	11
250	12	**6**	**6**	11	11	11
250	12	**6**	**6**	11	11	11
250	12	**6**	**6**	11	11	11
250	**6**	**6**	**6**	12	12	12
500	12	12	**6**	12	12	11
500	12	**6**	**6**	12	12	12
500	12	12	**6**	12	12	12
500	12	12	**6**	11	11	11
500	12	12	**6**	11	11	11
1000	12	**6**	**6**	12	12	12
1000	**6**	**6**	**6**	12	12	12
1000	12	**6**	**6**	12	12	12
1000	12	**6**	**6**	12	12	12
1000	12	**6**	**6**	12	12	12
2000	12	**6**	**6**	12	12	12
2000	12	12	**6**	12	12	12
2000	12	12	**6**	12	12	12
2000	**6**	**6**	**6**	12	12	12
2000	12	**6**	**6**	12	12	12
5000	9	12	**6**	12	12	12
5000	12	**6**	**6**	12	12	12
5000	12	12	**6**	12	12	12
5000	12	**6**	**6**	12	12	12
5000	12	9	**6**	12	12	12