Using model selection criteria to determine the optimal number of B-spline control points for areal deformation modelling

C. Harmening, H. Neuner
Department of Geodesy and Geoinformation, Research Group Engineering Geodesy,
TU Wien, Gusshausstr. 25-29, 1040 Vienna, Austria

Abstract. The increased use of areal measurement techniques in engineering geodesy requires the development of adequate areal analysis strategies. In this paper, an outline of a research project is presented which aims to develop a spatiotemporal continuous collocation in order to describe areal deformations.
The trend component of the collocation is modelled by estimated B-spline surfaces in this study. Among other form parameters B-spline surfaces are characterized by the number of estimated control points. Typically, the appropriate number of control points is set under consideration of parsimony by trial-and-error procedures. In this contribution the determination of the number of control points is regarded as a model selection problem. Two linear model selection criteria – the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) – are investigated: Although both criteria lead to a penalized maximum likelihood estimation, they are based on different principles: The AIC is an information-theoretic approach, which approximates the Kullback-Leibler distance, whereas the BIC is based on Bayes' theorem and approximates the marginal density of the respective likelihood. Both criteria are applied to simulated data sets and the results are analyzed, compared and evaluated. Based on the theoretical backgrounds as well as on the empirical evaluation, the criteria's usage is justified. Furthermore, nonlinear methods from the field of statistical learning theory are investigated.

Keywords. AIC, BIC, B-spline surfaces, deformation modelling, model selection, VC dimension

1 Introduction

The increased use of areal measurement techniques in engineering geodesy leads to the necessity of adequate areal analysis strategies. Commonly, the first step of an areal analysis is the modelling of the acquired point clouds by means of mathematical functions (Ohlmann-Lauber and Schäfer (2011)). Because of their flexibility and the associated ability to model even complex objects, freeform curves and surfaces like B-splines move more and more into focus concerning this matter (Koch (2010), Schmitt et al. (2013), Harmening and Neuner (2015b)).
A B-Spline curve $\boldsymbol{C}(u)$ (surface $\boldsymbol{S}(u,v)$) of degree p (and q) is defined by its $n+1$ ($(n+1)*(m+1)$) control points $\boldsymbol{P}_i$ ($\boldsymbol{P}_{ij}$), the B-spline basis functions $N_{i,p}(u)$ (and $N_{j,q}(v)$) as well as a knot vector $U = [u_0,\ldots u_r]$ (and $V = [v_0,\ldots v_s]$) (Piegl and Tiller (1997)):

$$\boldsymbol{C}(u) = \sum_{i=0}^{n} N_{i,p}(u)\boldsymbol{P}_i, \tag{1}$$

$$\boldsymbol{S}(u,v) = \sum_{i=0}^{n}\sum_{j=0}^{m} N_{i,p}(u)\, N_{j,q}(v)\boldsymbol{P}_{ij}, \tag{2}$$

$$\text{with: } u, v = [0,\ldots,1]. \tag{3}$$

When estimating a best-fitting B-Spline, a variety of unknown parameters has to be determined, leading to a nonlinear adjustment problem. In order to achieve a linear relationship between the observations $\boldsymbol{C}(u)$ ($\boldsymbol{S}(u,v)$) and the unknown control points, the remaining unknown parameters are typically excluded from the adjustment procedure and are determined a priori: The B-spline's degree is usually set to $p = 3$ ($q = 3$), which is generally accepted as a reasonable choice (Piegl and Tiller (1997)). In Harmening and Neuner (2015a) an approach to allocate appropriate surface parameters u and v to the observations is presented and Schmitt and Neuner (2015) propose a strategy to include the determination of the knot vector into the linear estimation of the control points. The remaining parameter type, which has to be determined, is the

number of control points $n+1$ (and $m+1$), whose influence on the estimation's result is immense, as it determines substantially the B-spline's complexity. Typically, this parameter is chosen quite arbitrary by using trial-and-error procedures. However, regarding this parameter's influence on the estimation's result, a justified choice of the control points' number is inevitable in order to improve the estimation of B-splines.

Gálvez et al. (2015) propose a clonal selection algorithm in order to determine the optimal number of control points and in Harmening and Neuner (2014) first investigations about the use of the model selection criterion AIC are introduced.

Model selection criteria like AIC or BIC were developed to solve tasks in which the optimal model has to be found. They are used in geodesy in order to find the best mathematical function to model ice mass variations (Baur et al. (2012)) or in order to determine the optimal structure of artificial neural networks (Neuner (2012)). In the present paper the investigations of Harmening and Neuner (2014) are carried on and the applicability of those model selection criteria to the determination of the optimal number of estimated control points is investigated.

The paper is structured as follows: In section 2 the fundamental principles of model selection and the mathematical basics concerning AIC and BIC are presented. Section 3 deals with the concrete problem of choosing the optimal number of control points using AIC and BIC. The results are analysed, compared and evaluated. Section 4 gives an outlook to an alternative way to evaluate the complexity of functions and in section 5 the results are summarized and an outlook is given.

2 Principles of model selection

2.1 Fundamental principles

A common task in engineering geodesy is to approximate a finite and noisy data set of size l by means of a mathematical function. Typically, the model structure is assumed to be known, and a fixed number of optimal parameters $\widehat{\boldsymbol{\theta}}$ has to be determined (Cherkassky and Mulier (2007)). This well studied task of parameter estimation is usually solved by means of the least squares method or the maximum likelihood (ML) theory (Koch (1999)).

However, the observed phenomena have in general physical and mathematical structures which are not known in all their diversity a priori. In order to avoid a failure of classical parameter estimation in cases when the assumed parametric structure is wrong, flexible learning methods are used: Those methods are built on wide sets of function classes $f(\boldsymbol{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which are able to approximate each continuous function with a certain amount of precision. The problem of finding an adequate functional description of the data is thus extended by choosing the optimal model complexity in addition to the optimal parameters (Cherkassky and Mulier (2007)).

Model selection deals with the problem of how to choose the optimal model from a given set of models, following the principle of parsimony, which states that a good model has to be as simple as possible while being a good approximation of the data (Burnham and Anderson (2002)). The compliance of this fundamental principle is identical with the finding of a trade-off between the function's bias and variance respectively: The more parameters are included into the estimation problem, the better the available training data is approximated, leading to a small approximation error. However, the function's variance increases, which leads to an overfitting. Consequently, the validation error of another realization of the same phenomenon would be large, as not only the observed phenomenon, but also the training data's noise is modelled. On the contrary, models which are too simple have a large bias, as they underfit the data and are not able to describe the observed phenomenon sufficiently well (Cherkassky and Mulier (2007)). The optimal model, which has to find a balance between simplicity and complexity, is chosen by means of model selection criteria, which sort the candidate models according to a certain score (Claeskens and Hjort (2008)). Those criteria can be categorised into two classes (McQuarrie and Tsai (1998)):

- Asymptotic efficient criteria are based on the quote of George Box, who stated that "all models are wrong, but some are useful": The phenomenon which underlies the data is assumed to have infinite dimension. As a consequence, it is not possible to find a model which is able to describe this phenomenon completely. Efficient criteria aim to find a model, which approximates the underlying truth

(Burnham and Anderson (2002)). The general approach of such criteria is to minimize a distance measure between each model and the truth. Using samples with infinite sample size, those criteria chose that model which produces the minimal quadratic error (McQuarrie and Tsai (1998)).

- Asymptotic consistent criteria assume an underlying truth of finite dimension which is included into the set of candidate models. Those criteria identify the correct model asymptotically with probability 1.

In the following the two most popular model selection criteria AIC and BIC are used as representatives for these two classes in order to investigate the applicability of model selection criteria to determine the optimal number of B-spline control points.

Both criteria are built on ML theory, which choses those model parameters $\widehat{\boldsymbol{\theta}}$ to be optimal, which maximize the conditional probability $\mathcal{L}(\widehat{\boldsymbol{\theta}}|\text{data})$. The principles of ML theory are not regarded in this paper, for further information regarding this topic please refer to Koch (1999)).

2.1 Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is an asymptotic efficient criterion. According to the definition of those criteria, a certain distance measure has to be minimized. The AIC interprets the model as well as the underlying truth as probability distributions g and f respectively and minimizes the Kullback-Leibler (KL) distance between these two distributions:

$$I(f,g) = \int f(x) \log\left(\frac{f(x)}{g(x|\boldsymbol{\theta})}\right) \mathrm{dx}. \qquad (4)$$

The KL-distance is a fundamental information-theoretical measure and can either be interpreted as the distance between two probability distributions or as the loss of information which is sustained, when the model g is used instead of the truth f.

As the truth is unknown, the KL-distance cannot be minimized directly; instead, the relative KL-distance is used. As a consequence, it is no longer evaluated, which model is the best in an absolute sense, but rather, which model is best compared to the other candidate models (Burnham and Anderson (2002)).

Akaike (1998) proved that the log-likelihood of the estimated parameters is a biased estimator of the expected relative KL-Distance and that the bias is identical to the number of estimated parameters K. These investigations lead to the AIC as a penalized log-likelihood criterion:

$$\mathrm{AIC} = -2\log\left(\mathcal{L}\left(\widehat{\boldsymbol{\theta}}|\text{data}\right)\right) + 2K. \qquad (5)$$

Although the derivation of the AIC is not based on the principle of parsimony, the criterion itself finds a trade-off between bias and variance: The log-likelihood attempts to choose that model producing the smallest approximation error, whereas the model's complexity in terms of the number of parameters is penalized by means of the second term in equation (5) (Claeskens and Hjort (2008)).

In case of normally distributed errors with constant variance, the l residuals $\hat{\epsilon}$ of a least-squares regression can directly be used to compute the AIC-scores (Burnham and Anderson (2002)):

$$\mathrm{AIC} = l\log(\hat{\sigma}^2) + 2K, \quad \text{with } \hat{\sigma}^2 = \frac{\sum \hat{\epsilon}^2}{l} \qquad (6)$$

In case of the least-squares estimation of a two dimensional B-spline curve, the two dimensions of the n+1 control points have to be determined. Additionally, the number of estimated parameters includes the variance factor σ^2 (Burnham and Anderson (2002)), which leads to the absolute number of estimated parameters

$$K = 2(n+1) + 1. \qquad (7)$$

However, the bias in equations (5) and (6) is only equal to the number of estimated parameters K, when the sample size is large compared to the number of estimated parameters. In cases, when the ratio l/K is small ($l/K < 40$), the use of the modified criterion AIC_C is recommended:

$$\mathrm{AIC}_C = \mathrm{AIC} + \frac{2K(K+1)}{l-K-1}. \qquad (8)$$

2.2 Bayesian Information Criterion (BIC)

The Bayesian/Schwarz Information Criterion (BIC/SIC) was introduced by Schwarz (1978) and is an asymptotic consistent criterion. It evaluates the posteriori probabilities of the candidate models M_j

(j = 0,…,n) and choses that one which seems to be the most likely according to the given data (Cavanaugh and Neath (1999)).
According to Bayes' theorem, the posteriori probabilities of the models are given by the prior probabilities of the models $P(M_j)$, the unconditional likelihood of the data f(data) and the marginal likelihood $\lambda_{n,j}$(data) (Claeskens and Hjort (2008)):

$$P(M_j|\text{data}) = \frac{P(M_j)}{f(\text{data})}\lambda_{n,j}(\text{data}) \tag{9}$$

$$\lambda_{n,j}(\text{data}) = \int f(\text{data}|\, M_j, \boldsymbol{\theta}_j)\, \pi(\boldsymbol{\theta}_j|M_j) d\boldsymbol{\theta}_j. \tag{10}$$

The latter is computed by means of the prior density $\pi(\boldsymbol{\theta}_j|M_j)$ of the parameters $\boldsymbol{\theta}_j$, given Model M_j, as well as the likelihood f(data$|\, M_j, \boldsymbol{\theta}_j$) = $\mathcal{L}_{n,j}(\boldsymbol{\theta}_j)$ of the data, given the jth model and its parameters.
In equation (9), f(data) is constant across all models. Usually, there is no information available concerning the models' prior probabilities, so that they are assumed to be equally likely. As a consequence, the models' marginal likelihoods $\lambda_{n,j}$(data) are the critical quantities to be evaluated:

$$\text{BIC} = 2\log\left(\lambda_{n,j}(\text{data})\right). \tag{11}$$

As a closed calculation of this quantity is in general impossible, the marginal likelihood is approximated by means of the Laplace approximation, which results in the final computation of BIC (Claeskens and Hjort (2008)):

$$\text{BIC} \approx -2\log\left(\mathcal{L}(\widehat{\boldsymbol{\theta}}|\text{data})\right) + \log(l)K. \tag{12}$$

Comparing equations (5) and (12) it can be seen, that – although the derivations are based on two completely different theories – the criteria themselves differ only slightly: AIC and BIC are both penalized log-likelihood criteria, with BIC imposing a stronger penalty on the models' complexity than AIC for $l \geq 8$.

3 Determination of the optimal number of control points using model selection criteria

The derivations of AIC and of BIC are both based on the assumption, that the data sets are independent identically distributed (iid), which cannot be taken for granted in engineering geodesy. For this reason the following basic investigations are not based on real data sets but on simulated ones which fulfil this assumption:
Starting points are a B-spline curve with six control points (n + 1 = 6) and a B-spline surface with 5 * 7 control points (n + 1 = 5, m + 1 = 7), which are superimposed by white noise. In the following the two criteria are investigated with regard to the repeatability and their behaviour in case of varying sample sizes. For this reason, five data sets with varying sample sizes are generated. For each sample size, the noise generating is repeated five times, resulting in 25 data sets which are different realizations of the same phenomenon. One of the realizations of the B-spline curve can be seen in Figure 1.
The sample sizes, which were used, vary between l_1=100 and l_5=2000 in case of the curve estimation and between l_1=900 and l_5=4900 in case of the surface estimation. These sample sizes are smaller than data sets resulting from terrestrial laser scanning commonly are. However, as they are uncorrelated, they are suitable to conclude from them to larger and correlated data sets.
B-spline curves and surfaces respectively are fitted through all those data sets while varying the number of control points in a range from n + 1 = [4,…,15] in case of the curve and n + 1 = m + 1 = [4,…,11] in case of the surface. The resulting residuals are used

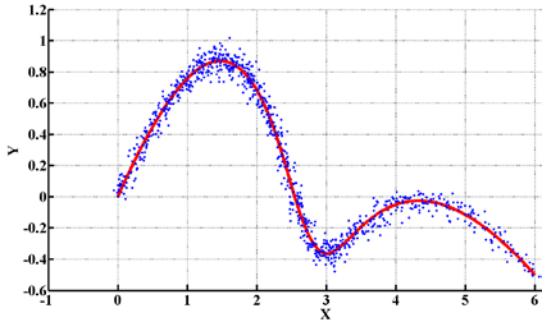

Fig. 1 A simulated B-spline curve

Table 1. Optimal number of curve control points according to AIC. (The curve was generated with n + 1 = 6.)

Sample size	100	250	500	1000	2000
Repetition 1	5	**6**	7	10	9
Repetition 2	5	**6**	**6**	9	15
Repetition 3	**6**	**6**	**6**	9	9
Repetition 4	**6**	7	**6**	7	9
Repetition 5	**6**	**6**	**6**	7	9

Table 2. Optimal number of curve control points according to BIC. (The curve was generated with n + 1 = 6.)

Sample size	100	250	500	1000	2000
Repetition 1	5	5	5	**6**	**6**
Repetition 2	5	5	**6**	**6**	**6**
Repetition 3	5	**6**	**6**	**6**	**6**
Repetition 4	5	5	**6**	**6**	**6**
Repetition 5	**6**	**6**	**6**	7	7

to compute the AIC- and BIC-scores (equations (6) and (12)) and the number of control points producing the smallest score is marked to be optimal.

As pointed out before, the determination of B-splines is much more complex than just estimating the control points. The remaining parameters are determined according to Piegl and Tiller (1997) (knot vector) and Harmening and Neuner (2015a) (curve and surface parameters) respectively.

The optimal numbers of control points for the curve estimation are listed in table 1 (AIC) and in table 2 (BIC): For sample sizes up to $l = 500$ AIC identifies the actual number of curve control points quite successfully: In at least three out of five cases the optimal number of control points is identical with the actual one and if it is not identical, the optimal number is close to the actual one. However, the larger the sample size is, the more unstable the AIC is concerning the repeatability and the larger the discrepancy between the optimal and the actual number of control points is. This behaviour is well known: AIC follows the notion that more details become visible when the sample size grows. Consequently more complex models have to be preferred (Aho et al. 2014). Altogether AIC has a tendency to overfit the data.

BIC shows a completely contradictory behaviour: For the sample size l=100 and l=250, BIC has a clearly visible tendency to underfit the data – however, the optimal number of curve control points is always close to the actual one. The larger the sample size becomes, the better BIC choses the optimal number of control points and – unlike to AIC – BIC is stabilizing at the actual complexity.

Table 3. Optimal number of surface control points according to AIC. (The surface was generated with $n+1$=5, $m+1$=7.)

Sample size	900	1600	2500	3600	4900
Rep. 1	8,11	10,11	10,11	6,10	7,9
Rep. 2	11, **7**	7,11	6,11	**5**,10	7,11
Rep. 3	**5**,11	11,10	9,9	11,11	11,11
Rep. 4	11,11	7,11	7,11	8,11	**5**,9
Rep. 5	**5**,9	6,11	9,**7**	11,10	**5**,11

Table 4. Optimal number of surface control points according to BIC. (The surface was generated with n+1 = 5, m+1 = 7.)

Sample size	900	1600	2500	3600	4900
Rep. 1	**5**,**7**	**5**,**7**	7,**7**	6,10	8,10
Rep. 2	**5**,**7**	**5**,**7**	**5**,9	**5**,**7**	**5**,9
Rep. 3	**5**,**7**	**5**,**7**	**5**,10	**5**,**7**	**5**,10
Rep. 4	**5**,**7**	6,**7**	7,**7**	7,**7**	**5**,9
Rep. 5	**5**,**7**	**5**,8	6,**7**	**5**,**7**	**5**,11

A similar behaviour can be observed in case of the surface estimation (table 3 and 4): For the major parts of the data sets, the number of control points chosen by AIC is significantly larger than the actual number of control points. Unlike to the curve estimation, this behaviour can be studied even if the sample size is comparatively small. Only few data sets exist, for which the AIC identifies at least one of the two parameters correctly; the correct combination, however, is never identified.

The results which are yielded by BIC are much more satisfying: Especially, for the sample sizes l=900 and l=1600, the number of parameters, which is chosen to be optimal, is identical to the actual one. However, unlike to the curve estimation, this result becomes instable with growing sample sizes. Nevertheless, in the major part of the data sets, at least one of the parameters is identified correctly.

Comparing the results of the curve and the surface estimation, it is obvious that it is much more difficult to identify the correct number of surface control points than those of a curve: The estimation of a B-spline surface requires the a priori computation of two knot vectors and two surface parameters for each observation, whereas only one of each is required in case of the curve estimation. Consequently, the uncertainty which is included into the adjustment because of the a priori computations is larger in case of the surface than in case of the

curve. This uncertainty is compensated by the inclusion of additional control points.

Altogether it can be noted, that BIC identifies the actual model complexity better than AIC does. However, AIC is based on an idea which seems to be much more suitable for the concrete problem: The curve (surface) estimation is not based on the curve (surface) parameters u (u and v) or on the knot vector U (U and V) which were used to simulate the data, but are determined independently. As the observations are overlaid with noise, the computed curve/surface parameters as well as the knot vectors differ from those, which were used to simulate the data sets. As a consequence, the truth is not contained in the set of candidate models and the concrete goal of model selection in this context should be to find a model which approximates the truth in an optimal manner (AIC) and not to identify the true model (BIC).
At first glance, the results presented above seem to contradict these basic ideas. However, it must be considered, that these basic ideas are valid for the asymptotic case, which is not relevant in practise, where sample sizes are always finite. As in the concrete context the performance properties for finite sample sizes are much more important than those for infinite sample size, BIC seems to be more suitable in order to choose the correct number of control points than AIC does, although the asymptotic properties might let one suggest something different.
As pointed out before, all investigations are based on simulated iid data. A straightforward adaption of the two criteria to correlated data is not possible: In this case the score-function does not have a local minimum and the criteria chose always the maximal possible number of control points, as additional control points are used in order to model the correlations. For this reason a decorrelation of the data is necessary, which requires the knowledge about the data's noise behaviour. First investigations concerning this task can be found in Kauker and Schwieger (2015).

4 An alternative way to describe the complexity of B-splines

In the section above a function's complexity was specified in terms of its number of free parameters K.

However, regarding the function $y = \sin(wx)$, it becomes obvious, that a function having a small number of parameters can be quite complex as well (Cherkassky and Mulier (2007)). For this reason, Vapnik (1998) introduced the Vapnik-Chervonenkis (VC) dimension as a complexity measure for function classes, which is independent of the number of free parameters. The VC dimension is an essential component of the statistical learning theory and forms the basis for structural risk minimization (SRM) which is an alternative to model selection criteria.
This section focuses on the fundamental question, if the SRM is at all suitable in order to determine the optimal number of parameters. In order to answer this question it is investigated, if the VC dimension of the respective B-spline reflects the number of control points in an appropriate manner. For reasons of simplicity, the investigations are limited to the two dimensional B-spline curve in the following.

4.1 Definition of the VC dimension

The VC dimension is a quantity which was originally used in classification theory and was primarily referred to indicator functions. For this reason, the descriptive definition of the VC dimension is given in terms of the classification problem although it can be expanded to real valued functions as well.
In order to define the VC dimension, the term of *shattering* has to be introduced: A class of indicator functions *shatters* a binary data set with sample size l, if it can split the data set in all 2^l possibilities. Figure 2 shows a line in 2D which shatters a data set consisting of three points.

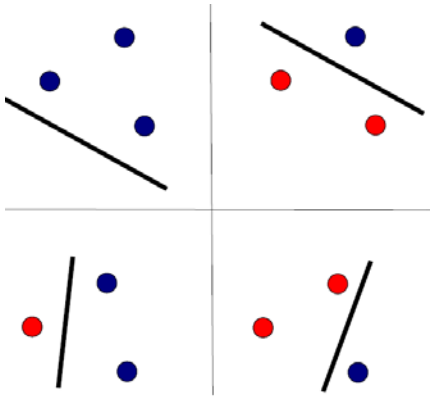

Fig. 2 A 2D line shatters a data set consisting of three points

The VC dimension of a class of functions $f(\boldsymbol{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the maximum number of samples, which can be shattered (Cherkassky and Mulier (2007)).
In case of the two dimensional line the VC dimension equals three, as there exists a data set consisting of three points, which can be shattered by the line (see Figure 2), whereas a data set consisting of four points cannot be shattered (see Figure 3).

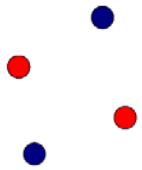

Fig. 3 A set of four points cannot be shattered by a 2D line.

4.2 Estimating the VC dimension of B-splines

Although this definition is quite simple and clear, analytical values for the VC dimension exist only for a few classes of functions. For this reason, Vapnik et al. (1994) propose a method to estimate the VC dimension empirically.
The method is based on a theoretically derived formula for the maximum deviation between the frequencies of errors $\xi(l)$ which is produced by a classifier on two randomly labeled data sets $\boldsymbol{Z}^1$ and $\boldsymbol{Z}^2$ of size l:

$$\xi(l) = \max(|\mathrm{Error}(\mathbf{Z}^1) - \mathrm{Error}(\mathbf{Z}^2)|) \qquad (13)$$

$$\xi(l) \leq \begin{cases} 1, & \text{if } (\tau < 0.5) \\ a\,\dfrac{\ln(2\tau)+1}{\tau-\kappa}\left(\sqrt{1+\dfrac{b(\tau-\kappa)}{\ln(2\tau)+1}}+1\right). \end{cases} \qquad (14)$$

The constants $a = 0.16$, $b = 1.2$ and $\kappa = 0.14928$ are empirically determined by Vapnik et al. (1994). With $\tau = l/h$, this formula is a function only of the VC dimension h and the sample size l. Consequently, the formula can be used to estimate the VC dimension. For this purpose empirical values for $\xi(l)$ are computed for different sample sizes $l_1, l_2 \ldots l_k$. (for a detailed instruction see (Cherkassky and Mulier (2007))). That integer value of h, which produces the best fit between the k values and the formula (14) is the wanted VC dimension.
In order to use a B-spline curve as a classifier, the basic idea of a linear classifier is extended: Given is a training data set consisting of l coordinate tuples $(x_1, y_1) \ldots (x_l, y_l)$ as well as the corresponding, randomly chosen labels $L_i \in [0, 1]$. Wanted is a decision boundary, which separates the training data with minimal training error:

$$\sum(L_i - \hat{L}_i)^2 = min. \qquad (15)$$

In case of a linear decision boundary, the estimated labels $\hat{L}_i$ are given by:

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{L} \qquad (16)$$

$$\hat{\boldsymbol{L}} = \boldsymbol{A}\hat{\boldsymbol{x}} \qquad (17)$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_l & y_l & 1 \end{bmatrix}. \qquad (18)$$

In order to expand this idea to a classification by means of a B-spline curve, the input space consisting of x- and y-coordinates is transformed into a high dimensional feature space as it is usual in classification theory (Cherkassky and Mulier (2007)). The linear decision boundary in this high dimensional feature space is a nonlinear decision boundary in the original input space. In order to achieve a B-spline curve as a decision boundary, the transformation is performed by means of the B-spline basis functions resulting into the classification problem (16) and (17) with:

$$\boldsymbol{A} = \begin{bmatrix} N_{0,p}(u_1) & N_{1,p}(u_1) & \ldots & N_{n,p}(u_1) \\ \vdots & \vdots & & \vdots \\ N_{0,p}(u_l) & N_{1,p}(u_l) & \ldots & N_{n,p}(u_l) \end{bmatrix}. \qquad (19)$$

Using this classifier, empirical values for $\xi(l_i)$ can be determined according to formula (13). In figure 4 those empirical values are exemplarily depicted for a B-spline curve with p=3 and n+1 = 6. That VC dimension h, which produces the best fit between these empirical values and the analytical formula (14) is the wanted VC dimension of the respective B-spline curve.
In this manner, the VC dimension of different B-spline curves can be determined (see table 5). As

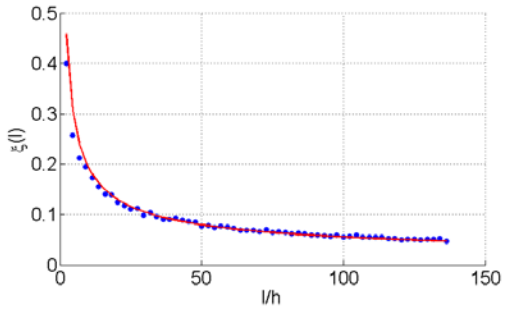

Fig. 4 Empirical values for $\xi(l_i)$ (blue points) for a B-spline curve with $p = 3$ and $n+1 = 6$ and the best fitting analytical curve (h = 10, red curve).

might be expected, with increasing number of control points the VC dimension increases.
However, it is not possible to derive an analytical functional relationship between the number of control points and the VC dimension.
Interestingly, the degree of the B-spline curve does not seem to have an influence on the VC dimension.
Altogether it can be noted, that the VC dimension is an alternative way to express the complexity of B-splines. Consequently, methods from statistical learning theory like structural risk minimization could be an alternative to model selection criteria like AIC and BIC.

Table 5. VC dimension *h* of B-spline curves with degree *p* and *n+1* control points

n + 1	4	5	6	7	8
p = 3	6	8	10	11	13
p = 4		8	10	11	13
p = 5			10	11	13

5 Summary and Outlook

In the present paper the applicability of the model selection criteria AIC and BIC to the determination of the optimal number of B-spline control points was investigated. Both, for simulated B-spline curves and for simulated B-spline surfaces with iid noise, the results of BIC were much more satisfying than those of AIC: Based on the simulated data, the BIC put out to be an appropriate method to determine the optimal number of control points.
However, in order to make the BIC applicable to real data sets, the concerning noise behavior has to be known in order to decorrelate the data (see Kauker and Schwieger (2015)).
Whereas BIC evaluates the complexity of a function by means of the free parameters, the VC dimension is a complexity measure, which is more or less independent of the number of parameters. First investigations showed, that the VC dimension of B-splines can be estimated, so that it has to be investigated, if methods from the statistical learning theory represent an alternative to model selection criteria.

Acknowledgements

This paper presents the results developed during the research project "Integrierte raumzeitliche Modellierung unter Nutzung korrelierter Messgrößen zur Ableitung von Aufnahme-konfigurationen und Beschreibung von Deforma-tionsvorgängen" (IMKAD) (1706–N29), which is funded by the Austrian Science Fund (FWF).

References

Aho, Ken; Derryberry, DeWayne; Peterson, Teri (2014): Model selection for ecologists: the worldviews of AIC and BIC. In *Ecology 95 (3)*, pp. 631–636.

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

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

Burnham, Kenneth P.; Anderson, David R. (2002): Model selection and multimodel inference. A practical information-theoretic approach. Second edition.

Cavanaugh, Joseph; Neath, Andrew (1999): Generalizing the derivation of the schwarz information criterion. In *Comm. in Stats. - Theory & Methods 28 (1)*, pp. 49–66.

Cherkassky, Vladimir S.; Mulier, Filip (2007): Learning from data. Concepts, theory, and methods. 2nd ed. Hoboken, N.J.: IEEE Press; Wiley-Interscience.

Claeskens, Gerda; Hjort, Nils Lid (2008): Model selection and model averaging. Cambridge, New York: Cambridge University Press (Cambridge series in statistical and probabilistic mathematics).

Gálvez, Akemi; Iglesias, Andrés; Avila, Andreina; Otero, César; Arias, Rubén; Manchado, Cristina (2015): Elitist clonal selection algorithm for optimal choice of free knots

in B-spline data fitting. In *Applied Soft Computing 26*, pp. 90–106.

Harmening, Corinna; Neuner, Hans (2014): Raumkontinuierliche Modellierung mit Freiformflächen. In Proceedings of the 139. DVW-Seminar Terrestrisches Laserscanning 2014, December 11.-12.2014, Fulda.

Harmening, Corinna; Neuner, Hans (2015a): A constraint-based parameterization technique for B-spline surfaces. In *Journal of Applied Geodesy 9 (3)*.

Harmening, Corinna; Neuner, Hans (2015b): Continuous modelling of point clouds by means of freeform surfaces. In *Österreichische Zeitschrift für Vermessung & Geoinformation 103 (2-3)*, pp. 121–129.

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

Koch, Karl-Rudolf (1999): Parameter estimation and hypothesis testing in linear models. 2nd. Berlin and New York: Springer.

Koch, Karl-Rudolf (2010): NURBS surface with changing shape. In *Allgemeine Vermessungsnachrichten (3)*, pp. 83–89.

McQuarrie, Allan D. R; Tsai, Chih-Ling (1998): Regression and time series model selection. Singapore, River Edge, N.J.: World Scientific.

Neuner, Hans (2012): Model selection for system identification by means of artificial neural networks. In *Journal of Applied Geodesy 6 (3-4)*.

Ohlmann-Lauber, Johannes; Schäfer, Thomas (2011): Ansätze zur Ableitung von Deformationen aus TLS-Daten. In *Proceedings of the DVW Seminar Terrestrisches Laserscanning 2011 mit TLS-Challenge (66)*, pp. 147–158.

Piegl, Les A.; Tiller, Wayne (1997): The NURBS book. 2nd ed. Berlin, New York: Springer (Monographs in visual communications).

Schmitt, Claudius; Neuner, Hans (2015): Knot estimation on B-Spline curves. In *Österreichische Zeitschrift für Vermessung & Geoinformation 103 (2-3)*, pp. 188–197.

Schmitt, Claudius; Neuner, Hans; Neumann, Ingo (2013): Strain detection on bridge constructions with kinematic laser scanning. In *Proceedings of the 2nd Joint international Symposium on Deformation Monitoring*.

Schwarz, Gideon (1978): Estimating the Dimension of a Model. In *The Annals of Statistics 6 (2)*, pp. 461–464.

Vapnik, Vladimir; Levin, Esther; Le Cun, Yann (1994): Measuring the VC-Dimension of a Learning Machine. In *Neural Computation 6 (5)*, pp. 851–876.

Vapnik, Vladimir Naumovich (1998): Statistical learning theory. New York: Wiley (Adaptive and learning systems for signal processing, communications, and control).