Czech University of Life Sciences in Prague
Faculty of Forestry and Wood Sciences
Department of Forest Management

METHODS OF STEM FORM MODELING

by

KAREL KUŽELKA

A thesis submitted to
the Faculty of Forestry and Wood Sciences, Czech University of Life Sciences in Prague,
in partial fulfilment of the requirements for the degree of Doctor.

PhD programme: Forest Management

Prague, June 2013

Thesis Supervisor:

doc. Ing. RÓBERT MARUŠÁK, Ph.D.
Department of Forest Management
Faculty of Forestry and Wood Sciences
Czech University of Live Sciences in Prague
Kamýcká 1176
165 21 Praha 6
Czech Republic

Author's Declaration

I declare that I am the sole author of the doctoral thesis METHODS OF STEM FORM MODELING. The elements in the doctoral thesis are not a copy of or similar to any other thesis submited elsewhere. The doctoral thesis was done on the basis of an original research work and only with the help of literature sources listed in Bibliography.

Prague, May 22, 2013. Karel Kuželka

Abstract and contributions

This doctoral thesis deals with the utilization of splines for stem curves representation.

In the first part, the thesis deals with possibilities of utilization of splines for modeling stem curves of individual stems. At first regular stems of coniferous tress are represented by splines. On the basis of multiple criteria performance of number of spline types is compared and advantages of individual spline types are discussed. For those splines that show the best pertinence for stem form modeling optimal distribution of input points is proposed to maximize the accuracy of the stem curve models simultaneously with minimizing the number of diameter measures required. Subsequently exploitation of splines for describing irregular stem curves of broadleaved trees, optimization of input point placement, and evaluation of suitability of particular spline types for this purpose is solved.

The second part deals with regression models of local typical stem curve. Possibilities of utilization of regression splines are evaluated. Accuracy of regression spline models is compared with the accuracy of taper models of polynomial, segmented polynomial and variable exponent forms. Methodology of utilization of regression splines is summarized.

In particular, the main contributions of the doctoral thesis are as follows:

1. Design of a methodology of a single stem curve representation by splines, readily feasible for the use in the software DendroScanner.
2. Design of a methodology of a regression model of typical stem using splines, readily feasible for the use in the software DendroScanner.

Keywords:

Stem curve, taper function, spline, regression, model.

Acknowledgements

First of all, I would like to express my gratitude to my dissertation thesis supervisor, doc. Ing. Róbert Marušák, PhD. He has been a constant source of encouragement and insight during my research and helped me with numerous problems and professional advancements.

I would like to thank to Dr. Daniel Zahradník and further also to Prof. Ken-chi Kamo and Prof. Atsushi Yoshimoto for giving me valuable advices regarding the statistical aspects of the thesis. I would like to thank to Dr. Ivana Linkeová for giving me a theoretical framework for spline techniques.

Special thanks go to the staff of the Department of Forest Management, who maintained a pleasant and flexible environment for my research. I would like to express special thanks to the department management for providing most of the funding for my research. I would like to express thanks to my colleagues from the Department of Forest Management, namely Ms. Helena Brožková for her valuable comments, and Ms. Markéta Žďárská, Ms. Radka Stolariková and Mr. Lubomír Tippmann for the help with data collection.

My research has also been supported by the project of the Ministry of Agriculture of the Czech Republic No. QI102A079 "Research on biomass of broadleaved species" and by the projects of the University Internal Grant Agency IGA No. 201043140024 "Verification of possibilities of obtaining mensurational magnitudes using terrestrial digital stereophotogrammetry", No. 4314013123132 "Stem curve modeling using spline functions", and No. 20124347 "Regression analysis of local stem form using spline functions".

Finally, my greatest thanks go to my family members, for their infinite patience and care.

Contents

List of Figures

List of Tables

Abbreviations

Common Mathematical Functions and Operators

$\mathbf{b}$	Vector $\mathbf{b}$
b_i	the i^{th} element of vector $\mathbf{b}$
$\|\|\mathbf{b}\|\|$	Norm of vector $\mathbf{b}$
$\dim \mathbf{b}$	Dimension of vector $\mathbf{b}$
$\mathbf{A}$	Matrix $\mathbf{A}$
$a_{i,j}$	Element of matrix $\mathbf{A}$ at the i^{th} row, and the j^{th} column
$\mathbf{A}^{-1}$	Inverse matrix to matrix $\mathbf{A}$
$\mathbf{A}^{\text{T}}$	Transposed matrix to matrix $\mathbf{A}$
x'	First derivative of x
x''	Second derivative of x
$\|x\|$	Absolute value of x
$\|\mathbf{CD}\|$	Distance between points $\mathbf{C}$ and $\mathbf{D}$

Taper model equations

D	Diameter at breast height
d	Diameter
H	Total stem height
h	Height
H_D	Breast heigt
T	Relative height, h/H
β_i	Model parameter

Spline computation

$(\mathbf{P}_i)_{i=0}^n$	Definition polygon consisting of definition points $\mathbf{P}_i$
$(\mathbf{Q}_i)_{i=0}^n$	Driving polygon consisting of driving points $\mathbf{Q}_i$
$C(t)$	Curve cordinates at parameter value t
p	Degree of curve
t	Parameter value
t_i	Parameter value at the i^{th} knot
T	Knot vector - sequence of parameter values at the knots
F_i	i^{th}Ferguson cubic basis function (i^{th} Hermite polynomial)
$F_i(t)$	Value of the i^{th}Ferguson cubic basis function at parameter value t
$N_{i,k}$	i^{th} B-spline basis function of degree k
$N_{i,k}(t)$	Value of the i^{th} B-spline basis function of degree k at parameter value t
$R_{i,k}$	i^{th} rational basis function of degree k
$R_{i,k}(t)$	Value of the i^{th} rational basis function of degree k at parameter value t
w_i	Weight of the i^{th} input point
$\mathbf{T}_i$	Partial driving polygon centers

Chapter 1

Introduction

1.1 Motivation

The research is motivated by an intention of the research project of the Ministry of Agriculture of the Czech Republic No. QI102A079 "Research on biomass of broadleaved species". One of the particular objectives of that research project is to improve software called DendroScanner. Dendroscanner is an application analyzing a terrestrial digital photograph of a tree taken by a common digital camera [Pretzsch et al., 2009]. The aim is to obtain higher accuracy of field-measured data and simultaneously savings of time spent by field work. A subsequent processing of the pictures allows obtaining measurational data, e.g. stem volume or stem profile.

The stem profile is represented by a polynomial curve defined by data points entered interactively by an user. The polynomials interpolate well the relatively simple stem curve of coniferous species; however they are not always suitable for a more complicated stem profile of broadleaved species. Moreover, single-segment polynomial functions are not convenient for interactive modeling, because position of each individual point influences the shape of the whole curve [Piegl and Tiller, 1996]. Therefore an appropriate spline function has been searched in order to replace polynomials in the software.

1.2 Problem Statement

The form and taper of tree stems has been studied for more than a century; the first attempts to express the stem form in terms of easily measured tree characteristics appear already in the nineteenth century [Metzger, 1894; Kunze, 1896].

An appropriate function describing the form of the tree stems has been searched for using various approaches. The objective is to find a mathematical expression of a continuous curve that predicts diameter of the stem in any height. Using taper functions volume of the stem can be precisely derived. The taper function is also an important basis of estimation of sorting of the stem and of optimal utilization of the wood mass.

Although the form of a tree may have a simple appearance, the construction of its mathematical model is a demanding task. The main problem is to construct a function of which only incomplete information is available [Lahtinen, 1988]. The functional values are known only at a few discrete points. In that case the function cannot be solved exactly and it must be replaced by a known function which agrees sufficiently with the known facts of the original function.

With bigger or lesser success many different mathematical functions from a wide range have been used for stem profile representation. Much work has been done on polynomial (e.g. Goulding and Murray 1976), logarithmic (e.g. Demaerschalk 1972), trigonometric (e.g. Thomas and Parresol 1991), sigmoidal (e.g. Biging 1984), segmented polynomial (e.g. Max and Burkhart 1976) or variable exponent (e.g. Fonweban et al. 2011) equations.

Spline is a special curve piecewise defined by mathematical functions that are linked together in points called knots in order to compose a smooth continual curve. The most frequently so called polynomial splines are used whose segments are defined by polynomial or rational functions usually of low degrees. From the polynomial splines mostly the cubic interpolation spline is used. By the spline approach a high forming ability of a curve is obtained with preservation of advantages of low degree polynomials at the same time. At the present time splines are widely used to interpolate or smooth experimental data.

The investigation of possibilities of splines in stem profile representation has been conducted since the beginning of the eighties of the last century [Liu, 1980]. As stated by Kublin et al. [2008] spline functions have been successfully applied to stem form modeling for 30 years. Also Kaufmann [2001] considers cubic interpolation spline to

be very suitable to describe stem profiles. Cubic splines have been used to model the whole stem profile Figueiredo-Filho et al. [1996a], to calculate volume of logs of different lengths [Biging, 1988] and also to predict the stem profile from lower-stem measurements[Möttönen and Nummi, 2002]. Cubic splines were also implemented in the inventory methods of e.g. Swiss National Forest Inventory [Kaufmann, 2001].

Majority of so far published studies concerning splines and stem form modeling exploit the cubic interpolation spline (mostly the special case of the cubic interpolation spline currently also called natural cubic spline). Only exceptionally quadratic spline [Lahtinen, 1988] or cubic smoothing splines [Liu, 1980] have been involved. However splines are a very wide class of functions varying both in origination and properties. Although the interpolation cubic spline was referred to have some undesirable characteristics, different spline types have not been involved in the studies. Moreover the past studies concern above all coniferous species with relatively uncomplicated stem profiles. Stems of broadleaved species and stems of irregular shape were involved by Lahtinen [1988].

An important factor influencing accuracy of the spline model is the number of input points (measured diameters) and their distribution along the stem. Smaltschinski [1983] shows that for interpolation by sufficiently accurate curve it is necessary to involve at least six input points. Figueiredo-Filho et al. [1996a] express the confidence that at least ten appropriately placed input points are needed to describe properly the stem profile using natural cubic spline.

The thesis deals with two major challenges. The first challenge is to find an appropriate spline function for describing with sufficient accuracy stem curves of individual stems of both coniferous and broadleaved species using points obtained from a digital photography interactively by human interpreter.

The study contains a comparison of the performance of natural cubic spline with performances of several other spline types. The accuracy of individual curves produced from several variants with different input point numbers is evaluated. Because the splines used in the study are different from one another, also the optimal positions of input points will differ. Therefore optimal positions of input points will be determined for each of the spline type and each of the input points counts. The main goals of the challenge are:

- To verify possibilities of spline curves utilization for modeling stem curves of individual stems of coniferous species and for modeling stem curves of straight parts of stems of broadleaf species.

- To compare the behavior of individual spline types and their suitability for stem curve modeling.
- To determine the optimal number of measured upper stem diameters as input points for spline computation and their distribution along the stem.
- To explore possibilities of utilization of individual spline types for stem form modeling of irregular stems of broadleaf species and to determine optimal distribution of input points in relation to stem irregularities.

The second challenge involves generalized regression model of local stem form. The stem form is a result of many factors including genetic influences, climatic factors, site quality, stand density, etc. [Muhairwe et al., 1994]. A huge number of the factors are stand specific; they influence almost all trees in a stand in the same way. Stem curves in a stand, or more generalized in a locality, tend to have identical shape and therefore they can be described by a local tree profile.

From a number of measured profiles a regression spline model is created. It allows to model a generalized stem form for a new stem of which DBH and height is measured. Applicability of regression (or smoothing) splines for creating a regression function of local tree profile is tested. Performance of several regression splines is compared and the optimal amount of smoothness is determined using several methods. The main goals of the challenge are:

- To verify possibilities of utilization of smoothing splines to generate a regression model of a local stem curve of coniferous and broadleaved species.
- To compare behavior of individual regression splines and their suitability for local stem curve modeling.
- To determine optimal values of parameters of regression splines for local stem curve modeling.
- To determine the optimal number of measured upper stem diameters as input points for spline computation and their distribution along the stem for creating the regression model.

1.3 Contributions of the Thesis

1. Design of a methodology of a single stem curve representation by splines, readily feasible for the use in the software DendroScanner.
2. Design of a methodology of a regression model of typical stem using splines, readily feasible for the use in the software DendroScanner.

1.4 Structure of the Thesis

The thesis is organized into chapters as follows:

1. *Introduction*: Describes the motivation behind our efforts together with our goals. There is also a list of contributions of this doctoral thesis.
2. *Background and Literature Review*: Introduces the reader to the necessary theoretical background and surveys the current state-of-the-art.
3. *Methods*: Describes data, software and methods used to achieve the goals of the doctoral thesis.
4. *Results and Discussion*: Presents the results of the main problems and discusses them.
5. *Conclusions*: Summarizes the results of our research and concludes the thesis.

Chapter 2

Background and Literature Review

2.1 Stem form representation and taper models

2.1.1 Basic concepts

Stem form

Stem form refers to the characteristic shape of a stem from ground level to the stem tip in terms of the dependence of diameter on height [Burkhart and Tomé, 2012]. There are several ways to describe the stem shape. The stem form can be expressed as a stem curve, as a stem profile, as form quotients or as form series. The stem form determines the volume of the stem and is of prime importance for timber quality [Väisänen et al., 1989].

Taper

Taper is defined as the rate of narrowing in diameter with respect to increase in height from ground level to the stem tip[Gray, 1956; Burkhart and Tomé, 2012].

Stem curve

Stem curve represents the stem shape in the form of continuous function [Fabrika and Pretsch, 2011].It is the most frequently utilized way to express the stem form. In fact, so called taper models most frequently express the stem form as a stem curve. The stem curve encloses the stem in space and rotation of the stem curve creates the surface of the

stem [Korf et al., 1972]. When the stem form is described using the stem curve, it is assumed, that the stem is a revolution solid. In fact, that is not always the case, but the inaccuracy sequent from the assumption is small enough to be neglected [Korf et al., 1972].

The advantage of stem form representation as a continuous function consists in the possibility to obtain directly the stem volume by integrating a function derived from the stem curve.

Stem profile
Stem profile expresses the stem form as a set of discrete points on the stem curve. The stem profile is a set of coordinate pairs, where abscissae are heights of the measurement points from the ground level and ordinates are appropriate thickness of the stem [Fabrika and Pretsch, 2011]. Stem curve can be derived from stem profile by fitting a function or by using interpolation techniques.

Form quotients
Form quotient is the ratio of the stem diameter in a certain relative height of the stem and the diameter in a defined comparative height [Korf et al., 1972]. The comparative height can be defined as an absolute height (most frequently the breast height) or as a relative height.

Form series
Form series are series of form quotients in uniform distances (defined relatively or absolutely) between the ground level and the stem tip [Korf et al., 1972].

2.1.2 Taper models: definition and classification

Many taper models of different forms have been developed for trees of a wide range of species and geographical areas. Most of them were intended to predict upper stem diameters and estimate the volume of the stem with little input data; usually diameter in breast height (DBH) and the tree height:

$$d = f(D, H, h). \tag{2.1.1}$$

This approach relies on assumption that taper is in accordance with the fundamental growth principles and the stem of a forest tree conforms to the dimensions of certain predesignated geometric solids. Such models provide reasonable estimates of taper and volume of a ordinary tree from little data.

The second approach doubts, that any preconceived general functional form can represent the taper properly [Grosenbaugh, 1966]. Stem taper is an unstable factor sensitive to many interactions in a dynamic forest system and individual trees seem capable to assume infinite variety of shapes as a consequence of several factors including climatic fluctuations, site quality, stand age and stand density [Muhairwe et al., 1994]. Such variation in stem forms makes difficult to formulate general mathematical rules readily applicable to a single species or even to all trees in a single stand [Larson, 1963; Perez et al., 1990]. Flewelling et al. [2000] proposes utilizing of multi-point profile prediction systems having as input together with diameter at breast height and total height also one or more upper stem diameter measurement. Unlike to model based on diameter at breast height and total height predicting profiles very well on the average throughout the area, multi-point taper systems providing additional information are more flexible and allow individual tree profile differences to be recognized. Moreover multi-point systems have the ability to greatly reduce biases. Liu [1980] supposes that more rational approach than fitting the taper to any preconceived functional form is to provide numerical technique that is capable of assuming various functional forms depending on the distribution of data points.

According to Diéguez-Aranda et al. [2006]; Burkhart and Tomé [2012] taper models can be classified in three groups:

1. Single taper models describe the entire profile of the stem using a single equation. Preliminary attempts used lower degree polynomials in terms of the relative height on the stem [Matte, 1949], later also higher polynomial were used [Bruce et al., 1968; Goulding and Murray, 1976]. Some single taper models use power functions, logarithmic functions or an equation based on trigonometric functions [Alder, 1978; Thomas and Parresol, 1991]. The single taper models can relatively accurate describe the general taper of trees. Although they may provide reasonable estimates in the mid-portion of the stem, they usually lack the ability to describe accurately the stem profile in the basal and upper stem segments [Brooks et al., 2008].

2. Segmented model approach assumes that a tree is made up of two or more segments with form being constant within a segment and different between segments. It is

generally assumed that a tree stem can be divided in three geometric shapes. The butt resembles a frustum of a neiloid, the central section resembles a frustum of a paraboloid or a cone and the top resembles a cone [Diéguez-Aranda et al., 2006]. The transition from one segment to the next is gradual and smooth. However, there are some exceptions. As reported by Valentine and Gregoire [2001] some species like slash pine (*Pinus elliottiii* Engelm.), ponderosa pine (*Pinus ponderosa* Doug. ex Laws.) and yellow poplar (*Liriodendron tulipifera* L.) do not conform to that segment shapes. Segmented taper models describe particular components of the stem with different equations, which are then mathematically joined to produce an overall segmented function. Segmented models were first introduced by Max and Burkhart [1976].

3. Variable form (variable exponent) taper models describe the stem shape with a changing exponent or variable from ground to top to represent the neiloid, paraboloid, conic, and several intermediate forms. This approach is based on the assumption that the stem form varies continuously along the length of a tree [Diéguez-Aranda et al., 2006]. The tree form varying from one point to another along the stem is expressed by a single continuous function.

Taper equations are also classified as compatible or non-compatible. Compatible taper equations, when integrated produce identical estimate of total volume to that given by existing volume equations [Demaerschalk, 1972; Cao et al., 1980]. Compatibility avoids discrepancies between volume estimates from a taper equation and those from commonly used volume equations. Compatible taper equations can be derived from existing volume equations [Demaerschalk, 1972]. Non-compatible taper equations are defined as taper equations that are not compatible.

No of the mentioned approaches is perfect in all respects. The models are created to describe the real taper as close as possible where accuracy and utility are sought at the expense of characteristics considered less important. As mentioned by Goodwin [2009] compatible taper models are constrained by entire stem volume at the expense of a diameter constraint, segmented models loose simplicity and constrain flexibility and variable exponent models sacrifice computational speed.

Accuracy is not the only demanded property of a taper model. Kozak and Smith [1993] and Goodwin [2009] are concerned with ideal model characteristics. According to Goodwin [2009], who summarizes the properties of the ideal taper model in ten items, the model should (briefly):

1. be accurate for its purpose,
2. be reliant on easily measurable or obtainable variables,
3. be fast algebraically integrable and invertible,
4. have continuity in the whole range of heights and diameters,
5. use inputs as constraints where possible so that the predictions are consistent with inputs,
6. should allow any pair of diameter and height as input and should not require the DBH as an input,
7. be regionalisable, i. e. be adaptable to form changes resulting from regional or genotypic factors,
8. be localisable, i. e. able to model additional between-tree variability,
9. be a single non-segmented equation in order to conveniently accommodate diameter constraints,
10. be applicable to a wide range of species and sizes.

The enumerated criteria may not be connected with a dispassionate selection of the model most accurate in diameter prediction. The requirements are related with the effort to construct a model applicable in practise for a given purpose and therefore are necessarily subjective.

2.1.3 Single taper models

Polynomial single taper models

In earliest attempts to define taper simple polynomial equation was used: [Matte, 1949]:

$$y^2 = Cx^r, \tag{2.1.2}$$

where the parameter r was known as form exponent. The next attempts brought a little more complicated rational equation:

$$y = \frac{x}{a + bx}, \tag{2.1.3}$$

where a and b are regression coefficients. In both equations x denotes the distance from tip to section as a proportion of the length form tip to breast height and y denotes the corresponding diameter as a proportion of DBH [Matte, 1949].

To describe the stem profile of Loblolly pine (*Pinus taeda* L.) above breast height Matte [1949] proposes a polynomial model expressed by equation 2.1.4:

$$\frac{d}{D} = x\sqrt{ax^2 + bx + c}, \tag{2.1.4}$$

where x is ratio of distance from tip to measurement point to total height above breast height:

$$x = \frac{H - h}{H - D_D}. \tag{2.1.5}$$

The coefficients a, b and c are found using least squares regression. Mentioned taper equation was succesfully applied also on balsam fir (*Abies balsamea* (L.) Mill.) and white pine (*Pinus strobus* L.). The equation should be suitable to describe stem profile of any coniferous species [Matte, 1949]. The weakness of the low degree polynomial models is the inability to describe the lower portion of a tree with significant basal swelling.

The polynomial form were further utilized by Bruce et al. [1968]. He described the stem profile of red alder (*Alnus rubra* Bong.) using polynomial regression equation with the square of relative diameter as dependent variable and the independent variable x expressed as

$$x = \frac{H - h}{H - 4.5ft}, \tag{2.1.6}$$

where h is the height of the predicted diameter and H is the total height of the stem. Trials of a range of numerical values of the power exponents led to the exponent values 3/2, 3 and 32 or 40. It is mentioned that the exponents 3/2 and 3 are satisfactory to describe the upper four fifths of the stem; the high exponents are required to describe the butt swell. The effect of the exponents 32 and 40 is similar and the exact exponent value is not critical. Further it is showed that expression of diameter relative to the diameter at some percentage of total height rather the at breast height simplified the fitting procedure eliminating a considerable part of the effect of tree size [Bruce et al., 1968]. The taper model was later used and modified by Hilt [1980] to describe taper of upland oaks.

An interesting functional form of a simple polynomial taper model was proposed by Laasasenaho [1982]. The model is in the form of polynomial utilizing exponents corresponding with the numbers of Fibbonaci series up to 34.

Polynomial models of different powers was proved also by Munro [1968], who compared numbers of different polynomial models where (d^2/D^2) was predicted using powers of (h/H). Although he achieved reasonable estimations of upper stem diameters from equations based on the assumption that the stem form is paraboloid, fifth degree polynomial taper equation proposed significantly more precise estimations. More complicated functions incorporating high exponentiation did not significantly improve the precision of estimates of stem diameters [Munro, 1968]. Goulding and Murray [1976] developed compatible taper equations using polynomial functions of 5^{th} degree.

A substantial number of new models and modifications of Munro's equations were tested by Kozak et al. [1969b] on 19 commercial species and species groups of British Columbia. The taper models were developed from the relationship of parabolic function. In case of few large, extremely rapidly tapering trees, the shape was found to be rather neiloidal than quadratic paraboidal. Bias could be eliminated using empirical percentage corrections. Instead of using empirical corrections Kozak *et al.* integrated additional conditions into the equations. A number of simple yet effective taper functions were derived and least squares regression constants for individual species were stated. None of the modifications was significantly better than the functions suggested by Munro [1968].

Although in later 70's the focus shifted to segmented-polynomial taper models and later to variable-exponent models, development of single polynomial models continued in that later time. As an example serves the two-parameter model of Amidon [1984] expressed by equation

$$d = \beta_1 \mathrm{D}(H-h)/(H-H_D) + \beta_2(H^2/h^2)(h-H_D)/H^2. \tag{2.1.7}$$

Model of Amidon [1984] considered only the stem form over breast height (bh). Excluding the most bottom part of the stem profile was not seldom used for the single models because consistently with Bruce [1972] excluding the profile below breast height enables to simplify a taper model and to increase precision of the remainder of the stem. The lower part of the stem was considered as a cylinder with some loss of accuracy for cubic volume [Amidon, 1984].

A recent single polynomial taper model applicable to a wide range of species proposed by Goodwin [2009] is derived from a cubic polynomial by adding constraints. The model can be fit using one or two upper stem diameters. None of them needs to be the DBH. For the single-diameter constraint the optimum diameter location is said to be at around centroid height (25-30 %) and the locations of measures for two-diameter constraints is recommended to be at 10 % and either 30 % or 60 % of the tree height. As reported,

the second diameter constraint substantially improves the accuracy of the model. The performance of the model was compared to five other models, among others the segmented models of Max and Burkhart [1976] and Candy [1989] and variable exponent models of Bi and Long [2001] and Eerikäinen [2001] and was considered to suit the best the criteria of ideal taper model stated by Goodwin [2009].

Non-polynomial single taper models

Non-polynomial taper models are based on number of approaches utilizing different non-polynomial functions. Most often logarithmic, trigonometric and sigmoidal functions have been used. A simple logarithmic function was developed by Höjer in 1903 [Matte, 1949]:

$$\frac{d}{D} = C \log \frac{c+l}{C}, \tag{2.1.8}$$

where d is diameter at distance l expressed as a percentage form the tree top, C and c are constants varying with form quotient [Wenger, 1984].

A different logarithmic taper equation was developed by Demaerschalk [1972]. A logarithmic volume equation he converted into a logarithmic taper equation

$$\log d = \beta_0 + \beta_1 \log D + \beta_2 \log (H - h) + \beta_3 \log H \tag{2.1.9}$$

by reasoning process based on premise that total volume estimates, based on integration of the taper equation, must be identical to those given by the original volume equation. Also a comparison was made between standard errors obtained by the logarithmic taper equation and the quadratic polynomial model used by Kozak et al. [1969a]. It was showed, that the logarithmic equations of Demaerschalk were for most of the species tested (16 out of 23) in closer agreement with the real taper curves than the quadratic taper equations. It is possible to derive in a similar way compatible taper equation from most commonly used volume equations. By derivation the best fit is achieved for volume and the fit for diameters can be optimized by the choice of the optimum values for some parameters [Demaerschalk, 1972].

An early taper function proposed in 1927 by Behre [Bruce, 1972; Amidon, 1984] utilized hyperbolical function. This model was later modified and applied by Bruce [1972].

The research on stem taper includes also taper models based on trigonometric functions. In that area belong the work of Alder [1978], who developed a model characterized by equation 2.1.10:

$$g_r = h_r + \beta(\cos (2\pi h_r) + 1), \tag{2.1.10}$$

where g_r is the sectional area at measurement point relative to sectional area in breast height and h_r is ratio of distance from tip to measurement point to total height above breast height. Alder [1978] indicates, that equation 2.1.10 was found to be simpler and more economical in practice than known polynomials models.

Another taper equation utilizing trigonometric function of the form:

$$\frac{d^2}{D^2} = \left(\beta_1(x-1) + \beta_2 \sin(c\pi x) + \beta_3 \cot\left(\frac{\pi}{2}x\right)\right) + \varepsilon, \tag{2.1.11}$$

where d is diameter inside bark a at given height, D is diameter inside bark at breast height, x is the corresponding relative height, c is a coefficient and β's are model parameters, was used to describe the taper of slash pine (*Pinus elliottii* Engelm. var. *elliottii*), willow oak (*Quercus phellos* L.), and sweet gum (*Liquidambar styraciflua* L.) [Thomas and Parresol, 1991; Thomas et al., 1995]. This simply expressible trigonometric model is flexible enough to fit both conifer and hardwood bole forms. It was also compared with the segmented polynomial model developed by Max and Burkhart [1976]. The trigonometric model performs equally well and has real advantages in term of parsimony [Thomas and Parresol, 1991].

Biging [1984] proposes a sigmoidal taper model to describe the stem form of six conifer species of Northern California including ponderosa pine (*Pinus ponderosa*), Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), white fir (*Abies concolor* (Gord. & Glend.) Lindl. (Iowiana [Gord.])), red fir (*Abies magnifica* A. Mutt.), sugar pine (*Pinus lambertina* Dougl.) and incense cedar (*Libocedrus decurrens* Torr.). The taper equation is derived from Chapman-Richards function and is fit to the stem profiles using nonlinear regression techniques. The model is proved to be able to fit well the stem form. The comparison with the Max and Burkhart's [1976] segmented polynomial model shows both models even accurate; the aspect of computing simplicity favors the two-parameter sigmoidal model against the more complicated six-parameter segmented polynomial model [Biging, 1984].

2.1.4 Segmented taper models

Splitting the stem into several parts representing basic geometric figures is not a new idea. Already in 1837 Smalian compared sections of tree stems to frustums of conoids with varying powers of exponents [Bruce and Max, 1990]. All that is new is the development of equations describing continuous curves with continuous derivatives and the use of integrals on these functions to estimate section volumes.

Ormerod [1973] developed a generalization of the model used by Bruce et al. [1968]. Using this generalized model upper stem diameters can be estimated by equation 2.1.12:

$$d = D\left[\frac{H-h}{H-k}\right]^{p}, p > 0, \tag{2.1.12}$$

where D is the measured diameter at height k and p is the fitted exponent. The model is conditioned to exactly predict observed DBH [Czaplewski and McClure, 1988]. Due to changes in the stem form along the length of the stem, the simple model of equation 2.1.12 may not provide an adequate description of the bole. Therefore Ormerod suggests to modify the equation to a step function and to fit separate exponents for each step:

$$d = (D_i - C_i)\left[\frac{H_i-h}{H_i-k}\right]^{p_i} + C_i, p_i > 0, \tag{2.1.13}$$

where C_i is the section diameter intercept. Ormerod further deals with an assumption based on an inspection of number of taper curves, that all boles have a single taper inflection point at 30 % of the total height. Separate exponents are then fitted for both sections below and above that point by regression of the logarithmic transformation of equation 2.1.13. Estimation of the volume is performed by integration of the equation.

Max and Burkhart [1976] described the taper curve using segmented polynomial regression models based on statistical methods of Fuller [1969] and Gallant and Fuller [1973]. The stem was partitioned and a different polynomial submodel was defined on each section of the partition. By connecting the submodels at the joint points a segmented polynomial model was obtained. The submodels are connected by imposing binding conditions in such a manner that the functional value is continuous and has continuous first or higher order derivatives. If compared to the general definition of splines piecewise polynomial function with continuous functional value and first k derivatives the segmented polynomial models can be considered as spline models although the utilized function is not called spline by the authors.

Max and Burkhart compared four models. The first model was a simple quadratic model. The second model a quadratic-quadratic model was composed of two quadratic functions with restrictions that functional value and the first derivative must be continuous. The quadratic-linear-quadratic model described the taper of the base and the top of the tree by two quadratic polynomials and the middle section by polynomial of first degree and has continuous functional value. The fourth model consisted of three quadratic submodels grafted together with the restriction of continuous first derivative. The described

models were tested on natural-stand and plantation loblolly pine trees and it was concluded, that the segmented models were better than the simple quadratic model and the three-segmented models were preferred over the two-segmented model. The quadratic-linear-quadratic model and the quadratic-quadratic-quadratic model showed no difference when applied to the plantation trees, but the quadratic-quadratic-quadratic model was significantly better when applied to the natural-stand data. The model is expressed by equation 2.1.14

$$\left(\frac{d}{D}\right)^2 = \beta_1(T-1) + \beta_2(T^2-1) + \beta_3(\alpha_1 - T)^2 I_1 + \beta_4(\alpha_2 - T)^2 I_2, \tag{2.1.14}$$

where $T = h/H$, α_1 and α_2 are the relative heights of the upper and lower join points, respectively, of the three segments, $\beta_1 \ldots \beta_2$ are the parameters and I_1 and I_2 are as follows:

$$I_1 = \begin{cases} 1 & \text{if } \alpha_1 - z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if } \alpha_2 - z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The Max and Burkhart's segmented polynomial models have been afterwards used as a ground for volume equations. Through integration of the Max and Burkhart's taper model consisting of three quadratic segments, Brooks et al. [2008] derived the volume equation and evaluated the volume equation on the sample data of three commercial tree species in Turkey, Brutian pine (*Pinus brutia* Ten.), Cedar of Lebanon (*Cedrus libani* A. Rich.), and Cilicica fir (*Abies cilicica* Carr.). The models explained approximately 97 %, 95 % and 98 % of the variation for predicting volume of the three tree species respectively and the average volume error was less than 0.009 m^3. All the parameters for each equation were found to be significant, which approves that the model consisting of three quadratic segments describes the stem curve better than the other Max and Burkhart's segmented models.

A different segmented polynomial taper equation was developed by Cao et al. [1980]. It consists of three submodels grafted at two join points. Each submodel was in the form of a modified compatible taper equation presented by Goulding and Murray [1976] using a quadratic functions instead of a fifth-degree polynomials. The model was proved to be approximately compatible. Itl was compared with several previously developed models. No single model was the best for all purposes. The results of the comparison idicate,

that a simple quadratic equation cannot adequately describe tree taper or predict volume; the quadratic model of Kozak and others [1969b] neither estimates well the volumes, nor predicts accurately tree diameters. Using segmented models consisting of three simple quadratic submodels definitely improves the taper predictions. The compatible fifth degree polynomial taper equation of Goulding and Murray [1976] as well as other compatible equations adequately estimates volumes to various height limits but poorly predicts tree diameters. The conclusion is that a compatible taper equation does not appear to be a good choice it the sole purpose is to describe tree taper. This is verified also by the results of the Cao's segmented polynomial model which estimates well the volume, but has poorer ability to predict diameters than the Max and Burkhart's [1976] non-compatible segmented model. For diameter prediction the Max and Burkhart's segmented polynomial model had the best ranking among all the models compared.

The Cao model was further modified by McClure and Czaplewski [1986]. A condition was added to the approximately compatible model to be exactly compatible with any volume equation. Application of the exact compatibility did not undesirably affect error in diameter prediction and the anomalies between volume estimates using the segmented model and those made using existing volume equations and tables were reduced.

Czaplewski and McClure [1988] supplemented the stem profile model of Max and Burkhart [1976] with a second upper stem measurement. The stem profile model was conditioned to exactly predict this available additional upper-stem measurement, which might further improve estimates, at least near this additional measurement. Additional measurement at 5.3 m was used. Variance of residuals were reduced, however, bias was approximately the same for both conditioned and unconditioned models [Czaplewski and McClure, 1988].

Another segmented taper model consisting of two functions linked together at the inflection point is proposed by Demaerschalk and Kozak [1977] and tested on 32 age and locality groupings of British Columbia species; a cubic-cubic segmented polynomial model was derived by Parresol et al. [1987]. A compatible segmented polynomial model was proposed by Candy [1989]. Three segment taper model was developed by Fang et al. [1999, 2000]; segment taper model together with volume equations were developed by Jiang et al. [2005] using models of Max and Burkhart [1976] and Clark et al. [1991].

2.1.5 Variable form and variable exponent taper models

An approach considering the presence not only the neiloid, paraboloid and conic forms but also several intermediate forms is the variable-exponent approach. The variable-exponent approach assumes a continuous change between individual geometric forms. The exponent determines whether the model resembles the profile of a neiloid, paraboloid or cone, or the transition from one geometric form to the next. It varies continuously throughout the stem profile as a function of height in order to compensate for the form of different tree sections [Perez et al., 1990; Valentine and Gregoire, 2001].

The variable-form taper model was first derived by Newberry and Burkhart [1986] from the model of Ormerod [1973]. Further it has been introduced by Newnham [1988] and Kozak [1988, 1997]. The model of Kozak [1988], which has become widely used in the US and Canada [Sharma and Zhang, 2004], is in the form shown in equation 2.1.15:

$$d = \beta_0 D^{\beta_1} \beta_2^D X^{\beta_3 T^2 + \beta_4 \ln(T+0.001) + \beta_5 T^{0.5} + \beta_6 e^T + \beta_7 (D/H)}, \tag{2.1.15}$$

where $X = (1 - T^{1/2})/(1 - p^{1/2})$ and p is the inflection point.

When compared to simple or segmented models the variable exponent models were found to be superior [Newnham, 1992; Fonweban et al., 2011] and considered to give unbiased and accurate estimates both for stem form and volume. The promising type of models was therefore dealt with by number of authors such as Perez et al. [1990]; Muhairwe et al. [1994] or Huang et al. [2000] who all applied the model of Kozak [1988]. As input data can serve only DBH and the total height [Flewelling and Raynes, 1993] or together with former magnitudes also one [Rustagi and Loveless Jr., 1991] or more [Flewelling, 1993] upper stem diameter measurements.

Taper model of simple form with only one parameter β (equation 2.1.16) was developed by Sharma and Oderwald [2001].

$$d^2 = D^2 \left(\frac{h}{h_D} \right)^{2-\beta} \left(\frac{H-h}{H-h_D} \right) \tag{2.1.16}$$

In spite of its simplicity its prediction of diameters were superior to the segmented-polynomial taper models up to 60% of total height, while in the higher part the bias was slightly higher. Later [Sharma and Zhang, 2004] the parameter β was expressed in terms of the relative height $\beta = f(T)$ which resulted in a variable taper equation. Sharma and Zhang [2004] examined several forms of functions of T (linear, quadratic, exponential and their combination) and the preferable variable-taper model with four parameters

$\beta_0 \ldots \beta_3$ expressed as equation 2.1.17.

$$\left(\frac{d}{D}\right)^2 = \beta_0 \left(\frac{h}{h_D}\right)^{2-(\beta_1+\beta_2 T+\beta_3 T^2)} \cdot \left(\frac{H-h}{H-h_D}\right) \tag{2.1.17}$$

This taper equation was utilized for the three most important commercial tree species grown in eastern Canada, jack pine (*Pinus banksiana* Lamb.), black spruce (*Picea mariana* (Mill.) B. S. P.) and balsam fir (*Abies balsamea* (L.) (Mill.)), and it was found to be superior to the model of Kozak [1988]. Moreover the equation was modified to incorporate effects of stand density and thinning.

Bi [2000] developed a variable form model based on logarithms of trigonometric functions (equation 2.1.18). Although most variable form models are species specific, model of Bi [2000] is referred to be flexible in terms of ability to fit data of different species and different stem forms. The model was successfully tested on 25 species of genus *Eucaliptus* and *Corymbia*.

$$\frac{d}{D} = \left(\frac{\log \sin\left(\frac{\pi}{2}\cdot T\right)}{\log \sin\left(\frac{\pi}{2}\cdot \frac{1.3}{H}\right)}\right)^{\beta_1+\beta_2\cdot \sin\left(\frac{\pi}{2}\cdot T\right)+\beta_3\cdot \cos\left(\frac{3\pi}{2}\cdot T\right)+\beta_4\cdot \sin\left(\frac{\pi}{2}\cdot T\right)/T+\beta_5\cdot D+\beta_6\cdot T\sqrt{D}+\beta_7\cdot T\sqrt{H}} \tag{2.1.18}$$

Because most of variable form models and variable exponent models are relatively complicated, undesirable effects of multicollinearity often occur Kozak [1997]; Rojo et al. [2005]. The last studies tend to develop a variable exponent model with low effect of multicollinearity. An example of such model is the nine-parameter model of Kozak [2004]:

$$d = \beta_1 \cdot D^{\beta_2} \cdot H^{\beta_3} \cdot \left(\frac{1-T^{1/3}}{1-p^{1/3}}\right)^{\beta_4\cdot T^4+\beta_5\cdot (1/e^{D/H})+\beta_6\cdot \left(\frac{1-T^{1/3}}{1-p^{1/3}}\right)^{0.1}+\beta_7\cdot \left(\frac{1}{D}\right)+\beta_8\cdot H^{1-(\frac{h}{H})^{1/3}}+\beta_9\cdot \left(\frac{1-T^{1/3}}{1-p^{1/3}}\right)}, \tag{2.1.19}$$

where $p = hi/H$; hi denotes the stem height of the inflection point where the taper curve changes from neiloid to paraboloid. However, the parameter p is estimated as a model parameter.

Number of recently developed or applied variable-exponent taper models during last years [Zakrzewski, 1999; Muhairwe, 1999; Eerikäinen, 2001; Bi and Long, 2001; Valentine and Gregoire, 2001; Lee et al., 2003; Bluhm et al., 2007; Yang et al., 2009a; Li et al., 2012] attests that variable-exponent taper models are still interesting and presently evolving topic.

2.1.6 Process-oriented models

It is necessary to mention that the approach of the empirical stem curve model is not the only one used. Number of studies have demonstrated that the stem form depends on external conditions like lateral forces on stems or changes in resources availability [Larson, 1963; Dean et al., 2002]. Sloboda et al. [1998] believed, that scientifically satisfying taper models must be based on process-oriented approaches, but such models were not mature to estimate yield and assortments with the required precision and therefore empirical models were in use. There are several theories that more or less successfully attempt to explain the stem form on the basis of physiological processes.

One of well-developed approaches is the uniform stress hypothesis. It is based on the idea that the cambium produces new wood in such a way as to equalize the distribution of stress along the outer surface of the stem. The stem form minimizes the chance of breakage by increasing cambial growth in regions with highest stress. The hypothesis was supported by results of experimental studies [Larson, 1965; Milne and Blackburn, 1989; Ezquerra and Gil, 2001; Dean et al., 2002]. On the base of the theory a taper model can be developed with accurate predictions of the stem curve for various species [Morgan and Cannell, 1994].

Another approach of process-oriented stem taper models is based on pipe-model theory, which was formulated by Shinozaki et al. [1964a; 1964b]. The theory rely on assumption that the sapwood cross-sectional increment per unit of foliage above, or foliage increment above, is constant over time. The physiological assumptions were validated by experimental studies, rules of stem growth were stated [Chiba, 1990, 1991] and subsequently several taper models were developed [Chiba and Shinozaki, 1994; Osawa et al., 1991]. Results of such taper model development [Mäkelä, 2002] indicate that the pipe model assumptions are capable of producing realistic predictions of the vertical distribution of stem and branch diameter in trees of different sizes in the stand.

Process-based models often are combined with empirical models, because some problems (e.g. height growth) are difficult to solve in process-based growth modelling [Sloboda and Pfreundt, 1989]. A example of such combined models is the stem form model based on carbon budget and carbon partitioning developed by Deleuze and Houllier [1995], which is controlled by an empirical site-dependent height growth curve.

2.1.7 Problems with fitting taper models and taper model comparsion

Most of segmented models and variable exponent models are species-specific; the diameter prediction accuracy depends on tree species Sharma and Zhang [2004]. The challenge to represent the stem shape by a model is even more complicated due to intraspecific variation. It has been long realized that trees of the same species grown at different stand densities do not necessarily have the same tree form [Sharma and Zhang, 2004; Sharma and Parton, 2009]. The form is affected also by thinning [Tasissa et al., 1997; Sharma et al., 2002] and prunning [Valenti and Cao, 1986], water availability [Wiklund et al., 1995] and the form also differs between trees grown in a plantation and a natural stand Sharma and Zhang [2004]. Fortunately due to their complexity the segmented models and variable-exponent models are generally able to conform the different shapes. There have been attempts at including variables like crown ratio, site index or age class into a segmented polynomial model [Burkhart and Walton, 1985] and a variable exponent model [Muhairwe et al., 1994; Tasissa and Burkhart, 1998]. In most cases efforts to incorporate such factors offered limited succes. The studies approvingly reported no significant improve of the goodness of fit. Amidon [1984] used even 20 independent variables as predictors and determined best subsets of the predictor variables from all possible combinations. He also reported little gain in precision with an usage of more than two independent variables. Li and Weiskittel [2010] also reports little impact of incorporation of corwn variables on stem form prediction. On the other hand the incorporation of crown variables substantially improved stem volume predictions [Li and Weiskittel, 2010].

Parameters determination of the majority of the taper models is based on ordinary least squares or nonlinear least squares methods. As reported by Amidon [1984], the least squares analysis assume several conditions which in reality are seldom present. Three assumptions are considered most important: the data are typical, the correct form of the equation was chosen and the observations of the dependent variable are not statistically correlated. But because data used to develop taper models have mostly multiple measurements taken along the stem of each sample tree, observation are likely to be correlated Yang et al. [2009a]. Therefore the assumption of identical and independent errors, which is the basis for the LS model fitting approach is not longer valid and effects of spatial autocorrelation between model residuals then occur. Although model parameters estimated by least squares methods are still unbiased, estimated variances are biased. Related hypothe-

sis testing and confidence interval estimation of model parameters are invalid [Gregorie, 1987; Yang et al., 2009a]. Therefore autocorrelation can have several adverse consequences in terms of the statistical inference at efforts to indicate statistically significant predictor variables. The resulting bias Amidon [1984] considers as a serious problem without any general solution both in forestry and other fields of work. To partially eliminate consequences of autocorrelation Amidon [1984] compares models on the basis of multiple criteria of fit, rather than on a single functional form.

The problem of spatial autocorrelation between model residuals in empirical modeling of stem profile curves was solved by later developed mixed-effect taper equation. Mixed-effect models are statistical regression models containing both fixed effects and random effect. Mixed-effect taper models take into account the correlation among multiple diameter measurements on an individual stem. Including random effects allows to account most of the residual autocorrelation[Trincado and Burkhart, 2006]. In contrast to traditional regression the mixed-effects modeling allows for both population-specific and subject-specific models. A population-specific model considers only fixed-effects parameters and a subject-specific model considers both fixed- and random-effects parameters. The Max and Burkhart [1976] segmented polynomial model was improved in mixed-effects modeling framework by Leites and Robinson [2004] and later by Trincado and Burkhart [2006] or Cao and Wang [2011]. In the last years the mixed-effect modeling approach was is also used for more complicated variable-exponent models [Lejeune et al., 2009; Yang et al., 2009a,b].

Another problem is the multicollinearity. Multicollinearity is defined as a high degree of correlation among several independent variables. Its presence is usual in overcomplicated models with too many variables being included. Therefore effects of multicollinearity may occur in models incorporating site variables like site index, thinning or class age that may be correlated with basic measurational variables (DBH, heigt). The existence of multicollinearity is not a violation of the assumptions underlying the use of regression. Therefore it does not seriously affect the predictive ability. However the presence of multicollinearity may cause that the variance of the predicted values is inflated and the standard errors of the regression coefficients have larger variance with consequent lack of statistical significance [Rojo et al., 2005].

To evaluate the presence of multicollinearity among variables in models, Rojo et al. [2005] developed *condition number*. The condition number is defined as the square root of the ratio of the largest to the smallest eigenvalue of the correlation matrix. It has

been found [Rojo et al., 2005] that very low effect of multicollinearity occurs in the three-parameter model of Riemer et al. [1995].

A number of studies deal with comparison of different models for various purposes. Most of recent works comparing taper models [Bi and Long, 2001; Rojo et al., 2005; Li and Weiskittel, 2010] report the variable-form taper models to have superior performance than the segmented polynomial models or single models. Their better performance is a result of their flexibility to depict changes in stem form between individual trees of different size [Bi and Long, 2001]. Rojo et al. [2005] carried out a comparison of total 31 taper models from all three groups of models (single, segmented and variable exponent models) that were fitted on diameter-height data of 203 trees of maritime pine (*Pinus pinaster* Ait.). He found important differences among the three groups of models, but not among the variable-form models that showed the best performance. Li and Weiskittel [2010] compared 10 segmented polynomial and variable-form models for the purpose of developing a regional taper equation for the primary conifer species of Acadian region of North America - balsam fir, red spruce and white pine. Although for red spruce and white pine stem prediction was the model of Kozak [2004] the most accurate, for all three species the Clark et al. [1991] segmented polynomial equation was proved to be the best. The equation of Clark et al. [1991] also outperformed other segmented polynomial models in the comparison carried out by Figueiredo-Filho et al. [1996b].

On the other hand by certain circumstances also single polynomial taper model was shown to perform better than both segmented polynomial and variable-form models [Goodwin, 2009]. Selecting the best equation is difficult and it depends on species, dataset and intended use [Li and Weiskittel, 2010].

2.1.8 Spline model forms

To describe the stem form entirely polynomial splines have been used. The most common polynomial spline is the cubic spline, which was considered to be very suitable to describe stem profiles [Kaufmann, 2001]. Despite many favorable properties a cubic spline does not preserve the monotony. Even for monotonically increasing or decreasing input points the resulting curve can due to its intrinsic properties produces an oscillation. The oscillations are mainly caused by the continuity of the second derivative [De Boor, 2001] and cannot therefore be totally removed [Lahtinen, 1988]. The most frequently utilized special case of cubic interpolation spline is the natural cubic spline, which was used in most

works dealing with the spline representation of stem curve [Liu, 1980; Hradetzky, 1981; Smaltschinski, 1983; Kirchner et al., 1991; Rios, 1997; Sloboda et al., 1998].

Except the cubic interpolating spline also a quadratic spline has been used [Lahtinen, 1988; Lahtinen and Laasasenaho, 1979]. The usual quadratic interpolating spline was found to be inferior to the interpolating cubic spline [Lahtinen and Laasasenaho, 1979]. However, the quadratic interpolating spline, in contrast to the cubic spline, can be transformed to a monocity preserving quadratic interpolating spline [Lahtinen, 1988, 1990; Schumaker, 1983]. which has an advantage of the preservation of monotony and thus elimination of oscillations. On the other hand, the quadratic spline is theoretically a little stiffer at the butt, which arises from the smaller degree of the polynomial pieces. Moreover the quadratic spline may have a more angular shape than the cubic spline. The reason for this is that the quadratic spline has only one continuous derivative while the cubic spline has two.

Smaltschinski [1983] worked with natural cubic spline curves on data of Douglas firs. He stated that 6 input points are enough to describe the taper curve properly. Additional input points made the curve more accurate, but from 10 input points the improvement of the accuracy was negligible. Three of the input points were represented by DBH, diameter at the relative height of 0.1 and the height of the stem. The other three input points were set by experiments to diameters at relative heights of 0.2, 0.5 and 0.7. The butt was extrapolated using free-end conditions. The achieved volume estimation had an error less than 1.5% and the interpolation error of the 8 cm diameter was less than 3%.

Lahtinen [1988] compared the performance of cubic the spline and a monotony preserving quadratic spline as a taper curve. He found, that if a tree has a regular shape then the monotony preserving quadratic spline and the cubic spline calculated from high number of input points (14 measured diameters and the fixed top diameter) give equal results. For other trees there were differences in favur of the quadratic spline. It was also shown that the monotony preserving taper spline suffers less than the cubic spline of the reduction of the number of interpolating points. The eight point monotony preserving quadratic spline gave the same total volumes and there were differences in partial volumes only at the butt. The differences in diameter were small. The five point quadratic spline offers satisfactory approximation of the taper; the total volume was identical, but there were differences in partial volumes. Outside the butt were the diameter differences reasonable small, but the spline cannot give a true shape to a singular tree. A spline with three measured diameters and the fixed top diameter can be used only for trees of regular shape. In that case it still

gives reasonable total volumes [Lahtinen, 1988].

Since the accuracy of taper curves based on spline functions depend both on the number and placement of the interpolating points, Figueiredo-Filho et al. [1996a] tested how many input points are necessary to adequately represent the stem taper by interpolate cubic spline and where are their best positions. He had totally 16 points available: two absolute points at 0.1 m, 1.3 m, thirteen relative points and the height of the stem. 96 combinations of input points consisting of 4 to 13 individual diameters were generated. All combinations included diameter at 0.1 and total height. The resulting curves were compared with the spline calculated from all 16 input points. Figueiredo-Filho et al. [1996a] stated that 10 or more well-distributed points can product a curve predicting partial volumes very well. Among 8 or more input points no individual point has a fundamental importance, but the points should be well distributed. When 7 or less points are used, the placement along the stem is very important. Poor placement can cause oscillations. Diameters near 10-15 % and 35-45 % seem to be fundamental in this case. 5 input points are enough to predict the stem form with a mean residual less than 2 %, if the topmost section of the stem is not important.

In consequence of irregularities of the stem and contamination of data points the monotocity of a taper curve is not always preserved in measured data. Therefore Liu [1980] attempts to use a smoothing procedure in lieu of interpolation. To portray stem taper of yellow poplar (*Liriodendron tulpifera* L.), Liu [1980] uses natural cubic spline and smoothing spline with application of the algorithm of Reinsch [1967]. To fit the spine curves 14 radii at 0, 2, 4, 6, 8, 10, 20, 30, 40, 50, 60, 70, 80, and 90 percent of the total tree were used as ordinates, while their associated positional heights were treated as abscissas. The resulting spline functions were evaluated at 51 points along the horizontal axis. The curve produced by the smoothing algorithm with the extent of smoothing recommended by Reinsch failed to trace the basal region curvature. Therefore Liu suggests that smoothing procedures should not be used for the derivation of a taper curve but for smoothing of isolated errors in measured data. The interpolation cubic spline gave a good result. There was no significant difference between the cubic spline calculated of 14 input points and the real measured taper curve. The volume difference of the entire stem varied from -3,3 % to 2,6 % with the mean value of -0.27 %. It is conceivable that better results can be achieved with more input points, distributed over the entire stem.

Cubic spline was used also by Trincado and Vidal [1999]; Trincado and Sandoval [2002] for volume estimation of stems of *Nothofagus pumilio*. Trincado and Vidal [1999]

modeled the stem form using four input points at the stump height, at the breast height, at 40 % of the total height and at the top of the stem. In spite of relatively high errors both in diameter predictions and sectional volume estimations Trincado and Vidal [1999] consider the cubic spline as a suitable instrument for estimation of measurational data and merchantable volumes. Trincado and Sandoval [2002] compares performance of cubic spline with the taper equation of Bruce et al. [1968] and reports that the spline interpolation obtains superior results. At the same time the relative height of 40 % is considered to be an important position for an input point.

Laasasenaho et al. [2005] utilizes cubic spline to forma a bark thickness curve based on several bark thickness measurement or predictions along the stem on Norway spruce trees. Also under bark diameters were interpolated using cubic splines. Resulting stem curves were used for under bark volume estimation.

Cubic smoothing spline was later employed to predict stem curves in order to determine the optimal cutting points along the stem for forest harvesters. Since it is not economical to measure the whole stem from butt to top before crosscutting, the goal was to predict the stem curve of which only a short bottom part has been measured. The idea of predicting stem curve based on cubic smoothing spline was introduced by Möttönen and Nummi [2002] and later modified by Nummi and Möttönen [2004]. Based on measured data of previously measured and crosscut stems, stem heights of three points where the stem diameter falls below a given limit are predicted. Subsequently a smoothing spline is fitted through the predicted points and the known stem curve measurements. To adjust the predicted stem curve with each new measurement Kalman filter [Kalman, 1960] is used. The approach was later modified by Koskela et al. [2006]. Only two stem points were predicted: the stem height where the stem diameter falls bellow 150 mm and the diameter at a height of 11 m. Accuracy of the stem curve prediction based on smoothing spline fitted to predicted stem points and measured diameters of the known bottom part was later compared with a stem curve prediction based on mixed-model [Lisky and Nummi, 1995] and also with model based taper equation of Kozak [1988]. It was shown that the spline-based prediction method provides highly accurate stem curve predictions. It outperforms the approach based on mixed-model and was also superior to Kozak's model-based method.

A taper and volume prediction method based on nonparametric regression techniques developed by Lappi [2006] is based on assumptions arising from results of Gaffrey et al. [1998], that show that there are very strong relationships between the diameters at breast

height and diameters at arbitrary relative height. Diameters in any relative height can be expressed as a functional value of DBH:

$$\hat{d}(h, H, D) = \alpha_1(T) \cdot D + \alpha_2(T) \quad T \in [0, 1], \tag{2.1.20}$$

where $\alpha_1(T)$ and $\alpha_2(T)$ are continuous, interpolating functions over T. Gaffrey et al. [1998] recommend to use a polynomial with the degree of six to express the functions $\alpha_1(T)$ and $\alpha_2(T)$; polynomials of degree of more than eight are not advisable because of the risk of overshooting.

The mehtod of Lappi is based on nonparametric regression functions between diameter value at a given height and diameter at breast height that were estimated with smoothing splines. At the regression estimaion stage regression functions (also called mean functions) are built for a fixed set of diameters - four diameters at absolute heights below breast height (0.1, 0.25, 0.45 and 0.75 m) and eight diameters at relative distances between the breast height and the top of the tree (0.1, 0.2 ... 0.7, 0.85) - as well as for the total height. Similar nonparametric regression functions of dependency on DBH were obtained also for variances of mean functions and for correlations between different basic dimension (diameters and height). From a known DBH or from a known DBH and total height as well as diameters in any height all basic dimension can be estimated. The set of estimated diameters is subsequently interpolated by cubic splines as well as the estimations of diameter variances at given heigths. The resulting stem model contains both the predicted stem profile curve and variance curves.

The approach was further developed by Kublin et al. [2008]. The model utilizes a semiparametric method implemented in the German national forest inventory methodology which predicts the mean diameter at location x from known diameters at 5 % and 30 % of the stem heigth:

$$E[d(x)|x, d_{0.05H}, d_{0.3H}, H] = a(x)d_{0.05H} + b(x)d^2_{0.05H} + c(x)d_{0.05H}d_{0.3H} + d(x)D_{0.05H}H, \tag{2.1.21}$$

where the regression coefficients $a(x)$,$b(x)$, $c(x)$, $d(x)$ are in the form of cubic splines. With separate population average regressions for diameter at 5 % and 30 % of height the model can be expressed as

$$E[d(x)|x, D, H] = \mu_0(x, \mu_1(D, H), \mu_2(D, H), H). \tag{2.1.22}$$

2.2 Spline functions

2.2.1 Interpolation techniques

To construct a function of which only functional values at a few some discrete points are known interpolation techniques are used. Interpolation techniques allow constructing new data points anywhere in the range of a discrete set of known data points. The techniques are based on fitting the set of data points (or its component) by an appropriate function and evaluate the functional value for any desired parameter value. For computer-aided interpolation Piegl and Tiller [1996] recommend that in the ideal situation the interpolation function belongs in a class of functions which

- are capable of precisely representing all the curves needed;
- are easily, efficiently, and accurately processed in a computer, in particular
  - the computation of points and derivatives on the curve is efficient,
  - numerical processing of the function is relatively insensitive to floating point round-of error,
  - the functions require little memory for storage and
- are simple and mathematically well understood.

Polynomial interpolation

The polynomials are a class of widely exploited functions and several interpolation techniques utilizing polynomials have been developed. Among them, the Lagrange polynomial or the Newton polynomial belong to the most widely known techniques [De Boor, 2001].

The Lagrange polynomial is a interpolation polynomial of the least degree that fits all data points in a given set. For given set of $n+1$ data points $(\mathbf{P}_i)_{i=0}^n$ and the parametrization vector expressed in general as $H = (h_i)_{i=0}^n$ is the Lagrange interpolation polynomial defined:

$$\mathbf{L}(t) = \sum_{i=0}^{n} L_i(t)\mathbf{P}_i, \quad t \in [h_0, h_n], \tag{2.2.1}$$

where L_i, $i = 0, \ldots, n$ are Lagrange polynomials of the n^{th} degree

$$L_i(t) = \prod_{j=0,\, j\neq i}^{n} \frac{t - h_j}{h_i - h_j}. \tag{2.2.2}$$

Polynomials satisfy well the last two criteria of the list mentioned above. The first criterion is not entirely fulfiled: there are a number of important curve types that cannot accurately be represented using polynomials.Moreover curves build of only one polynomial segment have the following disadvantages.

- To fulfill the given restrictions construction of polynomials of high degrees would often be necessary; e. g. fitting of n points requires polynomial of $n - 1^{\text{th}}$ degree. High degree polynomials are not capable for efficient computer processing. In addition interpolation using high-degree polynomials is often accompanied by so-called Runge's phenomenon, problem of undesirable strong oscillation at the edges of the interpolated interval.
- One-segment polynomial curves are not suitable for interactive shape design. The position of each single input point affects the shape of the whole curve. If the position of one input point is altered, the behavior of the curve is changed not only locally in the surrounding of the point, but the whole curve is influenced [Piegl and Tiller, 1996].

Piecewise interpolation

Both mentioned disadvantages of polynomial interpolation can be eliminated by utilizing piecewise polynomial curves compound of a number of lower-degree polynomial curves. The simplest method of piecewise interpolation is piecewise constant interpolation, where the desired functional value at parameter value t is assigned from the functional value of the nearest known data point. Linear interpolation, which utilizes two data points $A[x_a; y_a]$ and $B[x_b, y_b]$, determines the functional value $f(t)$:

$$f(t) = y_a + (y_b - y_a) \cdot \frac{t - x_a}{x_b - x_a}. \tag{2.2.3}$$

. Quadratic interpolation utilizes three nearest data points which are inerpolated by the Lagrange polynomial. As results from following definition of splines, piecewise interpolation is a form of splines. The piecewise constant interpolation correspods to spline interpolation of degree zero, linear interpolation corresponds to spline interpolation of degree ,1 etc.

2.2.2 Splines: definition and continuity

Spline functions were the first time defined by I. J. Schoenberg in 1943 [Najzar, 2006]. The term spline is a general tag for a very wide class of functions that are utilizable to

interpolate or smooth multivariate data. The spline is a special curve defined piecewise by polynomial functions usually of low degrees (the most often polynomials of 3^{th} degree are used) or non-polynomial functions [Späth, 1973]. Usually polynomial splines are used, since they are easiest to handle. By junction of a number of low-degree polynomials the shaping ability of high-degree polynomials can be achieved together with preservation of all advantages of low-degree polynomials. The degree of spline function corresponds to the degree of polynomials used.

Component polynomial segments of a spline curve are connected in join points called knots. The requirement that the resulting curve is continuous and smooth can be mathematically expressed as three conditions:

- The functional value must be continuous at the knot,
- the direction of tangent vectors must be continuous at the knot and
- the curvature vector must be continuous at the knot both in magnitude and direction [Manning, 1974].

These conditions are generally sufficient to ensure the curve appears smooth in knots. The eye is not able to detect the position of the knot on the resulting curve. A discontinuity of the functional value appears as a break of the resulting curve at the knot. A discontinuity of the tangent vector can be observed as a corner. A discontinuity of the curvature vector can be detected by practiced eye as a step change of torsion of the curve.

The tree conditions are equivalent with the requirement, that the functional value, the first derivative and the second derivative of the curve are continuous. The third derivative of the curve is not required to be continuous, because a discontinuity in the rate of change of curvature of a curve is not visible [Manning, 1974].

Parametric continuity of k^{th} order (expressed also as C^k continuity) indicates that at the knots vectors of first k derivatives of both functions are identical [Linkeová, 2007]. The above requirement can also be expressed that the curve should be C^2 continuous. Still there are some spline functions that have C^1 continuity only; their second derivative is not continuous. Step changes of the curvature at the knots can be observed. Such spline curve lacks the smoothness of C^2 curves.

2.3 Utilized types of splines

Spline is a general name for a very wide class of functions with a huge number of different types of spline functions [Najzar, 2006]. In accordance to the interpretation of input data two essential approaches to the construction of the mathematical model of the curve can be distinguished: interpolation and approximation [Linkeová, 2007]. In the case of interpolation the resulting interpolation spline curve fits all points in the given sequence of input points. The result of approximation is an approximation spline that does not necessarily fit the input points, nevertheless the curve follows the geometry of the input points.

2.3.1 Ferguson cubic interpolation

Ferguson cubic is a cubic curve formed of a single segment interpolating two known input points [Linkeová, 2008]. If two input points $\mathbf{P}_i$ and $\mathbf{P}_{i+1}$, a tangent vector $\mathbf{p}_i$ at the point $\mathbf{P}_i$ and a tangent vector $\mathbf{p}_{i+1}$ at the point $\mathbf{P}_{i+1}$ are given, the vector equation of Ferguson's cubic is expressed as a linear combination of basis functions $F_0, \ldots, F_3$, where as the particular weights act the coordinates of the input points and the tangent vectors in both input points. Ferguson cubic in parameter value t is given by

$$\mathbf{C}(t) = F_0(t)\mathbf{P}_i + F_1(t)\mathbf{P}_{i+1} + F_2(t)\mathbf{p}_i + F_3(t)\mathbf{p}_{i+1}, \quad t \in [0,1], \tag{2.3.1}$$

where the basis functions

$$\begin{aligned} F_0(t) &= 2t^3 - 3t^2 + 1, \\ F_1(t) &= -2t^3 + 3t^2, \\ F_2(t) &= t^3 - 2t^2 + t, \\ F_3(t) &= t^3 - t^2, \quad t \in [0,1], \end{aligned} \tag{2.3.2}$$

are the Hermite polynomials of 3^{th} degree [Linkeová, 2008].

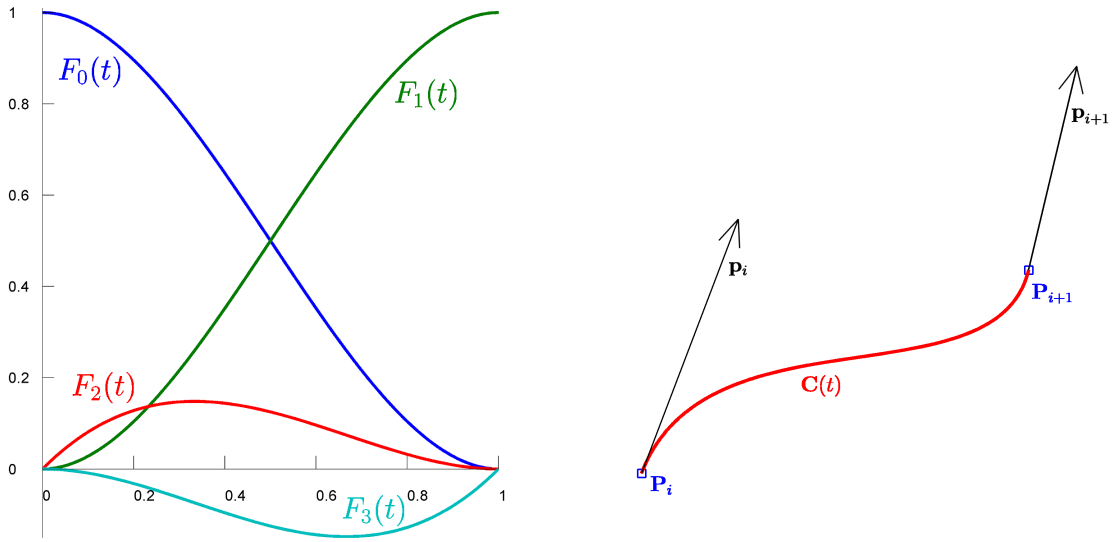


Figure 2.1: Basis function of Fergusons's cubic curve, also called Hermite polynomials, and a Fergusons's cubic defined by points $\mathbf{P}_i$ and $\mathbf{P}_{i+1}$ and tangent vectors $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$

The functional value $F_0(t = 0)$ is equal to one, while all other Hermite polynomials are zero at the parameter value $t = 0$. The same is valid for the basis function $F_3(t = 1)$. Therefore the curve starts at the coordinates of the point $\mathbf{P}_i$ and ends exactly at the coordinates of the point $\mathbf{P}_{i1}$.

2.3.2 Cardinal splines and Catmull-Rom spline

The term *cardinal splines* denotes a general class of interpolating cubic splines compound of Ferguson cubics where the tangent vector $\mathbf{p}_i$ is defined as a multiple of the chord connecting the previous and the next interpolated point:

$$\mathbf{p}_i = \tau(\mathbf{P}_{i+1} - \mathbf{P}_{i-1}), \tag{2.3.3}$$

where the constant τ known as *"tension"* affects how sharply the curve bends at the interpolated points. As a consequence of identity of tangent vectors in the endpoint of each segment and in the startpoint of the following segment the C^1 continuity in knot points between two segments is assured. From the way tangent vectors are computed results the matter of fact that in the first and the last input point no tangent vector is defined. The cardinal spline does not interpolate the first and the last input point. That

problem can be solved by duplicating the first and last point or equivalently by setting the tangent vector in the first point to $\mathbf{p}_0 = \tau(\mathbf{P}_1 - \mathbf{P}_0)$ and analogously in the last point.

A particular example of the class of cardinal splines is Catmull-Rom spline [Catmull and Rom, 1974] for which the value of parameter τ is $1/2$. The tangent vector for Catmull-Rom spline is calculated:

$$\mathbf{p}_i = \frac{1}{2}(\mathbf{P}_{i+1} - \mathbf{P}_{i-1}), \tag{2.3.4}$$

which is simply the average of the previous chord $\mathbf{P}_i - \mathbf{P}_{i-1}$ and the following chord $\mathbf{P}_{i+1} - \mathbf{P}_i$ [Kochanek and Bartels, 1984].

2.3.3 Interpolating Ferguson curve

Interpolation method piecewiese interpolating input points utilizing Ferguson cubics is proposed by Linkeová [1999, 2001]. Tangent vectors are computed from coordinates of input points and through the use of forming parameters corrections of the length as well as the direction of tangent vectors are applied. Tangent vectors $\mathbf{p}_0, \mathbf{p}_1 \ldots, \mathbf{p}_n$ in input points $\mathbf{P}_0, \mathbf{P}_1 \ldots, \mathbf{P}_n$ are determined by the help of Lagrange polynomial interpolation curves $\mathbf{L}_i(t)$ compuded always for three points $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2; \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3; \ldots \mathbf{P}_{n-2}, \mathbf{P}_{n-1}, \mathbf{P}_n$. Lagrangian tangent vector $\mathbf{L}'_0$ for three points $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2$ is expressed as linear combination of input point coordinations and first derivatives $L'_i(t)$, $i = 0, 1, 2$ of Lagrangian polynomials:

$$\mathbf{L}'_0 = L'_0(0)\mathbf{P}_0 + L'_1(0)\mathbf{P}_1 + L'_2(0)\mathbf{P}_2. \tag{2.3.5}$$

The first derivatives $L'_i(t)$, $i = 0, 1, 2$ of Lagrangian polynomials are given by equations:

$$\begin{aligned} L'_0(t) &= \frac{2t - t_1 - t_2}{(t_0 - t_1)(t_0 - t_2)}, \\ L'_1(t) &= \frac{2t - t_0 - t_2}{(t - 1 - t_0)(t_1 - t_2)}, \\ L'_2(t) &= \frac{2t - t_0 - t_1}{(t_2 - t_0)(t_2 - t_1)}, \end{aligned} \tag{2.3.6}$$

where $t_i, i = 0, 1, 2$ denotes parameter value at input points $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2$.

Resulting tangent vectors $\mathbf{p}_i(t)$ determining Ferguson cubics at input points $\mathbf{P}_i(t)$ are given

by following equation:

$$\mathbf{p}_i(t) = \begin{cases} L'_i(t = \mathbf{P}_i) & i = 0 \\ L'_i(t = \mathbf{P}_i) + L'_{i-1}(t = \mathbf{P}_i) & i \in [1, n-2] \\ L'_{i-1}(t = \mathbf{P}_i) + L'_{i-2}(t = \mathbf{P}_i) & i = n-1 \\ L'_{i-2}(t = \mathbf{P}_i) & i = n \end{cases} \tag{2.3.7}$$

Tangent vectors set in this way are well determined in case of no strong irregularities in input point distribution; irregular point distribution can cause overshoots. Therefore corrections both for directions and lengths of the tangent vectors are established. The direction correction is applied in the event that a change of driving polygon direction exceeds a preset value of the forming parameter k_d (empirically set to $k_d = 30°$). In such case the resulting tangent vector is given by only one of its components (equation 2.3.7) reducing the direction alternation.

The forming parameter k_l for length correction is expressed as equation:

$$k_{d_i} = 1 + \frac{k_{d_{\max}} - k_{d_{\min}}}{180°}(180° - \gamma_i), \tag{2.3.8}$$

where $k_{d_{\max}}$ and $k_{d_{\min}}$ are empirically set to values $k_{d_{\max}} = 3$; $k_{d_{\min}} = 1$ and γ_i is the angle between vectors $\mathbf{P}_{i-1} - \mathbf{P}_i$ and $\mathbf{P}_i - \mathbf{P}_{i+1}$.

2.3.4 Natural cubic spline

Interpolating cubic spline is the most widely used interpolation spline due to number of its advantageous properties. The effectiveness of interpolation using cubic spline is a consequence of its strong convergence property and best approximation property [Liu, 1980]. It has been proved that a continously differentiable fuction can be interpolated by a cubic spline with any prescribed accuracy by using a sufficient high number of interpolating points in contrast to an interpolating polynomial which does not allways have this property [De Boor, 2001; Lahtinen, 1988]. Interpolating cubic spline has also a minimum curvature property; it has minimal curvature expressed as functional

$$\Phi(u) = \int_a^b (u''(x))^2 \, \mathrm{d}x \tag{2.3.9}$$

among twice differentiable interpolating functions [Marčuk, 1987; Prautzsch et al., 2002].

Cubic spline is defined by usage of cubic polynomials on interspaces between individual knots and by continuity of the first and second derivatives in all the points of the curve. Specification of the input point positions does not determine the curve uniquely; for unicity of determination of spline function specification of two more conditions is required. Generally values of first of second derivative at the end points are set; most frequently the so-called free boundary condition $g''(a) = 0$, $g''(b) = 0$ is chosen which means that the second derivative of the spline function g at the beginning point a and at the end point b are set to zero [Ježek, 2000]. Zero curvature at the edge points causes a "natural" look of the resulting curve, which gives the curve its name.

Natural cubic spline is given by equation 2.3.10 [Marčuk, 1987]:

$$g(x) = m_{i-1}\frac{(x_i - x)^3}{6h_i} + m_i\frac{(x - x_{i-1})^3}{6h_i} + \left(f_{i-1} - \frac{m_{i-1}h_i^2}{6}\right)\frac{x_i - x}{h_i} + \left(f_i - \frac{m_i h_i^2}{6}\right)\frac{x - x_{i-1}}{h_i}, \tag{2.3.10}$$

where f_i denotes functional value at the point x_i, $h_i = x_i - x_{i-1}$ and unknowns m_i are calculated from a system of linear equations 2.3.11:

$$\mathbf{Am} = \mathbf{Hf}. \tag{2.3.11}$$

The square tridiagonal symmetric matrix $\mathbf{A}$ is defined by equation 2.3.12, the rectalgle tridiagonal matrix $\mathbf{H}$ is given by equation 2.3.14 and the vector $\mathbf{f}$ is the vector of length n containing known funcional values.

$$\mathbf{A} = \begin{pmatrix} \frac{h_1 + h_2}{3} & \frac{h_2}{6} & 0 & \cdots & 0 & 0 \\ \frac{h_2}{6} & \frac{h_2 + h_3}{3} & \frac{h_3}{6} & \cdots & 0 & 0 \\ 0 & \frac{h_3}{6} & \frac{h_2 + h_3}{3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{h_{n-1}}{6} & \frac{h_{n-1} + h_n}{3} \end{pmatrix} \tag{2.3.12}$$

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_{n-1} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \tag{2.3.13}$$

$$\mathbf{H} = \begin{pmatrix} \frac{1}{h_1} & \left(-\frac{1}{h_1} - \frac{1}{h_2}\right) & \frac{1}{h_2} & \cdots & 0 & 0 \\ 0 & \frac{1}{h_2} & \left(-\frac{1}{h_2} - \frac{1}{h_3}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \left(-\frac{1}{h_{n-1}} - \frac{1}{h_n}\right) & \frac{1}{h_n} \end{pmatrix} \tag{2.3.14}$$

2.3.5 B-spline

Regarding the presence of random errors of measurements that occur at the determination of input point coordinates also approximation splines can be used. The class of B-splines is the most important representative of approximation splines. In case of approximation, the curve does not need to fit all the given input points, but follows the geometry of the input points. Therefore the input points $\mathbf{Q}_0, \mathbf{Q}_1 \ldots, \mathbf{Q}_n$ are called driving points and the ordered set of driving points $(\mathbf{Q}_i)_{i=0}^n = (\mathbf{Q}_0, \mathbf{Q}_1 \ldots, \mathbf{Q}_n)$ is called the driving polygon [Piegl and Tiller, 1996; Linkeová, 2007].

The B-spline is allways smoother than its driving polygon [MacCallum and Zhang, 1986] and whole B-spline curve allways lie within the convex hull of the driving polygon. Therefore maximum deviations of the curve from the driving points are limited. Connections of the individual segments satisfies the conditions of C^{p-1} continuity, where p is the degree of the curve. The letter B in the term B-spline comes from the initial letter of the word basis, because B-spline is upon so called basis functions.

Parameter values in that segments of B-spline curve are joined are called knots. The ordered non-decreasing sequence of $(m+1)$ real numbers $T = (t_i)_{i=0}^m = (t_0, t_1 \ldots, t_m)$ is called knot vector. In B-spline the knot vector is allways uniform which means that the spacing between knots is of equal length, $t_{(i+1)} - t_i = \Delta t = \text{constant}, i = 0, 1, \ldots, m-1$. On a given knot vector T the B-spline basis functions $N_{i,p}(t), i = 0, \ldots, n$ of the degree p are defined by recursion formula 2.3.15.

$$N_{i,0}(t) = \begin{cases} 1, & t \in [t_i, t_{i+1}), \\ 0, & t \notin [t_i, t_{i+1}), \end{cases} \tag{2.3.15}$$

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k} - t_i} N_{i,k-1}(t) + \frac{t_{i+k+1} - t}{t_{i+k+1} - t_{i+1}} N_{i+1,k-1}(t), k = 1, \ldots, p.$$

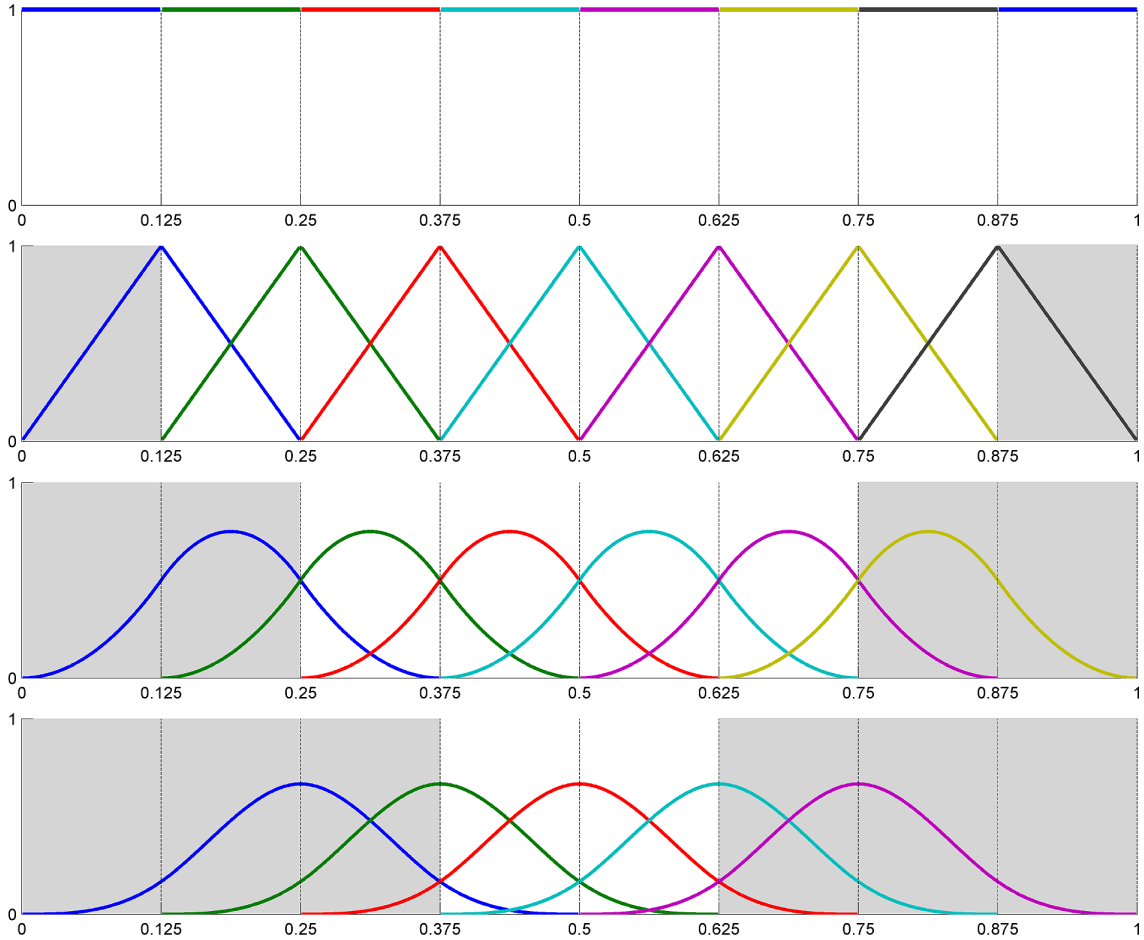


Figure 2.2: B-spline basis functions of degree zero (at the top) to three (at the bottom) for an open B-spline defined by an uniform knot vector $T = (0, 1/8, 2/8, \cdots, 8/8)$. Segments without the full support property (passive segments) are filled with grey colour. B-spline basis functions can be apprehended as courses of weihgts for individual driving points. The B-spline coordinates at parameter value t is then apprehended as an average of driving point coordinates weighted by basis function values $N_{i,p}(t)$.

The sum of B-spline basis functions in active B-spline segments is constant and equal to one. The B-spline curve $C(t)$ of degree p is then constructed as linear combination of

driving point coordinates and basis functions of degree p based on equation 2.3.16.

$$C(t) = \sum_{i=0}^{n} N_{i,p}(t)\mathbf{Q}_i, \tag{2.3.16}$$

In respect to relation between the driving polygon and the knot vector three types of B-spline are distinguished. Open B-spline does not fit any of the driving points. Its knot vector is formed by non-decreasing sequence of knots:

$$T = (t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_m). \tag{2.3.17}$$

The full support property (active B-spline segments) is not fulfilled in the whole knot range $t \in [t_0, t_m]$, but only on the interval $[t_p, t_{m-p}]$, where exactly $p+1$ basis function gather non-zero values (Figure 2.2). The pasive segments belonging to parameter values in intervals $[t_0, t_p)$ and$(t_{m-p}, t_p]$ conect the outer points of active B-spline curve with the coordinate basic origin (the outer points of the whole open curve have coordinates [0,0], because all basis function at knots $t = t_0$ and $t = t_p$ are zero) and are ignored. Therefore the parameter range of open curves is only the interval $[t_p, t_{m-p}]$.

Clamped B-splines exactly fit the first and the last point of the driving polygon. The knot vector contains at the beginning and at the end multiple knots of $(p+1)$ multiplicity and the inner knots generate again a non-decreasing sequence:

$$T = (t_0 = t_1 = \cdots = t_p \leq t_{p+1} \leq t_{p+2} \leq \ldots \leq t_{m-p-1} \leq t_{m-p} = \cdots = t_{m-1} = t_m). \tag{2.3.18}$$

As a consequence of multiple knots with $(p+1)$ multiplicity p knot spans of zero length are generated both at the begining and the end of the knot vector. Therefore the passive segments are reduced to a point (concretely to the driving point P_0 at the beginnig of the curve and to the driving point P_n at the end of the curve) and the clamped curve contains active segments only. Basis functions associated to the outer points of the driving polygon are at the outer values of the knot vector equal to one and the clamped curve always interpolate these points.

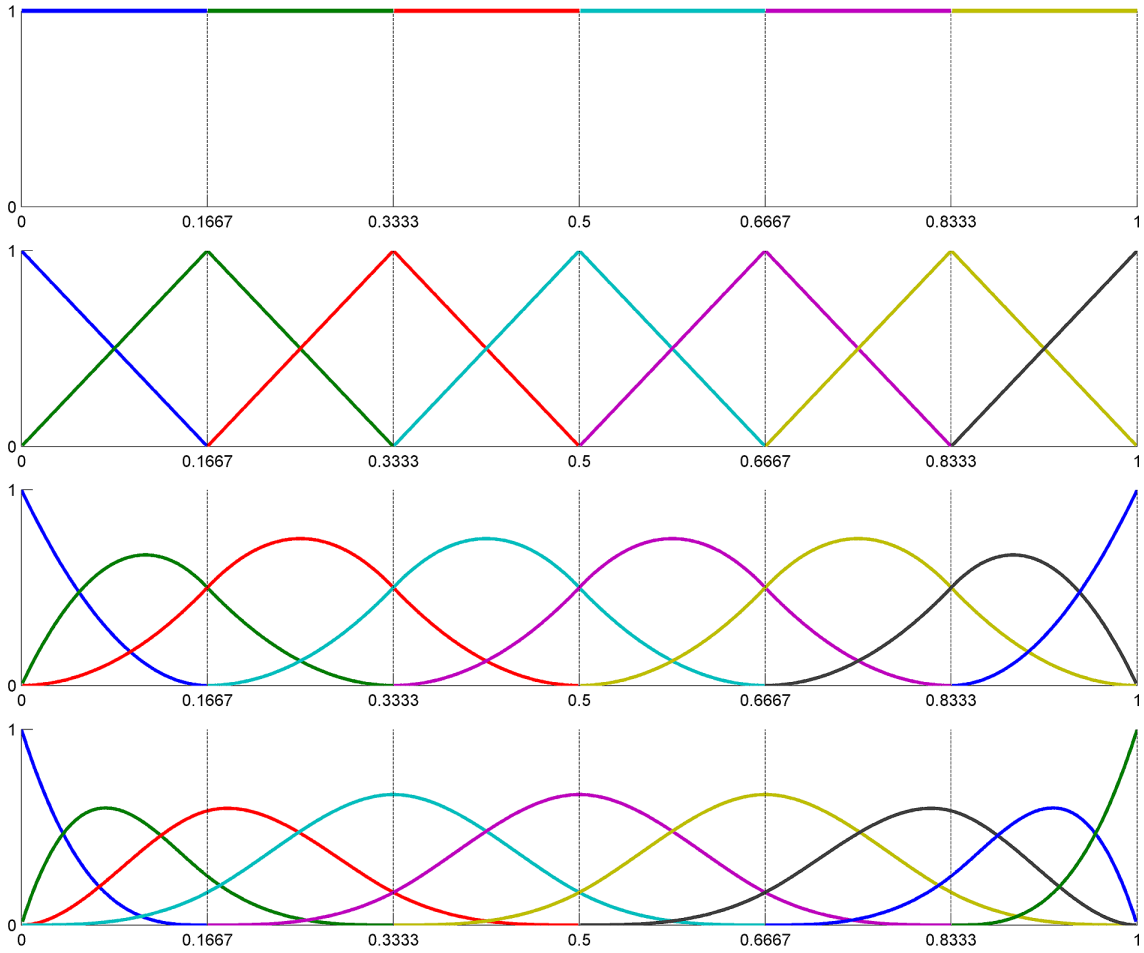


Figure 2.3: B-spline basis functions of degree zero to three for a clamped B-spline curve with an uniform knot vector $T = (0, 1/6, 2/6, \cdots, 6/6)$. All segments of clamped B-splines are active with $p+1$ non-zero basis function and with sum of basis functions equal to one. Basis functions related to the first and to the last driving points gather the value of one at both endpoints. This indicates that the full weight is given to the end point coordinates and the curve exactly fits the point.

Closed B-spline is such curve that has the first driving point identical with the last driving point whereas the continuity of the curve in that point corresponds to the continuity of the whole curve. In order to create such curve, p initial driving points have to be repeated at the end of the driving polygon:

$$\mathbf{Q}_{n-p+1} = \mathbf{Q}_0, \mathbf{Q}_{n-p+2} = \mathbf{Q}_1, \ldots, \mathbf{Q}_n = \mathbf{Q}_{p-1}. \tag{2.3.19}$$

The knot vector is then formed by non-decreasing sequence of knots:

$$T = (t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_m). \tag{2.3.20}$$

As well as the open curve does, the closed curve has an active range at the interval $[t_p, t_{m-p}]$ and the passive segments are ignored [Piegl and Tiller, 1996; Linkeová, 2007].

2.3.6 NURBS (Non-Uniform Rational B-spline)

Non-uniform rational B-splines are a generalization of B-splines. The firs step of generalization is the non-uniformity of knot vector. The distance between individual knots is not constant; the knots can be defined in any distances. As a consequence the basis functions are modificated and the curve shape differs from the B-spline. The second step of generalization of B-splines to NURBS is the racional property. Each driving point $\mathbf{Q}_i$ has assigned a weight w_i expressing the relative importance of the concrete driving point on the shape of the resulting curve compared to the importance of the other driving points. As a consequence, instead of being piecewisely compond of polynomial functions, the NURBS are in general composed of rational functions.

NURBS curve $C(t)$ of the degree p is given by equation:

$$C(t) = \sum_{i=0}^{n} R_{i,p}(t)\mathbf{Q}_i, \tag{2.3.21}$$

where $R_{i,p}(t)$ are rational basis functions defined by following:

$$R_{i,p}(t) = \frac{N_{i,p}(t)w_i}{\sum_{j=0}^{n} N_{j,p}(t)w_j}. \tag{2.3.22}$$

The consequence of weight implementation is the rational property of the curve; individual segments can be analytically expressed as a proportion of two polynomials. This property allows the NURBS to exactly form even general curves that cannot be expressed by polynomials, e. g. conic sections [Procházková and Procházka, 2007].

Contrary to B-spline, the shape of NURBS can be modified in several ways. Except coordinations of driving points themselves the weights of driving points can be used to emphasize the relative improtance of concrete driving points. The weights can be set interactively by user or can be generated automatically from the configuration of the driving point using several optimization methods. One of the most common methods of automatical weight

optimization is the averaging method (equation 2.3.23). The averaging algorithm designates weight distribution so that the resulting curve passes all input points at relatively equal distances [Linkeová, 2007]:

$$\begin{aligned} w_i &= \frac{1}{i+1}\sum_{j=0}^{i}|\mathbf{Q}_i\mathbf{T}_j|, \quad i < p, \\ w_i &= \frac{1}{p+1}\sum_{j=i-p}^{i}|\mathbf{Q}_i\mathbf{T}_j|, \quad p \le i \le n-p, \\ w_i &= \frac{1}{n-i+1}\sum_{j=i-p}^{n-p}|\mathbf{Q}_i\mathbf{T}_j|, \quad i > n-p, \end{aligned} \tag{2.3.23}$$

where the partial driving polygon centers $\mathbf{T}_i$ is given by

$$\mathbf{T}_i = \frac{1}{p+1}\sum_{j=i}^{i+p}\mathbf{Q}_j, \quad i = 0, \ldots, n-p. \tag{2.3.24}$$

The averaging weihgth w_i of the driving point $\mathbf{Q}_i$ is then equal to the average distance of the driving point from the partial driving polygon centers of all segments in which the point $\mathbf{Q}_i$ takes part on the construction of the curve [Linkeová, 2007].

The shape of the curve is influenced also by the degree of spline. With lower degrees of splines each segment is defined by a few driving points and each driving point has a height importance on the shape of the curve. The curve then follows the driving polygon very closely. In contrast, sements of higher degree curves are defined by higher number of driving points and therefore each of the driving points has relatively low impact on the curve shape. Higher degree curves are smoother [Piegl and Tiller, 1996; Linkeová, 2007]. The maximum degree that can be chosen for a curve is equal to the number of spans of the driving polygon. In that case, the curve is compound of only one rational segment.

The third magnitude influencing the shape of the resulting curve is the knot vector. The same way as the weights the knot spacing can be set manually or by optimizing algorithms. The most usually the uniform knot vector, the averaging method [Linkeová, 2007] or the chord method [Linkeová, 2007] are used. Implementation of multiple knots causes decrease of degree of continuity in corresponding points. By implementing multiple knots also sharp edges can be modeled [Procházková and Procházka, 2007]. The uniform

knot vector is given by equation 2.3.25.

$$\begin{aligned} t_i &= 0, \quad i = 0, \ldots, p, \\ t_i &= \frac{i-p}{m-2p}, \quad i = p+1, \ldots, m-p-1, \\ t_i &= 1, \quad i = m-p, \ldots, m. \end{aligned} \tag{2.3.25}$$

The averaging knot vector is given by equation 2.3.26.

$$\begin{aligned} t_i &= 0, \quad i = 0, \ldots, p, \\ t_i &= \frac{1}{L} \sum_{j=1}^{i-p} l_j, \quad i = p+1, \ldots, m-p-1, \\ t_i &= 1, \quad i = m-p, \ldots, m, \end{aligned} \tag{2.3.26}$$

where the partial driving polygon centers $\mathbf{T}_i$ are given by

$$\begin{aligned} \mathbf{T}_0 &= \mathbf{Q}_0, \\ \mathbf{T}_i &= \frac{1}{p+2} \sum_{j=i-1}^{i+p} \mathbf{Q}_j, \quad i = 1, \ldots, n-p, \\ \mathbf{T}_{n-p+1} &= \mathbf{Q}_n, \end{aligned} \tag{2.3.27}$$

the distance of two subsequent partial driving polygon centers l_i is

$$l_i = |\mathbf{T}_{i-1} - \mathbf{T}_i|, \quad i = 1, \ldots, n-p+1, \tag{2.3.28}$$

and the sum of all these distances L

$$L = \sum_{i=1}^{n-p+1} l_i. \tag{2.3.29}$$

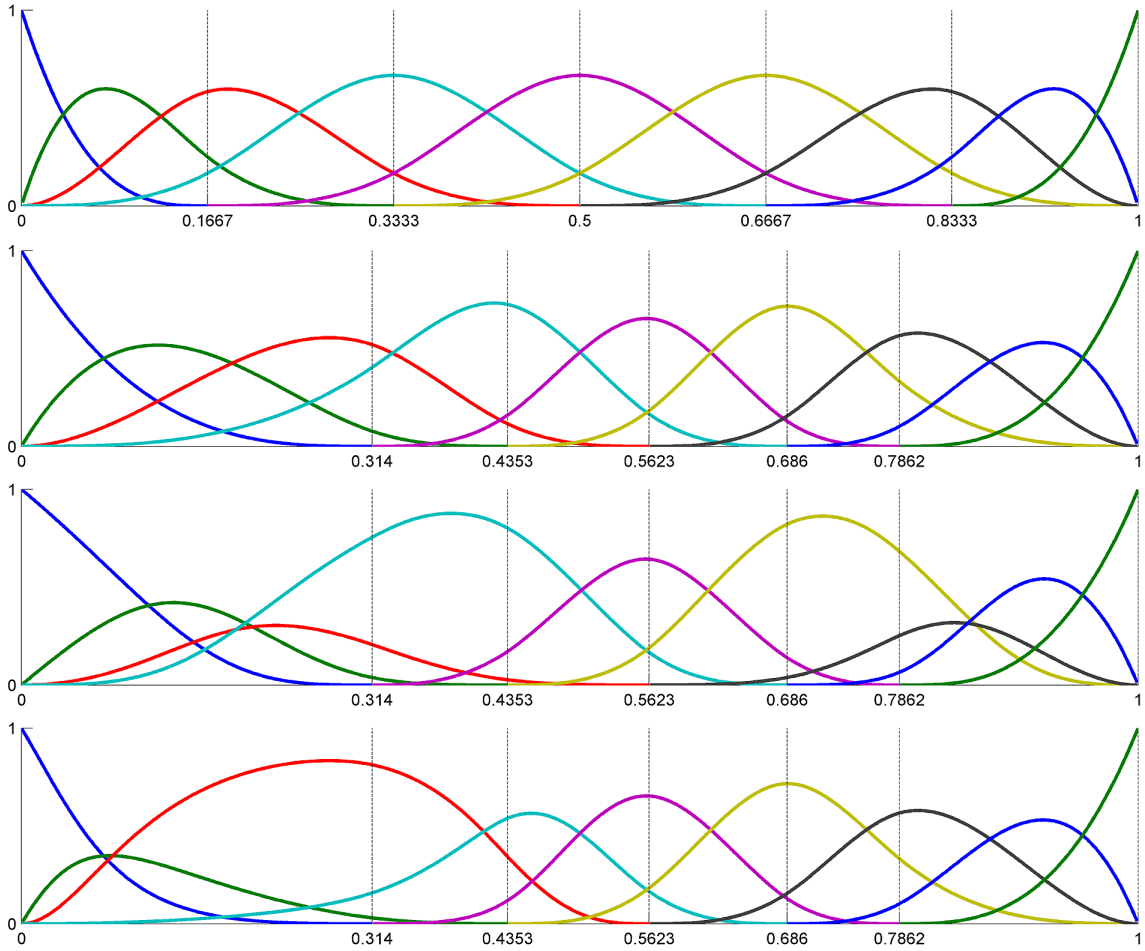


Figure 2.4: 3^{rd} degree rational basis functions of a clamped NURBS with different knot vectors and weight distribution. From above: Uniform knot vector $T = (0, 0, 0, 0, 1/6, 2/6, \ldots, 1, 1, 1, 1)$ and uniform weight distribution $W = (1, 1, \ldots, 1)$. Averaging knot vector and uniform weight distribution. Averaging knot vector and averaging weight distribution. Averaging knot vector and interactively set weights $W = (1, 1, 4, 1, \ldots, 1)$.

2.3.7 Interpolation B-spline

B-splines and NURBS can serve not only as approximation curves but also as interpolation curves [Linkeová, 2007; Eberly, 2008]. In that case the curve exactly fits all points in the given sequence.

In order that B-spline or NURBS curves, which are by their definition approximation

splines, can interpolate a sequence of $\tilde{n}+1$ given definition points $\mathbf{P}_i$ a sequence of yet unknown $n+1$ driving points $\mathbf{Q}_i$ must be found in that way that B-spline or NURBS approximating the new driving polygon exactly fits all the orginal definition points. This is achieved by arranging an equation system whose solution are coordinates of the new driving polygon. That is an inverse problem to the problem of constructing the approximation curve [Linkeová, 2007].

Two types of interpolation curves are distinguished. In the first case the construction of an interpolation curve is coming out of the condition that the number newly found driving points $\tilde{n}+1$ is identical to the number of the given definition points $n+1$. Such interpolation curve is called simple interpolation curve.

The second approach sets the condition that the number of active segments of the resulting interpolating curve is identical to the number of spans in the definitiol polygon $\tilde{n}$ and that the knots of the resulting curve correspond with the definition points. Each segment of the interpolation curve then fits two following points as well as the segments of cubic interpolation splines. Such interpolation curve is called knot interpolation curve. Additional conditions must be defined so that the knot interpolation curve is uniquely defined. As additional conditions most frequently the vectors of derivatives of the curve in the edge points are used.

As the first step in construction of the interpolation spline a parametrization vector must be defined. The parametrization vector H is a non-decreasing sequence of $\tilde{n}+1$ real numbers representing the parameter values of the curve at each definition point:

$$H = (h_0 \leq h_1 \leq h_2 \leq \ldots \leq h_{\tilde{n}}). \tag{2.3.30}$$

There are several methods to obtain the parametrization vectors. The uniform parametrization vector $H = (h_i)_{i=0}^{\tilde{n}}$ with constant distance of parameter values is given by

$$h_i = \frac{i}{\tilde{n}}, \quad i = 0, \ldots, \tilde{n}. \tag{2.3.31}$$

Frequenty used is the chord length parametrization vectror that has the distances of parameter values proportional to the chord lengths of the definition polygon:

$$\begin{aligned} h_0 &= 0, \\ h_i &= \frac{\sum_{j=1}^{i} |\mathbf{P}_{j-1}\mathbf{P}_j|}{\sum_{j=1}^{\tilde{n}} |\mathbf{P}_{j-1}\mathbf{P}_j|}, \quad i = 1, \ldots, \tilde{n}. \end{aligned} \tag{2.3.32}$$

Similar to the chord length parametrization vector is the centripetal parametrization vector. The distances between parameter values are proportional to the square roor of the chord lengths of the definition polygon. This method should eliminate undesirable overshoots of the curve in cases of sharp angles in the definition polygon:

$$\begin{aligned} h_0 &= 0, \\ h_i &= \frac{\sum_{j=1}^{i} \sqrt{|\mathbf{P}_{j-1}\mathbf{P}_j|}}{\sum_{j=1}^{\tilde{n}} \sqrt{|\mathbf{P}_{j-1}\mathbf{P}_j|}}, \quad i = 1, \ldots, \tilde{n}. \end{aligned} \tag{2.3.33}$$

The simple interpolate B-spline $\mathbf{C}(t)$ is then given by (the simple interpolation NURBS is constructed by analogy):

$$C(t) = \sum_{i=0}^{\tilde{n}} N_{i,p}(t)\mathbf{Q}_i, \tag{2.3.34}$$

where $\mathbf{P}_i, i = 0, \ldots, \tilde{n}$ are the driving points. Coordinates of the driving points are obtained as a solution of equation system

$$C(h_i) = \sum_{j=0}^{\tilde{n}} N_{j,p}(h_i)\mathbf{Q}_j = \mathbf{P_i}, \quad i = 0, \ldots, \tilde{n}, \tag{2.3.35}$$

which can be expressed as matrices:

$$\begin{pmatrix} N_{0,p}(h_0) & N_{1,p}(h_0) & \ldots & N_{\tilde{n},p}(h_0) \\ N_{0,p}(h_1) & N_{1,p}(h_1) & \ldots & N_{\tilde{n},p}(h_1) \\ \vdots & \vdots & \ddots & \vdots \\ N_{0,p}(h_{\tilde{n}}) & N_{1,p}(h_{\tilde{n}}) & \ldots & N_{\tilde{n},p}(h_{\tilde{n}}) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{Q}_0 \\ \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_{\tilde{n}} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_{\tilde{n}} \end{pmatrix} \tag{2.3.36}$$

2.3.8 Iterative non-uniform B-spline

Another method to construct an interpolation curve using B-splines was presented by Lin et al. [2004]. Using the so-called iterative non-uniform B-spline a given ordered set of points can be fitted by a non-uniform B-spline of degree 3 without solving a linear equation system.

From the given input point set $\mathbf{P}_i, i = 1, 2, \ldots, n$ the initial control polygon is derived as follows:

$$\mathbf{P}_i^0 = \mathbf{P}_i, \quad i = 1, 2, \ldots, n; \quad \mathbf{P}_0^0 = \mathbf{P}_1^0, \mathbf{P}_{n+1}^0 = \mathbf{P}_n^0, \tag{2.3.37}$$

and an initial non-uniform B-spline curve $C^0(t)$ is constructed on the knot vector $t_i, i = 0, 1, \ldots, n+5$ defined as 2.3.38:

$$t_i = \sum_{j=2}^{i-2} \|\mathbf{P}_j - \mathbf{P}_{j-1}\|, \quad i = 4, 5, \ldots, n+2, \quad t_0 = t_1 = t_2 = t_3 = 0, \quad t_{n+2} = t_{n+3} = t_{n+4} = t_{n+5}. \tag{2.3.38}$$

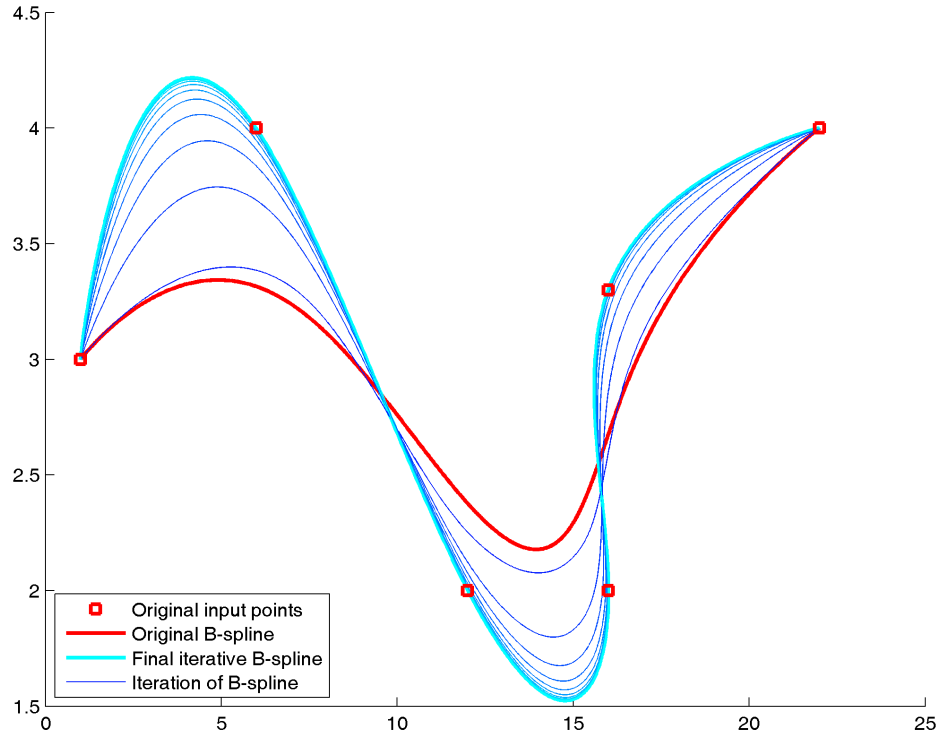


Figure 2.5: Iterative non-uniform B-spline with 10 iterations.

By gradually adjusting the positions of the control points with iterative formula, the B-spline curve approaches the given point set step by step (Figure 2.5). The iterative shift of j^{th} control point in k^{th} iteration is given by adjusting vector:

$$\mathbf{\Delta}_j^k = \mathbf{P}_j^k - C^k(t_{j+2}), \quad j = 1, 2, \ldots, n. \tag{2.3.39}$$

Then the control points of the iterative non-uniform B-spline after $(k+1)^{\text{th}}$ iteration are given by

$$\mathbf{P}_i^{k+1} = \mathbf{P}_i^k + \mathbf{\Delta}_j^k, \quad i = 1, 2, \ldots, n; \quad \mathbf{P}_0^{k+1} = \mathbf{P}_1^{k+1}, \mathbf{P}_{n+1}^{k+1} = \mathbf{P}_n^{k+1}. \tag{2.3.40}$$

Each curve of the sequence has the same non-uniform B-spline basis functions defined on the same knot vector. Only the control points are adjusted according to the iterative

format. Lin et al. [2004] proved that the limit curve of the iterative sequence interpolates the given point set, when the iterative time tends to infinity.

2.4 Regression splines and smoothing splines

Smoothing methods and nonparametric regression techniques serve the purpose of detecting a overall tendency in empiric data, finding a regression model and prediction using the nonparametric regression models. The basic aim is to estimate an unknown smooth function $f \in C^2[x_1, x_n]$ of unspecified form [Aydin, 2007].

Nonparametric regression techniques are mostly utilized for describing data sets too complicated to be fully modeled with parametric models. As reported by Eilers and Marx [1996], the name *nonparametric* does not always correspond with the principle of the methods. While moving statistic and kernel smoothing have no parameters, spline smoothers are in fact described by a huge number of parameters.

2.4.1 Simple nonparametric regression methods

The most simple smoothing methods are the group of moving statistics [Rodríguez, 2001; Aydin, 2007]. From that group the moving average is the basic method. The regression value at predictor value x_i is obtained by averaging the observation values y corresponding to predictor values x in a neighborhood of x_i:

$$f(x_i) = \sum_{j \in N(x_i)} (y_j/n_i), \tag{2.4.1}$$

for a neighborhood $N(x_i)$ with n_i observations. Usually a symmetric neighborhood consisting of the nearest $2k + 1$ points is chosen:

$$N(x_i) = \{\max(i-k, 1), \ldots, i-1, i, i+1, \ldots, \min(i+k, n)\} \tag{2.4.2}$$

The moving average can be used only for equally spaced points and near endpoints it shows high bias. Therefore rather utilization of moving line is recommended. The moving line consists of series of local lines estimated by least squares:

$$f(x_i) = \hat{\alpha}_i + \hat{\beta}_i x_i, \tag{2.4.3}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the least squares estimates based on points in a neighborhood $N(x_i)$ of x_i.

Kernel smoothing corresponds to weighted moving average with weights declining with rising distance from the target value in x-space. A weighting scheme [Aydin, 2007] sets a weight associated with observation y_j for prediction at x_i:

$$w_{ij} = \frac{K\Big(\frac{x_i - x_j}{\lambda}\Big)}{\sum_{j=1}^{n} K\Big(\frac{x_i - x_j}{\lambda}\Big)} = \frac{K(u)}{\sum_{j=1}^{n} K(u)}, \tag{2.4.4}$$

where $K(u)$ is a decreasing function of distance u from the target value and λ is the bandwith, also called the smoothing parameter. It has been shown [Aydin, 2007], that the selection of smoothing parameter is much more important for the performance of the kernel regression than the selection of the kernel function. The kernel function is supposed to be symmetric (even), non-negative and twice differentiable. Some popular kernel functions are shown in Table 2.1.

Table 2.1: Kernel function examples for kernel smoothing

Kernel	Equation	
Gaussian kernel	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2}$,	$u \in [-\infty, \infty]$
Triangular kernel	$K(u) = (1 - \lvert u \rvert)$,	$u \in [-1, 1]$
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)$,	$u \in [-1, 1]$
Minimum var	$K(u) = \frac{3}{8}(3 - 5u^2)$,	$u \in [-1, 1]$

A combination of kernel smoothing and polynomial regression was proposed by Cleveland [1979] and later improved by Cleveland and Devlin [1988]. The method is called LOESS or LOWESS (locally weighted scatterplot smoothing) and is also known as locally weighted polynomial regression. At each point in the data a low-degree polynomial is fitted to the subset of the data located in the neighbourhood of the point. The range of the neighbourhood is determined analogicaly to kernel smoothing by value of smoothing parameter λ, which is a function of the chosen degree of local polynomial.

2.4.2 Smoothing cubic spline

Smoothing cubic spline was proposed by Reinsch [1967]. It is based on the method of interpolate cubic spline. Instead of minimizing the functional 2.3.9 it is demanded to minimize functional 2.4.5

$$\Phi_1(u) = \int_a^b (u''(x))^2 \,\mathrm{d}x + \sum_{k=0}^{n} p_k (u(x_k) - \tilde{f}_k)^2. \tag{2.4.5}$$

This functional relates the requirement of minimal curvature of the function with the requirement of minimal residual sum of squares. Given positive numbers p_k are weight coefficients influencing the rate of minimal distance of the smoothing function from given points against the minimal curvature requirement. The first term denotes the curvature of the function and it penalizes its roughness; the second term denotes the residual sum of squares weighted by coefficients p_k and it penalizes lack of fit.

The rate of importance of minimization of residual sum of squares and minimization of curvature of the function can also be expressed by smoothing paramerer λ penalizing the curvature of the function [Aydin, 2007] in equation 2.4.6

$$\Phi_1(u) = \lambda \int_a^b (u''(x))^2 \,\mathrm{d}x + \sum_{k=0}^{n} (u(x_k) - \tilde{f}_k)^2 \tag{2.4.6}$$

To find the coefficients of smoothing spline functions it is necessary to determine the vector $\mathbf{m}$ by solving equation 2.4.8, where the equation system matrix $(\mathbf{A} + \mathbf{H}\mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}})$ is fivediagonal and symmetric. Square diagonal matrix $\mathbf{P}$ contains the weight cefficients p_k (2.4.7).

$$\mathbf{P} = \begin{pmatrix} p_0 & 0 & \cdots & 0 \\ 0 & p_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{pmatrix} \tag{2.4.7}$$

$$(\mathbf{A} + \mathbf{H}\mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}})\mathbf{m} = \mathbf{H}\tilde{\mathbf{f}} \tag{2.4.8}$$

The next step is to determine the vector μ of the functional values of the smoothing spline at knots by solving equation 2.4.9. After the functional values are known, the resulting smoothing spline function is easily determined using equation 2.3.10.

$$\mu = \tilde{\mathbf{f}} - \mathbf{P}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{m} \tag{2.4.9}$$

2.4.3 P-splines

P-splines [Eilers and Marx, 1996; Marx, 2010], or B-splines with difference penalties, are based on the approach of O'Sullivan [1988], who used B-splines with second derivative penalty to smooth data. To make the B-spline smooth the data, O'Sullivan [1986] uses a relatively high number of knots and introduces a penalty on the second derivative of the fited curve. The derived least squares objective function, which should be minimized, is:

$$S = \sum_{i-1}^{m} \Big(y_i - \sum_{j=1}^{n} \alpha_j N_j(x_i) \Big)^2 + \lambda \int_{x_{\min}}^{x_{\max}} \Big(\sum_{j=1}^{n} \alpha_j N_j''(x) \Big)^2 \mathrm{d}x \tag{2.4.10}$$

for regression of m data points $\mathbf{P}_i = (x_i, y_i)$ on a B-spline consinsing of n B-spline basis functions with α_j being the B-spline coefficients. Instead of the penalty of secont derivative, P-splines use penalties of k^{th}-order finite differencies:

$$S = \sum_{i-1}^{m} \Big(y_i - \sum_{j=1}^{n} \alpha_j N_j(x_i) \Big)^2 + \lambda \sum_{j=k+1}^{n} (\Delta^k \alpha_j)^2. \tag{2.4.11}$$

For example of $k = 2$ can be the difference expressed as:

$$\Delta^2 \alpha_j = \Delta\Delta\alpha_j = a_j - 2\alpha_{j-1} + \alpha_{j-2}. \tag{2.4.12}$$

To the most important properties of P-splines as a regression technique belongs no boundary effect, in contrast to many kernel smoothers. Polynomial data can be fitted exactly by P-splines and the limit of P-splines fit with strong smoothing (large values of λ) is a polynomial.

2.4.4 Amount of smoothnig

Most smoothing techniques allow the user to influece easily the amount smoothness of fitted curve by changing the velue of the parameter λ. A major problem of any smoothing technique is the choice of the optimal amount of smoothing [Eilers and Marx, 1996]. The problem of optimal smoothing was being solved by Craven and Wahba [1979]; Utreras [1981]. Eilers and Marx [1996] propose utilizing the Akaike information criterion.

The Akaike information criterion (AIC) is originally proposed to determine the relative goodness of fit of a statistical model in comparison to alternative statistical models. The general form of AIC is given by:

$$AIC = -2 \log L(\hat{\theta}) + 2k, \tag{2.4.13}$$

where $L(\hat{\theta})$ is the maximized likelihood function and k is the number of free parameters in the model. For the least squares method the AIC is given by:

$$AIC = n \log\left(\frac{RSS}{n}\right) + 2k, \tag{2.4.14}$$

where RSS denotes the residual sum of squares. When comparing models he mode with each other, thel with lowest value of AIC is considered as the best. To determine the optimal smoothing

Another technique that allows to determine the omptimal amounth of smoothing is the cross-validation method (CV) [Stone, 1974] and the generalized cross-validation (GCV) method [Craven and Wahba, 1979; Eilers and Marx, 1996]. In general approach the CV is a technique for assesing how a statistical model will generalize to an independent data set using the same data that were used to fit the model. The cross-validation method involves partitioning sample of data into complementary subsets, performing the analysis on one subset, and validating the analysis on the other subset. There are several types of cross-validation. In repeated random sub-sampling cross-validation method are data repeatedly randomly splited into training and validation data and the results of validation are then averaged. The K-fold CV randomly partitions the data into K subsamples of which a single subsample serves as validation data, while the others are used as training data to fit the model. The process is then repeated K-times so that each of the subsamples is used exactly once as the validation data. A special case of K-fold CV is the 2-fold CV, where the data are divided into two equal-sized parts. Marx [2010] recommends the usage of leave-one-out cross validation approach (LOOCV), which is a special case of K-fold CV, where K corresponds with the number of observations. In that approach a single observation serves as validation data, while all the other observations are used to fit the model:

$$\text{LOOCV} = \sum_{i=1}^{n} \left(y_i - \mu_i(\hat{\theta}_{[-i]})\right)^2, \tag{2.4.15}$$

where $\hat{\theta}_{[-i]}$ is the estimator obtained by removing the i-th sample.

2.5 Volume estimation

There are number of different log volume estimation techniques available. Three of the most used are the equations of Smalian, Huber and Newton.

Smalian's equation is given as

$$V = (A_1 + A_2)L/2, \tag{2.5.1}$$

Huber's equation as

$$V = (A_3)L, \tag{2.5.2}$$

and Newton's equation as

$$V = (A_1 + 4A_3 + A_2)L/6, \tag{2.5.3}$$

where V is the log volume (m^3), A_1 and A_2 are the cross-sectional areas of the ends of the log (m^2), A_3 is the cross-sectional area of the midpoint of the log (m^2) and L is the log length (m).

Huber's and Smalian's equations assume that the log form is parabolic. If the geometric solid is an exact frustum of paraboloid, both formulae yield exact results [Biging, 1988]. If the log form departs form parabolic form, both equations become biased. In contrast, Newton's equation is based on assumption that the taper can be expressed as a third-degree polynomial [Wensel, 1977]. In that case the equation of Newton provide exact result.

Comparison of accuracy of log volume estimation techniques has been performed, mostly by comparing computed volumes with results obtained from xylometry measurements [Goulding, 1979; Biging, 1988; Figueiredo-Filho and Schaaf, 1999; Figueiredo-Filho et al., 2000; Özçelik et al., 2008; Machado and Nadolny, 1991; Machado et al., 2006]. Results proposed Biging [1988] indicate that the Huber's and Newton's equations offer the most accurate results for logs of length 8 feet or longer. For shorter logs there were found no differences. Also the study of Figueiredo-Filho et al. [2000] comparing traditional formula with cubic spline method, centroid sampling [Wood et al., 1990] and over ng bolts [Bailey, 1995] consider Huber's method as the most accurate for longer sections, while for sections shorter than 2 m all methods give satisfactory results. That is also consistent with the conclusions of Machado and Nadolny [1991]; Machado et al. [2006]. By contrast Özçelik et al. [2008] consider the Huber's and Smalian's equation to be unsatisfactory and recommend to use center-of-gravity [Lynch et al., 1994] or centroid method. Biging [1988] qwrns that the results does not need to be unquestionable, since the xylometry method is not perfectly precise itself because of absorptivity of the wood, necessity of holding the log under the water surface by a machine and other influences. Schreuder et al. [1993] recommend to use Huber equation in case of uniform length of the logs; otherwise the Smalian's method is recommended for the sections not longer than 3 m.

When the stem profile is expressed by a taper equation, the volume can be easily and precisely obtained by integration of the taper equation [Bruce et al., 1968; Goulding and Murray, 1976; McClure and Czaplewski, 1986]. This is also the approach for developing compatible taper equations. Analytical integration can be used for taper equations with single explicit functional form. For segmented models the volume must be calculated piecewise by integrating each segment individually [Cao et al., 1980]. The same way would be applicable for spline representation of stem profile but the process would be complicated. Therefore numerical integration is usually applied.

Numerical integration methods represent an approach of getting an approximation of the integral of a function. Numerical integration is used in cases that it is not possible to find the primitive function $F(x)$ of the function $f(x)$ whose integral is wanted to be calculated, or the function $f(x)$ is too complex [Děmidovič and Maron, 1966]. In practice the function is sometimes defined by a list of discrete values and in that cases the term primitive function looses its sense.

Numerical integration methods evaluate the finite integral by splitting the interval of integration into a number of very small subintervals $< a, b >$ and to approximate the each of the areas S limited by the subinterval $< a, b >$ and the function $f(x)$. A class of numerical integration methods is based on interpolating functions. The simplest method is to let the interpolating function be a piecewise constant function which passes through the point $[(a+b)/2, f((a+b)/2)]$. The area under the function is splitted into number n of rectangles. Therefore the method is called the rectangle rule. The integral is approximated by the sum of the rectangle areas:

$$\int_a^b f(x)\mathrm{d}x \approx (b-a) \cdot f\left(\frac{a+b}{2}\right). \tag{2.5.4}$$

The rectangle rule corresponds to the volume estimation using Huber's function with a high number of very short sections.

Another simple method is called the trapezoidal rule. The interpolating function is a piecewise linear function which passes through the points $[a, f(a)]$ and $[b, f(b)]$. The area of under the function $f(x)$ is divided into n trapeze sections. The approximation of the sectional area is given by:

$$\int_a^b f(x)\mathrm{d}x \approx (b-a) \cdot \frac{f(a)+f(b)}{2}. \tag{2.5.5}$$

This trapezoidal rule corresponds to the volume estimation using Smalian's equation. With shrinking the subinterval $< a, b >$ the results of both methods tend to approach each other. In the limit case $(b-a) \to 0$ and $n \to \infty$ both methods are identical.

For computing the volume of stem represented by spline mostly trapezoidal rule (or Smalian's formula) was used[Trincado and Sandoval, 2002; Goulding, 1979]. As has been shown [Biging, 1988; Schreuder et al., 1993], there is no difference in volume estimation for short sections between both equations. Numerical integration works with extremely short sections. In this case predominate the advantage of Smalian's equation of its computational simplicity. Moreover, some spline types do not assure uniform length of the intervals between their discrete values, which practically disallows the usage of Huber's equation.

2.6 Accuracy evaluation

From statistical point of view it is sufficient and correct approach to use a single criteria for model comparisons, such as sum of squared residuals, AIC or CV. But because the taper models should be applicable for a particular purpose with a number of special demands, it is reasonable to use multiple criteria expressing different demands on the desired taper function. Amidon [1984] uses multiple evaluation criteria to overcome the consequences of autocorellation.

To compare number of taper model Cao et al. [1980] computed volumes both outside and inside bark at 10-percent intervals of total height and compared to predicted volumes from each of the models. Also predicted diameters both inside and outside bark at the points of 10-percent height intervals were compared to actual diameters. Three criteria were then applied to evaluate the models: bias (the mean of the differences), mean absolute difference and standard deviation of differences. Liu [1980] evaluated accuracy of cubic spline curves at 51 points along the height axis. He also compared the predicted and observed volume for the entire stem and for the first, second and third 5-meter sections. Smaltschinski [1983] considered the spline curve as accurate, if it fulfilled three criteria. The sum of the deviations at n measurement points has to be smaller than $0.5 \cdot n$, the volume error has to be lower than 0.5% and the interpolation error at the 8 cm diameter (merchantable thickness limit) is not higher than 4 mm.

Lahtinen [1988] has three criteria of suitability of a spline model. The most important one is that the taper curve accurately estimates the volume of the stem or of any part of it. Volume differences for total volume and for each of seven parts are evaluated. The second criterion is the magnitude of diameters error. Each taper curve is tested by evaluating the maximal diameter difference. The third criterion of suitability is the form of the taper

curve and how natural the shape is. This is a important property, but difficult to measure.

Parresol et al. [1987] and Figueiredo-Filho et al. [1996a] have consistent approach of evaluating diameters at selected fixed heights. Also the statistics are very similar. Figueiredo-Filho et al. [1996a] used eleven relative points along the stem to analyze the accuracy of the cubic spline stem form prediction. The points were purposely chosen at the midpoints of the sections between measured points. As comparative data were considered diameters predicted by spline using all 16 measured points. Differences between diameters predicted by the spline and the comparative diameters were used to calculate following statistics (table 2.2). The same statistics were used to analyze the total and partial (10 sections along the stem) volume estimations.

Table 2.2: Statistics used to evaluate predicted diameters and volumes by Figueiredo-Filho et al. [1996a]

Statistics	Calculation
Bias	$(\sum \text{Diff}_i)/N$
Mean absolute deviation	$(\sum \text{abs}(\text{Diff}_i))/N$
Standard deviation of differences	$[\sum(\text{Diff}_i - \text{Bias})^2/(N-1)]^{0.5}$
Sum of squared relative residuals	$\sum(\text{Diff}_i/d_i)^2$
Percentage of residuals	$\sum(\text{Diff}_i/d_i)100/N$

To evaluate the precision and accuracy of diameter estimates of different taper models Rojo et al. [2005] use three statistical criteria obtained from residuals: bias $\overline{E}$ (equation 2.6.1), mean square error MSE (equation 2.6.2) and adjusted coefficient of determination R^2_{adj} (equation 2.6.3), which in contrast to common coefficient of determination penalizes the statistics as extra variables are included in the model.

$$\overline{E} = \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{n}, \tag{2.6.1}$$

$$\text{MSE} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-p}, \tag{2.6.2}$$

$$R^2_{\text{adj}} = 1 - (n-1) \cdot \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-p} \cdot \sum_{i=1}^{n} (y_i - \overline{y}_i)^2, \tag{2.6.3}$$

where y_i, $\hat{y}_i$ and $\overline{y}_i$ are the measured, predicted and average values of the dependent variable, which in most cases differs from the diameter. n is the total number of observation used to fit the model and p is the number of model parameters. Moreover Rojo

ET AL. [2005] used the Akaike information criterion (AIC) [KITAGAWA, 2008] and a cross-validation approach for all three mentioned statistics. Similar criteria were used by CAO ET AL. [1980]; BI AND LONG [2001] and LI AND WEISKITTEL [2010], who used mean absolute bias, root mean squared error and mean percentage of bias.

[KOZAK AND SMITH, 1993] proposed a methodology for comparing taper functions based on the calculation of three statistics for each estimate. GOODWIN [2009] extended the approach and created an index of predictive accuracy called the MSS18 Index. For each tree in the data set, a measurement point and associated log is randomly selected. For each point or log, six predictions are made of known quantities with each taper model. Prediction errors are assimilated across size categories into three statistical measures of precision and bias, giving a total of 18 statistics for each model. Synthesis of the statistics into the MSS18 Index provides a simple basis for comparison. The six predictions are defined as follows:

1. Diameter in given height,
2. height of given diameter,
3. entire stem volume from ground to tip,
4. log volume between two heights,
5. log volume between a bottom height and top diameter,
6. log volume between a bottom diameter and top diameter.

The six predictions are then evaluated using three statistics.

1. Mean absolute bias indicates the evenness of bias across size categories.
2. Mean standard error of estimate indicates overall prediction accuracy because it incorporates both bias and precision.
3. One minus coefficient of determination is an overall measure of prediction.

Each of 18 statistics is scaled so that its mean value is 1. The MSS18 Index is then calculated as a mean of all scaled statistics. Model with the lowest MSS18 index is considered the best.

Statistical methods for determining the sample size for fitting taper models were developed by DEMAERSCHALK AND KOZAK [1974] and KITIKIDOU AND CHATZILAZAROU [2008].

Chapter 3

Methods

3.1 Data

The research uses data both of coniferous and broadleaved trees. The data of coniferous trees come from 85 Norway spruce trees (*Picea abies* [L.] Karst.). The trees were from three 50 to 100-year-old stands located in the School Forest Enterprise Kostelec nad Černými lesy, Czech Republic. In order to cover the shape variability in the stands dominant trees as well as suppressed trees were selected for the analysis. The diameter at breast height (DBH) of the trees ranges from 88 mm to 438 mm (mean 204 mm), and tree heights range from 10.6 m to 37.1 m (mean 21.3 m).

The data of broadleaved trees include two species, common beech (*Fagus sylvatica* L.) and mountain oak (*Quercus petraea* (Matt.) Liebl.). The data come from the protected landscape area Bílé Karpaty, Czech Republic. The data collection was carried out in cooperation with the research project of the Ministry of Agriculture of the Czech Republic No. QI102A079 'Research on biomass of broadleaved species' and in cooperation with the company Lesy Komňa, s. r. o. The broadleaved data consist of 48 trees of both species. The DBH of the trees ranges from 270 mm to 577 mm, and tree heights range from 14.6 m to 33.4 m.

The data sampling design is the same for both coniferous and broadleaved species. Diameters outside bark is measured on the felled trees from the tree base to the top at 0.1-m intervals so that the ordered pairs of values represent coordinates of points representing the stem profile. The distances from the tree base (abscissae) are measured using a steel

tape with 0.01-m precision, and the diameters (ordinates) are measured and recorded with an electronic caliper with 1-mm precision. Thus, more than several hundred diameters describing the stem form were obtained from each tree and were compared with the diameters predicted by splines, as described in the following sections.

3.2 Software

3.2.1 MATLAB® R2012b

MATLAB® is a programming language and interactive environment for numerical technical computation, visualization, algorithm development, data analyses and programming [The MathWorks, Inc., 2010]. MATLAB was created at the end of the 1970s; in 1984 it was launched into the market by the company The MathWorks, Inc. The software originally designated for mathematical purposes was gradually extended and presently it is utilized in a wide spectrum of applications. Nowadays, it is a world standard in the field of technical computation and simulation in the sphere of research and development, industry and education.

MATLAB is the main tool utilized for the elaboration of the thesis. All the necessary computing was performed in the MATLAB environment. MATLAB was used above all to design the routines for particular splines computation, for taper function fitting and statistical evaluation of the models. Nearly all figures presented in the thesis were obtained using MATLAB.

Possibilities of exploitation of MATLAB in particular fields are extended by the help of application libraries and packets, in MATLAB terminology called toolboxes and blocksets, containing related functions and tools. Following toolboxes were used for the purposes of the thesis.

Statistics Toolbox™

Statistics Toolbox™ is a set of functions and tools for organizing, analyzing and modeling data [The MathWorks, Inc., 2003]. The toolbox includes tools for statistical hypothesis testing, correlation and regression analysis, data classification, multivariate statistics etc. and allows presenting the results in both numerical and graphical outputs. To facilitate data processing, the toolbox also includes special data types for organizing and hetero-

geneous data, which allowes storing numeric data, text data and metadata in a single variable.

For the purposes of the thesis the Statistics Toolbox was used especially for statistical processing of computed stem curves and their residuals. Functions for cross-validation evaluation approach as well as statistical tests included in the toolbox were used to compare the performance of particular spline types. Non-linear regression analysis tools were utilized to fit taper models of polynomial, segmented polynomial or variable-exponent forms.

Curve Fitting Toolbox™
Curve Fitting Toolbox™ provides tools and functions for fitting curves and surfaces to data [The MathWorks, Inc., 2002]. The toolbox contains tools for conducting regression analyses and comparison of candidate models as well as libraries for optimizing solvers for curve and surface fitting. Also non-parametric fitting techniques, such as splines, interpolation and smoothing are included.

The functions of Curve Fitting Toolbox were utilized in preliminary analyses of the data, for verification of the created programs for spline fitting and for fitting and verification of regression models.

3.3 Individual stem curve modeling

3.3.1 Primary comparison of spline types for modeling stem curves of coniferous trees

3.3.1.1 Specification of selected splines

The first step in evaluating abilities of individual spline types to represent the stem form and selecting the best candidate was a simple comparison of particular splines. All splines evaluated in the primary comparison, together with their specifications (if needed) and abbreviations are listed in Table 3.1.

Ten representatives of interpolation splines were selected for the analysis. The first two representatives are cubic splines with first degree continuity, based on Ferguson's cubics: Catmull-Rom spline and Ferguson interpolation curve proposed by Linkeová [1999]. Further six representatives of interpolation B-splines with different degree and parameteriza-

Table 3.1: Spline types used in the study

	Abbrev.	Spline	Specification
Interpolation	CRS	Catmull-Rom spline	
	LFC	Linkeova Ferguson curve	
	IBSu2	Interpolation B-spline	degree 2; uniformly spaced parameterization
	IBSu3	Interpolation B-spline	degree 3; uniformly spaced parameterization
	IBSu4	Interpolation B-spline	degree 4; uniformly spaced parameterization
	IBSc2	Interpolation B-spline	degree 2; centripetal parameterization
	IBSc3	Interpolation B-spline	degree 3; centripetal parameterization
	IBSc4	Interpolation B-spline	degree 4; centripetal parameterization
	NCS	Natural cubic spline	
	IterBS	Iterative NUBS	degree 3; number of iterations: 50
Approximation	BS2	B-spline	degree 2
	BS3	B-spline	degree 3
	BS4	B-spline	degree 4
	NUBS2	Non-uniform B-spline	degree 2; averaging knot vector
	NUBS3	Non-uniform B-spline	degree 3; averaging knot vector
	NUBS4	Non-uniform B-spline	degree 4; averaging knot vector
	NURBSdbh2	NURBS	degree 2; averaging knot vector; weight in DBH = 3
	NURBSdbh3	NURBS	degree 3; averaging knot vector; weight in DBH = 3
	NURBSdbh4	NURBS	degree 4; averaging knot vector; weight in DBH = 3
	NURBSav2	NURBS	degree 2; averaging knot vector; averaging weight distr.
	NURBSav3	NURBS	degree 3; averaging knot vector; averaging weight distr.
	NURBSav4	NURBS	degree 4; averaging knot vector; averaging weight distr.

tion presets were selected. The last two representatives of interpolation splines are cubic splines with second degree continuity: the widely used natural cubic spline and the iterative B-spline proposed by Lin et al. [2004].

The group of selected approximation splines is built of representatives of B-splines and NURBS (non-uniform rational B-splines). For all approximation splines, clamped curves, which exactly fit the first and last control points, were used. Four approximation splines (all variants of the second, third, and fourth degree) were tested: BS, NUBS; NURBSdbh, and NURBSav. BS denotes the B-spline with uniform knot spacing. NUBS means non-uniform B-spline with the knot vector determined by averaging method, which optimizes the knot

spacing when input points are unevenly distributed. The third and the fourth variants were represented by NURBS with different weight distributions, both with averaging knot vector. In NURBSdbh, the relative importance of the input point representing the DBH was increased by setting its weight to 3, while the weights of the remaining points were set to 1; this should improve the fit of the butt swell curvature. In NURBSav, the weights were set with an averaging algorithm that optimizes weight distribution so that the resulting curve passes all input points at relatively equal distances [Linkeová, 2007].

3.3.1.2 Specification of input data points

Splines were compared using several input point distribution. By input point is meant a coordinate pair of a point measured on the stem surface where the abscissa is the above-ground height and the ordinate is the appropriate thickness. Because the spline models were fitted in terms of the stem curve (rotation of the stem curve produces the stem surface), the ordinates are expressed as the appropriate radii of the stem (half of the measured diameters).

To compare the performance of individual splines, nine input point sets with different point numbers were used. An input point set consists of two subsets of points.

The first subset contains four fixed input points determined by the above-ground height:

1. h = 0 m; thickness of the base of a tree,
2. h = 0.3 m; thickness at the stump height,
3. h = 1.3 m; DBH,
4. h = H; the height of a tree, zero thickness.

These positions may not be the optimal to improve accuracy of the curve [Smaltschinski, 1983; Goodwin, 2009], but have to be necessarily involved. Both the stem foot and the top must be involved to obtain the curve of the whole stem. The stump diameter is required for the proper description of the butt swell. DBH is included because it is a conventional parameter, and its value is always measured. All the positions are well defined and the diameters are easy to gain.

The second subset contains a different numbers of points, from 2 to 10. The points are defined by percentage relative height and represent the optimal input point distributions

for given numbers of input points, as defined by Figueiredo-Filho et al. [1996a]. The combinations of input point placement are shown in Table 3.2.

Table 3.2: Optimal combinations of input points for particular point numbers [Figueiredo-Filho et al., 1996a]

Combination no.	**Relative heights [%]**									
1	15	35								
2	15	45	75							
3	10	25	45	65						
4	10	25	45	65	85					
5	10	25	35	45	65	85				
6	10	15	25	45	65	85	95			
7	10	15	25	35	45	65	85	95		
8	10	15	25	35	45	55	65	85	95	
9	10	15	25	35	45	55	65	75	85	95

3.3.1.3 Evaluation of spline models

For each input point combination, stem curves were predicted using all splines. Residuals (differences between predicted and measured diameters) were assessed for each position of measured diameters for all trees. The accuracy of each predicted curve was described with five statistics (Table 3.3).

For each tree one value of each statistic is obtained. All the statistics are designed so that they are comparable among stems with different number of evaluation points. Diameter bias (DB) indicates whether a modeled curve systematically under- or overestimates stem thickness. The mean absolute residual (MAR) reflects the average distance between the predicted and the original diameters. The standard deviation of residuals (SDR) detects heterogeneity in residual values; large values may signify oscillations, which are undesirable. The mean squared residual (MSR) value reveals locally high deviations in the curve; high MSR values relative to the MAR value usually signifies oscillations. The total volume difference (TVD) calculated for the whole stem expresses the absolute error of volume if the real stem form is replaced by the spline function.

The volume of spline models was calculated as the sum of the volumes of log sections using Smalian's equation. This is recommended for short sections when the section lengths may

Table 3.3: Statistics used for evaluating the accuracy of the models

Abbreviation	**Statistics**	**Equation**
DB	Diameter Bias	$\sum_{i=1}^{N} \text{Diff}_i/N$
MAR	Mean absolute residual	$\sum_{i=1}^{N} \text{abs}(\text{Diff}_i)/N$
SDR	Standard deviation of residuals	$\sum_{i=1}^{N} [(\text{Diff}_i - \text{DB})^2/(N-1)]^{0.5}$
MSR	Mean squared residual	$\sum_{i=1}^{N} \text{Diff}_i^2/N$
TVD	Total volume difference	$\text{Vol}_{spl} - \text{Vol}_{orig}$

$\text{Diff}_i = d_i - \hat{d}_i$

$\hat{d}_i$ - measured diameter

d_i - predicted diameter

N - number of measured diameters

$\text{Vol}_{orig} = \sum_{i=1}^{N} l_i(\pi \hat{d}_i^2 + \pi \hat{d}_{i+1}^2)/2$ - volume computed from measured diameters

$\text{Vol}_{spl} = \sum_{i=1}^{N} l_i(\pi d_i^2 + \pi d_{i+1}^2)/2$ - volume computed form predicted diameters

l_i - distance between i^{th} and $(i+1)^{\text{th}}$ diameter

not be equal [Schreuder et al., 1993], although Huber's equation is generally considered to be more accurate [Biging, 1988; Figueiredo-Filho et al., 2000; Özçelik et al., 2008].

Accuracy of splines in terms of individual statistics was compared individually for each input point combination through the use of one-way analysis of variance and Tukey-Kramer honestly significant difference test for multiple comparisons. To examine the influence of input point number on the accuracy of the splines, the individual input point combinations were compared with the use of two-way analysis of variance. The influence of adding more input points was tested by comparing mentioned input point combination with an additional combination containing 22 input points. The combination consists of four fixed points and 18 points placed to the relative heights 10 %, 15 %, . . . , 95 %.

3.3.2 Determination of the optimal distribution of input points for modeling stem curves of coniferous trees

3.3.2.1 Selection of spline types

By virtue of the results of the previous comparison the number of splines was lowered. Splines regarded as the best representatives of four spline classes were used. A representative of first degree continual splines consisting of cubic polynomials, the Catmull-Rom spline, showed the best pertinence for stem profile modeling. The best representative of cubic splines (second degree continuous splines consisting of cubic segments) is the natural cubic spline; it has he minimal curvature among twice continuously differentiable interpolating curves and therefore it produces the least pronounced oscillations among interpolants in this class. As showed earlier, the accuracy of interpolation B-splines declines with rising degree of the curve. The second degree interpolation B-spline was selected from the class. The accuracy declines with rising degree also for approximation B-splines and NURBS. Adding weights does not improve the accuracy. Therefore second degree B-spline with uniform weights was selected from the class of approximation splines.

Regarding different properties and behavior of individual spline types it is useful to presuppose that the optimal input point distribution is different for particular spline types. Therefore each spline type is treated separately.

3.3.2.2 Input point positions

Smaltschinski [1983] considers the minimum number of input points for spline representation of stem curve to be 6; Lahtinen [1988] mentions that five point splines gives satisfactory approximation of the stem curve. Lower numbers of point are not sufficient. As results from primary analysis (Chapter 4.1.2) and some earlier works [Smaltschinski, 1983; Figueiredo-Filho et al., 1996a], it is not meaningful to add more than 9 points. Therefore optimal combinations containing 5-9 points were searched for.

The input point combinations are built in the same way as mentioned in chapter 3.3.1.2. An combination contains a subset of four fixed input points (0 m, 0.3 m, 1.3 m and H) and a subset of additional input points. The subset of additional input points contains from 1 to 5 points defined as radii in relative heights. The relative heights are chosen from the set {10 %, 15 %, ..., 95 % } (multiples of 5 % from the interval between 10 % and 95 %) For each number of additional input points all possible combinations were generated and

evaluated. Thus for each spline 18 models were evaluated for 5 input points, 153 models for 6 input points, ... and 8 568 models for 9 input points.

For each model all stems were evaluated. Diameter bias (DB), mean absolute residual (MAR), standard deviation of residuals (SDR), mean squared residual (MSR) and total volume difference (TVD) (all defined in Table 3.3) were calculated for the whole stem as well as for ten uniformly spaced height sections (0-10 %, 10-20 %, ...).

3.3.2.3 Aggregate objective function

Individual input combinations were compared using aggregate objective function (AOF) approach. Aggregated objective function (or target function) method is a optimization technique utilized in multi-criterion optimization. To each criteria a weight is assigned, expressing the relative importance of that criteria. The aggregate objective function value is calculated as the weighted sum of the normalized criteria values. The alternative with the lowest (for minimization problem) or highest (for maximization problem) is considered as the best solution.

Before the target function is calculated, all values X_i of each criteria X must be normalized in order to take the value between 0 and 1. Because diameter bias and total volume difference can take both positive and negative values, absolute values in equation 3.3.1 assure non-negative values of the normalized criteria, so that the criteria are comparable.

$$X_{i_{\text{norm}}} = \frac{|X_i| - \min(|X|)}{\max(|X|) - \min(|X|)} \tag{3.3.1}$$

The weight distribution for the criteria is shown in Table 3.4. The criteria and their weights were chosen so that the average accuracy of the curves is well balanced with their reliability. The mean value of the error statistics must be as low as possible so that the predicted curves are in average as accurate as possible. Simultaneously, the variance of all the error statistics is minimized and therefore the stem curve models are as reliable as possible.

The third of the total weight is given to magnitudes controlling the shape of the curve - MAR, SDR and MSR, where the weights are distributed relatively as 12.5 %, 12.5% and 25 % respectively for the three magnitude medians and the same weight distribution for their variations. The second third is given to the statistics controlling systematic shift of the curves - DB and TVD. The weights are distributed equally among their means and

Table 3.4: Distribution of weights assigned to individual criteria for the aggregate objective function

Characteristic			Weight (%)		
Shape accuracy	MAR	median	33.33	8.33	4.17
		variance			4.17
	SDR	median		16.67	8.33
		variance			4.17
	MSR	median		16.67	8.33
		variance			4.17
Systematic error	DB	median	33.33	16.67	8.33
		variance			8.33
	TVD	median		16.67	8.33
		variance			8.33
Significance of prediction error	DB	section CI	33.33	16.67	5.56
		total CI			11.11
	TVD	section CI		16.67	5.56
		tota CIl			11.11

variations. The last third of the total weight is given to magnitudes expressing if the total DB and TVD and the DB and TVD of individual sections are systematically shifted from zero, i. e. if they are systematically over- or underestimate. The magnitude labelled as Section CI in the table 3.4 expresses the number of sections with DB and TVD values respectively significantly different from zero. The values were considered to be significantly different from zero, if the 95% confidence interval did not contain zero. The magnitude labeled as Total CI is a binary magnitude expressing if the total DB or TVD significantly differs from zero.

3.3.3 Comparison of performance of individual splines

Comparison of suitability of particular spline types for representing stem curves of coniferous trees with the use of different numbers of input points is carried out. In order to satisfy demands of individual splines resulting from their different properties, the stem curves are modeled on the base of input points distributed according to the optimal positions, that were ascertained in chapter 4.1.3.

The comparison is carried out separately for each number of input points. Stem curve of each measured tree is modeled by four different splines (Catmull-Rom spline, natural cubic spline, interpolation B-spline, B-spline; for the details about spline selection see chapter 3.3.2.1). Spline models of particular stems are evaluated using five criteria (Table 3.3). All the criteria values are calculated both for the whole stem and for ten uniformly spaced height sections (0-10 %, 10-20 %, ...). The values of criteria calculated for the whole stem are compared among particular spline types. For the reason of different dispersion of the criteria values among the spline types Kruskall-Wallis test was used to compare the means of the criteria values. For DB and TVD (these criteria can obtain both positive and negative values) one-sample t-tests are used to detect if their means are significantly different from zero.

3.3.4 Spline representation of irregular stem curves of broadleaved trees

Utilizing splines for stem curve representation of broadleaved species may be more beneficial than for coniferous species due to higher complicacy of stem curve of broadleaved species compared to coniferous species and due to high percentage of stems with irregular shape. Irregularities in broadleaved stem curves are represented mainly by sudden diameter drop due to bifurcation of the main stem or large branching. Moreover such large branchings are usually preceded by local enlargement of the stem diameter, which violates the monotony of the stem curve and makes the profile even more complicated. Stem curves of such irregular stems are difficult or even impossible to be described by common taper functions usually restricted to create a smooth monotone curve. In this study data from two broadleaved species, European beech and Sessile oak were used (see chapter 3.1 for more information about the data).

The selection of splines for the study is based on results of previous analyses. Three candidates were selected for further utilization: Catmull-Rom spline with tangent vector corrections, natural cubic spline and B-spline of 2^{nd} degree.

The basic set of input points for spline computation contained six input points. Four fixed input points represented diameters conventional heights (base diameter, stump diameter, DBH and the height). Another two input points interactively added to describe the location of the most significant malformation. A position of the seventh input point was optimized to maximize the accuracy of the curve.

It is obvious, that the optimal position of the next input point depends on the location of the two inserted points. If they are located in the lower stem, it is necessary to place the point so that it controls the upper part of the curve. With shifting the inserted points upwards also the additional point must be shifted so that the input points are evenly distributed along the stem. Above certain threshold height it is not useful to specify more the curve above the inserted points; the accuracy of the curve is improved by placing the point rather to the lower part of the stem.

To evaluate the accuracy of splines, residuals at each position of measured diameters were determined and five statistics were computed (Table 3.3). The optimal positions of 7^{th} input point were determined for each stem individually using multi-criteria optimizing method of aggregate objective function with ten criteria. The first five criteria were the medians of above mentioned statistics with weights set as 0.05, 0.05, 0.05, 0.1 and 0.25, respectively. The second five criteria were variances of the same magnitudes with identical weight distribution. Dependency of the three best positions of seventh input point on the position of inserted points was searched using linear regression.

Suitability of additional input point placement according to the linear regression was reviewed. Several options of the seventh input point were compared. Firstly, the seventh point was placed to the optimal position determined individually for each stem. Secondly, the point was placed according to the linear regression. Thirdly, the point was placed in the center of the largest input point interspace. Fourthly (in tables labeled as Randomly I), the point was placed randomly near the center of the largest input point interspace (normal distribution with mean value equal to the center of the largest interspace and with such variance that the interspace covers the random point with probability 99.9 %). Finally (in tables labeled as Randomly II), the point was placed randomly into one of the two largest interspaces (uniform distribution). As the control variant served splines computed from six points only. Finally, the performance of particular splines was compared.

3.4 Local regression model of typical stem form

3.4.1 Data

For development of the local regression model of typical stem form regular stems of coniferous trees were considered. Therefore the spruce data described in chapter 3.1 were utilized for this purpose.

The generalized local model results from a number of spline models of individual stems from a stand or locality. When spline models for several individual stems are defined, as many height-diameter pairs as needed can be generated.

Data rescaling

In order to have identical scale the height-diameter data must be normalized before fitting the regression model. The height components of all data points are divided by the height of the respective stem so that the normalized height ranges from 0 (ground level) to 1 (stem height).

$$h_{\text{norm}} = T = \frac{h}{H} \tag{3.4.1}$$

The diameter cannot be normalized into the range from 0 to 1 by reason of the pronounced variability in the shape of butt swell. Therefore the diameter is expressed as the ratio either to a diameter in a fixed absolute height (most frequently the DBH) or to a diameter in a relative height. Because the stem form is more related to the relative proportions than to the absolute measures, the diameter in 10 % of the total height would be the best to be chosen as the factor. The normalized diameters would be:

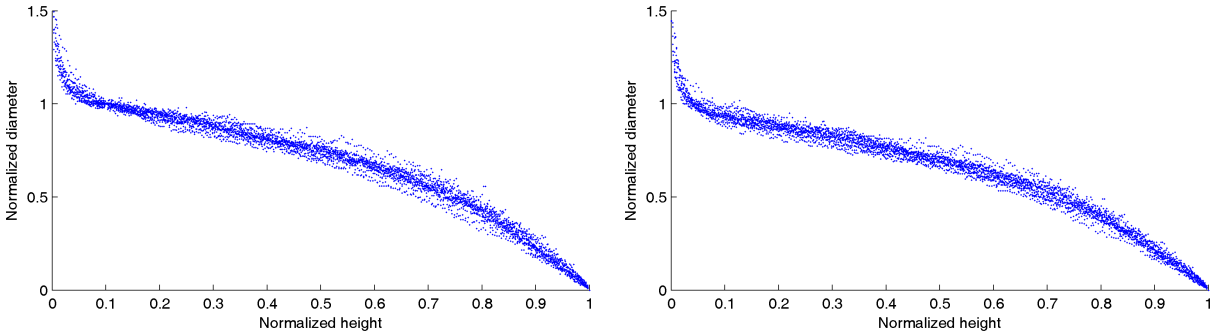
$$d_{\text{norm}} = \frac{d}{d_{0.1}}. \tag{3.4.2}$$

However, most taper models utilize the DBH as the factor. This approach is more effective from the practical point of view. The value of the factor diameter is necessary to be known both for normalizing the stem diameters and for the reverse operation of computing the real diameters of a new stem derived from the model. The conventional DBH is easier to obtain in comparison of the diameter in the relative height of 10 %. Moreover, it is more favorable that the diameter at breast height is kept less biased. Therefore the DBH is utilized as the diameter factor. The normalized diameters are:

$$d_{\text{norm}} = \frac{d}{D}. \tag{3.4.3}$$

Figure 3.1 shows the difference between both mentioned approaches of data normalization.

With increasing number of stems incorporated in computing the regression spline model of typical stem form, the point cloud of normalized diameters becomes too dense and the fit of the regression spline becomes computationally demanding. In order to lower down the number of points the original data set containing diameters measured with interspaces 0.1 m was reduced to stem profiles with interspaces 2 m. Thus stem diameters at 0 m, 1 m, 1.3 m, 3 m and at each next odd meter together with the stem height were used as input points.



(a) Diameter at relative height ($d_{0.1}$) (b) Diameter at breast height (D)

Figure 3.1: Normalized diameters and heights of 15 spruce stems. Diameter can be normalized using two different factors: a diameter at a relative height or the DBH.

3.4.2 Spline regression models and parameter optimizing

The normalized data are fitted to construct the non-parametric regression model. Two different regression splines are used to construct the models; the ordinary smoothing spline and the P-spline.

For both methods the optimal amount of smoothing must be determined. This is carried out using the leave-one-out cross-validation criteria. According to the LOOCV approach the best λ is the value that minimizes CV. For smoothing spline, the only parameter to be changed is the smoothing parameter λ (lambda). For P-splines the amount of smoothing depends on two parameters: the smoothing parameter λ and the number of segments.

Optimization of amount of smoothing for smoothing spline

The smoothing parameter λ can take any real value from the close range $< 0, 1 >$. In order to find out the development of CV criterion in dependence on λ the whole range must be covered. It is important to cover closely above all both extremities of the range. Therefore behavior of smoothing spline was examined with twenty λ values

$$\lambda \in \{10^{-6}, 10^{-5}, \ldots, 10^{-2}, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1 - 10^{-2}, 1 - 10^{-3}, \ldots, 1 - 10^{-10}\}$$

covering sufficiently the range $< 0, 1 >$.

In order to determine the dependency of the optimal amount of smoothing on number of input points, the λ value was optimized for different numbers of stems. The λ value was optimized subsequently for 1 to 50 stem profiles consisting of 17 points in average; thus

the number of smoothed points varied from 15 to 845. The input point density can be also expressed as different height interspaces between diameters. Therefore the λ value was also optimized for several numbers of stem profiles, where the interspaces between diameters were subsequently set to 0.1 m, 0.2 m, 0.3 m, 0.4 m and 0.5 m.

All diameters of all stems were treated as one data set. To compute the CV criterion for a given λ the smoothing spline was fitted n times whereas each of the n points served just once as the validation data. In order to eliminate the influence of different point numbers in data sets the CV criterion was divided by number of input points. Therefore the LOOCV denotes the mean squared residual of the validation points.

$$\text{LOOCV} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu_i(\hat{\theta}_{[-i]})\right)^2, \tag{3.4.4}$$

where $\hat{\theta}_{[-i]}$ is the estimator obtained by removing the i-th sample.

Optimization of amount of smoothing for P-spline

The smoothing parameter λ of P-spline can take any non-negative value. The set of λ values utilized for describing the behavior of P-spline smoothing consists of powers of two with exponents gradually changing from -10 to 12:

$$\lambda \in \{2^{-10}, 2^{-9}, \ldots, 2^{0}, \ldots, 2^{11}, 2^{12}\}.$$

Because the P-spline regression method is based on B-splines, the amount of smoothing is influenced also by another parameter, the number of segments (as indicated in the description of B-spline properties; chapter 2.3.6). When P-spline is fitted to data, number of segments is to be chosen as well as the smoothing parameter. Therefore the smoothing parameter must be optimized depending on the chosen number of P-spline segments. The set of the parameter *number of segments (nseg)* values that were used for λ optimization consists of powers of two with exponents from 1 to 9:

$$\text{nseg} = \{2, 4, 8, 16, \ldots, 512\}.$$

The CV criterion in dependence on λ value is calculated for all values of *nseg*. In order to find the dependency of the optimal λ also on the number of input points, the development of CV with different λ values was observed for different numbers of trees. The calculation of CV criterion is performed the same way as in the case of smoothing spline.

3.4.3 Comparison of spline regression models and taper functions

The performance of spline regression models is compared with conventional parametric taper equations. Based on the taper model comparison carried out by Rojo et al. [2005] the following models are selected for the comparison. As the best model of polynomial form the model of Cervera was selected. The best rerpresentative of the group of segmented polynomial models is considered the original model of Max and Burkhart [1976]. Because the variable exponent models are designated as the most accurate models two of them were selected for the comparison; the model of Bi [2000] and the model proposed by Lee et al. [2003].

Table 3.5: Taper equation selected for comparison with regression splines

Taper model	Equation
Cervera (1979)	$\frac{d}{D} = b_1 + b_2 \cdot X + b_3 \cdot X^2 + b_4 \cdot X^3 + b_5 \cdot X^4$
Max and Burkhart [1976]	$\left(\frac{d}{D}\right)^2 = \beta_1(z-1) + \beta_2(z^2/1) + \beta_3(\alpha_1 - z)^2 I_1 + \beta_4(\alpha_2 - z)^2 I_2,$
Bi [2000]	$\frac{d}{D} = \left(\frac{\log\sin(\frac{\pi}{2}\cdot T)}{\log\sin(\frac{\pi}{2}\cdot\frac{1.3}{H})}\right)^{b_1+b_2\cdot\sin(\frac{\pi}{2}\cdot T)+b_3\cdot\cos(\frac{3\pi}{2}\cdot T)+b_4\cdot\sin(\frac{\pi}{2}\cdot T)/T+b_5\cdot D+b_6\cdot T\sqrt{D}+b_7\cdot T\sqrt{H}}$
Lee et al. [2003]	$d = b_1 \cdot D^{b_2} \cdot (1-T)^{b_3 T^2 + b_4 T + b_5}$

For fitting the spline models and the selected conventional taper models the same data set was used as for smoothing optimization. 85 spruce stem profiles with 2 m long interspaces served for this purpose.

Before finding the regression parameters of the taper functions, the original data had to be transformed in order to correspond to the independent and dependent variables featuring in individual taper models. The taper functions were fitted using the least squares method. For fitting the non-linear functions of variable-exponent taper models, the Levenberg-Marquardt algorithm was used.

The comparison of individual regression techniques with optimal amount of smoothing was carried out using the LOOCV approach. When the models are fitted using all accessible data, a single stem is retained as validation data, while all other stems are used to compute a regression spline or to fit a taper model. Residuals, i.e. differences between the measured diameters of the validation stem and the diameters predicted by the model, are assessed for each position of measured diameters of the validation stem. The residual values of each

validation stem are evaluated using the statistics listed in Table 3.3. This procedure is repeated for all stems, so that every single stem serves as validation data exactly once.

The models were also fitted using lower numbers of stems: 5, 10, 20, 40 and 60. For these cases the appropriate number of stems was selected randomly from the whole data set. From the remaining stems, one was randomly selected as the validation stem and the real diameters were compared to the diameters predicted by the model. This procedure was repeated 400 times.

Because the variances of the evaluative statistics were not equal in all cases, the Kruskal-Wallis test was used to test the equality of mean values of the statistics among taper models.

Chapter 4

Results and Discussion

4.1 Individual stem curve modeling

4.1.1 Tangent vector corrections for Catmull-Rom spline

The Catmull-Rom spline interpolates well input point sequences with approximately uniform chords, i.e. distances between successive points. In such case the length of the tangent vectors approximately corresponds to the length of the chords and the Catmull-Rom spline generates a smooth interpolation curve. Because the tangent vector length at the point $\mathbf{P}_i$ is derived from the distance of points $\mathbf{P}_{i-1}$ and $\mathbf{P}_{i+1}$, in case of markedly unequal distances between input points may the length of the tangent vector exceed several fold the length of the chord of the interpolated segment. The curve is then forced to produce oscillation or even loops (Figure 4.1).

This effect can be eliminated by correcting the lengths of the tangent vectors for short curve segments adjacent to a long curve segment. Therefore corrections of the tangent vectors lengths are implemented into the algorithm. The tangent vectors are corrected not concerning individual input points, but in relation to curve segments. In the following equation, $\mathbf{p}_{i_{\text{out}}}$ denotes the tangent vector which originates in the input point $\mathbf{P}_i$ and drives the i^{th} curve segment in the closeness of the input point $\mathbf{P}_i$. The tangent vector $\mathbf{p}_{i+1_{\text{in}}}$ drives the i^{th} curve segment in the closeness of the input point $\mathbf{P}_{i+1}$

Using the notation used in chapter 2.3.1 the corrected tangent vectors can be expressed:

$$\mathbf{p}_{i_{\text{out}}} = \begin{cases} k_f \cdot c_i / c_{i-1} \cdot (\mathbf{P}_{i+1} - \mathbf{P}_{i-1}) & \text{if } c_{i-1} > k_c \cdot c_i \\ \tau \cdot (\mathbf{P}_{i+1} - \mathbf{P}_{i-1}) & \text{otherwise} \end{cases} \tag{4.1.1}$$

$$\mathbf{p}_{i+1_{\text{in}}} = \begin{cases} k_f \cdot c_i / c_{i+1} \cdot (\mathbf{P}_{i+2} - \mathbf{P}_{i}) & \text{if } c_{i+1} > k_c \cdot c_i \\ \tau \cdot (\mathbf{P}_{i+2} - \mathbf{P}_{i}) & \text{otherwise} \end{cases}$$

where c_i denotes the length of the i^{th} chord:

$$\begin{aligned} c_i &= ||\mathbf{P}_{i+1} - \mathbf{P}_i|| \\ c_{i-1} &= ||\mathbf{P}_i - \mathbf{P}_{i-1}||. \end{aligned} \tag{4.1.2}$$

Values of the parameters k_c - the correction condition - and k_f - the correction factor - are to be set empirically. For given purposes proved good the values $k_c = 5$, $k_f = 2$.

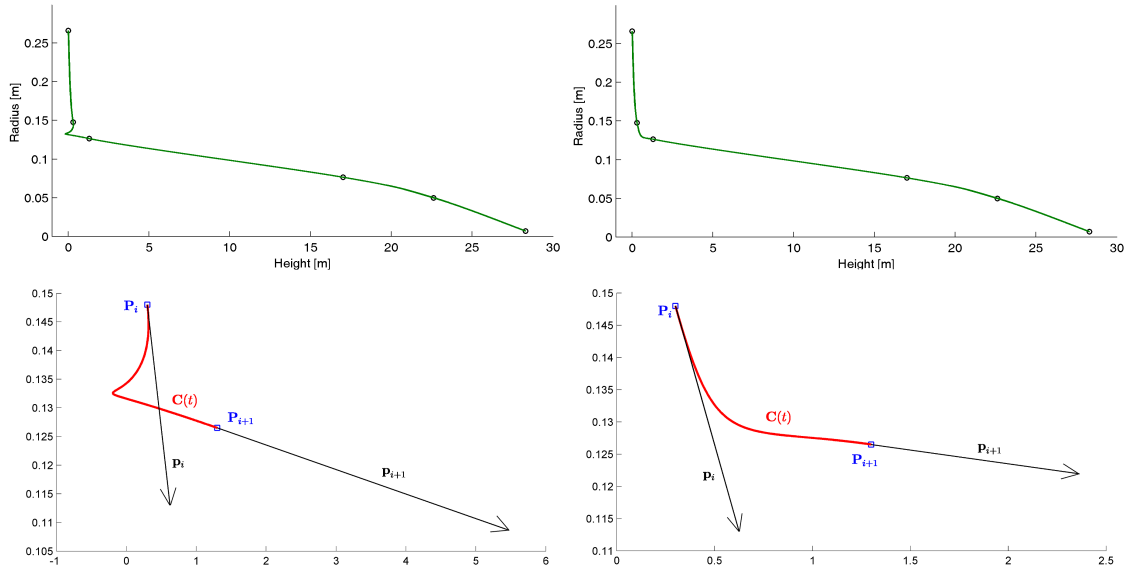


Figure 4.1: Tangent vector length corrections for Catmull-Rom spline. Left: A stem profile fitted by the original Catmull-Rom spline and the problematic spline segment. Right: Catmull-Rom spline with tangent vector length correction.

The corrections adjust the length of the tangent vectors in case of markedly different subsequent chord lengths. The important thing is that although the lengths of tangent

vectors entering and leaving a particular input point may not be equal, the direction of the tangent vectors is kept. Such corrections do not infract the first degree continuity of the curve.

4.1.2 Primary comparison of spline types

4.1.2.1 Lowering the number of evaluated spline types

Based on preliminary analyses, all interpolation B-splines with centripetal parameterization (IBSc2, IBSc3, IBSc4) were excluded because they generated uncontrollable oscillations. For the same reason, interpolation B-splines with uniformly spaced parameterization of degree 3 and 4 (IBSu3, IBSu4) were also excluded. Approximation splines of degree 4 (BS4, NUBS4, NURBSdbh4, NURBSav4) were removed from subsequent analysis because of their low forming ability. The Ferguson interpolation curve (LFC) was also excluded from further analysis because it was unable to conform to the lower part of the stem curve; this failure to conform is caused by tangent vector directions derived from Lagrange interpolation polynomials. The mentioned splines are not suitable for this special purpose and do not figure in the analysis anymore.

4.1.2.2 The influence of input point combination

The accuracy of spline stem curve models improved with an increase in the number of input points up to a threshold number. As illustrated in Table 4.1, near maximum accuracy was obtained in almost all cases with nine input points (four input points at fixed positions and five additional points at relative heights); the use of more than nine input points did not result in a statistically significant improvement in accuracy. No significant improvement in accuracy was observed even with the use of the full set of 22 input points (absolute heights 0 m, 0.3 m, 1.3 m and relative heights 10 %, 15 % ... 95 %, 100 %). The effect of numbers of input points is demonstrated with an example based on MAR values obtained with the Catmull-Rom spline (Figure 4.2).

Table 4.1: Numbers of input points sufficient for accurate description of the stem form in terms of the indicated statistic

Spline	DB	MAR	SDR	MSR	TVD
CRS	9	9	9	7	7
IBSu2	9	7	7	7	8
NCS	8	8	8	8	8
IterBS	9	8	8	8	8
BS2	9	9	9	7	9
BS3	9	9	7	7	9
NUBS2	9	9	9	7	9
NUBS3	9	9	9	9	9
NURBSdbh2	9	9	7	7	9
NURBSdbh3	9	9	9	9	9
NURBSav2	9	9	9	9	9
NURBSav3	11	9	9	9	11

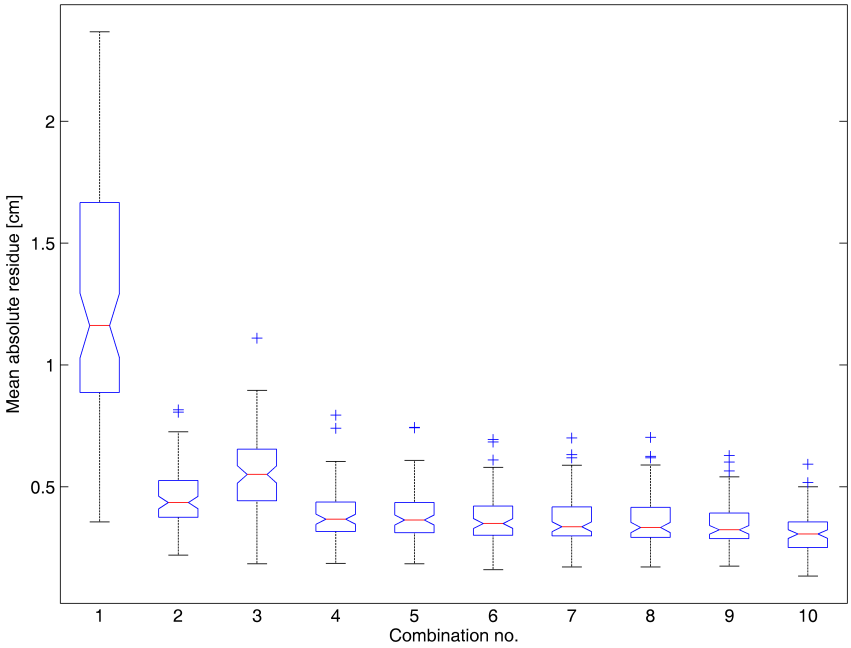


Figure 4.2: Relationship between mean absolute residual (MAR) and number of input points for the Catmull-Rom spline. Combination no. 10 contains 22 input points, as idicated in the text. Boxes indicate medians and 25th and 75th percentiles; notches indicate comparison intervals for medians; whiskers correspond to 99 % coverage of the data; crosses denote outlier values.

Both Figure 4.2 and Tables A.2 and A.3 indicate, that the input point combination no. 3 (eight input points) paradoxically produces curves with higher errors than the previous combination containing seven points. This efect is caused by a lack of points located in the upper parts of the stem. Combination no. 2 contains an input point at 75 % of the stem height, while the most upperpoint of combination no. 3 is placed to 65 % of the stem height. Therefore the upper stem curvature is not well described and diameters of the last third of the stem are underestimated. If the point at 65 % would be replaced by a point at 75 % of the stem height, the errors of splines based on point combination no. 3 would be much lower, as indicated in Figure 4.3.

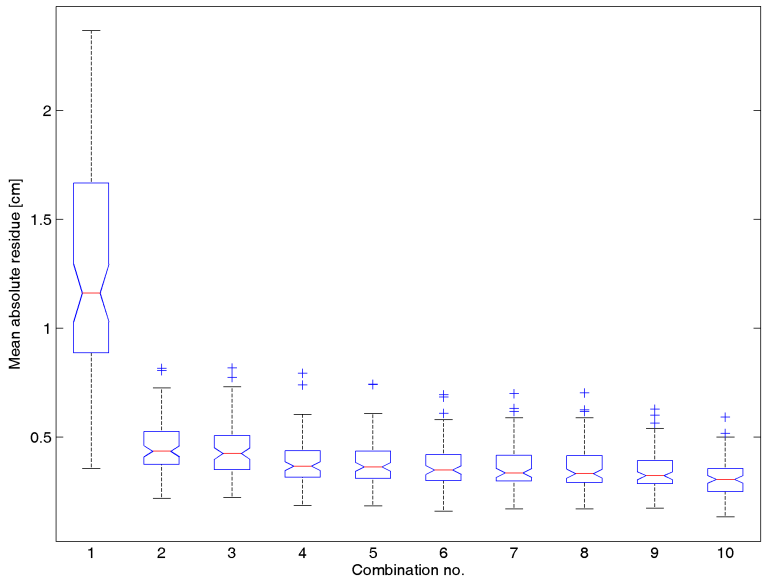


Figure 4.3: Developement of MAR values with corrected input point combination no 3. See explanation in the text.

The number of diameters required to describe the stem curve properly well corresponds with the results of Smaltschinski [1983]. He states that from 10 points the contribution of an added point is negligible which is in accordance with the results obtained in this study (Table 4.1). Figueiredo-Filho et al. [1996a] considers the accuracy of the whole curve sufficient not until 10 input points, which is not a big difference.

Comparison of Figures 4.2 and 4.3 well illustrates the importance of the input point distribution. This is in accordance with later works, e.g. Smaltschinski [1983] and Figueiredo-Filho et al. [1996a]. The input point combinations proposed by Figueiredo-Filho et al. [1996a] are regarded as the optimal point distribution for natural cubic

spline. But obviously they are not optimal in case of using Catmull-Rom spline. Considering those two facts it is in evidence that different splines have different optimal distribution of input points. That topic is being solved in chapter 4.1.3.

4.1.2.3 Comparison of spline types

The comparisons of individual spline types with 6 (combination no. 1; the least number investigated) and 9 (combination no. 4; the highest number meaningful for accuracy improvement) input points are shown in Tables 4.2 and 4.3. The complete set of tables illustrating results of comparisons of splines based on 6 to 10 points (combinations 1 to 5) are inserted in Appendix A. Because a further increase in point number did not affect the accuracy, results for combinations with more input points (combinations 6 to 9 in Table 3.2) are not shown. To illustrate no significant improve of accuracy with more points, Appendix A also contains Table A.6 with the results of the comparison of splines computed from the full set of points. Stem curves modeled by all interpolation splines and two representatives of approximation splines (BS2 and NURBSav2) based on six to nine input points are showed in Figures B.1-A.6 in Appendix A.

Table 4.2: Comparison of splines based on 6 input points (combination no. 1). For each statistic mean value and standard deviation is shown. Values in a column followed by the same letter indicate no significant difference between spline types. Stars in columns DB and TVD indicate mean values significantly different from zero.

Spline	**DB (10^{-2} m)**			**MAR (10^{-2} m)**			**SDR (10^{-2} m)**			**MSR (10^{-3} m^2)**			**TVD (%)**		
CRS	-1.03	0.52	a*	1.30	0.54	a,b	1.64	0.69	a,b	0.32	0.26	a	-8.22	3.69	a*
IBS2u	-0.92	0.49	a*	1.14	0.47	a	1.46	0.61	a	0.25	0.21	a	-8.14	3.81	a*
NCS	0.45	1.03	c*	1.76	1.33	c	2.11	1.49	c	0.67	0.99	b	6.27	11.01	b*
IterBS	0.49	1.11	c*	2.26	1.74	d	2.73	1.97	d	1.13	1.61	c	7.12	12.89	b*
BS2	-1.16	0.61	a,b*	1.57	0.61	b,c	1.93	0.75	b,c	0.43	0.33	a,b	-9.24	4.01	a*
BS3	-1.15	0.61	a,b*	1.67	0.64	b,c	2.02	0.77	b,c	0.47	0.35	a,b	-8.85	4.17	a*
NUBS2	-1.24	0.63	a,b*	1.61	0.63	b,c	1.95	0.76	b,c	0.45	0.34	a,b	-10.34	4.09	a,c*
NUBS3	-1.28	0.65	a,b*	1.71	0.66	b,c	2.04	0.79	b,c	0.48	0.37	a,b	-10.83	4.31	a,c*
NURBSdbh2	-1.17	0.64	a,b*	1.68	0.63	b,c	2.08	0.77	b,c	0.50	0.36	a,b	-8.99	4.36	a*
NURBSdbh3	-1.27	0.68	a,b*	1.83	0.68	c	2.20	0.80	c	0.55	0.39	a,b	-10.30	4.94	a,c*
NURBSav2	-1.44	0.75	b*	1.93	0.72	c,d	2.30	0.85	c,d	0.61	0.44	b	-12.33	4.90	c*
NURBSav3	-1.47	0.73	b*	1.90	0.73	c,d	2.22	0.86	c	0.57	0.43	a,b	-12.98	4.76	c*

Table 4.3: Comparison of splines based on 9 input points (combination no. 4)

Spline	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
CRS	0.00	0.15	a,b	0.39	0.11	a	0.54	0.17	a	0.03	0.02	a	0.25	2.17	a
IBS2u	0.02	0.15	a,b	0.43	0.12	a,b	0.66	0.23	a,b	0.05	0.04	a,b	-1.27	2.52	b*
NCS	-0.16	0.19	c*	0.70	0.32	c	1.07	0.53	c	0.14	0.15	c	-1.68	2.51	b,c*
IterBS	-0.26	0.23	d*	0.95	0.48	d	1.52	0.82	d	0.30	0.33	d	-2.89	2.77	c*
BS2	-0.01	0.17	a,b	0.52	0.13	b,e	0.83	0.24	b	0.07	0.04	a,b	1.43	2.34	a,d,e*
BS3	-0.02	0.19	a,b	0.58	0.15	e	0.93	0.29	b,c	0.09	0.06	b,c	1.63	2.44	d,e*
NUBS2	-0.02	0.17	a,b	0.51	0.12	b,e	0.80	0.23	b	0.07	0.04	a,b	1.22	2.31	a,d*
NUBS3	-0.04	0.20	a*	0.59	0.16	c,e	0.94	0.28	b,c	0.10	0.06	b,c	1.39	2.47	a,d*
NURBSdbh2	0.05	0.19	b	0.58	0.14	e	1.03	0.35	c	0.12	0.08	b,c	2.64	2.77	d,e*
NURBSdbh3	0.02	0.22	a,b	0.66	0.17	c,e	1.17	0.42	c	0.15	0.11	c	2.72	3.05	d,e*
NURBSav2	0.00	0.20	a,b	0.62	0.15	c,e	1.07	0.35	c	0.12	0.08	b,c	2.24	2.79	d,e*
NURBSav3	-0.08	0.24	a,c*	0.69	0.2	c	1.06	0.34	c	0.12	0.08	b,c	1.45	2.71	a,d,e*

Interpolation splines

The most accurate results were obtained with the Catmull-Rom spline. While both diameter prediction and volume estimation were underestimated for 6 and 8 points, with 9 and more points the Catmull-Rom spline was the only spline that did not show systematic errors neither in diameter prediction nor in volume estimation. Moreover, the Catmull-Rom spline had the lowest values of MAR, SDR, and MSR, which indicates that the Catmull-Rom spline closely follows the original data and with well-balanced diameter errors. No oscillations or locally increased diameter errors were detected with the Catmull-Rom spline.

The second degree interpolation B-spline also had low residual value; in most cases its errors were not significantly higher than those of Catmull-Rom spline. Unfortunately, the interpolation B-spline systematically underestimates the total volume even for higher numbers of input points.

The natural cubic spline and the iterative B-spline had positive values of diameter bias and overestimated the stem volume when six points were used (combination 1 in Table 3.2),. With seven or more input points, both second-degree continual cubic splines systematically underestimated the diameters as well as the total volumes. With six or seven input points, these two functions had the highest SDR and MSR values among all splines, indicating a chance of oscillations; deviations with the iterative B-spline were always higher. With more than seven input points, residuals and their square values were most pronounced with the

iterative B-spline while oscillations generated by the natural cubic spline were reduced.

The propensity of the natural cubic spline for oscillations has been described earlier [De Boor, 2001; Lahtinen, 1988] and is attributed to the continuity of its second derivative. In agreement with Figueiredo-Filho et al. [1996a], oscillation in the current study most often occurred with smaller numbers of input points; as the number of input points increased, the oscillations decreased.

The oscillations produced by the iterative B-spline were even higher than those of the natural cubic spline and are easily explained. The natural cubic spline is defined as a curve having minimal curvature among all second-order continuous curves interpolating a given set of points Liu [1980]; Lahtinen [1988]. Because the iterative B-spline has been proven to approach the interpolation curve [Lin et al., 2004] and has second-order continuity, its curvature must necessarily be higher than that of natural cubic spline, which results in more pronounced oscillations. For the same reasons, oscillations were observed in this study for all interpolation B-splines of the third and fourth degree, which have second and even third, respectively, degree continuity.

Because there is no reason to assume that the stem curve is necessarily second-degree continuous, it is better to avoid the risk of oscillation by not using second-degree continuous interpolation splines. This agrees with Lahtinen [1988], who achieved better results using a quadratic spline with only first-degree continuity.

Approximation splines

The behavior of all approximation splines was very similar. With an increase in the number of input points, MAR, SDR, and MSR for all the approximation splines decreased, indicating that the fit of the splines to the real stem curve improved. On the other hand, bias values and total volume estimates increased with increasing input points. For input combinations 1-3 (six to eight input points), bias was negative, while for more than 11 input points, bias became positive. The TVD value was negative for the smallest number of input points; for more than eight input points, the total volume was significantly overestimated.

Plain second degree B-spline was the best representative of the group. With the approximation splines, residual values tended to increase (in many cases with statistical significance) with an increase in the spline degree. In addition, residual values were higher for NURBS than for B-splines.

Residual values were higher for approximation splines of the third degree than for those of the second degree, and were even higher for approximation splines of the fourth degree. The poor performance of approximation splines of higher degrees can be explained according to Linkeová [2007]. The greater the degree of the B-spline, the more points drive each segment of the curve. Because B-splines are defined as a linear combination of input point coordinates and basis functions, an increase in the number of points describing each segment decreases the relative influence of each input point position on the shape of the segment and therefore increases the distance between the resulting curve and the input points.

Also weight implementation affects the curve accuracy. NURBS that emphasized DBH and that had weights set by the averaging method had higher residual values than the plain B-spline with uniform weight distribution. Weights of the input points in NURBS express the relative importance of the input points. The negative consequence of improving the approach of the curve to the emphasized point is a reduction in how the other points affect the curve. Splines with averaging weight distribution suffer in the same matter. BS2 with uniform knot vector and NUBS2 with averaging knot vector produce in all cases very similar results. Therefore, the knot vector does not appear to be crucial.

B-splines and NURBS are always smoother than the driving polygon [MacCallum and Zhang, 1986] and lie within its convex hull. Therefore, the lower part of a stem is always overestimated while the upper part is underestimated. With few input points, all of the splines examined here underestimated the diameters in average and total volume. An increase in the number of input points improved the representation of the upper stem but not of the butt swell curvature. When the numbers of points were increased, therefore, the upper stem diameter was no longer underestimated, and the total values were influenced by the positive deviation at the lower stem. Bias and the volume estimation would be more balanced if more points were also included at the bottom part of the stem.

4.1.3 Determination of the optimal input point distribution

It was found that for proper fit of the lower stem curvature it is useful to place an input point approximately to 10 % of the stem height. For lower trees this is satisfied by the point at breast height. Therefore the data set was split into two height classes where the threshold value was 20 m and input point placement was optimized for each class separately.

For reasons of readability, the demonstration of the dependence of standardized criteria values on input point positions in Figure 4.4 displays the optimization of a position of a single input point in five-point Catmull-Rom spline. For the optimization of a single point there are 18 positions to be evaluated. For two points there are 153 different combinations and for next numbers of input points there are even more possible combinations to evaluate. Such numbers of combinations cannot be legibly shown in a chart.

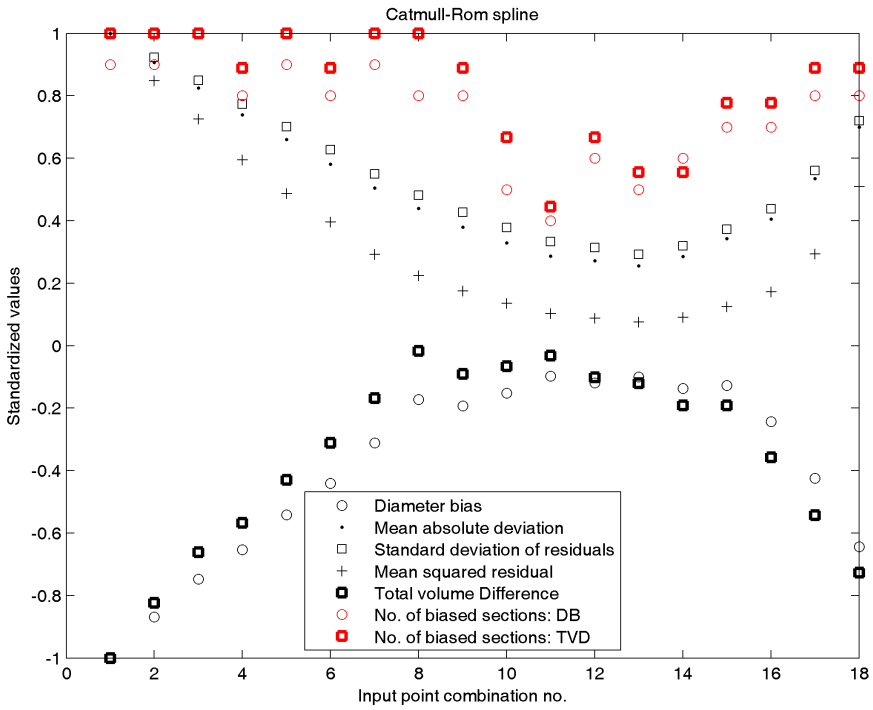


Figure 4.4: Developement of standardized criteria values for different input point combinations; Catmull-Rom spline; 1 additional input point

As indicates Figure 4.4, the standardized values of criteria are closest to zero in the optimal point position. With growing distance from the optimal position the values of the criteria increase. The criteria have mostly their minimum close to each other, but apparently the minima of the cirteria do not meet exactly. The optimal position is then depentend on the importance of particular criterion, which is expressed as the weight of the aggregate objective function. The higher weight of a criterion, the closer to its minimum is the final optimal position.

The chart visualizing the development of standardized criteria values (Figure 4.4) and also tables showing properties and details of twelve combinations with the lowest value of the aggregate objective function were generated for each spline and each input point number.

Table 4.4: Overview of best positions for 5th input point; Catmull-Rom spline (AOF = value of aggregate objective function; Comb. = rank of the combination; Pos. = positon of the 5th input point)

	AOF	**Comb.**	**Pos.**	**DB**	**MAR**	**SDR**	**MSR**	**TVD**
1	0.148	11	60	-0.16	0.53	0.71	0.06	-0.40
2	0.165	10	55	-0.26	0.61	0.80	0.08	-1.20
3	0.248	13	70	-0.16	0.47	0.61	0.04	-2.25
4	0.255	12	65	-0.20	0.51	0.66	0.05	-1.87
5	0.257	8	45	-0.30	0.83	1.03	0.13	-0.23
6	0.273	14	75	-0.23	0.53	0.67	0.05	-3.75
7	0.300	9	50	-0.33	0.71	0.91	0.10	-1.78
8	0.380	16	85	-0.44	0.75	0.92	0.09	-7.37
9	0.396	7	40	-0.55	0.95	1.18	0.17	-3.58
10	0.460	15	80	-0.21	0.64	0.79	0.07	-3.70
11	0.461	17	90	-0.78	0.98	1.18	0.16	-11.43
12	0.481	6	35	-0.78	1.09	1.35	0.22	-6.75

Table 4.5: Sectional and total diameter bias for 12 best input point combinations for CRS. Yellow background of a cell indicates that the mean significantly differs from zero.

	No. of biased sections	**Section**										**Total**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
1	4	0.04	0.24	0.08	0.03	-0.03	0.05	-0.09	-0.44	-0.90	-0.76	-0.16
2	5	0.04	0.27	0.16	0.13	0.02	-0.08	-0.45	-0.84	-1.19	-0.86	-0.26
3	5	0.04	0.16	-0.15	-0.33	-0.46	-0.31	-0.08	0.06	-0.27	-0.49	-0.16
4	6	0.04	0.18	-0.07	-0.21	-0.31	-0.17	-0.08	-0.22	-0.66	-0.67	-0.20
5	8	0.04	0.43	0.56	0.57	0.19	-0.35	-0.92	-1.28	-1.49	-0.96	-0.30
6	6	0.04	0.11	-0.27	-0.53	-0.71	-0.59	-0.33	0.03	0.01	-0.27	-0.23
7	8	0.04	0.31	0.27	0.25	0.04	-0.28	-0.77	-1.14	-1.40	-0.93	-0.33
8	7	0.05	0.03	-0.49	-0.87	-1.18	-1.15	-0.94	-0.47	0.06	0.28	-0.44
9	9	0.04	0.43	0.52	0.32	-0.33	-0.93	-1.43	-1.67	-1.73	-1.04	-0.55
10	7	0.04	0.10	-0.33	-0.62	-0.84	-0.73	-0.45	0.02	0.34	0.17	-0.21
11	8	0.05	-0.04	-0.67	-1.17	-1.6	-1.68	-1.55	-1.15	-0.58	0.11	-0.78
12	8	0.05	0.45	0.42	-0.07	-0.86	-1.44	-1.85	-1.99	-1.92	-1.10	-0.78

Table 4.6: Sectional and total volume difference for 12 best input point combinations for CRS. Yellow background of a cell indicates that the mean significantly differs from zero.

	No. of biased sections	**Section**										**Total**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
1	4	0.40	3.30	1.39	0.62	-0.26	1.08	-0.96	-9.39	-26.91	-45.82	-0.40
2	6	0.41	3.61	2.33	1.86	0.31	-1.29	-8.03	-18.85	-35.53	-52.2	-1.20
3	5	0.43	2.22	-1.89	-4.86	-7.37	-5.93	-2.22	0.84	-8.84	-28.7	-2.25
4	6	0.42	2.55	-0.80	-3.00	-4.89	-3.36	-1.81	-5.03	-19.89	-39.87	-1.87
5	9	0.4	5.72	7.92	8.61	3.3	-5.39	-16.2	-28.41	-43.8	-58.08	-0.23
6	5	0.45	1.70	-3.40	-7.47	-11.03	-10.5	-6.61	0.36	1.50	-10.26	-3.75
7	8	0.39	4.13	3.77	3.48	0.41	-5.01	-14.4	-26.05	-41.71	-56.45	-1.78
8	7	0.47	0.72	-6.09	-12.06	-17.6	-19.44	-18.31	-12.1	0.74	31.61	-7.37
9	9	0.41	5.66	7.40	4.79	-4.9	-15.36	-25.97	-36.95	-50.37	-62.44	-3.58
10	7	0.45	1.56	-3.88	-8.32	-12.04	-11.79	-7.8	2.14	14.6	30.06	-3.70
11	8	0.51	-0.13	-8.40	-15.98	-23.23	-27.28	-29.09	-27.47	-21.18	8.55	-11.43
12	8	0.47	5.89	5.89	-0.86	-12.7	-23.19	-33.18	-43.29	-55.41	-66.19	-6.75

The first table (demonstration on the example of Catmull-Rom spline, 5 input points is shown in Table 4.4) shows the details of the best combinations and values of the five statistics characterizing accuracy of curves based on those input point combinations. Sectional diameter bias and total diameter bias with the indication of bias values significantly different from zero are shown in Table 4.5. Similarly the sectional and total volume differences are visualized (Table 4.6).

Twelve combinations considered as best in terms of the aggregate objective function were evaluated concerning the requirement of stability and reliability of the combination. A input point combination was selected as optimal, if small shift of the point positions (up to 5 % of the stem height) does not markedly affect the accuracy of the curve. Therefore input combinations generating curves with low errors by chance are eliminated.

Due to different behavior of individual splines optimal input point positions also differ. With natural cubic spline the input points are added preferably to the lower third of the stem in order to reduce oscillations emerging above all in the lower third. On the other hand with B-spline the points are placed preferably to the surroundings of 70 % of the height so that the approximation spline is able to describe the major change of direction of the upper tree profile. With Catmull-Rom spline and interpolation B-spline the points are

Table 4.7: Optimal positions of additional points for trees under 20 m. Relative heights (%)

CRS					NCS					IBS2					BS2				
60					35					55					70				
50	85				15	45				20	60				65	85			
15	50	85			15	40	80			20	40	75			25	65	85		
15	35	55	85		15	25	35	95		20	40	65	85		10	50	60	85	
15	25	35	55	85	10	15	20	65	95	20	40	50	70	90	10	15	50	60	85

Table 4.8: Optimal positions of additional points for trees over 20 m. Relative heights (%)

CRS					NCS					IBS2					BS2				
65					25					50					65				
50	80				15	45				25	65				60	85			
10	50	80			15	35	80			15	30	65			30	50	80		
10	40	60	85		10	15	35	80		15	30	55	75		15	50	65	80	
10	20	50	75	95	10	15	25	45	85	15	30	50	70	85	10	40	50	65	80

distributed more evenly along the stem. The final positions of input points for lower and higher stems are shown in Tables 4.7 resp 4.8. These positions were then used to model stem curves for spline comparison.

The optimal input point positions for natural cubic spline found in this study differ from those stated by SMALTSCHINSKI [1983] and also from those stated by FIGUEIREDO-FILHO ET AL. [1996a]. Reasons are following. Neither of them included the stem foot in the spline. SMALTSCHINSKI [1983] avoided the demanding curvature of the stem butt by starting the spline at 1.3 m; FIGUEIREDO-FILHO ET AL. [1996a] started the stem profile at the height 0.1 m. Moreover, they both used only the two edge points as fixed points (at 1.3 m and 0.1 m, respectively, and the top of the stem); the other point positions were optimized to minimize errors. In this study the stem is modeled from the very bottom and except the foot and the top of the stem two other points were fixed at conventional positions of stump height and breast height, while fewer positions were optimized. As stated by SMALTSCHINSKI [1983], the conventional measuring height 1.3 m is not favorable concerning the accuracy of the resulting curve, but the model is expected to reflect the conventional measuring point. What makes this study yet more exacting is a more pronounced but swell of spruce trees causing higher propensity of the curves for oscillating in comparison with the more moderate but swell of loblolly pine trees used by FIGUEIREDO-

Filho et al. [1996a]. This difference of the shape of the butt swell is clearly visible when comparing figures in this thesis with the figures of Figueiredo-Filho et al. [1996a].

4.1.4 Comparison of performance of individual splines

Results of the evaluation of performance of splines based on the optimal input point distributions for five to nine input points is shown in Table 4.9. The table contains criteria values calculated for the whole stem curve and their statistical comparison. Sectional diameter biases and volume differences for splines with five to nine input points are shown in Tables B.1 to B.5 (Appendix B).

A reliable curve with well-balanced error is produced by Catmull-Rom spline. For all input point numbers the Catmull-Rom spline gives unbiased estimates of total volume with the mean total volume difference less than 1 % .The overall diameter prediction is slightly underestimated (less than 2 mm) for five input points; for more input points the prediction is unbiased (Table 4.9). The mean absolute residual moves around 0.5 cm for five points and less for more points. The low values of SDR and MSR for all input point numbers illustrate the evenness of the error distribution along the stem. When only five input points are used the spline does not represent well the two major direction changes of the stem profile. Therefore the second section is slightly overestimated and the three topmost sections are underestimated both for diameter and volume predictions. With six input points only the second section is biased; with more input points the spline gives diameter and volume predictions without any, neither sectional nor total, systematic deviations.

The oscillations of natural cubic spline are strongly pronounced with lower numbers of input points. With rising number of input points the oscillation is reduced; with eight and nine points its extent is limited only to the lowermost third of the stem. However the oscillation is not completely eliminated even with nine input points, as illustrated by significantly higher values of MAR, SDR and MSR throughout all the input point numbers. Although the total volume estimation and overall diameter prediction are not significantly biased, the high sectional diameter and volume errors also show the unsuitability of this spline for the given purpose.

A reasonable representation of the stem curve produced by interpolation B-spline is evidenced by low values of MAR, SDR and MSR for all numbers of input points. With 5 input points the two lowermost sections and the topmost third are biased and the total volume estimate is overestimated. With 6 and more input points the total volume and

the overall diameter prediction are accurate. Splines based on seven and more points give unbiased predictions of diameter and volume both for the sections and the entire stem.

Approximation B-spline suffers by systematic errors in both main curvatures. For all numbers of input points it overestimates the lower part and underestimates the topmost sections (tables). The unbiased total diameter and volume estimates emerging in some cases are consequences of compensation of both systematic errors. With increasing number of input points, values of MAR, SDR and MSR tend to decrease in contrast to the volume error, which grows and decreases in turns.

Table 4.9: Comparison of splines based on optimal distribution of input points for different numbers of input points. For each statistics mean and standard deviation is shown. Values in a column followed by the same letter indicate no significant difference between spline types. Stars in columns DB and TVD indicate mean values significantly different from zero.

	Spline	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
5 points	**CRS**	-0.18	0.32	a*	0.58	0.20	a	0.75	0.26	a	0.06	0.05	a	-0.91	4.86	a,b
	NCS	-5.35	5.06	b*	8.36	6.45	b	9.49	7.32	b	14.43	20.23	b	4.45	25.49	a,b
	IBS2u	0.03	0.30	a	0.72	0.26	a	1.34	0.58	c	0.21	0.18	c	1.32	3.95	a*
	BS2	-0.49	0.37	c*	1.09	0.36	c	1.36	0.43	c	0.20	0.13	c	-2.36	5.02	b*
6 points	**CRS**	-0.02	0.25	a	0.48	0.16	a	0.63	0.22	a	0.04	0.03	a	0.20	3.67	a
	NCS	-0.11	0.41	a*	2.03	1.42	b	2.49	1.66	b	0.89	1.18	b	-1.31	5.25	b*
	IBS2u	-0.01	0.26	a	0.51	0.18	a	0.76	0.29	a	0.07	0.06	a	-0.07	3.75	a,b
	BS2	-0.08	0.29	a*	0.73	0.22	c	1.02	0.30	c	0.11	0.07	c	0.92	4.64	a
7 points	**CRS**	-0.01	0.18	a	0.43	0.14	a	0.58	0.20	a	0.04	0.03	a	-0.24	2.57	a
	NCS	-0.11	0.38	b*	1.59	0.91	b	2.02	1.14	b	0.53	0.60	b	-1.47	5.37	a*
	IBS2u	-0.01	0.20	a,b	0.44	0.14	a	0.63	0.21	a	0.04	0.03	a	-0.29	2.70	a
	BS2	-0.05	0.20	a,b	0.63	0.17	c	0.94	0.26	c	0.10	0.05	c	1.11	2.67	b*
8 points	**CRS**	-0.03	0.15	a	0.40	0.12	a	0.56	0.18	a	0.03	0.03	a	-0.16	2.06	a
	NCS	-0.03	0.41	a	0.81	0.30	b	1.19	0.47	b	0.16	0.13	b	0.29	5.02	a,b
	IBS2u	-0.03	0.20	a	0.44	0.14	a	0.63	0.22	a	0.04	0.03	a	-0.07	2.56	a
	BS2	-0.03	0.17	a	0.54	0.12	c	0.86	0.24	c	0.08	0.04	c	0.84	2.47	b*
9 points	**CRS**	-0.03	0.15	a	0.39	0.11	a	0.54	0.18	a	0.03	0.03	a	-0.26	1.91	a,b
	NCS	-0.13	0.39	b*	0.75	0.33	b	1.16	0.54	b	0.16	0.20	b	-1.04	8.60	a
	IBS2u	0.01	0.12	a	0.40	0.13	a	0.60	0.22	a	0.04	0.03	a	-0.12	1.95	b
	BS2	-0.01	0.16	a	0.53	0.12	c	0.85	0.24	c	0.08	0.04	a	1.18	2.25	c*

The B-spline is always smoother than the driving polygon and therefore the lower stem is always overestimated, while the upper stem underestimated. With adding more input points the curve can be considerably improved, but in the curvatures it will never entirely reach the input points. When more points are added to the upper part, the upper curve is put more precisely and the underestimation of the upper stem is reduced. That is exactly what happens with adding the seventh and the ninth point. Although the curve is more accurate, the plus error of the lower stem dominates and paradoxically the absolute value of total volume error grows. The volume error is again lowered with 8-point spline. The 8-point set contain an input point in relative height 10 %, which reduces the plus error of the lower stem.

Values of the evaluation criteria for the natural cubic spline in this study are higher as compared to the values presented by Figueiredo-Filho et al. [1996a]. Except the differences between both studies, which were mentioned in chapter 4.1.3 and which explain the differences between the optimal distribution of input points, one reason more takes place. Values of error statistics were determined from deviations between the predicted curve and measured diameters, while Figueiredo-Filho et al. [1996a] used a smooth curve as a reference.

With the exception of natural cubic spline all the splines selected for this study have only the first degree continuity. Therefore they do not suffer from oscillations and their errors are lower than the errors of the natural cubic spline. This is with agreement with Lahtinen [1988] who reported that the quadratic spline, which is only once continuously differentiable, was superior to the cubic spline and also with Goulding [1979] who recommends infracting the second degree continuity of the curve in cases of malformed stems or in cases of unevenly long intervals between input points in order to avoid oscillations. Well workable cubic segments and interruption of second order continuity in knots are two important properties of Catmull-Rom spline that cause the ability to represent accurately the stem without the risk of oscillations.

4.1.5 Spline representation of irregular stem curves of broadleaved trees

With respect to the assumed type of dependence between the optimal positions of additional input point and the positions of the inserted points describing the stem irregularity, the resulting point set was split into two parts and linear regression was computed sepa-

rately for each part (Figure 4.5). The threshold value was determined in such way that the residual sum of squares is minimized. Expected relation between optimal position of the seventh point and positions of the inserted points was found only for Catmull-Rom spline. Therefore regression was fitted only for Catmull-Rom spline (Figure 4.5). Coefficients of determination of both lines are very low (0.29 and 0.13 respectively), but the regression parameters are statistically proved to be significantly different from zero.

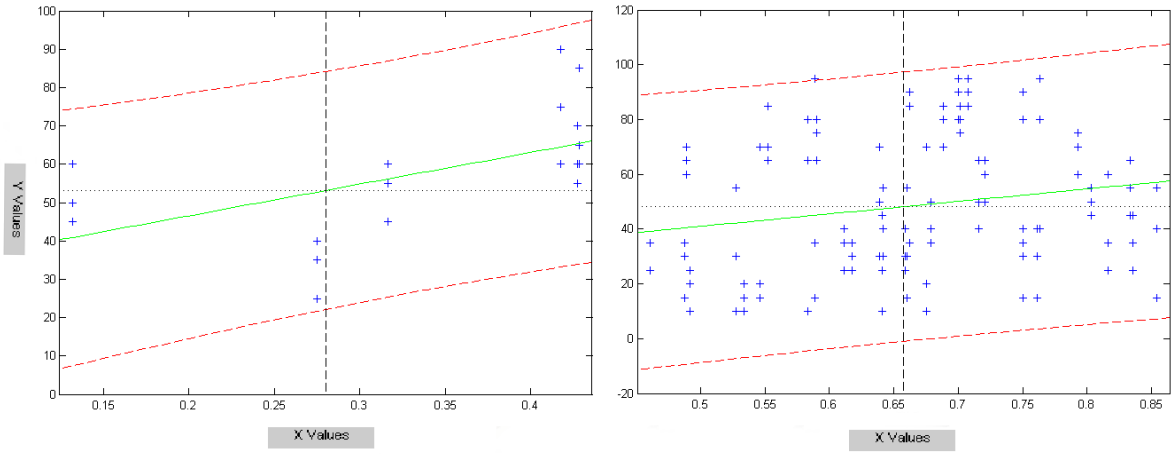


Figure 4.5: Regression lines for 7^{th} point placement for Catmull-Rom spline. X values denote the relative height of the inserted points describing the stem irregularity. Y values are the optimal position of seventh input point. Solid green line is the regression line; dashed red lines denote borders of 95% confidence region.

With natural cubic spline and B-spline the location of the optimal input points is in principle independent on the positions of the inserted input points and it does not allow building the regression. Therefore general optimal position of the seventh point for all stems was calculated. It was found that both for natural cubic spline and B-spline the curve is the most accurate if the seventh point is placed in 20 % of the stem height. For evaluating different methods of placing the seventh point was the regression replaced by situating the additional point to the relative height 20 % of the stem height. The dependence of the optimal input point positions on the positions of the inserted points is shown in Figure C.1 (Appendix C).

For all splines, according to expectations the lowest error values are found in case that the point is placed to its optimal position. But rarely the errors significantly differ from errors of other variants of point placement.

Table 4.10: Comparison of variants of the seventh point placement for Catmull-Rom spline. Different indexes in columns indicate statistically significant differences among variants.

Variant of placement	DB (10^{-2} m)	MAR (10^{-2} m)	SDR (10^{-2} m)	MSR (10^{-3} m^2)	TVD (%)
Optimal position	0.001 a,b	0.394 a	0.523 a	0.033 a	0.204 a
Regression	-0.020 a,b	0.465 b	0.585 a,b	0.045 a,b	-0.051 a
Center of largest intersp.	0.039 a	0.467 b	0.576 a,b	0.046 a,b	1.076 a
Randomly I (near center)	-0.045 a,b	0.483 b	0.600 b	0.050 a,b	-0.305 a
Randomly II	-0.042 a,b	0.527 b,c	0.634 b,c	0.063 b,c	-0.357 a
6-points spline	-0.089 b	0.570 c	0.665 c	0.071 c	-0.924 a

As shows Table 4.10, the errors of Catmull-Rom spline (Figure C.2 in Appendix C) with the additional point placed according to the linear regression, into the center of largest interspace or randomly near the center are approximately equal. Statistical tests indicate that there mostly are no significant differences between placing the seventh point to its optimal position, according to the regression and to the center of the largest interspace. On the contrary, considerably worse are variants with only six points or with the seventh point placed randomly.

Table 4.11: Comparison of variants of the seventh point placement for natural cubic spline.

Variant of placement	DB (10^{-2} m)	MAR (10^{-2} m)	SDR (10^{-2} m)	MSR (10^{-3} m^2)	TVD (%)
Optimal position	-0.765 a	2.411 a	2.873 a	1.354 a	5.431 a
20 % of stem height	-0.507 a,b	2.572 a	3.070 a	1.573 a	15.207 a
Center of largest intersp.	1.005 b,c	2.952 a	3.797 a	2.453 a	47.234 b,c
Randomly I (near center)	0.938 b,c	3.037 a,b	3.822 a,b	2.418 a	49.829 b
Randomly II	2.424 c	4.176 b	4.961 b	5.671 a	99.684 c
6-points spline	4.911 d	7.243 c	7.782 c	14.169 b	209.835 d

In natural cubic spline the seventh input point markedly reduces the high errors (Table 4.11) caused by uncontrollable oscillations (Figure C.3). The lowest error values are obtained by placing the seventh point to 20 % of the stem height. But for all variants the oscillation is still very strong.

For B-spline (Figure C.4) the results of the comparison (Table 4.12) show that placing the additional point to the relative height 20 % significantly lowers the diameter bias and the total volume difference if compared with other variants of the additional point placement. On the contrary, values of MAD, SDR and MSR are higher, in some cases significantly.

Table 4.12: Comparison of variants of the seventh point placement for B-spline.

Variant of placement	DB (10^{-2} m)	MAR (10^{-2} m)	SDR (10^{-2} m)	MSR (10^{-3} m^2)	TVD (%)
Optimal position	0.098 a,b	0.553 a	0.760 a,b	0.070 a	0.973 a,b
20 % of stem height	0.033 a	0.621 b	0.811 b	0.086 a,b	-0.004 a
Center of largest intersp.	0.241 c	0.610 a,b	0.756 a	0.079 a,b	3.743 c
Randomly I (near center)	0.204 c	0.618 b	0.774 a,b	0.082 a,b	2.848 c
Randomly II	0.181 a,b	0.654 b	0.807 b	0.089 b,c	2.684 b,c
6-points spline	0.163 b,c	0.718 c	0.857 b	0.104 c	2.506 b,c

While placing the point to the relative height 20 % improves the shape of the curve in the lower stem, which is the most important for volume estimation, placing the point into the largest interspace between existing points results in more even distribution of input points and decrease of the deviations of the modeled curve from the real data.

Differences between individual spline types are also significant. The lowest errors are produced by Catmull-Rom spline, the highest by the natural cubic spline.

Oscillations of natural cubic spline caused by its second degree continuity have been already many times reported Goulding [1979]; Lahtinen [1988]. Natural cubic spline may be suitable for describing smooth curves of coniferous trees, but is not able to cope with complicated profiles of broadleaved trees. By this reason Goulding [1979] does not recommend utilization of cubic splines for modeling stem curves of malformed stems. Also Lahtinen [1988] recommends lowering the degree of continuity for more complicated stems.

Also B-spline due to its approximation property is not suitable for complicated stem curves, although profiles of coniferous trees were well represented using B-spline. It smoothes the driving polygon and therefore it is not able to fit the rapid diameter drop following branching. Good choice is the utilization of Catmull-Rom spline.

Optimal input point combinations proposed by Smaltschinski [1983] or Figueiredo-Filho et al. [1996a] are not applicable for malformed profiles. Determination of optimal position individually for each stem in practice is out of the question, because it requires measurement of many diameters from which one is selected as the best. Utilization of regression seems to be useless regarding identical results obtained by placing the additional input point around the center of the longest interspace. A comparison shows, that the average distance between the location determined by the regression and the center of

the largest interspace is only 6 % of the stem height (median value is only 2 % of the stem height), which is a negligible distance. At the same time it is proved, that input point position determined in this way is not significantly worse than the best position of individual stems.

It can be concluded, that for improving the accuracy the position of the seventh point is important. It should not be placed randomly, but equalizing the point distribution by placing the point in the center of the longest input point interspace can be a reasonable way.

4.2 Local regression model of typical stem form

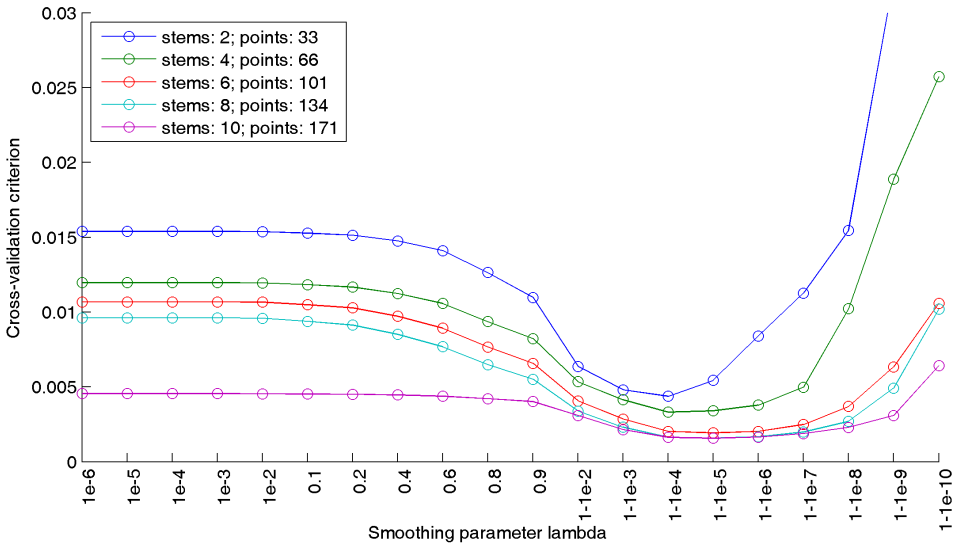
4.2.1 Optimal amount of smoothing for smoothing spline

The smoothing parameter λ can take any values from the close range $< 0, 1 >$. If $\lambda = 0$, the amount of smoothing is maximum and the spline becomes the least-squares regression line fitted to the data. With rising λ the amount of smoothing declines. For the other extreme $\lambda = 1$, the smoothing property of the curve disappears and the smoothing cubic spline becomes interpolation cubic spline.

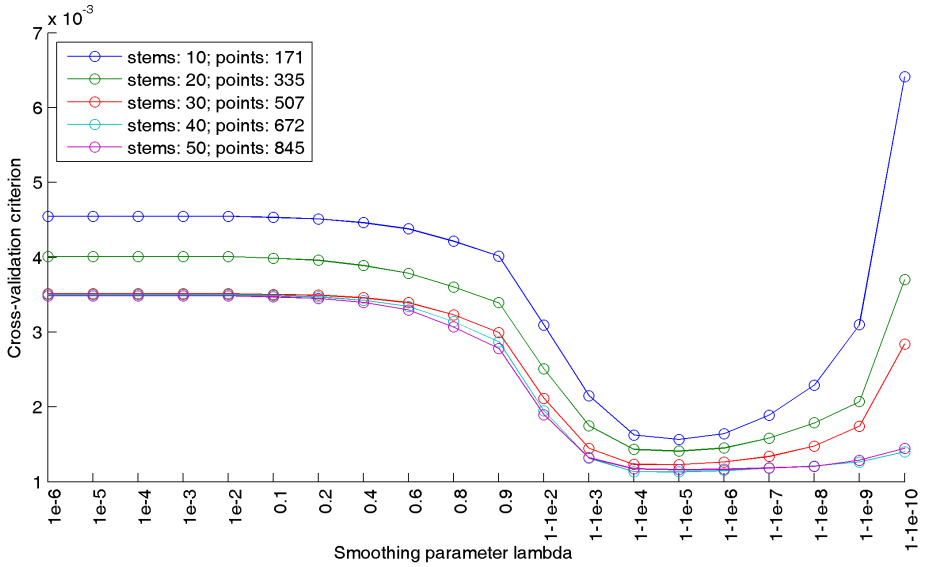
The development of CV criteria in dependence on the value of λ is shown in Figures 4.6 and 4.7. Separate lines in Figure 4.6 indicate the optimal amount of smoothing for different input point density expressed in term of numbers of trees used to fit the regression spline. The development of CV criterion is very similar for all point densities. As indicates Subfigure 4.6a for low numbers of trees (2-4 trees) the CV criterion is minimized with $\lambda = 0.9999$. For higher input point densities the minimum of CV criterion moves to $\lambda = 1 - 10^{-4} = 0.99999$. However, in the range between $1 - 10^{-4}$ and $1 - 10^{-6}$ the change of CV criterion is negligible. Outside that range the value of CV steeply increases. This tendency is observable in all input point densities; with higher densities the rise of CV criterion value with growing λ is less rapid and the tolerance of suitable λ value is extended in the direction towards lower amount of smoothing.

The same results are obtained in case of expressing the point density in terms of the length of input point interspaces (shown in Figure 4.7 on the example of five and twenty stems with different point density). Because the development of CV criterion in dependence on λ value was observed in several discrete points only, it can be concluded, that the optimal

amount of smoothing is achieved with λ ranging between $1 - 10^{-4}$ and $1 - 10^{-6}$. The best choice for λ is $1 - 10^{-5}$ or 0.99999.

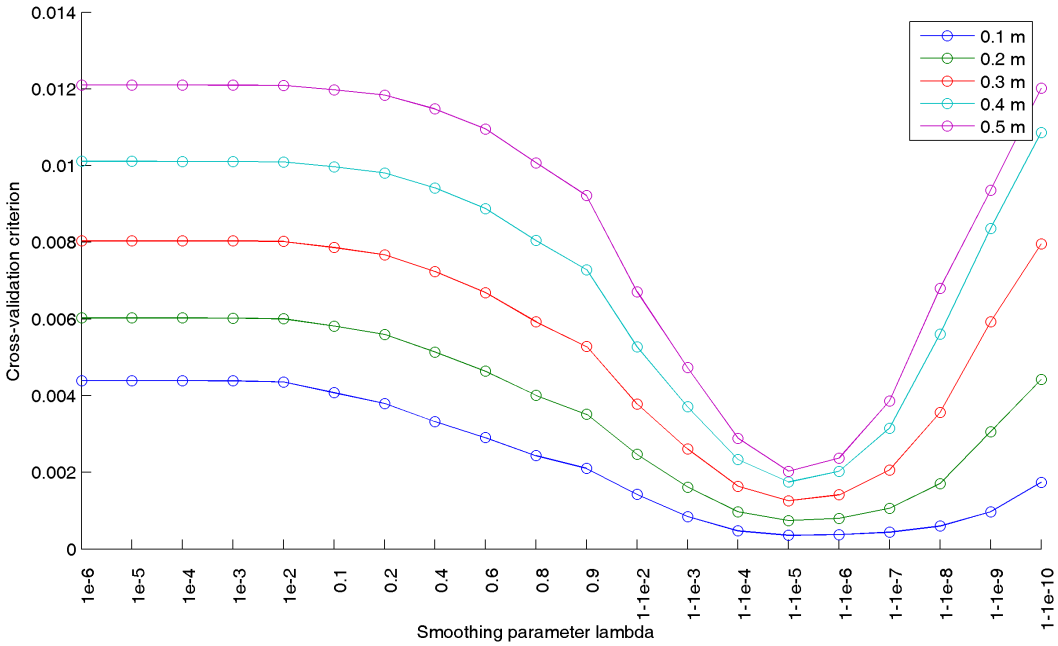


(a) Number of trees: 2, 4, 6, 8, 10

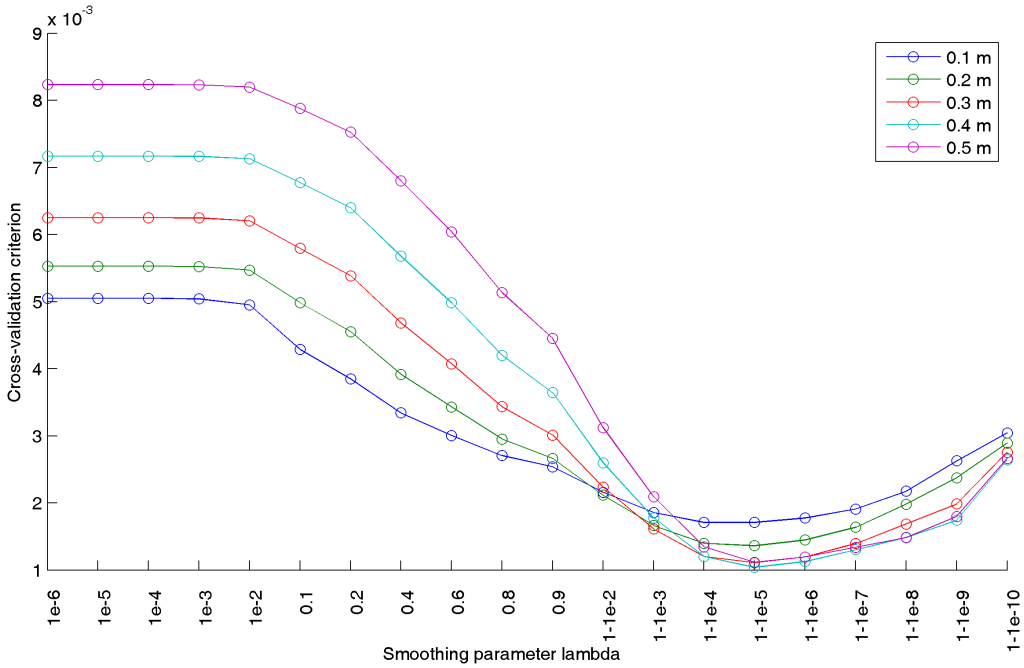


(b) Number of trees: 10, 20, 30, 40, 50

Figure 4.6: Cross-validation values in dependence on smoothing parameter λ for smoothing spline. Separate lines show the development of CV criterion for different density of input points expressed as number of trees.



(a) Number of trees: 5



(b) Number of trees: 20

Figure 4.7: Cross-validation values in dependence on smoothing parameter λ for smoothing spline. Separate lines show the development of CV criterion for different density of input points expressed as the interspace of input points in an individual stem.

It is difficult to compare the optimal value of λ resulting from this study with the results of other authors utilizing the smoothing spline for stem curve modeling or prediction. Liu [1980] utilized a different form of smoothing spline. Neither Nummi and Möttönen [2004] nor Kublin et al. [2008] included the investigation into optimal amount of smoothing in their work.

4.2.2 Optimal amount of smoothing for P-spline

For P-spline the smoothing parameter λ can take any non-negative value. With $\lambda = 0$ the P-spline becomes a polynomial fit. With rising value of λ also rises the amount of smoothing. As λ value approaches infinity, the P-spline becomes a linear regression function.

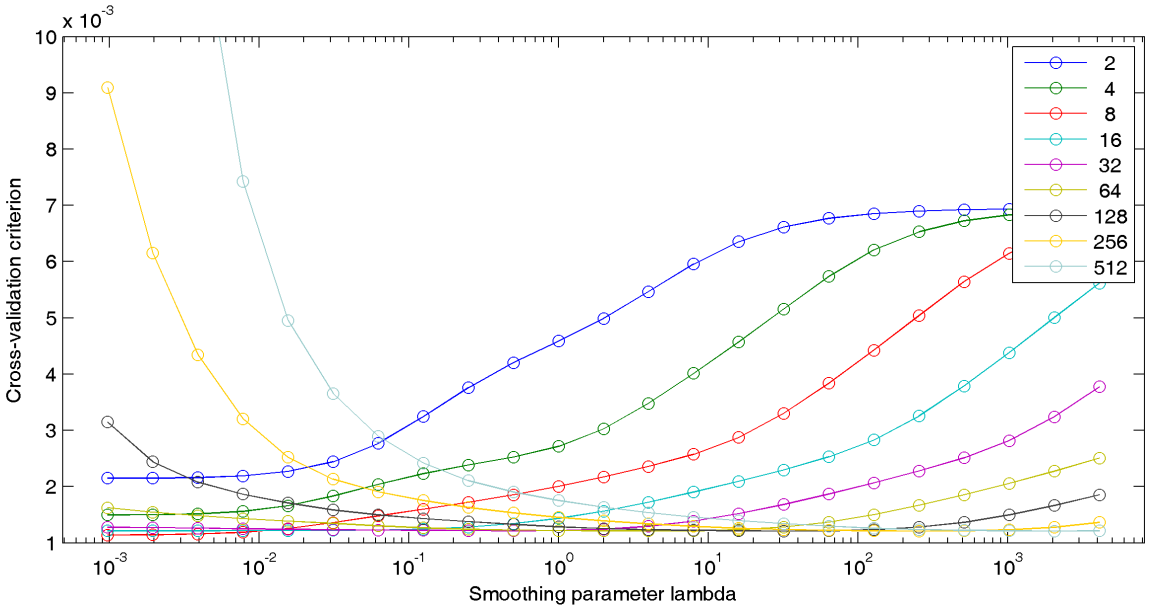


Figure 4.8: Cross-validation values in dependence on smoothing parameter λ for P-spline. Separate lines show the development of CV criterion for different number of segments.

The results of optimizing the smoothing parameter value in dependence on number of P-spline segments by the help of leave-one-out cross-validation are shown in Figure 4.10 and in Table 4.13. It is obvious, that the development of CV criterion with varying λ is strongly dependent on the number of segments. With low numbers of segments the optimal values of λ, having the lowest values of CV criterion, are also low. For rising number of segments,

the optimal λ also increases.

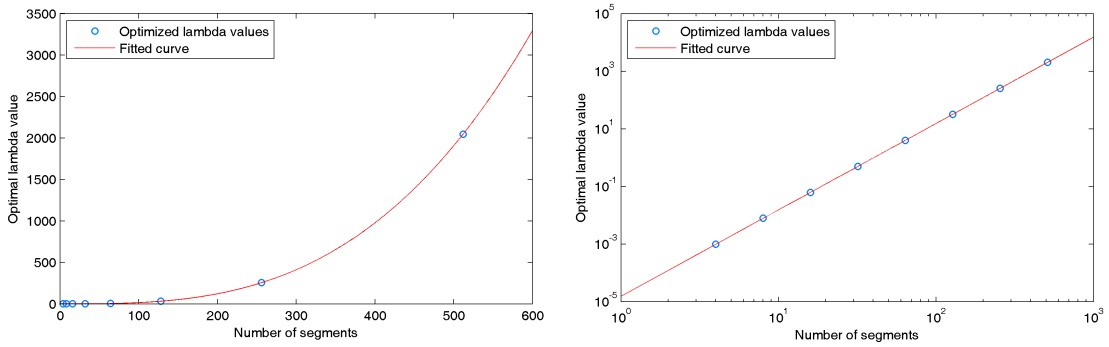
The dependence of the optimal amount of smoothing on the number of segments can be explained with the knowledge in B-spline properties. The lower is the number of P-spline segments, the more input points influence the shape of the segment and the lower is the relative effect of a position of each point. Moreover the cubic segments, which the P-spline consists of, have at the most one inflection point. P-splines consisting of low numbers of segments are smooth by themselves; only a little "additional smoothing" is required.

Table 4.13: Cross-validation values in dependence on smoothing parameter λ and number of segments (CV·10^{-3})

		Smoothing parameter λ											
		$\mathbf{2^{-10}}$	$\mathbf{2^{-8}}$	$\mathbf{2^{-6}}$	$\mathbf{2^{-4}}$	$\mathbf{2^{-2}}$	$\mathbf{2^{0}}$	$\mathbf{2^{2}}$	$\mathbf{2^{4}}$	$\mathbf{2^{6}}$	$\mathbf{2^{8}}$	$\mathbf{2^{10}}$	$\mathbf{2^{12}}$
Number of segments	**2**	2.15	2.16	2.27	2.77	3.76	4.59	5.47	6.35	6.77	6.90	6.93	6.94
	4	1.50	1.52	1.66	2.04	2.38	2.71	3.48	4.57	5.74	6.53	6.83	6.92
	8	1.14	1.16	1.26	1.48	1.72	2.00	2.36	2.88	3.84	5.04	6.15	6.70
	16	1.21	1.22	1.22	1.23	1.28	1.44	1.72	2.09	2.53	3.25	4.38	5.62
	32	1.28	1.26	1.24	1.23	1.22	1.22	1.29	1.52	1.87	2.28	2.82	3.77
	64	1.62	1.48	1.39	1.30	1.24	1.22	1.21	1.23	1.37	1.67	2.06	2.51
	128	3.15	2.08	1.71	1.50	1.38	1.29	1.23	1.22	1.21	1.28	1.50	1.85
	256	9.10	4.35	2.52	1.91	1.63	1.45	1.34	1.26	1.22	1.21	1.23	1.36
	512	34.17	12.35	4.95	2.90	2.11	1.76	1.54	1.39	1.30	1.24	1.22	1.21

In order to describe the dependency between the optimal value of the smoothing parameter on number of P-spline segments a regression analysis was carried out. The λ values with the lowest CV criterion are plotted against their respective number-of-segments. A regression power function fits nearly exactly all the data points (coefficient of determination $r^2 = 1$, root mean square error $rmse = 1.87 \cdot 10^{-13}$, p-value of F-test of the model fit $p = 0$). The form of the regression model together with the coefficients are indicated in equation 4.2.1; the regression is visualized in Figure 4.9.

$$\begin{aligned} \lambda &= \beta_1 \cdot \mathrm{nseg}^{\beta_2} \\ \beta_1 &= 1.526 \cdot 10^{-5} \\ \beta_2 &= 3 \end{aligned} \tag{4.2.1}$$



(a) Linear scales on both axes (b) Logarithmic scales on both axes

Figure 4.9: Dependence of optimal value of smoothing parameter λ on number of P-spline segments fitted by regression power function

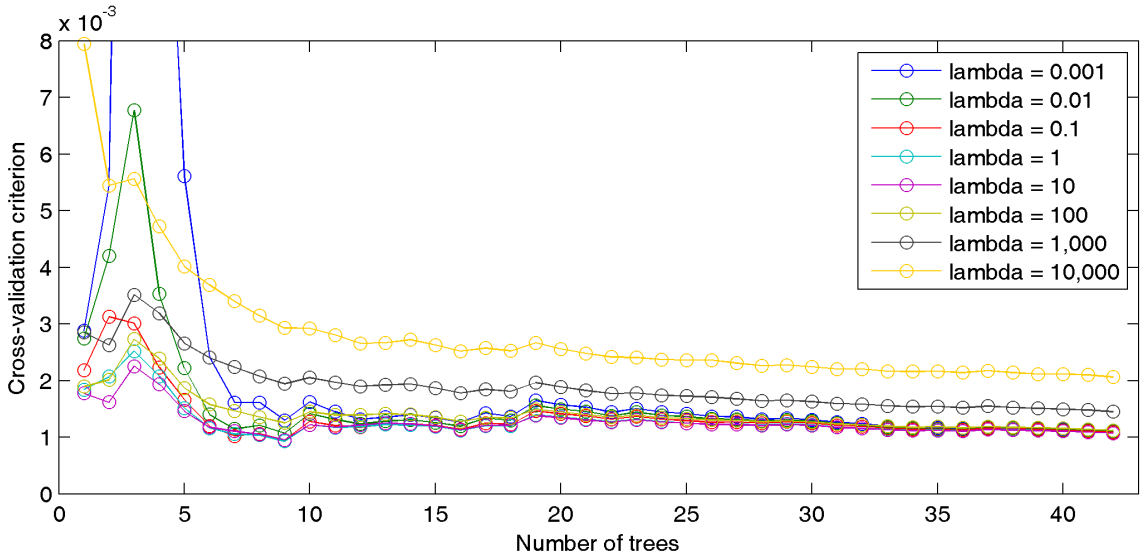
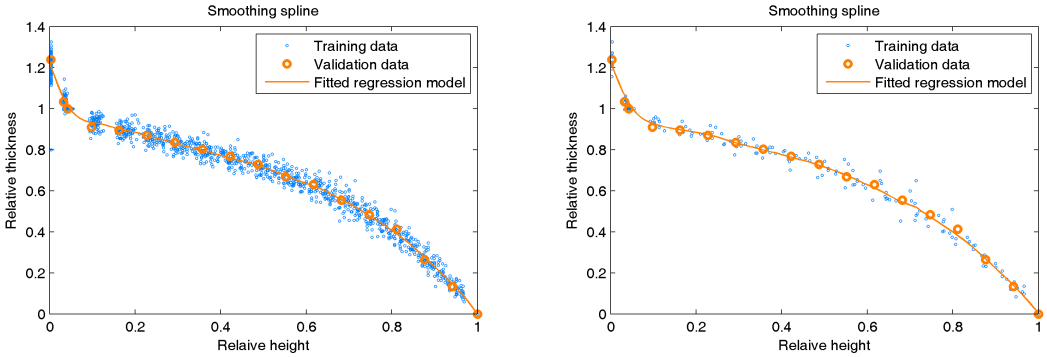


Figure 4.10: Cross-validation values in dependence on number of smoothed points. Separate lines show the development of CV criterion for different values of smoothing parameter λ.

The development of cross-validation criterion with gradually increasing number of input points, expressed in terms of number of stems, was observed in order to determine the influence of input point number on the optimal λ value. The example presented in Figure 4.9 shows the CV criterion development for P-spline with 64 segments. Separate lines denote different λ values.



(a) 84 stems (b) 10 stems

Figure 4.11: Smoothing spline regression curve derived from relative stem profiles of 84 and 10 stems and the validation stem fitted by the regression curve

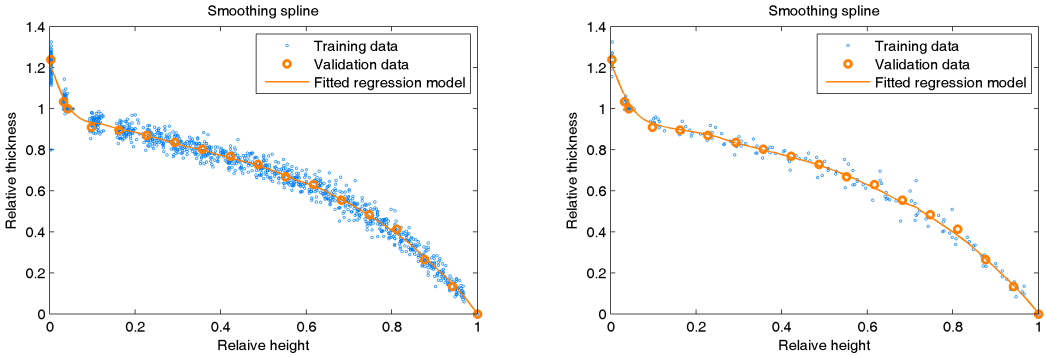
When more than 10 stems (171 points) are used, the CV criterion shows a steady moderate decrease with rising number of input points. The lines of different λ values converge, but their order does not change markedly. For a given number of knots the optimal value of λ is stable for different numbers of input points.

4.2.3 Comparison of spline regression models with selected taper models

Both regression spline models were fitted using the normalized data, that were obtained using the transformation expressed in the equations 3.4.1 and 3.4.3. The normalized height-diameters pairs of all measured stems and ten randomly selected stems are showed together with the regression splines in Figure 4.11 (smoothing spline) and Figure 4.12 (P-spline). In both figures are showed also data points representing a validation stem that was not used for model fitting.

All models retransformed to fit a particular validation stem in real proportions are showed in Appendix D. Figure D.2 and Figure D.3 show the fitted models derived from 84 stems (all measured stems with the exception of one validation stem). Models derived from measured data of 10 randomly selected stems are shown in Figures D.4 and D.5.

Figure 4.13 shows the development of the evaluative criteria (MAR and MSR) in dependence on rising number of stems incorporated in regression models. The particular example showed in the figure represents residuals of P-spline models; developments of the evaluative



(a) 84 stems (b) 10 stems

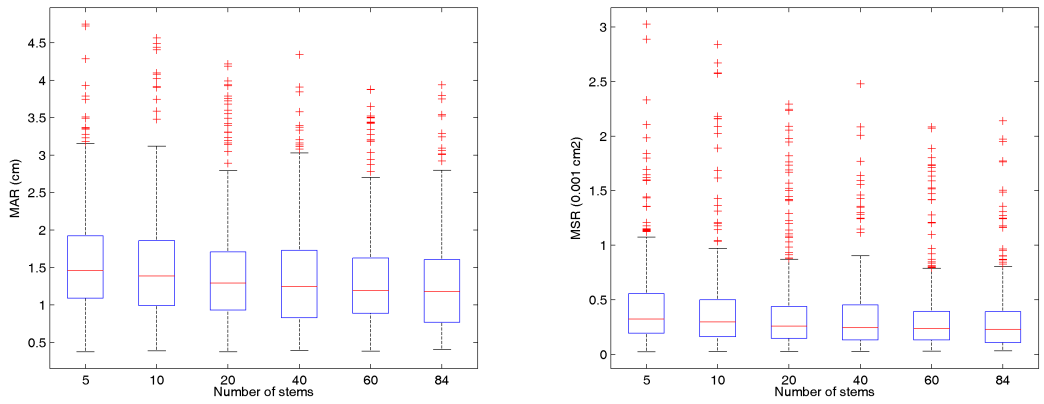
Figure 4.12: P-spline regression curve derived from relative stem profiles of 84 and 10 stems and the validation stem fitted by the regression curve

criteria for other taper models are similar. The development of total volume difference and absolute volume difference of P-spline regression models is showed in Figure D.1.

With rising numbers of stems the values of absolute diameter errors as well as absolute volume differences decline. With the help of two-way analysis of variance, the decline of the error values both for diameters and volume proved significant. For the model of Bi [2000] the accuracy drop with lower number of stems is very pronounced. While with 84 stems its accuracy is very good in comparison with other models, for five stems only the performance of the model is very poor. On the other hand the model of Lee et al. [2003] has the lowest accuracy among all models with high number of stems used, while for low number of stems the accuracy of its predictions was comparable with the other models. The accuracy of the model is not so strongly dependent on the number of stems.

When all measured stems were used to derive the regression model (Table 4.14), both splines represented the mean function of typical stem curve very well. There was no systematic error in diameter prediction nor in volume estimation. The mean errors for both diameter (less than 2 mm) and volume (less than 1 %) prediction were very low. However, the mean absolute residuals achieve quite high values, and also the variances of DB and TVD are high, what corresponds with the high values of mean absolute volume differences (Figure D.1 in Appendix D).

With lowering the number of stems, the diameter predictions of nearly all models become significantly biased. No significant diameter bias was only found with the model of Bi [2000]. This is caused above all by high variance of the prediction accuracy. Although the variances are not significantly different, the standard deviations of evaluative criteria



(a) Mean absolute residual

(b) Mean squared residual

Figure 4.13: Development of accuracy of stem curve prediction with rising number of stems incorporated in the P-spline regression model

in Tables D.1 to D.5 are usually the highest with the model of Bi [2000]. With only five stems used to derive the model parameters, variances of the criteria are extremely high (Table D.5). On the other hand, when the model was parameterized with the use of all stems, the accuracy of the model was high.

Concerning the criteria expressing the quality of fit of the curve (MAR, SDR, MSR) two groups of models according to the accuracy can be distinguished. For high number of stems used for model parameterization the segmented polynomial model of Max and Burkhart [1976], the variable-exponent model of Bi [2000], and both spline models show better results than the single polynomial model of Cervera and the variable-exponent model of Lee et al. [2003]. With lower number of trees the most accurate models are the segmented polynomial model together with the P-spline model; lower accuracy was observed with the smoothing spline models. The single polynomial model and both variable-exponent models showed significantly higher errors.

Table 4.14: Comparison of taper models based on all 85 stems. For each statistic mean value and standard deviation is shown. Values in a column followed by the same letter indicate no significant difference between spline types. Stars in columns DB and TVD indicate mean values significantly different from zero.

Taper model	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
Cervera	0.26	1.27	a	1.58	0.60	a	1.88	0.56	a	0.44	0.33	a	-1.22	6.61	a
Max-Burkhart	0.18	1.31	a	1.32	0.68	b	1.52	0.67	b	0.34	0.35	b	-0.24	6.73	a
Bi	0.01	1.09	a	1.31	0.60	b	1.55	0.62	b	0.32	0.27	b	-0.95	5.69	a
Lee et al.	0.22	1.10	a	1.56	0.50	a	1.93	0.48	a	0.43	0.25	a	0.09	5.70	a
Smoothing spline	0.19	1.34	a	1.37	0.68	a,b	1.56	0.65	b	0.35	0.34	b	0.59	6.75	a
P-spline	0.13	1.31	a	1.31	0.67	b	1.50	0.66	b	0.33	0.34	b	-0.29	6.69	a

Almost all tested models are able to represent well the mean function of the typical stem curve. With rising number of stems incorporated in the regression model the models approach the real mean function (the typical stem curve) and the new trees are represented with lower error. The problem, which the models are not able to cope with, is the between-tree stem form variability. If the shape of a particular stem differentiates from the mean stem curve function, the so-called fixed-effect taper models are not able to model accurately the stem curve. Therefore so-called mixed effect taper models were introduced in last years [Leites and Robinson, 2004; Lejeune et al., 2009; Cao and Wang, 2011; Fonweban et al., 2011]. The mixed effect models consist of the fixed mean function describing the average stem curve and the random effect describing the error structure and allowing to specify the stem curve for each invividual tree on the base of an additional information, e.g. an upper stem diameter Sharma and Parton [2009].

The regression models compared are the fixed effect models describing only the mean function an are not able to consider the between-tree variability and to conform to each individual stem. That is the reason why the variability of diameter and volume prediction is such high, although the mean error is very low. The method is suitable for predicting the volume of a group of trees or volume of a stand. In such case the mean stem curve will approach the regression model and the mean diameter and volume errors will approach zero. If volume of a single stem is estimated, the real stem curve can deviate from the model and the absolute volume error will be approximately 5 to 6 % in average, as shown in Figure D.1.

It was stated many times [Max and Burkhart, 1976; Newnham, 1992; Jiang et al., 2005], that the single polynomial models are too rigid to conform the complicated shape

of stem curve. This fact proved true also in this comparison, where the model considered being the best among the single polynomial models [Rojo et al., 2005] was outperformed by other models.

The rather complicated variable-exponent model of Bi [2000] is able to produce accurate predictions if the model parameter values are derived from a high number of stem profiles. In the comparison of Rojo et al. [2005] the models were parameterized using stem profiles of 203 stems. The variable-exponent taper model has to be fit using non-linear least squared fitting methods, such as Levenberg-Marquardt algorithm, that are not able to produce the unique best solution and are not abe to assure the optimality of the result. With high number of parameters in the model and few data point to be fitted, the methods for model parameterization can be unstable and give inaccurate results [Kublin et al., 2008].

An important result of the comparison is that the P-spline regression taper model performed at least as well as the models regarded [Rojo et al., 2005] as the best representatives of three main groups of taper models: single polynomial models, segmented models, and variable-exponent models. The performance of the smoothing spline regression taper model was comparable with the performance of P-spline in most aspects; however in some rare cases the values of the evaluative error statistics were higher with statistically significant difference. As stated by Koskela et al. [2006], for regression spline models the choice of the smoothing parameter λ is crucial. Regarding the studies performed to optimize the smoothing parameter value under variable conditions it can be assumed, that the λ value used approximated the optimal amount of smoothing.

The taper models utilizing regression splines, specifically the smoothing spline and the P-spline, can be used to represent the mean function of the local typical stem curve. Their performance is comparable with the performance of variable-exponent models, generally considered to be the most accurate type of taper models. Parameterization of non-linear variable exponent models with usually high number of parameters is done using computationally expensive and numerically instable algorithms. On the other hand, simple parametric models are usually too rigid to depict the stem curve appropriately. The non-parametric spline regression models combines the advantages of both approaches; they provide enough flexibility and fitting algorithms are inexpensive and numerically stable.

Chapter 5

Conclusions

5.1 Summary

The main goal of the doctoral thesis was to explore possibilities and suitability of spline functions for stem curve representation.

The first part of the thesis summarizes the basic concepts in stem form description and gives an overview of basic methods and approaches for stem curve modeling. Also a theoretical background for splines is provided. The rest of the thesis describes utilized methods and results of two major challenges that were solved. The first challenge was to find an appropriate spline function for describing with sufficient accuracy stem curves of individual stems of both coniferous and broadleaved species using points obtained from a digital photography interactively by human interpreter. The second challenge was to find a methodology for developing a regression model of typical local stem curve through the use of splines.

The research was processed on data of 85 Norway spruce stems and 48 stems broadleaved trees. The spruce data come from 50 to 100 years old stands located in the School Forest Enterprise Kostelec nad Černými lesy. The broadleaved stems come from the protected landscape area Bílé Karpaty and include two species, common beach and mountain oak. The data of both coniferous and broadleaved trees contain closely measured stem profiles with the diameter interspaces 0.1 m.

Individual stem curve modeling

On the basis of literature research several different spline techniques were selected and programs for spline computation were written. The number of splines for further analyses was reduced by virtue of results of preliminary analyses examining pertinence of individual spline techniques for stem form modeling. Spline techniques not suitable for the particular purpose were eliminated from subsequent analyses.

The performance of individual splines was compared using nine input point sets with different point numbers. Input point positions were chosen according to Figueiredo-Filho et al. [1996a], who considered those positions as optimal for spline interpolation. As a result of the primary comparison of the splines, four candidates were selected for further work: the natural cubic spline, the Catmull-Rom spline, 2^{nd} degree interpolation B-spline and 2^{nd} degree approximation B-spline.

The next work dealt with the optimization of input point positions. The input point placement can be crucial for accuracy of the stem curve representation. It was found that due to different properties of individual splines the optimal input point positions also differ. For mentioned splines optimal sets of input point positions were stated for five to nine input points. Less than five input points cannot ensure reasonable accuracy of the model. Adding more than nine input points has very little effect on accuracy improvement.

The four most promising spline types were once more compared with the use of individual optimal input point distribution for each spline. The stem curves are best represented by Catmull-Rom spline. The spline has no systematic bias in both diameter and volume prediction and produces a reliable curve with low and well-balanced residuals. A reasonable stem curve representation is produced also by interpolation B-spline, but the errors are significantly higher. The approximation B-spline can represent well the stem curve with the use of higher numbers of input points. However due to the approximation property of the spline, the diameters will always be overestimated in the convex part of the curve and underestimated in the concave part. The natural cubic spline, which has been utilized many times for stem curve representation, showed high propensity for oscillation and therefore it was outperformed by the other splines in terms of stem curve model accuracy. With rising number of input points the risk of oscillation decreases, but it does not disappear even with high numbers of input points.

The Catmull-Rom spline gives very good results in case of approximately uniform distribution of input points. If the points are distributed markedly unevenly, the curve can

produce oscillation or even loops. To avoid this effect, a method of tangent vector correction is presented.

Following recommendations result from the study. Catmull-Rom spline is a reasonable interpolation method applicable to the problem of stem curve modeling from a set of measured diameters along the stem. To maximize the accuracy the input point should be distributed as indicated in the thesis. To avoid unexpected behavior of the curve in case of markedly unequal distances between input points, the tangent vector length corrections should be used.

Local regression model of typical stem form

Possibilities of non-parametric regression techniques were investigated. For the purpose two spline regression techniques were selected: smoothing spline and P-splines. Both techniques were used to represent the mean function expressing the dependence of relative diameter on relative height.

For both techniques the optimal amount of smoothing was investigated. The smoothing parameter λ was optimized in dependence on number of stems incorporated in the model and on density of input points. The optimization was carried out using leave-one-out cross-validation method. For smoothing spline, the optimal value of the smoothing parameter was approximately 0.99999, independently on number of stems. For P-splines, the optimal value of the smoothing parameter is also independent on number of stems, but it is determined by number of P-spline segments.

The stem curve models represented by optimally smoothed smoothing spline and P-spline were compared with stem curves modeled by best representatives of three main groups of parametric taper models: a polynomial model, a segmented polynomial model and two variable-exponent models. The comparison was carried out using cross-validation approach, where repeatedly one stem was retained as validation data and the other stems served as data for model fitting.

Both regression splines showed good results. Their performance was significantly better that the performance of the polynomial model and one of the variable-exponent models. The accuracy of stem curves represented by the second variable-exponent model and the segmented polynomial model was comparable with the accuracy of stem curves represented by spline models. The advantage of spline models in contrast to variable-exponent models is the simplicity and numeric stability of the model computation. With decreasing number of stems incorporated in the regression model the accuracy decline for all models; however

with spline models the accuracy drop is not as strong as with some of the parametric models, especially the variable-exponent models.

Following findings result from the study. Both regression splines can be used to model the mean function of typical stem form. The accuracy of P-spline models is higher than the accuracy of smoothing spline models. For both splines the amount of smoothing must be optimized: for smoothing spline the optimal λ proved 0.99999, for P-spline the optimal λ is determined by number of P-spline segments.

5.2 Contributions of the Thesis

1. Design of a methodology of a single stem curve representation by splines, readily feasible for the use in the software DendroScanner.
2. Design of a methodology of a regression model of typical stem using splines, readily feasible for the use in the software DendroScanner.

5.3 Future Work

The author is aware that the topic of spline models of stem curves of individual stems and spline regression models of a typical local stem curve is not exhaustively developed in the doctoral thesis. The future work extending the scope of the doctoral thesis could cover the following:

- It would be interesting to explore possibilities of even more non-parametric regression methods involving splines, e.g. smoothing by using B-splines.
- The spline regression models utilized in the thesis are the mean functions of the typical stem curve. The error structure of the regression is neglected. The methodology of regression spline taper models could be further improved by implementation of random effect allowing the model to describe individual stems.
- There are several other non-parametric stem curve prediction methods, that have been proposed by different authors. It would be interesting to include such methods in the investigation and explore their performance in comparison with performance of the models used in the thesis.

Bibliography

Alder, D. (1978). Pymod: A forecasting model for conifer plantations in the tropical highlands of eastern Africa. In Fries, J., Burkhart, H. E. and Max, T. A., editors, *Growth models for long term forecasting of timber yields*, Blacksburg. Virginia Polytechnic Institute and State University.

Amidon, E. L. (1984). Notes: A general taper functional form to predict bole volume for five mixed-conifer species in California. *Forest Science*, 30(1):166–171.

Aydin, D. (2007). A comparison of the nonparametric regression models using smoothing spline and kernel regression. *Engineering Sciences*, 2(2):253–257.

Bailey, R. L. (1995). Upper stem volumes from stem analysis data: an overlapping bolts method. *Canadian Journal of Forest Research*, 25(1):170–173.

Bi, H. (2000). Trigonometric variable-form taper equations for australian eucalypts. *Forest Science*, 46(3):397–409.

Bi, H. and Long, Y. (2001). Flexible taper equation for site-specific management of Pinus radiata in New South Wales, Australia. *Forest Ecology and Management*, 148(1-3):79 – 91.

Biging, G. S. (1984). Taper equations for second-growth mixed conifers of Northern California. *Forest Science*, 30:1103–1117.

Biging, G. S. (1988). Estimating the accuracy of volume equations using taper equations of stem profile. *Canadian Journal of Forest Research*, 18:1002–1007.

Bluhm, A. A., Garber, S. M. and Hibbs, D. E. (2007). Taper equation and volume tables for plantation-grown red alder. Technical report, U.S. Dept. of Agriculture, Forest Service, Pacific Northwest Research Station, Portland.

Brooks, J. R., Jiang, L. and Ozçelik, R. (2008). Compatible stem volume and taper equations for Brutian pine, Cedar of Lebanon, and Cilicica fir in Turkey. *Forest Ecology and Management*, 256(1-2):147 – 151.

Bruce, D. (1972). Notes: Some transformations of the Behre equation of tree form. *Forest Science*, 18(2):164–166.

Bruce, D., Curtis, R. O. and Vancoevering, C. (1968). Development of a system of taper and volume tables for red alder. *Forest Science*, 14:339–350(12).

Bruce, D. and Max, T. A. (1990). Use of profile in tree volume estimation. In LaBau, V. J. and Cunia, T., editors, *USDA Forest Service general technical report PNW-GTR*, pages 213–227. Pacific Northwest Research Station.

Burkhart, H. and Tomé, M. (2012). *Modeling Forest Trees and Stands*. Springer.

Burkhart, H. E. and Walton, S. B. (1985). Notes: Incorporating crown ratio into taper equations for loblolly pine trees. *Forest Science*, 31(2):478–484.

Candy, S. G. (1989). Compatible tree volume and variable-form stem taper models for Pinus radiata in Tasmania. *New Zealand Journal of Forestry Science*, 19:97–111.

Cao, Q. V., Burkhart, H. E. and Max, T. A. (1980). Evaluation of two methods for cubic-volume prediction of loblolly pine to any merchantable limit. *Forest Science*, 26(1):71–80.

Cao, Q. V. and Wang, J. (2011). Calibrating fixed- and mixed-effects taper equations. *Forest Ecology and Management*, 262(4):671 – 673.

Catmull, E. and Rom, R. (1974). A class of local interpolating splines. In Barnhill, R. and Riesenfeld, R., editors, *Computer aided geometric design*, pages 317–326, San Diego. Academic Press.

Chiba, Y. (1990). Plant form analysis based on the pipe model theory I. A statical model within the crown. *Ecological Research*, 5(2):207–220.

Chiba, Y. (1991). Plant form based on the pipe model theory II. Quantitative analysis of ramification in morphology. *Ecological Research*, 6(1):21–28.

Chiba, Y. and Shinozaki, K. (1994). A simple mathematical model of growth pattern in tree stems. *Annals of Botany*, 73(1):91–98.

Clark, A., Souter, R. and Schlaegel, B. (1991). *Stem profile equations for southern tree species*. Research paper SE. Southeastern Forest Experiment Station.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:859–836.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83:596–610.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31:377–403. 10.1007/BF01404567.

Czaplewski, R. L. and McClure, J. P. (1988). Conditioning a segmented stem profile model for two diameter measurements. *Forest Science*, 34(2):512–522.

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

Dean, T., Roberts, S., Gilmore, D., Maguire, D., Long, J., O'Hara, K. and Seymour, R. (2002). An evaluation of the uniform stress hypothesis based on stem geometry in selected North American conifers. *Trees*, 16(8):559–568.

Deleuze, C. and Houllier, F. (1995). Prediction of stem profile of Picea abies using a process-based tree growth model. *Tree Physiology*, 15(2):113–120.

Demaerschalk, J. P. (1972). Converting volume equations to compatible taper equations. *Forest Science*, 18(3):241–245.

Demaerschalk, J. P. and Kozak, A. (1974). Suggestions and criteria for more effective regression sampling. *Canadian Journal of Forest Research*, 4(3):341–348.

Demaerschalk, J. P. and Kozak, A. (1977). The whole-bole system: a conditioned dual-equation system for precise prediction of tree profiles. *Canadian Journal of Forest Research*, 7(3):488–497.

Diéguez-Aranda, U., Castedo-Dorado, F., Álvarez-González, J. G. and Rojo, A. (2006). Compatible taper function for scots pine plantations in northwestern Spain. *Canadian Journal of Forest Research*, 36(5):1190–1205.

Děmidovič, B. P. and Maron, I. A. (1966). *Základy numerické matematiky*. Státní nakladatelství technické literatury, Praha.

Eberly, D. (2008). Least-squares fitting of data with B-spline curves. Technical report, Geometric Tools, LLC.

Eerikäinen, K. (2001). Stem volume models with random coefficients for Pinus kesiya in Tanzania, Zambia, and Zimbabwe. *Canadian Journal of Forest Research*, 31(5):879–888.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Ezquerra, F. J. and Gil, L. A. (2001). Wood anatomy and stress distribution in the stem of *Pinus pinaster* ait. *Investigación agraria. Sistemas y recursos forestales*, 10:165–177.

Fabrika, M. and Pretsch, H. (2011). *Analýza a modelovanie lesných ekosystémov*. Technická univerzita vo Zvolene, Zvolen.

Fang, Z., Borders, E. and Bailey, L. (1999). Tree volume and upper-stem diameters predictions for planted lobolly and slash pine based on compatible volume-taper system with segmented-stem form factors. Technical report, University of Georgia. Warnell School of Forest Resources.

Fang, Z., Borders, E. and Bailey, L. (2000). Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *Forest Science*, 46(1):1–12.

Figueiredo-Filho, A., Borders, B. E. and Hitch, K. L. (1996a). Number of diameters required to represent stem profiles using interpolated cubic splines. *Canadian Journal of Forest Research*, 26(7):1113–1121.

Figueiredo-Filho, A., Borders, B. E. and Hitch, K. L. (1996b). Taper equations for Pinus taeda plantations in Southern Brazil. *Forest Ecology and Management*, 83(1-2):39 – 46.

Figueiredo-Filho, A., Machado, S. A. and Carneiro, M. R. A. (2000). Testing accuracy of log volume calculation procedures against water displacement techniques (xylometer). *Canadian Journal of Forest Research*, 30(6):990–997.

Figueiredo-Filho, A. and Schaaf, L. B. (1999). Comparison between predicted volumes estimated by taper equations and true volume obtained by the water displacement technique (xylometer). *Canadian Journal of Forest Research*, 29(4):451–461.

Flewelling, J. W. (1993). Variable-shape stem-profile predictions for western hemlock. Part II. Predictions from DBH, total height, and upper stem measurements. *Canadian Journal of Forest Research*, 23(3):537–544.

Flewelling, J. W., Ernst, R. L. and Raynes, L. M. (2000). Use of three-point taper systems in timber cruising. In Hansen, M. and Burk, T., editors, *Integrated tools for natural resources inventories in the 21st century*, pages 364–371. U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.

Flewelling, J. W. and Raynes, L. M. (1993). Variable-shape stem-profile predictions for western hemlock. Part I. Predictions from DBH and total height. *Canadian Journal of Forest Research*, 23(3):520–536.

Fonweban, J., Gardiner, B., Macdonald, E. and Auty, D. (2011). Taper functions for scots pine (Pinus sylvestris L.) and sitka spruce (Picea sitchensis (Bong.) Carr.) in Northern Britain. *Forestry*, 84(1):49–60.

Fuller, W. A. (1969). Grafted polynomials as approximating functions. *Australian Journal of Agricultural Economics*, 13(01).

Gaffrey, D., Sloboda, B. and Matsumura, N. (1998). Representation of tree stem taper curves and their dynamic, using a linear model and the centroaffine transformation. *Journal of Forest Research*, 3:67–74.

Gallant, A. R. and Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association*, 68(341):144–147.

Goodwin, A. N. (2009). A cubic tree taper model. *Australian forestry*, 72(2):87–98.

Goulding, C. J. (1979). Cubic spline curves and calculation of volume of sectionally measured trees. *New Zealand Journal of Forest Science*, 9(1):89–99.

Goulding, C. J. and Murray, J. C. (1976). Polynomial taper equations that are compatible with tree volume equations. *New Zealand Journal of Forest Science*, 5:313–322.

Gray, H. R. (1956). The form and taper of forest-tree stems. *University of Oxford. Imperial Forestry Institute Paper*, (32).

Gregorie, T. G. (1987). Generalized error structure for forestry yield models. *Forest Science*, 33(2):423–444.

Grosenbaugh, L. R. (1966). Tree form: Definition, interpolation, extrapolation. *The Forestry Chronicle*, 42(4):444–457.

Hilt, D. (1980). *Taper-based system for estimating stem volumes of upland oaks.* Forest Service research paper. Department of Agriculture, USDA Forest Service, Northeastern Forest Experiment Station.

Hradetzky, J. (1981). Spline-Funktionen und ihre Anwendung in der forstlichen Forschung. *Forstwissenschaftliches Centralblatt*, 100:45–59.

Huang, S., Price, D., Morgan, D. and Peck, K. (2000). Kozak's variable-exponent taper equation regionalized for white spruce in Alberta. *Western Journal of Applied Forestry*, 15(2):75–85.

Ježek, F. (2000). *Geometrické a počítačové modelování, pomocný učební text.* ZČU, Plzeň.

Jiang, L., Brooks, J. R. and Wang, J. (2005). Compatible taper and volume equations for yellow-poplar in West Virginia. *Forest Ecology and Management*, 213(1-3):399 – 409.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering, Transactions of the ASME*, (82 (Series D)):35–45.

Kaufmann, E. (2001). Estimation of standing trimber, growth and cut. In Brassel, P. and Lischke, H., editors, *Swiss National Forest Inventory: Methods and Models of the Second Assessment*, Birmensdorf. WSL Swiss Federal Research Institute.

Kirchner, F., Figueiredo-Filho, A., Scolforo, J., Machado, S. and Mitishita, E. (1991). O uso de funções spline no cálculo de volume de árvores. *Revista Floresta*, 19:116–122.

Kitagawa, G. (2008). Contributions of professor Hirotogu Akaike in statistical science. *Journal of Tha Japan Statistical Society*, 38(1):119–130.

Kitikidou, K. and Chatzilazarou, G. (2008). Estimating the sample size for fitting taper equations. *Journal of Forest Science*, 54(4):176–182.

Kochanek, D. H. U. and Bartels, R. H. (1984). Interpolating splines with local tension, continuity, and bias control. *SIGGRAPH Comput. Graph.*, 18:33–41.

Korf, V., Hubač, K., Šmelko, Š. and Wolf, J. (1972). *Dendrometrie*. Státní Zemědělské Nakladatelství, Praha.

Koskela, L., Nummi, T., Wenzel, S. and Kivinen, V.-P. (2006). On the analysis of cubic smoothing spline-based stem curve prediction for forest harvesters. *Canadian Journal of Forest Research*, 36(11):2909–2919.

Kozak, A. (1988). A variable-exponent taper equation. *Canadian Journal of Forest Research*, 18(11):1363–1368.

Kozak, A. (1997). Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. *Canadian Journal of Forest Research*, 27(5):619–629.

Kozak, A. (2004). My last words on taper equations. *The Forestry Chronicle*, 80(4):507–515.

Kozak, A., Munro, D. D. and Smith, J. H. G. (1969a). More accuracy required. *Truck Logger*, (December):20–21.

Kozak, A., Munro, D. D. and Smith, J. H. G. (1969b). Taper functions and their application in forest inventory. *The Forestry Chronicle*, 45(4):278–283.

Kozak, A. and Smith, J. H. G. (1993). Standards for evaluating taper estimating systems. *The Forestry Chronicle*, 69(4):438–444.

Kublin, E., Augustin, N. and Lappi, J. (2008). A flexible regression model for diameter prediction. *European Journal of Forest Research*, 127:415–428.

Kunze, M. (1896). *Die absoluten Formzahlen der gemeinen Kiefer*. Schönfeld, Dresden.

Laasasenaho, J. (1982). Taper curve and volume functions for pine, spruce and birch. *Communications Instituti Forestalis Fenniae*, 108:1–74.

Laasasenaho, J., Melkas, T. and Aldén, S. (2005). Modelling bark thickness of Picea abies with taper curves. *Forest Ecology and Management*, 206(1-3):35 – 47.

Lahtinen, A. (1988). On the construction of monotony preserving taper curves. *Acta Forestalia Fennica*, (203):1–34.

Lahtinen, A. (1990). Shape preserving interpolation by quadratic splines. *Journal of Computational and Applied Mathematics*, 29(1):15 – 24.

Lahtinen, A. and Laasasenaho, J. (1979). On the construction of taper curves by using spline functions. *Communicationes Instituti Forestalis Fenniae*, 95(8):1–63.

Lappi, J. (2006). A multivariate, nonparametric stem-curve prediction method. *Canadian Journal of Forest Research*, 36(4):1017–1027.

Larson, P. R. (1963). Stem form development of forest trees. *Forest Science*, 9(Supplement 5):1–42.

Larson, P. R. (1965). Stem form of young Larix as influenced by wind and pruning. *Forest Science*, 11(4):412–424.

Lee, W.-K., Seo, J.-H., Son, Y.-M., Lee, K.-H. and von Gadow, K. (2003). Modeling stem profiles for Pinus densiflora in Korea. *Forest Ecology and Management*, 172(1):69 – 77.

Leites, L. P. and Robinson, A. P. (2004). Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. *Forest Science*, 50(2):204–212.

Lejeune, G., Ung, C.-H., Fortin, M., Guo, X., Lambert, M.-C. and Ruel, J.-C. (2009). A simple stem taper model with mixed effects for boreal black spruce. *European Journal of Forest Research*, 128:505–513. 10.1007/s10342-009-0300-8.

Li, R., Weiskittel, A., Dick, A. R., Kershaw, J. A. and Seymour, R. S. (2012). Regional stem taper equations for eleven conifer species in the Acadian Region of North America: Development and assessment. *Northern Journal of Applied Forestry*, 29(1):5–14.

Li, R. and Weiskittel, A. R. (2010). Comparison of model forms for estimating stem taper and volume in the primary conifer species of the North American Acadian Region. *Annals of Forest Science*, 67(3):302p1–302p16.

Lin, H., Wang, G. and Dong, C. (2004). Constructing iterative non-uniform B-spline curve and surface to fit data points. *Science in China Series F: Information Sciences*, 47(3):315–331.

Linkeová, I. (1999). *Konstrukce, výroba a měření obecných tvarových ploch.* PhD thesis, Czech Technical University in Prague, Faculty of Electrical Ingeneering.

Linkeová, I. (2001). Determination of tangent vectors of Ferguson interpolation curves and surfaces. *Acta Polytechnica*, 40(5-6):27–32.

Linkeová, I. (2007). *NURBS křivky. NeUniformní Racionální B-Spline křivky.* Nakladatelství ČVUT, Praha. 208 pg.

Linkeová, I. (2008). *Základy počitačového modelování křivek a ploch.* Nakladatelství ČVUT, Praha. 145 pg.

Lisky, E. P. and Nummi, T. (1995). Prediction of tree stems to improve efficiency in automatized harvesting of forest. *Scandinavian Journal of Statistics*, 22(2):255–269.

Liu, C. J. (1 September 1980). Log volume estimation with spline approximation. *Forest Science*, 26(9):361–369.

Lynch, T. B., Wiant Jr., H. V. and Patterson, D. W. (1994). Comparison of log volume estimates using formulae for log center of gravity and center of volume. *Canadian Journal of Forest Research*, 24(1):133–138.

MacCallum, K. J. and Zhang, J.-M. (1986). Curve-smoothing techniques using B-splines. *The Computer Journal*, 29(6):564–571.

Machado, S. A. and Nadolny, M. C. (1991). Comparação de metodos de cubagem de árvores e diversos comprimetnos de seção. In *Anais do III Congresso Florestal e do Meion Ambiente do Paraná*, volume 1, pages 89–104. Curitiba, Brazil: Associação Parananense de Engenheiros Florestais.

Machado, S. a. D. A., Téo, S. J., Urbano, E., Figura, M. A. and Silva, L. C. R. (2006). Comparação de métodos de cubagem absolutos com o volume obtido pelo xilômetro para bracatinga (*Mimosa scabrella* Bentham). *Young*, 12:239–253.

Mäkelä, A. (2002). Derivation of stem taper from the pipe theory in a carbon balance framework. *Tree Physiology*, 22(13):891–905.

Manning, J. R. (1974). Continuity conditions for spline curves. *The Computer Journal*, 17(2):181–186.

Marčuk, G. (1987). *Metody numerické matematiky*. Academia.

Marx, B. (2010). P-spline varying coefficient models for complex data. In Kneib, T. and Tutz, G., editors, *Statistical Modelling and Regression Structures*, pages 19–43. Physica-Verlag HD.

Matte, L. (1949). The taper of coniferous species with special reference to loblolly pine. *The Forestry Chronicle*, 25(1):21–31.

Max, T. A. and Burkhart, H. E. (1 September 1976). Segmented polynomial regression applied to taper equations. *Forest Science*, 22:283–289(7).

McClure, J. P. and Czaplewski, R. L. (1986). Compatible taper equation for loblolly pine. *Canadian Journal of Forest Research*, 16(6):1272–1277.

Metzger, K. (1894). Die absoluten Schaftformzahlen der Fichte. *Mündener forstliche Hefte*, 6:87–93.

Milne, R. and Blackburn, P. (1989). The elasticity and vertical distribution of stress within stems of Picea sitchensis. *Tree Physiology*, 5(2):195–205.

Morgan, J. and Cannell, M. G. R. (1994). Shape of tree stemsa re-examination of the uniform stress hypothesis. *Tree Physiology*, 14(1):49–62.

Möttönen, J. and Nummi, T. (2002). Scots pine stem-curve predictions. In Kellogg, L., Spong, B. and Licht, P., editors, *Proceedings of Wood for Africa 2002, Forest Engineering Conference: Forest Engineering Solutions for Achieving Sustainable Forest Resource Management: An International Perspective*. Oregon State University, Department of Forest Engineering, Corvallis.

Muhairwe, C. K. (1999). Taper equations for Eucalyptus pilularis and Eucalyptus grandis for the north coast in New South Wales, Australia. *Forest Ecology and Management*, 113(2-3):251 – 269.

Muhairwe, C. K., LeMay, V. M. and Kozak, A. (1994). Effects of adding tree, stand, and site variables to Kozak's variable-exponent taper equation. *Canadian Journal of Forest Research*, 24(2):252–259.

Munro, D. D. (1968). *Methods for describing distribution of soundwood in mature western hemlock trees*. PhD thesis, University of British Columbia, Dep. of Forestry.

Najzar, K. (2006). *Základy teorie splinů*. Nakladadatelství Karolinum, Praha. 204 pg.

Newberry, J. D. and Burkhart, H. E. (1986). Variable-form stem profile models for loblolly pine. *Canadian Journal of Forest Research*, 16(1):109–114.

Newnham, R. (1988). *A variable-form taper function*. Information report. Petawawa National Forestry Institute, Forestry Canada.

Newnham, R. (1992). Variable-form taper functions for four Alberta tree species. *Canadian Journal of Forest Research*, 22(2):210–223.

Nummi, T. and Möttönen, J. (2004). Prediction of stem measurements of scots pine. *Journal of Applied Statistics*, 31(1):105–114.

Ormerod, D. W. (1973). A simple bole model. *The Forestry Chronicle*, (49):136–138.

Osawa, A., Ishizuka, M. and Kanazawa, Y. (1991). A profile theory of tree growth. *Forest Ecology and Management*, 41(1-2):33 – 63.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Sciences*, 1:505–527.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, 9(2):363–379.

Özçelik, R., Wiant, H. V. and Brooks, J. R. (2008). Accuracy using xylometry of log volume estimates for two tree species in Turkey. *Scandinavian Journal of Forest Research*, 23(3):272 – 277.

Parresol, B. R., Hotvedt, J. E. and Cao, Q. V. (1987). A volume and taper prediction system for bald cypress. *Canadian Journal of Forest Research*, 17(3):250–259.

Perez, D. N., Burthart, H. E. and Stiff, C. T. (1990). Notes: A variable-form taper function for Pinus oocarpa Schiede in Central Honduras. *Forest Science*, 36(1):186–191.

Piegl, L. and Tiller, W. (1996). *The NURBS Book*. Springer-Verlag, Berlin, second edition. 646 pg.

Prautzsch, H., Boehm, W. and Paluszny, M. (2002). *Bézier and B-Spline Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Pretzsch, H., Klemmt, H.-J., Tauber, R., Stary, M., Machacek, P. and Mansfeld, V. (2009). 3d a 2d technologie pro snímání inventarizačních parametrů. *Lesnická práce*, 88(3):22–24.

Procházková, J. and Procházka, D. (2007). An implementation of NURBS curve derivatives in engineering practice. In *WSCG 2007 Poster Papers Proceedings*, pages 5–8, Plzeň. UNION Agency - Science Press.

Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische Mathematik*, (10):177–183.

Riemer, T., von Gadow, K. and Sloboda, B. (1995). Ein Modell zur Beschreibung von Baumschäften. *Allgemeine Forst und Jagdzeitung*, 166(7):144–147.

Rios, M. (1997). *A eficiência das funções polinomiais, da função spine cúbica e razões de volume para representar o perfil da árvore e estimar os sortimentos de Pinus elliottii*. PhD thesis, Universidade Federal de Lavras.

Rodríguez, G. (2001). Smoothing and non-parametric regression. Technical report, Princeton University.

Rojo, A., Perales, X., Sánchez-Rodríguez, F., Álvarez González, J. and von Gadow, K. (2005). Stem taper functions for maritime pine (Pinus pinaster Ait.) in Galicia (Northwestern Spain). *European Journal of Forest Research*, 124(3):177–186.

Rustagi, K. P. and Loveless Jr., R. S. (1991). Compatible variable-form volume and stem-profile equations for douglas-fir. *Canadian Journal of Forest Research*, 21(2):143–151.

Schreuder, H., Gregoire, T. and Wood, G. (1993). *Sampling Methods for Multiresource Forest Inventory*. Wiley.

Schumaker, L. L. (1983). On shape preserving quadratic spline interpolation. *SIAM J. Numer. Anal.*, 20:854–864.

Sharma, M. and Oderwald, R. G. (2001). Dimensionally compatible volume and taper equations. *Canadian Journal of Forest Research*, 31(5):797–803.

Sharma, M., Oderwald, R. G. and Amateis, R. L. (2002). A consistent system of equations for tree and stand volume. *Forest Ecology and Management*, 165(1-3):183 – 191.

Sharma, M. and Parton, J. (2009). Modeling stand density effects on taper for jack pine and black spruce plantations using dimensional analysis. *Forest Science*, 55(3):268–282.

Sharma, M. and Zhang, S. (2004). Variable-exponent taper equations for jack pine, black spruce, and balsam fir in eastern Canada. *Forest Ecology and Management*, 198(1-3):39 – 53.

Shinozaki, K., Yoda, K., Hozumi, K. and Kira, T. (1964a). A Quantitative Analysis of Plant Form - The Pipe Model Theory I. Basic Analyses. *Japanese Journal of Ecology*, 14(3):97–105.

Shinozaki, K., Yoda, K., Hozumi, K. and Kira, T. (1964b). A Quantitative Analysis of Plant Form - The Pipe Model Theory II. Further Evidence of the Theory and its Application in Forest Ecology. *Japanese Journal of Ecology*, 14(4):133–139.

Sloboda, B., Gaffrey, D. and Matsumura, N. (1998). Erfassung individueller Baumschaftformen und ihrer Dynamik durch Spline-Funktionen und Verallgemeinerung durch lineare Schaftformmodelle. *Allgemeine Forst und Jagdzeitung*, 169(2):29–39.

Sloboda, B. and Pfreundt, J. (1989). Tree and stand growth: A system analytical spatial model with consequences for test planing for thining and single tree development. In Burkhart, H. E., Rauscher, H. M. and Johann, K., editors, *Artificial intelligence and growth models for forest management decision. Proceedings of a meeting. Vienna, Austria. September 18-22, 1989.*, pages 119–154. Virginina Polytechnic Institute and State University.

Smaltschinski, T. (1983). Individuelle Baumschaftform und cubische Spline Interpolation. *Allgemeine Forst und Jagdzeitung*, 155(7/8):193–197.

Späth, H. (1973). *Spline-Algorithmen zur Konstruktion gratter Kurven und Flächen*. R. Oldenbourg, München. 134 pg.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147.

Tasissa, G. and Burkhart, H. E. (1998). An application of mixed effects analysis to modeling thinning effects on stem profile of loblolly pine. *Forest Ecology and Management*, 103(1):87 – 101.

Tasissa, G., Burkhart, H. E. and Amateis, R. L. (1997). Volume and taper equations for thinned and unthinned loblolly pine trees in cutover, site-prepared plantations. *Southern Journal of Applied Forestry*, 21(3):146–152.

The MathWorks, Inc. (2002). *Curve Fitting Toolbox User's Guide*. The MathWorks, Inc., Natick, Massachusetts.

The MathWorks, Inc. (2003). *Statistics Toolbox User's Guide*. The MathWorks, Inc., Natick, Massachusetts.

The MathWorks, Inc. (2010). *MATLAB ® 7: Getting Started Guide*. The MathWorks, Inc., Natick, Massachusetts.

Thomas, C. E. and Parresol, B. R. (1991). Simple, flexible, trigonometric taper equations. *Canadian Journal of Forest Research*, 21:1132–1137.

Thomas, C. E., Parresol, B. R., Le, K. H. N. and Lohrey, R. E. (1995). Biomass and taper for trees in thinned and unthinned longleaf pine plantations. *Southern Journal of Applied Forestry*, 19(1):29–35.

Trincado, G. and Burkhart, H. E. (2006). A generalized approach for modeling and localizing stem profile curves. *Forest Science*, 52(6):670–682.

Trincado, G. V. and Sandoval, V. (2002). Algoritmos para la estimación de volúmenes comerciales. *Quebrancho. Revista de Ciencias Forestales*, (9):106–114.

Trincado, G. V. and Vidal, J. B. (1999). Aplicación de interpolación "spline" cúbica en la estimación de volumen. *Bosque (Valdivia)*, 20(2):3–8.

Utreras, F. D. (1981). Optimal smoothing of noisy data using spline functions. *SIAM Journal on Scientific and Statistical Computing*, 2(3):349–362.

Väisänen, H., Kellomäki, S., Oker-Blom, P. and Valtonen, E. (1989). Structural development of pinus sylvestrs stands with varying initial density: A preliminary model for quality of sawn timber as affected by silvicultural measures. *Scandinavian Journal of Forest Research*, 4(1-4):223–238.

Valenti, M. A. and Cao, Q. V. (1986). A comparison of the effects of one-step and two-step pruning on loblolly pine stem form. *Southern Journal of Applied Forestry*, 10(4):251–253.

Valentine, H. T. and Gregoire, T. G. (2001). A switching model of bole taper. *Canadian Journal of Forest Research*, 31(8):1400–1409.

Wenger, K. F., editor (1984). *Forestry handbook*. John Wiley & Sons, Inc., 2 edition.

Wensel, L. C. (1977). A generalized prismoidal log volume equation. *Biometric Note*, (5). Department of Forestry and Resource Management, University of California, Berkeley.

Wiklund, K., Konôpka, B. and Nilsson, L.-O. (1995). Stem form and growth in Picea abies (L.) Karst, in response to water and mineral nutrient availability. *Scandinavian Journal of Forest Research*, 10(1-4):326–332.

Wood, G. B., Wiant, H. V., Loy, R. J. and Miles, J. A. (1990). Centroid sampling: A variant of importance sampling for estimating the volume of sample trees of radiata pine. *Forest Ecology and Management*, 36(2-4):233 – 243.

Yang, Y., Huang, S. and Meng, S. X. (2009a). Development of a tree-specific stem profile model for white spruce: a nonlinear mixed model approach with a generalized covariance structure. *Forestry*, 82(5):541–555.

Yang, Y., Huang, S., Trincado, G. and Meng, S. (2009b). Nonlinear mixed-effects modeling of variable-exponent taper equations for lodgepole pine in Alberta, Canada. *European Journal of Forest Research*, 128:415–429.

Zakrzewski, W. T. (1999). A mathematically tractable stem profile model for jack pine in Ontario. *Northern Journal of Applied Forestry*, 16(3):138–143.

Publications of the Author

[A.1] Kuželka, K. and Marušák, M. (2012). Spline representation of irregular and malformed stem profiles of broadleaved tree species in White Carpathian Mountains. *Beskydy*, 5(2): 111-120.

[A.2] Kuželka, K. (2011). Stem curve representation using spline functions. In: *SIGA 2011. Splines and IsoGeometric Analysis.* February 9, 2011, Prague.

[A.3] Kuželka, K. (2011). Stem curve modeling using spline functions. In: *Bulletin of Szent István University, Gödöllő 2011, Hungary, (special issue).* April 28$^{\text{th}}$ - 29$^{\text{th}}$, 2011, ISSN 1586-4502, p. 211.

[A.4] Kuželka, K. (2011). Stem curve modeling using spline functions. In: *ICFFI News. International Centre of Forestry and Forest Industries.* 1(13). November 9$^{\text{th}}$-11$^{\text{th}}$, 2011, p. 70.

Appendix A

Primary comparison of spline types

Table A.1: Comparison of splines based on 6 input points (combination no. 1). For each statistic mean value and standard deviation is shown. Values in a column are followed by the same letter sighnify no significant difference between spline types. Stars in columns DB and TVD indicate mean values significantly different from zero.

Spline	**DB (10^{-2} m)**			**MAR (10^{-2} m)**			**SDR (10^{-2} m)**			**MSR (10^{-3} m^2)**			**TVD (%)**		
CRS	-1.03	0.52	a*	1.30	0.54	a,b	1.64	0.69	a,b	0.32	0.26	a	-8.22	3.69	a*
IBS2u	-0.92	0.49	a*	1.14	0.47	a	1.46	0.61	a	0.25	0.21	a	-8.14	3.81	a*
NCS	0.45	1.03	c*	1.76	1.33	c	2.11	1.49	c	0.67	0.99	b	6.27	11.01	b*
IterBS	0.49	1.11	c*	2.26	1.74	d	2.73	1.97	d	1.13	1.61	c	7.12	12.89	b*
BS2	-1.16	0.61	a,b*	1.57	0.61	b,c	1.93	0.75	b,c	0.43	0.33	a,b	-9.24	4.01	a*
BS3	-1.15	0.61	a,b*	1.67	0.64	b,c	2.02	0.77	b,c	0.47	0.35	a,b	-8.85	4.17	a*
NUBS2	-1.24	0.63	a,b*	1.61	0.63	b,c	1.95	0.76	b,c	0.45	0.34	a,b	-10.34	4.09	a,c*
NUBS3	-1.28	0.65	a,b*	1.71	0.66	b,c	2.04	0.79	b,c	0.48	0.37	a,b	-10.83	4.31	a,c*
NURBSdbh2	-1.17	0.64	a,b*	1.68	0.63	b,c	2.08	0.77	b,c	0.50	0.36	a,b	-8.99	4.36	a*
NURBSdbh3	-1.27	0.68	a,b*	1.83	0.68	c	2.20	0.80	c	0.55	0.39	a,b	-10.30	4.94	a,c*
NURBSav2	-1.44	0.75	b*	1.93	0.72	c,d	2.30	0.85	c,d	0.61	0.44	b	-12.33	4.90	c*
NURBSav3	-1.47	0.73	b*	1.90	0.73	c,d	2.22	0.86	c	0.57	0.43	a,b	-12.98	4.76	c*

Table A.2: Comparison of splines based on 7 input points (combination no. 2)

Spline	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
CRS	-0.09	0.23	a,b*	0.46	0.13	a	0.62	0.18	a	0.04	0.03	a	-0.31	3.54	a,b
IBS2u	-0.05	0.22	a	0.46	0.14	a	0.66	0.21	a	0.05	0.04	a	-1.00	3.61	a,b*
NCS	-0.52	0.42	c*	1.47	0.98	c	1.98	1.37	c	0.57	0.81	b	-5.31	4.53	c*
IterBS	-0.85	0.57	d*	2.28	1.55	d	3.10	2.18	d	1.43	1.95	c	-7.6	5.01	d*
BS2	-0.21	0.28	a,b,e*	0.70	0.20	a,b	0.99	0.28	a,b	0.11	0.06	a	-0.23	3.65	a,b
BS3	-0.24	0.30	b,e*	0.82	0.26	b	1.15	0.35	b	0.14	0.09	a	-0.32	3.73	a,b
NUBS2	-0.25	0.29	b,e*	0.71	0.21	b	0.97	0.27	b	0.10	0.06	a	-0.71	3.63	a,b
NUBS3	-0.35	0.33	e,f*	0.89	0.29	b,e	1.20	0.37	b	0.16	0.10	a	-1.55	3.75	a,e*
NURBSdbh2	-0.18	0.30	a,b,f*	0.77	0.22	b	1.16	0.35	b	0.15	0.09	a	0.57	3.95	b
NURBSdbh3	-0.30	0.35	e*	0.95	0.29	b,e	1.38	0.45	b	0.21	0.14	a	-0.48	4.16	a,b
NURBSav2	-0.32	0.36	e*	0.90	0.28	b,e	1.29	0.40	b	0.18	0.11	a	-0.52	4.17	a,b
NURBSav3	-0.50	0.38	f*	1.13	0.38	e	1.43	0.47	e	0.23	0.15	a	-3.18	4.02	e*

Table A.3: Comparison of splines based on 8 input points (combination no. 3)

Spline	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
CRS	-0.23	0.21	a*	0.57	0.16	a	0.78	0.23	a	0.07	0.04	a	-1.00	2.38	a*
IBS2u	-0.20	0.20	a*	0.55	0.14	a	0.80	0.23	a	0.07	0.05	a	-2.34	2.66	a,b*
NCS	-0.30	0.22	a,b*	0.78	0.31	b	1.14	0.5	b	0.16	0.15	b	-2.67	2.66	b*
IterBS	-0.45	0.25	b,c*	1.02	0.45	c	1.58	0.79	c	0.31	0.32	c	-4.12	2.91	c*
BS2	-0.28	0.27	a,b*	0.75	0.22	b	1.08	0.32	b	0.13	0.07	a,b	-0.48	2.72	a,d
BS3	-0.29	0.28	a,b*	0.82	0.25	b,d	1.17	0.36	b	0.15	0.09	b	-0.44	2.83	a,d
NUBS2	-0.31	0.28	a,b*	0.76	0.23	b	1.07	0.32	b	0.13	0.08	a,b	-0.86	2.75	a,d*
NUBS3	-0.36	0.31	b*	0.87	0.28	b,d	1.21	0.38	b,d	0.16	0.1	b	-1.24	2.91	a,b*
NURBSdbh2	-0.24	0.29	a,b*	0.83	0.24	b,d	1.27	0.39	b,d	0.18	0.11	b	0.55	3.17	d
NURBSdbh3	-0.29	0.33	a,b*	0.94	0.29	c,d	1.41	0.46	c,d	0.22	0.14	b	0.09	3.41	a,d
NURBSav2	-0.38	0.37	b,c*	0.96	0.31	c,d	1.40	0.45	c,d	0.21	0.14	b	-0.75	3.50	a,d
NURBSav3	-0.52	0.38	c*	1.08	0.37	c	1.42	0.47	c,d	0.22	0.15	b	-2.75	3.36	b,c*

Table A.4: Comparison of splines based on 9 input points (combination no. 4)

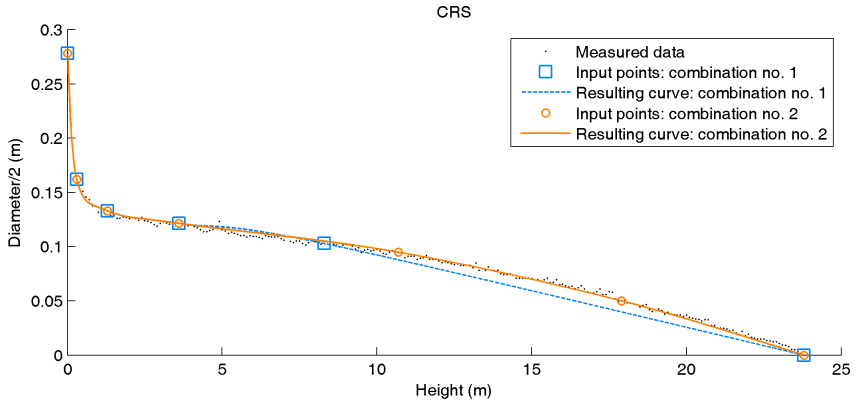
Spline	**DB (10^{-2} m)**			**MAR (10^{-2} m)**			**SDR (10^{-2} m)**			**MSR (10^{-3} m^2)**			**TVD (%)**		
CRS	0.00	0.15	a,b	0.39	0.11	a	0.54	0.17	a	0.03	0.02	a	0.25	2.17	a
IBS2u	0.02	0.15	a,b	0.43	0.12	a,b	0.66	0.23	a,b	0.05	0.04	a,b	-1.27	2.52	b*
NCS	-0.16	0.19	c*	0.70	0.32	c	1.07	0.53	c	0.14	0.15	c	-1.68	2.51	b,c*
IterBS	-0.26	0.23	d*	0.95	0.48	d	1.52	0.82	d	0.30	0.33	d	-2.89	2.77	c*
BS2	-0.01	0.17	a,b	0.52	0.13	b,e	0.83	0.24	b	0.07	0.04	a,b	1.43	2.34	a,d,e*
BS3	-0.02	0.19	a,b	0.58	0.15	e	0.93	0.29	b,c	0.09	0.06	b,c	1.63	2.44	d,e*
NUBS2	-0.02	0.17	a,b	0.51	0.12	b,e	0.80	0.23	b	0.07	0.04	a,b	1.22	2.31	a,d*
NUBS3	-0.04	0.20	a*	0.59	0.16	c,e	0.94	0.28	b,c	0.10	0.06	b,c	1.39	2.47	a,d*
NURBSdbh2	0.05	0.19	b	0.58	0.14	e	1.03	0.35	c	0.12	0.08	b,c	2.64	2.77	d,e*
NURBSdbh3	0.02	0.22	a,b	0.66	0.17	c,e	1.17	0.42	c	0.15	0.11	c	2.72	3.05	d,e*
NURBSav2	0.00	0.20	a,b	0.62	0.15	c,e	1.07	0.35	c	0.12	0.08	b,c	2.24	2.79	d,e*
NURBSav3	-0.08	0.24	a,c*	0.69	0.2	c	1.06	0.34	c	0.12	0.08	b,c	1.45	2.71	a,d,e*

Table A.5: Comparison of splines based on 10 input points (combination no. 5)

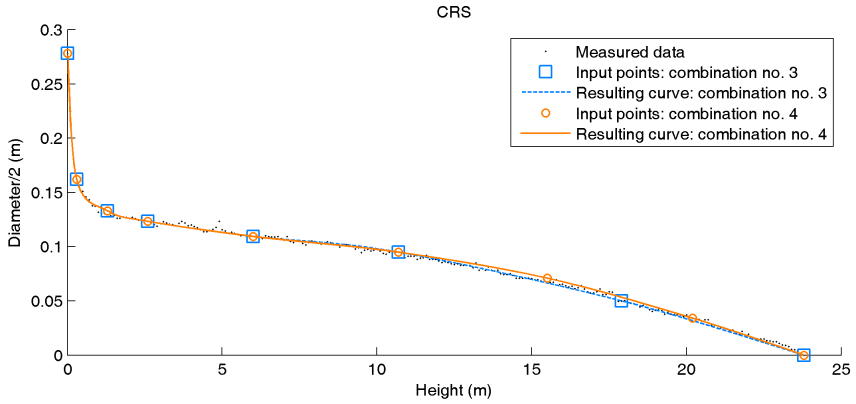
Spline	**DB (10^{-2} m)**			**MAR (10^{-2} m)**			**SDR (10^{-2} m)**			**MSR (10^{-3} m^2)**			**TVD (%)**		
CRS	-0.01	0.13	a,b	0.38	0.11	a	0.54	0.17	a	0.03	0.02	a	0.14	1.95	a
IBS2u	0.02	0.14	a,b	0.43	0.12	a,b	0.68	0.23	a,b	0.05	0.05	a,b	-1.69	2.57	b*
NCS	-0.19	0.20	c*	0.64	0.26	c	1.03	0.50	c	0.13	0.13	c	-2.31	2.68	b*
IterBS	-0.32	0.26	d*	0.84	0.38	d	1.45	0.77	d	0.27	0.29	d	-3.86	3.13	c*
BS2	0.01	0.15	a,b	0.51	0.13	b,e	0.82	0.24	b,e	0.07	0.04	a,b	1.58	2.02	d*
BS3	0.00	0.17	a,b	0.57	0.15	c,e	0.92	0.29	c,e	0.09	0.06	b,c	1.86	2.10	d,e*
NUBS2	-0.01	0.15	a,b	0.50	0.12	b,e	0.80	0.23	b,e	0.07	0.04	a,b	1.39	2.01	d*
NUBS3	-0.02	0.18	a,b	0.60	0.16	c,e	0.96	0.3	c,e	0.10	0.06	b,c	1.76	2.18	d,e*
NURBSdbh2	0.06	0.17	a*	0.58	0.14	c,e	1.03	0.35	c	0.12	0.08	b,c	2.81	2.46	e*
NURBSdbh3	0.04	0.20	a*	0.66	0.18	c	1.19	0.44	c	0.16	0.12	c	3.05	2.79	e*
NURBSav2	0.01	0.19	a,b	0.62	0.15	c	1.06	0.35	c	0.12	0.08	b,c	2.40	2.48	d,e*
NURBSav3	-0.07	0.22	b*	0.70	0.21	c	1.08	0.35	c	0.13	0.09	b,c	1.63	2.40	d*

Table A.6: Comparison of splines based on 22 input points (0 m, 0.3 m, 1.3 m, 10 %, 15 %, ..., 95 %, 100 %)

Spline	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
CRS	0.01	0.09	a	0.31	0.08	a	0.48	0.15	a	0.03	0.02	a	0.23	1.32	a
IBS2u	0.01	0.09	a	0.36	0.10	a,b	0.63	0.22	a,b	0.04	0.04	a	-1.39	2.14	b*
NCS	-0.08	0.12	b*	0.46	0.16	c	0.86	0.41	c,e	0.09	0.10	b	-0.76	1.54	b*
IterBS	-0.16	0.16	c*	0.57	0.22	d	1.21	0.62	d	0.19	0.20	c	-1.60	1.92	b*
BS2	0.13	0.10	d*	0.40	0.09	b,c	0.75	0.23	b,c	0.06	0.04	a,b	2.55	1.48	c*
BS3	0.17	0.11	d,e*	0.44	0.11	c	0.85	0.29	c,e	0.08	0.06	a,b	3.13	1.60	c,d*
NUBS2	0.14	0.10	d*	0.41	0.09	b,c	0.76	0.24	b,c	0.06	0.04	a,b	2.63	1.52	c*
NUBS3	0.19	0.11	d,e*	0.46	0.11	c	0.89	0.31	c,e	0.09	0.06	a,b	3.47	1.72	c,d,e*
NURBSdbh2	0.22	0.13	e*	0.48	0.12	c	1.03	0.39	e	0.12	0.10	b	4.12	2.14	d,e*
NURBSdbh3	0.25	0.15	e*	0.52	0.14	c,d	1.13	0.46	d,e	0.15	0.13	b,c	4.70	2.39	e*
NURBSav2	0.19	0.12	e*	0.47	0.11	c	0.97	0.35	e	0.11	0.08	b	3.78	2.02	d,e*
NURBSav3	0.22	0.13	e*	0.49	0.13	c	0.95	0.35	e	0.10	0.08	b	3.94	1.81	d,e*

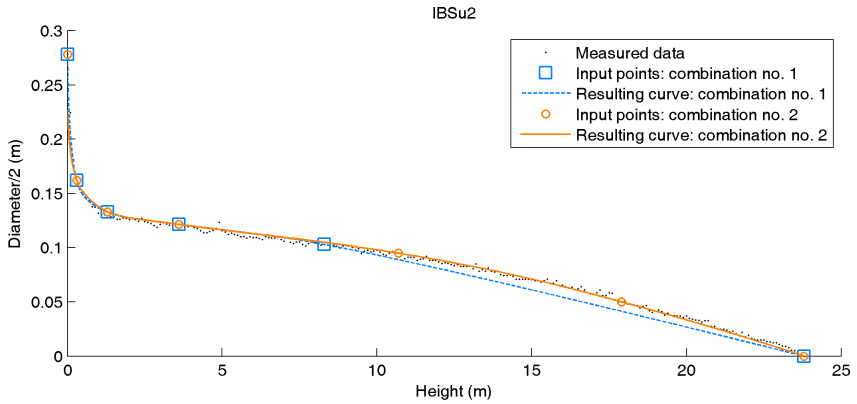


(a) Combinations no. 1 and 2

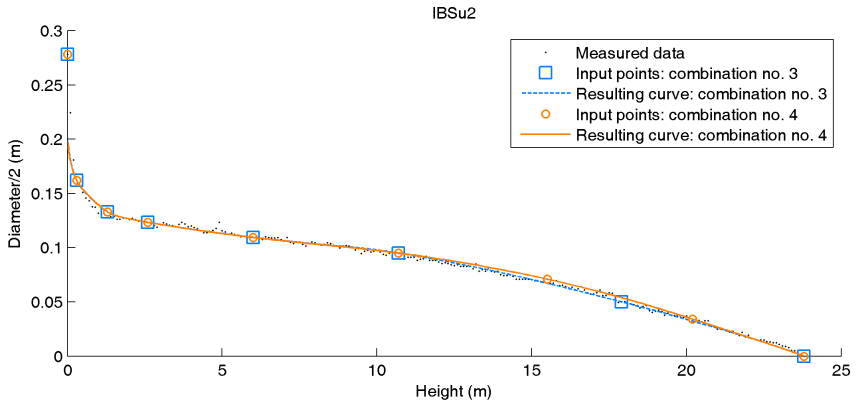


(b) Combinations no. 3 and 4

Figure A.1: Stem curves represented by Catmull-Rom spline

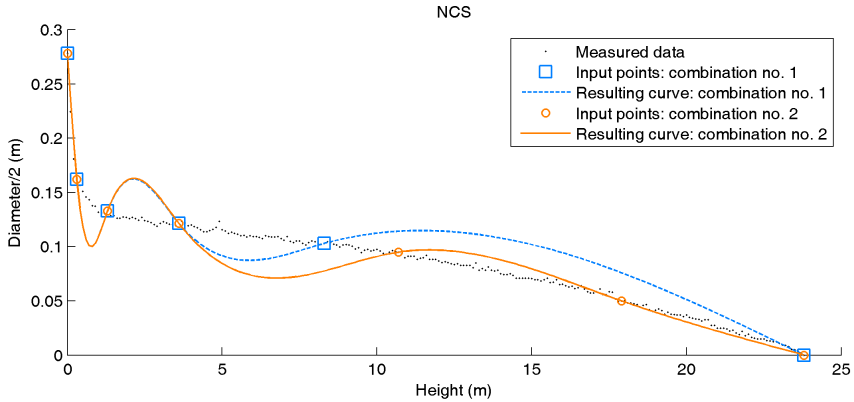


(a) Combinations no. 1 and 2

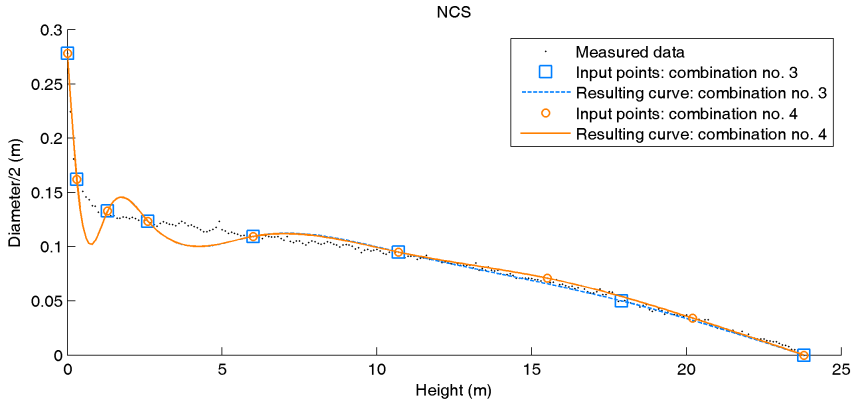


(b) Combinations no. 3 and 4

Figure A.2: Stem curves represented by intepolation B-spline

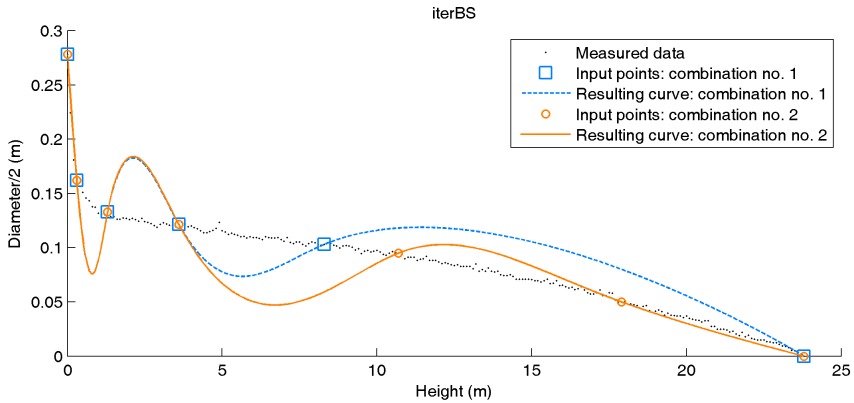


(a) Combinations no. 1 and 2

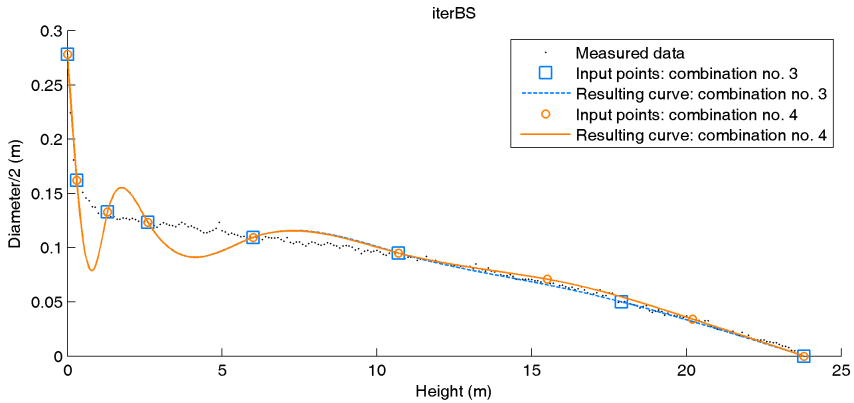


(b) Combinations no. 3 and 4

Figure A.3: Stem curves represented by natural cubic spline

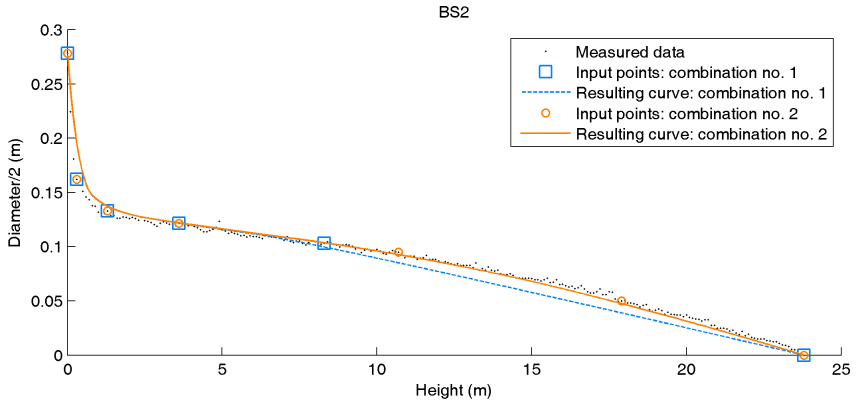


(a) Combinations no. 1 and 2

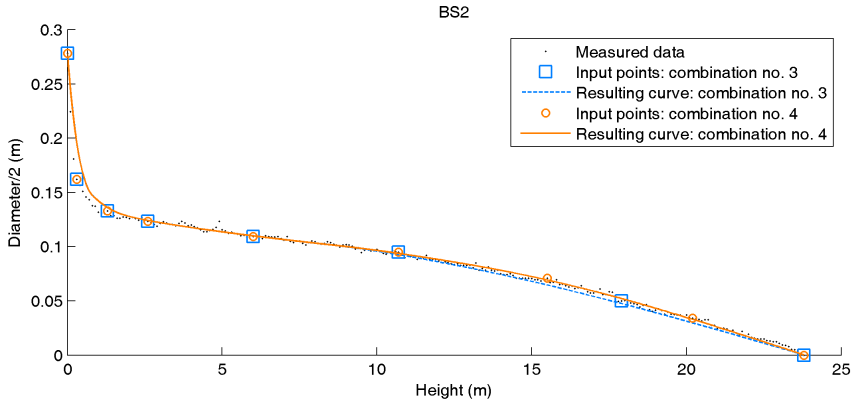


(b) Combinations no. 3 and 4

Figure A.4: Stem curves represented by iterative B-spline

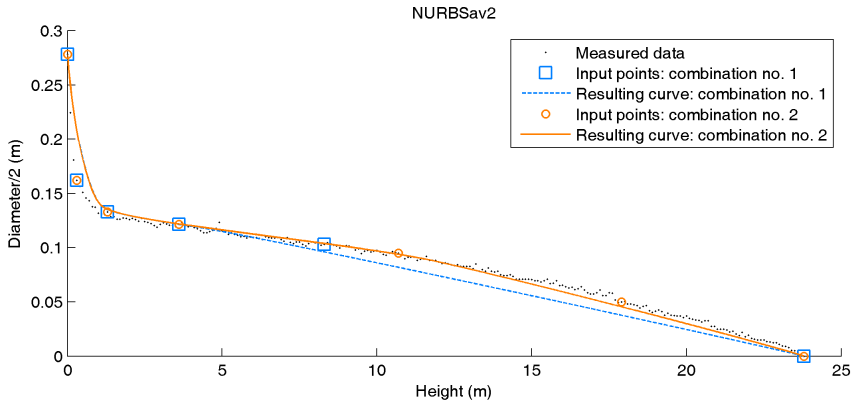


(a) Combinations no. 1 and 2

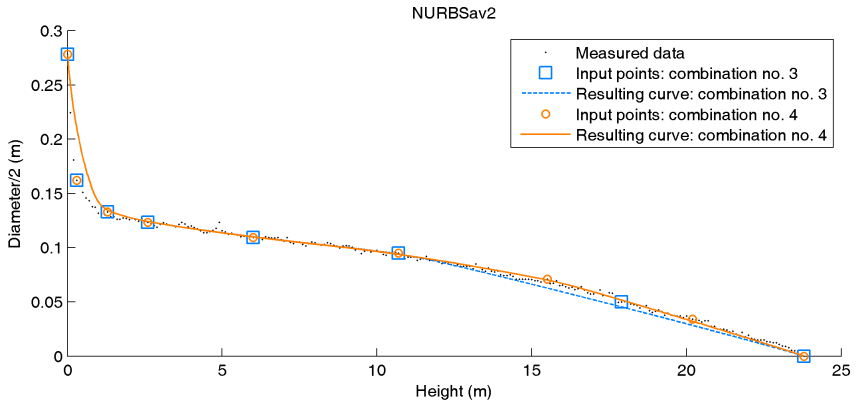


(b) Combinations no. 3 and 4

Figure A.5: Stem curves represented by B-spline



(a) Combinations no. 1 and 2



(b) Combinations no. 3 and 4

Figure A.6: Stem curves represented by NURBS (NURBSav2)

Appendix B

Comparison of performance of splines

Table B.1: Sectional diameter bias (DB) and relative volume difference (VD) for 5-point splines. Stars indicate values significantly different from zero.

	Spline	Section									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB (10^{-2} m)	**CRS**	-0.01	0.19*	-0.05	-0.12	-0.10	-0.03	-0.02	-0.33*	-0.73*	-0.66*
	NCS	0.46	8.22*	2.75*	-5.21*	-10.69*	-13.39*	-13.61*	-11.83*	-8.35*	-3.35*
	IBS	1.18*	0.23*	0.02	0.04	0.05	-0.04	-0.17*	-0.36*	-0.53*	-0.42*
	BS	1.52*	0.57*	-0.11	-0.48*	-0.75*	-1.08*	-1.35*	-1.54*	-1.49*	-0.84*
VD (%)	**CRS**	-0.20	2.33*	-0.09	-0.95	-1.08	0.05	0.08	-6.13*	-19.73*	-36.37*
	NCS	7.45*	105.40*	46.41*	-31.80*	-61.38*	-60.63*	-50.60*	-44.17*	-43.29*	-48.48*
	IBS	7.81*	2.58*	0.23	0.22	0.11	-0.70	-2.14	-5.51*	-12.36*	-20.75*
	BS	13.92*	6.07*	-1.00	-5.46*	-9.58*	-14.70*	-20.85*	-28.76*	-38.91*	-48.52*

Table B.2: Sectional diameter bias (DB) and relative volume difference (VD) for 6-point splines. Stars indicate values significantly different from zero.

	Spline	Section									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB (10^{-2} m)	**CRS**	-0.01	0.22*	0.03	0.01	0.03	-0.01	-0.01	0.03	0.03	-0.18*
	NCS	-0.04	0.17*	-3.86*	-2.82*	-0.07	1.68*	2.08*	1.53*	0.55*	-0.14
	IBS	0.08	0.19*	0.01	-0.15*	-0.19*	-0.09	0.06	0.09	-0.02	-0.13*
	BS	1.51*	0.57*	-0.07	-0.39*	-0.55*	-0.59*	-0.47*	-0.37*	-0.51*	-0.44*
VD (%)	**CRS**	-0.21	2.49*	0.40	-0.09	-0.19	-0.55	-0.28	0.40	0.27	-5.31
	NCS	1.98	3.06*	-33.92*	-27.12*	0.14	26.56*	40.96*	39.64*	25.89*	6.52
	IBS	0.14	2.17*	0.24	-1.44	-2.30*	-0.77	1.67	2.82	0.58	-1.68
	BS	13.86*	6.06*	-0.56	-4.44*	-7.17*	-8.26*	-7.37*	-6.74*	-14.17*	-25.22*

Table B.3: Sectional diameter bias (DB) and relative volume difference (VD) for 7-point splines. Stars indicate values significantly different from zero.

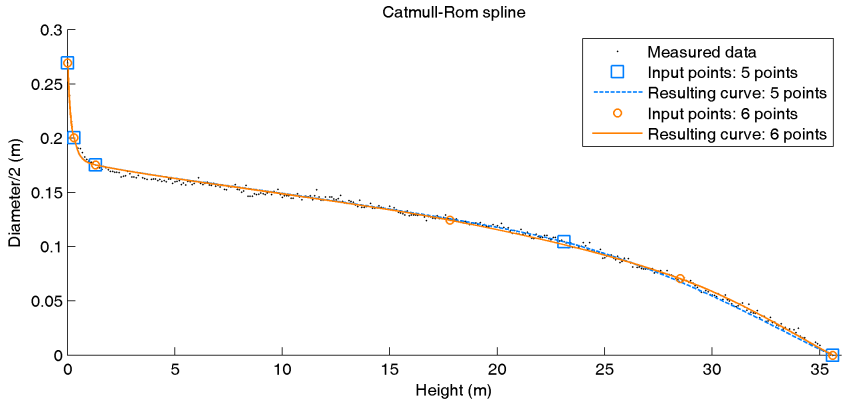
	Spline	Section									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB (10^{-2} m)	**CRS**	0.03	0.05	-0.06	-0.02	0.03	-0.01	-0.01	0.02	0.03	-0.18*
	NCS	-0.51*	-1.48*	-2.89*	-0.78*	1.24*	1.93*	1.58*	0.66*	-0.23*	-0.45*
	IBS	-0.01	0.06	0.01	0.00	-0.11*	-0.14*	-0.02	0.06	0.06	-0.01
	BS	1.39*	0.40*	0.01	-0.19*	-0.28*	-0.38*	-0.39*	-0.43*	-0.61*	-0.50*
VD (%)	**CRS**	0.21	0.46	-0.77	-0.54	-0.30	-0.51	-0.36	0.25	0.25	-5.31
	NCS	-2.38*	-10.84*	-27.43*	-9.72*	14.99*	29.17*	29.24*	16.14*	-5.68*	-24.19*
	IBS	-0.64	0.95	0.28	0.16	-1.43	-1.94*	-0.20	1.53	2.35	6.39
	BS	12.88*	4.07*	-0.16	-2.72*	-4.48*	-5.95*	-6.39*	-8.05*	-17.04*	-28.71*

Table B.4: Sectional diameter bias (DB) and relative volume difference (VD) for 8-point splines. Stars indicate values significantly different from zero.

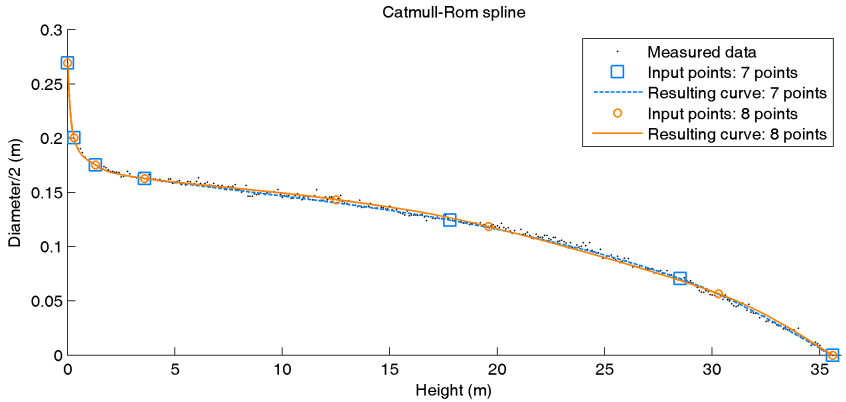
	Spline	Section									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB (10^{-2} m)	**CRS**	0.03	0.06	-0.03	-0.01	0.07	-0.01	-0.17*	-0.16*	0.02	-0.09*
	NCS	-0.55*	0.17*	0.28*	0.01	-0.23*	-0.10	0.19	0.30*	0.07	-0.20*
	IBS	0.08	0.07	0.05	0.03	-0.04	-0.04	0.09*	0.07	-0.13*	-0.25*
	BS	1.13*	0.06	-0.18*	-0.24*	-0.1*9	-0.07*	-0.09*	-0.19*	-0.51*	-0.49*
VD (%)	**CRS**	-0.06	0.04	-0.01	0.01	0.0*9	0.00	-0.10*	-0.10*	-0.0	1 0.00
	NCS	-0.15	0.09*	0.31*	0.02	-0.12	-0.06	0.09	0.11*	0.02	-0.02*
	IBS	0.03	0.02	0.02	0.01	0.00	0.00	0.05*	-0.01	-0.0*6	-0.03*
	BS	1.08*	0.06	-0.11*	-0.14*	-0.08*	-0.04*	-0.05*	-0.08*	-0.14*	-0.05*

Table B.5: Sectional diameter bias (DB) and relative volume difference (VD) for 9-point splines. Stars indicate values significantly different from zero.

	Spline	Section									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DB (10^{-2} m)	**CRS**	0.04	0.04	-0.09	-0.05	0.02	0.00	-0.05	-0.01	-0.03	-0.17*
	NCS	-0.58*	-0.02	-0.01	-0.27	-0.27	-0.12	0.10	0.21*	0.07	-0.17*
	IBS	0.04	0.06	-0.01	0.07	0.07	-0.01	0.00	0.03	-0.02	-0.12*
	BS	1.13*	0.06	-0.13*	-0.17*	-0.10*	-0.05	-0.09*	-0.19*	-0.51*	-0.49*
VD (%)	**CRS**	-0.04	0.01	-0.07	-0.02	0.06	0.01	-0.03	-0.03	-0.02	-0.01
	NCS	-0.19*	-0.03	0.00	-0.20*	-0.15*	-0.07	0.05	0.07*	0.01	-0.01*
	IBS	-0.03	0.01	-0.03	0.05	0.10*	0.02	0.00	0.00	-0.01	-0.01
	BS	1.08*	0.05	-0.09*	-0.11*	-0.02	-0.03	-0.05*	-0.08*	-0.14*	-0.05*

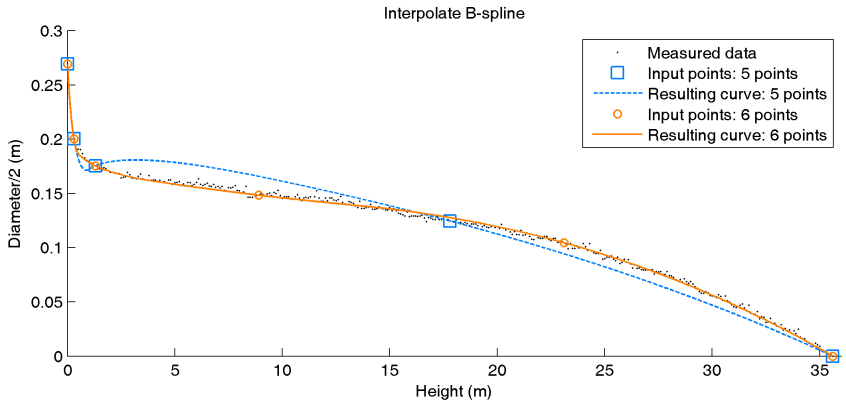


(a) Combinations no. 1 and 2

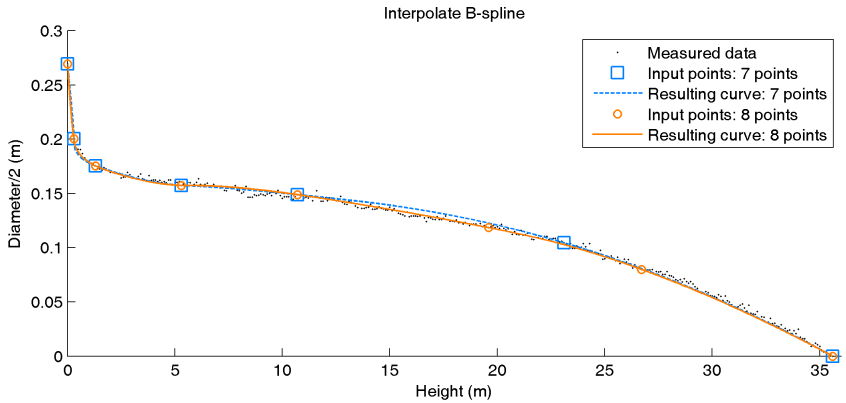


(b) Combinations no. 3 and 4

Figure B.1: Stem curves represented by Catmull-Rom spline

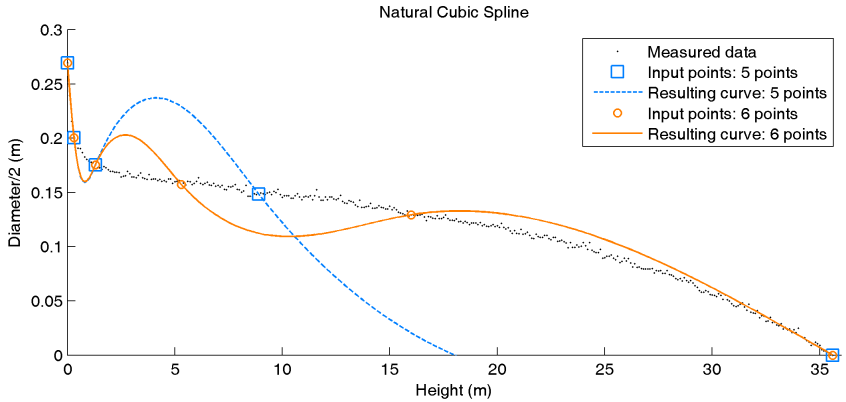


(a) Combinations no. 1 and 2

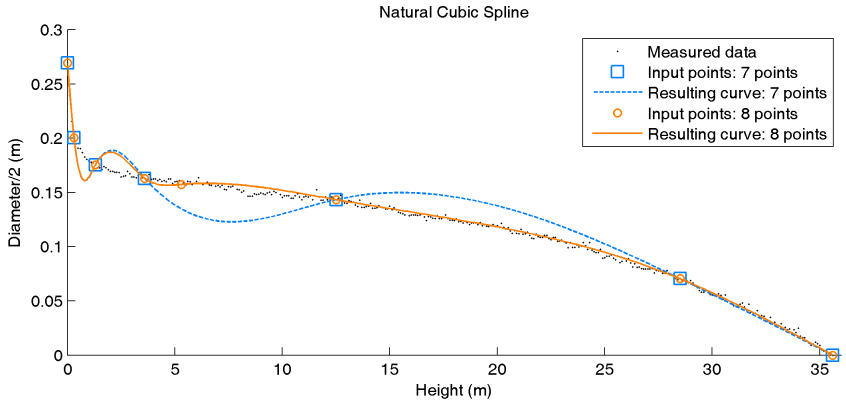


(b) Combinations no. 3 and 4

Figure B.2: Stem curves represented by intepolation B-spline

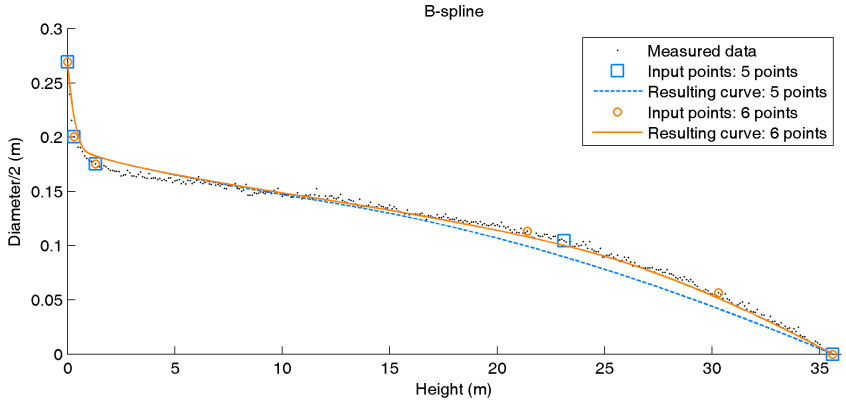


(a) Combinations no. 1 and 2

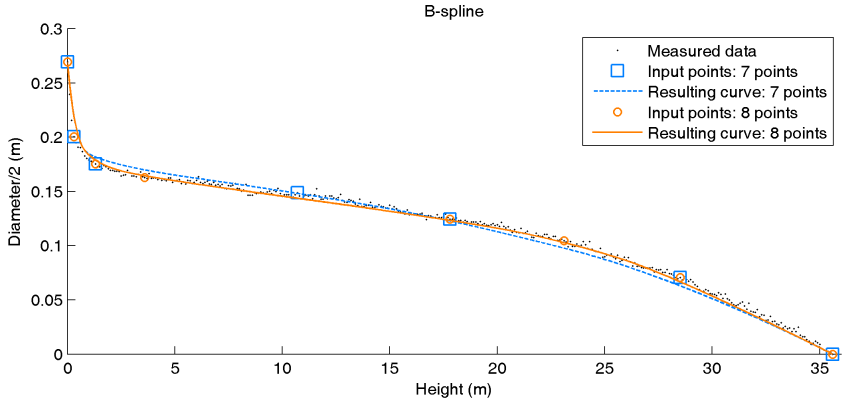


(b) Combinations no. 3 and 4

Figure B.3: Stem curves represented by natural cubic spline



(a) Combinations no. 1 and 2

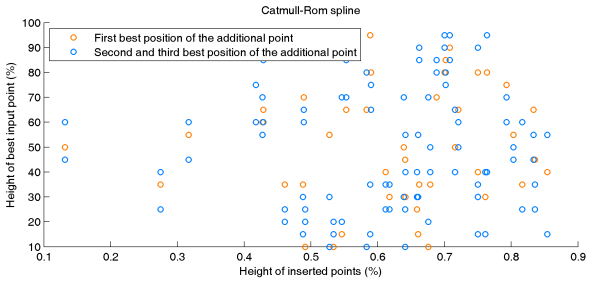


(b) Combinations no. 3 and 4

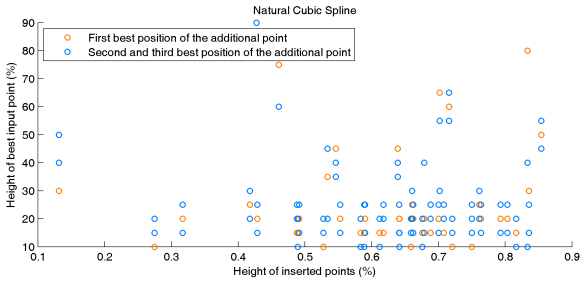
Figure B.4: Stem curves represented by B-spline

Appendix C

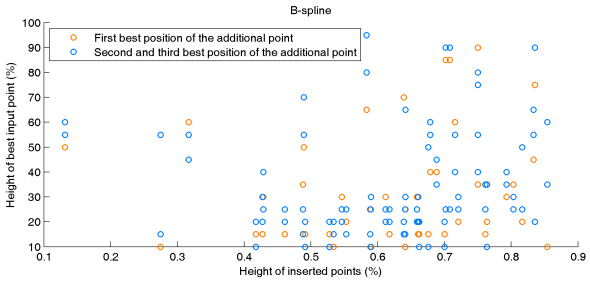
Irregular stem curve spline representation



(a) Catmull-Rom spline

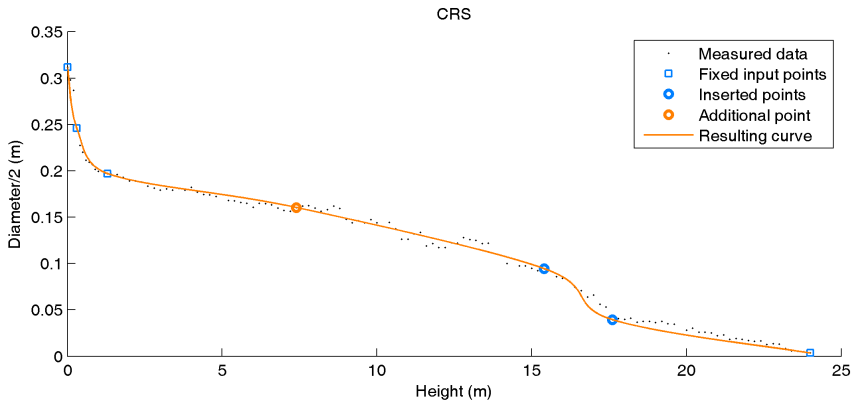


(b) Natural cubic spline

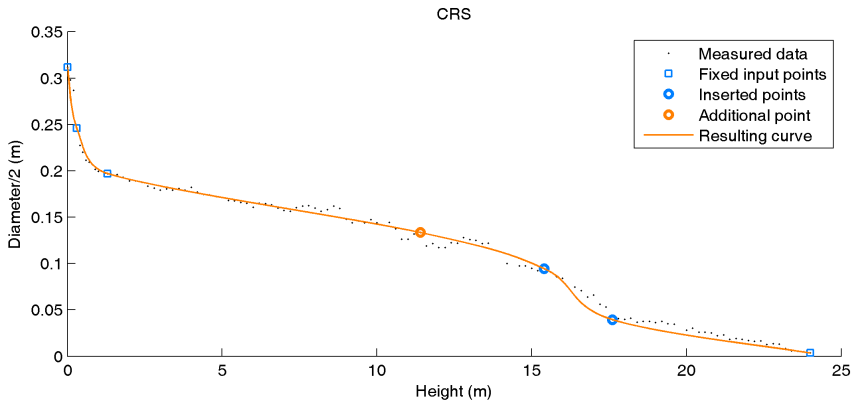


(c) B-spline

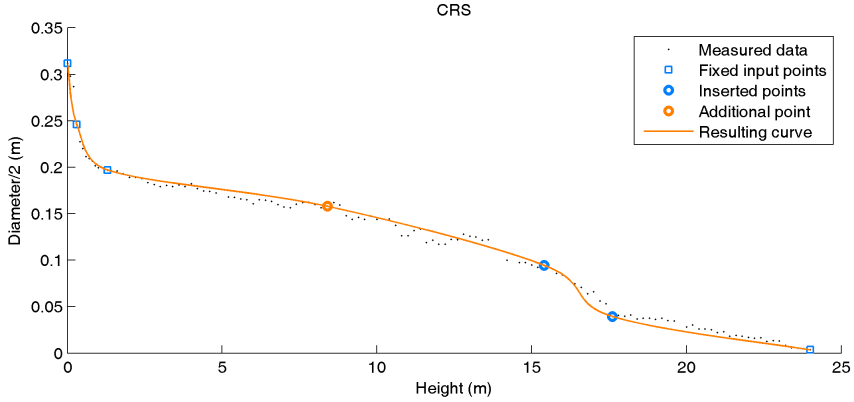
Figure C.1: Dependece of the optimal input point position on the position of the inserted points



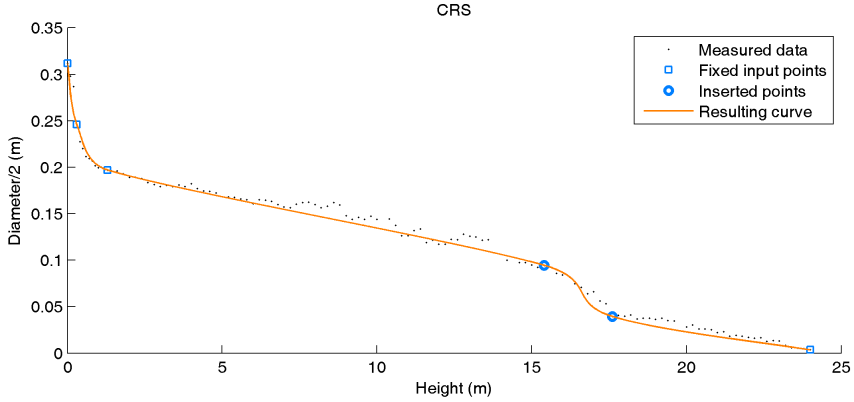
(a) Optimal position



(b) Accroding to regression

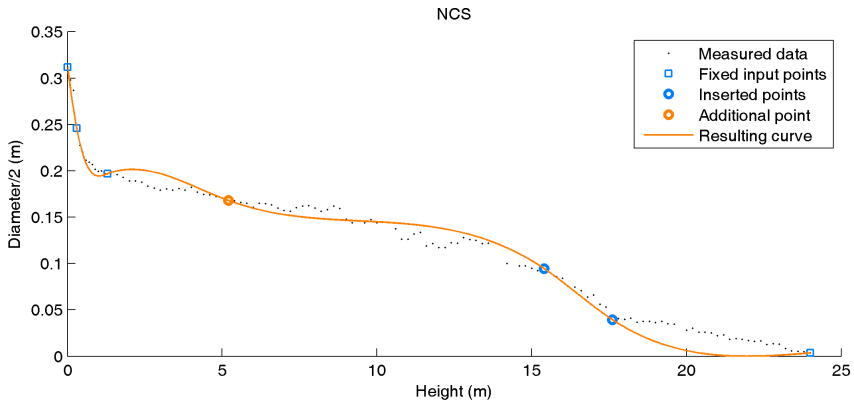


(c) Center of interspace

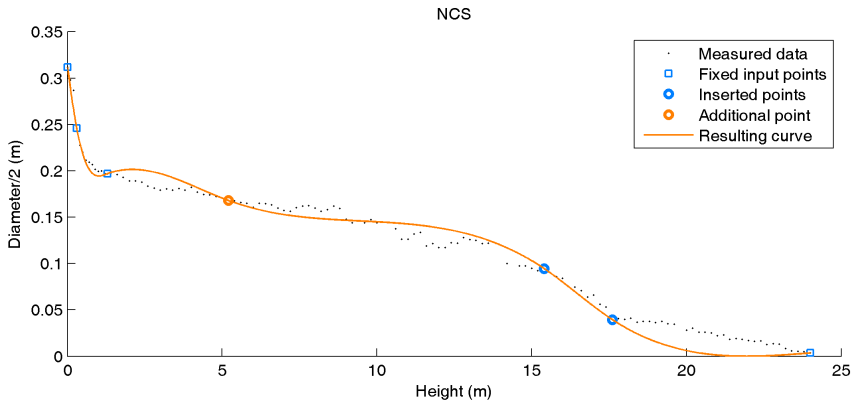


(d) Six points only

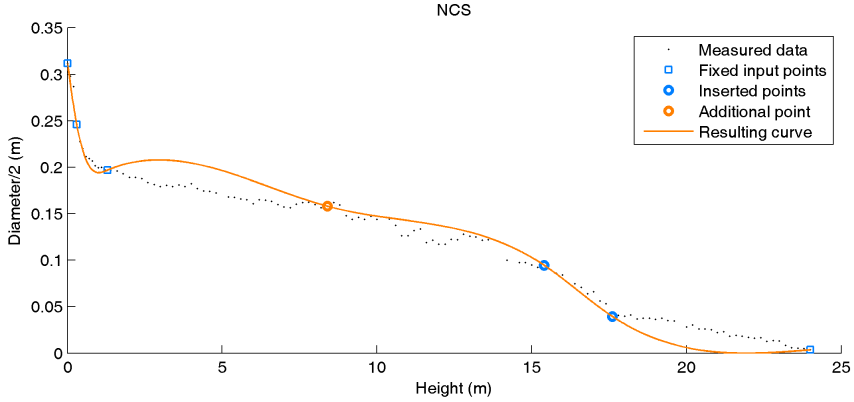
Figure C.2: Irregular stem curves represented by Catmull-Rom spline with different 7^{th} point placement; example no. 1



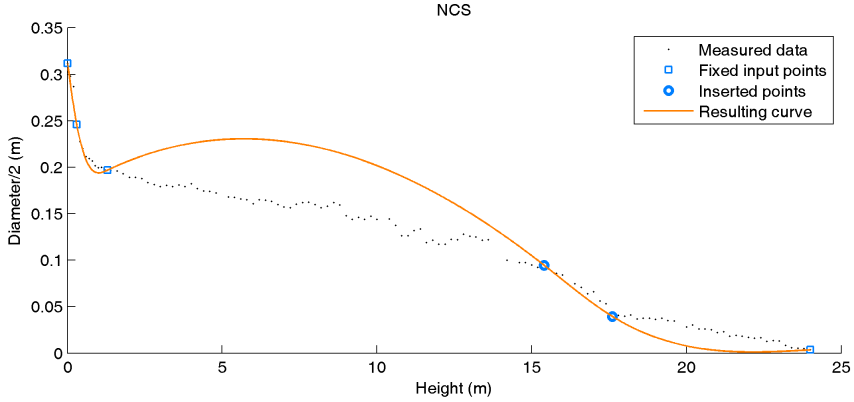
(a) Optimal position



(b) 20 % of the stem height

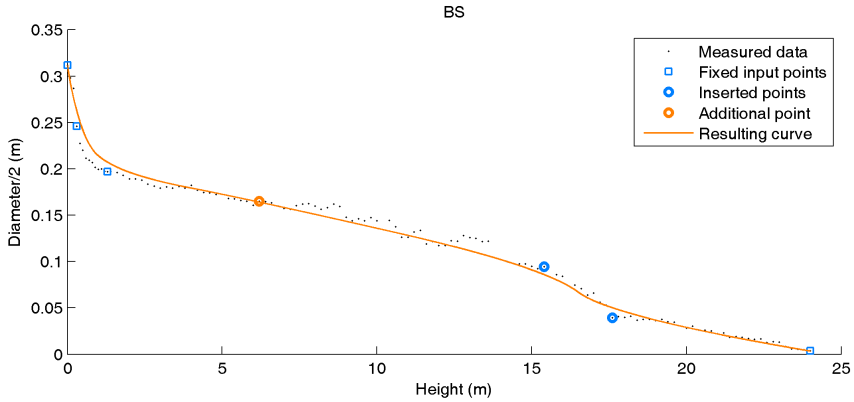


(c) Center of interspace

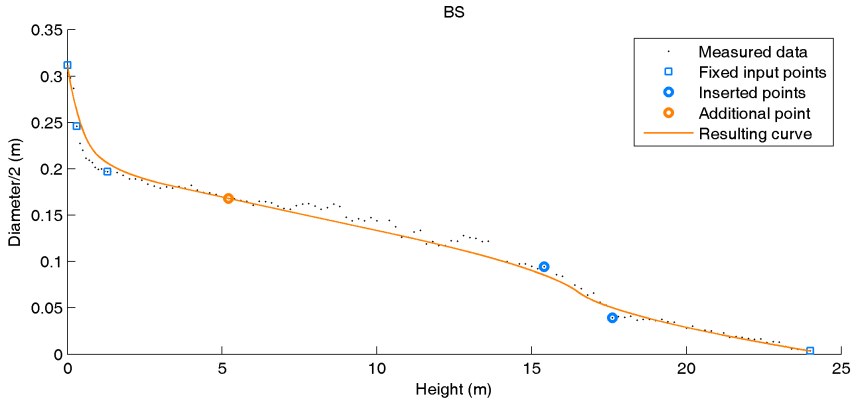


(d) Six points only

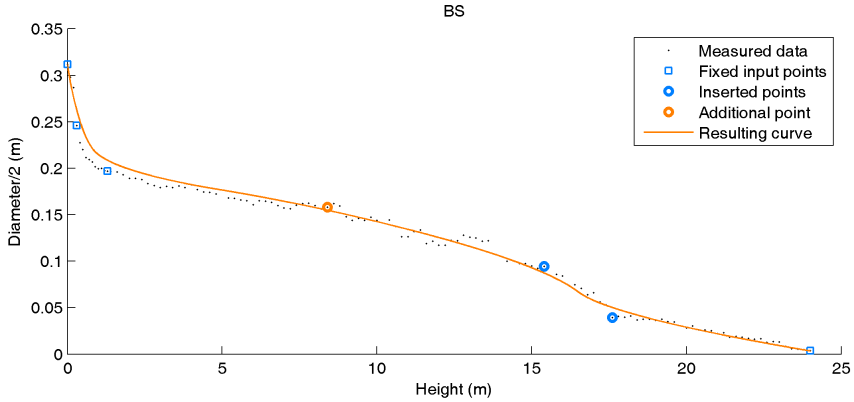
Figure C.3: Irregular stem curves represented by natural cubic spline with different 7^{th} point placement



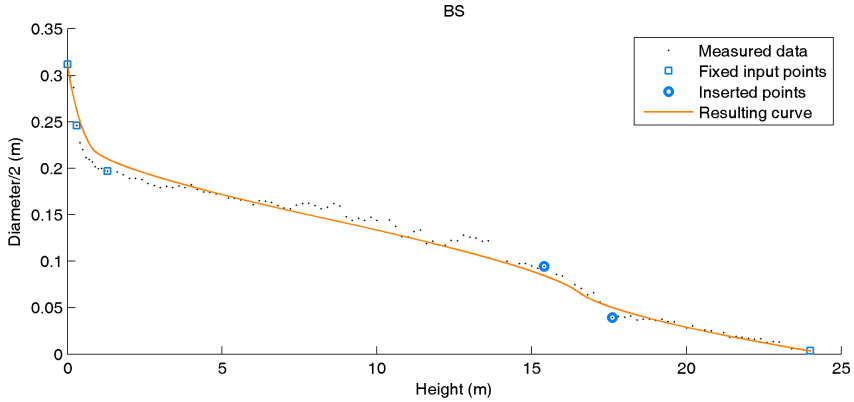
(a) Optimal position



(b) 20 % of the stem height



(c) Center of interspace

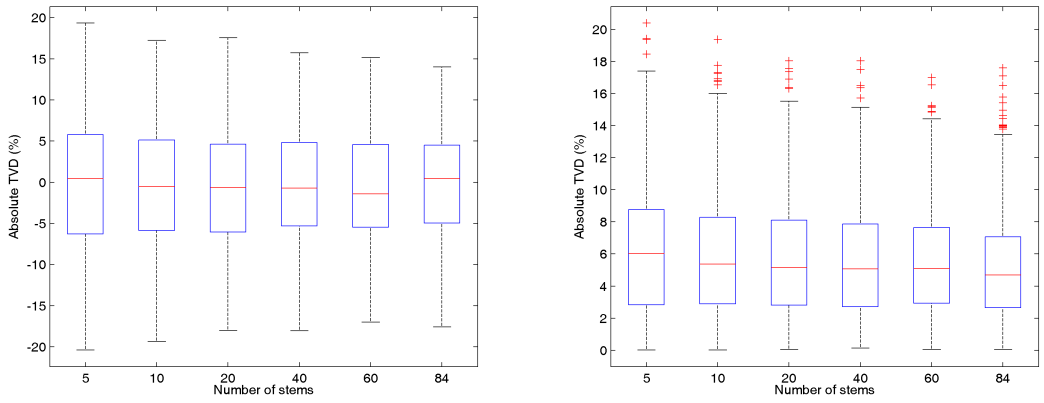


(d) Six points only

Figure C.4: Irregular stem curves represented by B-spline with different 7^{th} point placement

Appendix D

Comparison of taper regression models



(a) Total volume difference (b) Absolute total volume difference

Figure D.1: Development of accuracy of stem volume prediction with rising number of stems incorporated in the P-spline regression model

Table D.1: Comparison of taper models based on 60 stems. For each statistic mean value and standard deviation is shown. Values in a column followed by the same letter indicate no significant difference between spline types. Stars in columns DB and TVD indicate mean values significantly different from zero.

Taper model	**DB (10^{-2} m)**	**MAR (10^{-2} m)**	**SDR (10^{-2} m)**	**MSR (10^{-3} m^2)**	**TVD (%)**
Cervera	0.20 1.29 a*	1.65 0.61 a	1.99 0.57 a	0.48 0.34 a	-1.56 6.53 a*
Max-Burkhart	0.29 1.37 a*	1.38 0.72 b,c	1.58 0.71 b,c	0.37 0.38 b,c	0.22 6.80 b
Bi	0.07 1.33 a	1.49 0.81 b	1.74 0.83 d	0.43 0.54 a,b	-0.69 6.30 a,b*
Lee et al.	0.24 1.18 a*	1.68 0.54 a	2.07 0.51 a	0.50 0.27 a	0.25 5.99 b
Smoothing s.	0.14 1.38 a*	1.45 0.70 b,c	1.65 0.66 b,d	0.39 0.36 b,c	0.26 6.80 b
P-spline	0.15 1.28 a*	1.32 0.67 c	1.51 0.66 c	0.33 0.34 c	-0.30 6.42 a,b

Table D.2: Comparison of taper models based on 40 stems

Taper model	**DB (10^{-2} m)**	**MAR (10^{-2} m)**	**SDR (10^{-2} m)**	**MSR (10^{-3} m^2)**	**TVD (%)**
Cervera	0.21 1.34 a,b*	1.66 0.62 a	1.99 0.57 a	0.49 0.34 a	-1.48 6.82 a*
Max-Burkhart	0.15 1.47 a,b*	1.47 0.75 b,c	1.68 0.74 b,c	0.42 0.41 a,b	-0.27 7.27 a,b,c
Bi	0.00 1.35 a	1.52 0.78 b,d	1.78 0.83 b	0.45 0.49 a	-0.90 6.34 a,b*
Lee et al.	0.38 1.15 b*	1.64 0.53 a,d	2.02 0.47 a	0.48 0.26 a	0.98 6.12 c*
Smoothing s.	0.09 1.32 a	1.43 0.66 b,c	1.64 0.61 c	0.37 0.33 b	-0.06 6.54 b,c
P-spline	0.10 1.30 a	1.38 0.64 b	1.58 0.64 c	0.35 0.33 b	-0.63 6.39 a,b

Table D.3: Comparison of taper models based on 20 stems

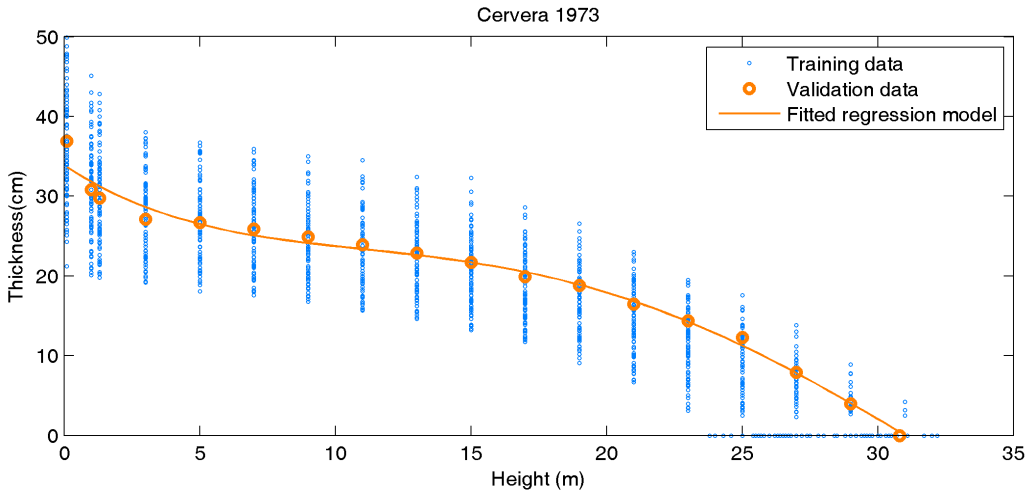
Taper model	**DB (10^{-2} m)**	**MAR (10^{-2} m)**	**SDR (10^{-2} m)**	**MSR (10^{-3} m^2)**	**TVD (%)**
Cervera	0.25 1.34 a*	1.69 0.63 a	2.01 0.59 a	0.50 0.37 a	-1.40 6.77 a*
Max-Burkhart	0.20 1.42 a*	1.42 0.73 b	1.64 0.72 b	0.40 0.39 a,b	-0.08 6.83 a,b
Bi	0.10 1.67 a	1.68 1.10 a	1.94 1.18 a,c	0.61 0.99 c	-0.47 7.34 a,b
Lee et al.	0.23 1.14 a*	1.66 0.53 a	2.05 0.51 a	0.48 0.28 a	0.02 5.88 b
Smoothing s.	0.23 1.51 a*	1.59 0.79 a	1.81 0.73 c	0.47 0.44 a,b	0.38 7.33 b
P-spline	0.25 1.35 a*	1.41 0.69 b	1.61 0.66 b	0.37 0.37 b	-0.14 6.64 a,b

Table D.4: Comparison of taper models based on 10 stems

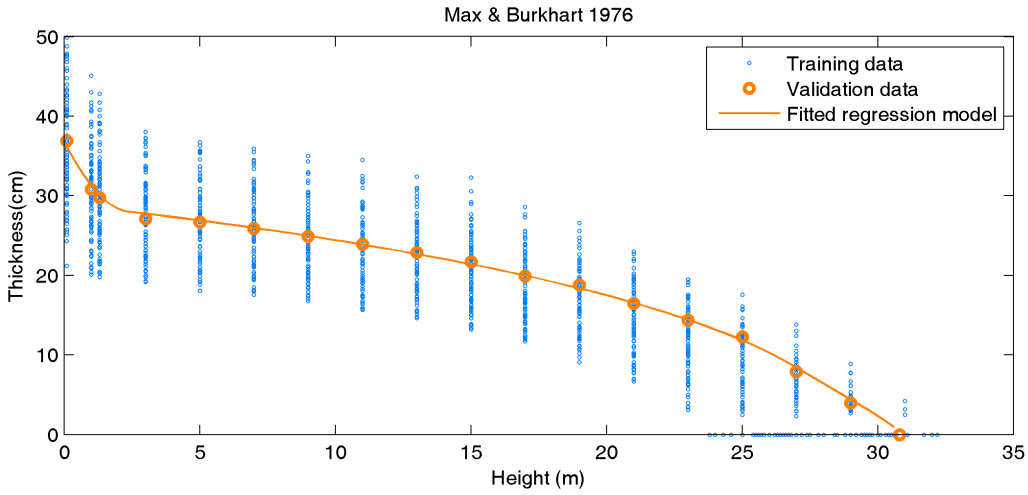
Taper model	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
Cervera	0.36	1.38	a*	1.69	0.67	a,b	2.01	0.62	a	0.51	0.41	a	-0.73	6.92	a*
Max-Burkhart	0.21	1.33	a*	1.37	0.68	c	1.58	0.67	b	0.36	0.36	a	-0.03	6.75	a,b
Bi	0.23	2.07	a*	1.85	1.52	a	2.17	1.79	a	0.94	3.08	b	-0.02	8.82	a,b
Lee et al.	0.22	1.21	a*	1.70	0.54	a,b	2.09	0.52	a	0.51	0.30	a	-0.09	6.02	a,b
Smoothing s.	0.29	1.43	a*	1.59	0.71	b	1.82	0.67	c	0.45	0.38	a	0.79	7.10	b*
P-spline	0.18	1.41	a*	1.52	0.70	b,c	1.74	0.66	b,c	0.42	0.37	a	-0.58	6.89	a,b

Table D.5: Comparison of taper models based on 5 stems

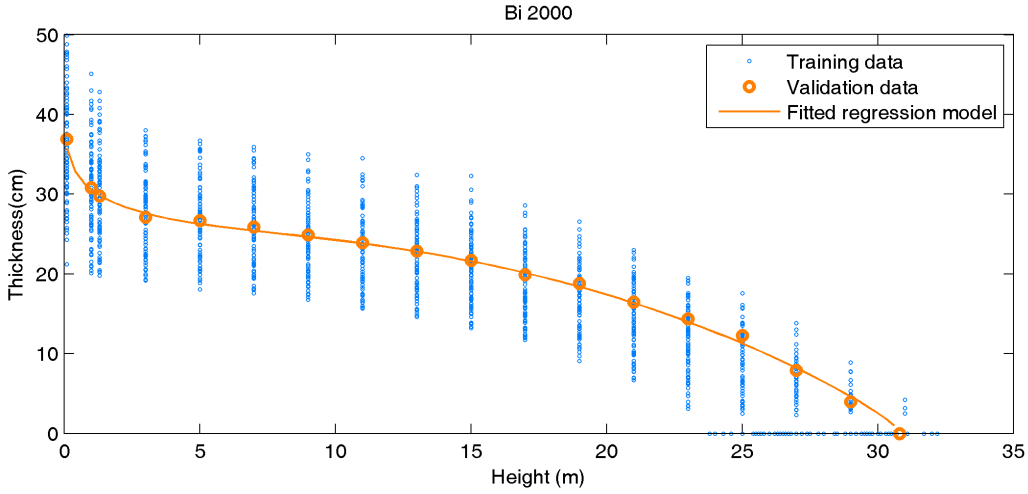
Taper model	DB (10^{-2} m)			MAR (10^{-2} m)			SDR (10^{-2} m)			MSR (10^{-3} m^2)			TVD (%)		
Cervera	0.22	1.46	a*	1.78	0.67	a	2.09	0.62	a	0.55	0.39	a	-1.55	7.41	a*
Max-Burkhart	0.18	1.46	a*	1.52	0.70	b	1.74	0.68	b	0.43	0.38	b	-0.16	7.48	a,b
Bi	0.60	8.09	a	3.38	7.53	c	4.34	14.08	a	23.41	293.06	c	59.57	1115.70	a,b
Lee et al.	0.38	1.40	a*	1.79	0.70	a	2.16	0.65	a	0.58	0.55	a,c	0.74	7.31	b*
Smoothing s.	0.31	1.53	a*	1.70	0.76	a	1.94	0.68	c	0.51	0.43	a,d	0.76	7.53	b*
P-spline	0.30	1.53	a*	1.65	0.77	a,b	1.85	0.70	b,c	0.48	0.41	b,d	-0.07	7.67	a,b



(a) Cervera (1973)

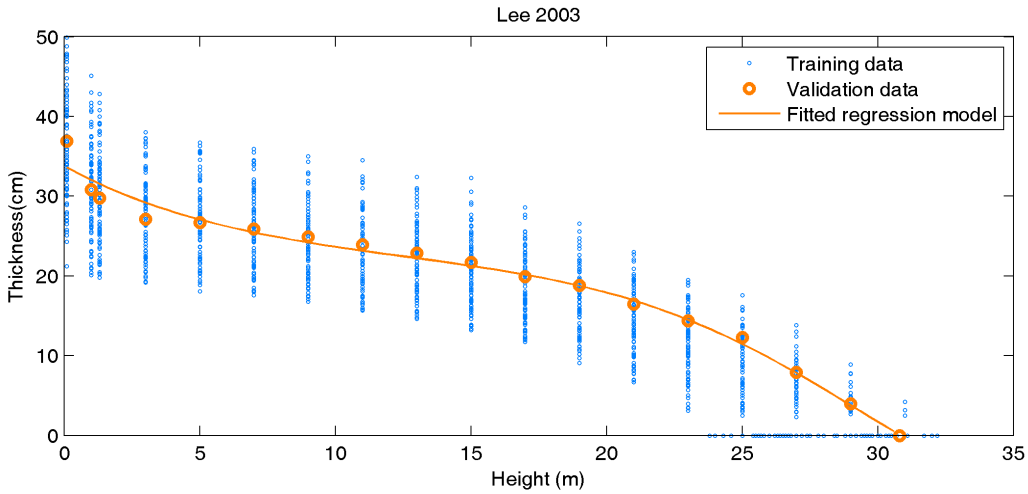


(b) Max and Burkhart [1976]

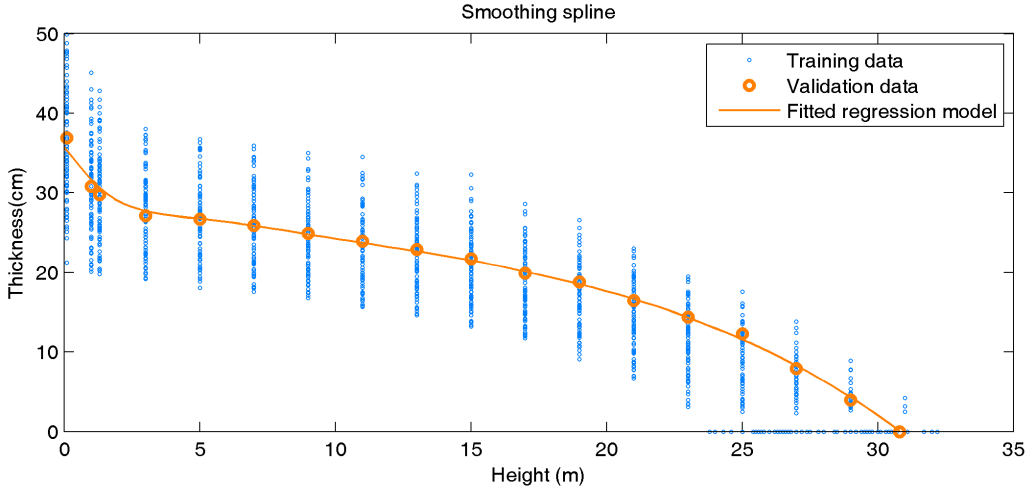


(c) Bi [2000]

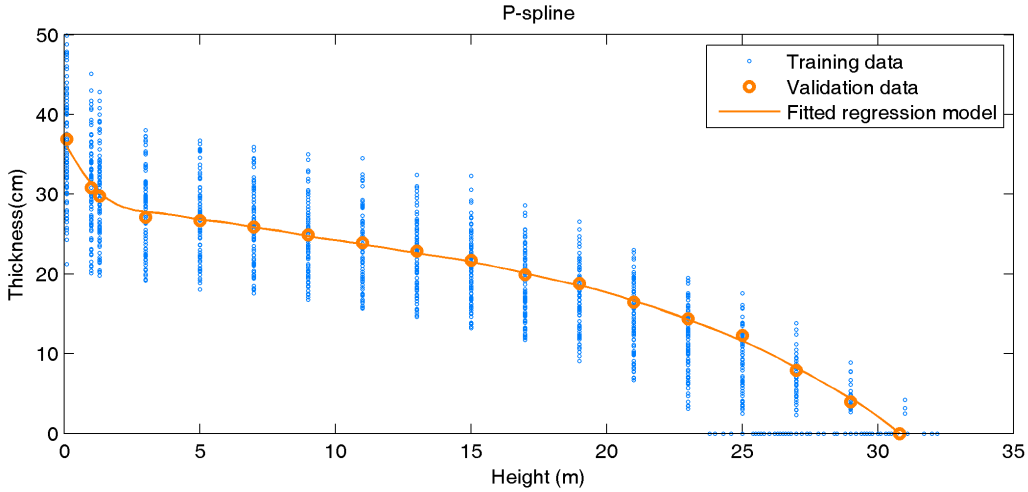
Figure D.2: Selected taper models based on 84 stems and the validation stem fitted by the model



(a) Lee et al. [2003]

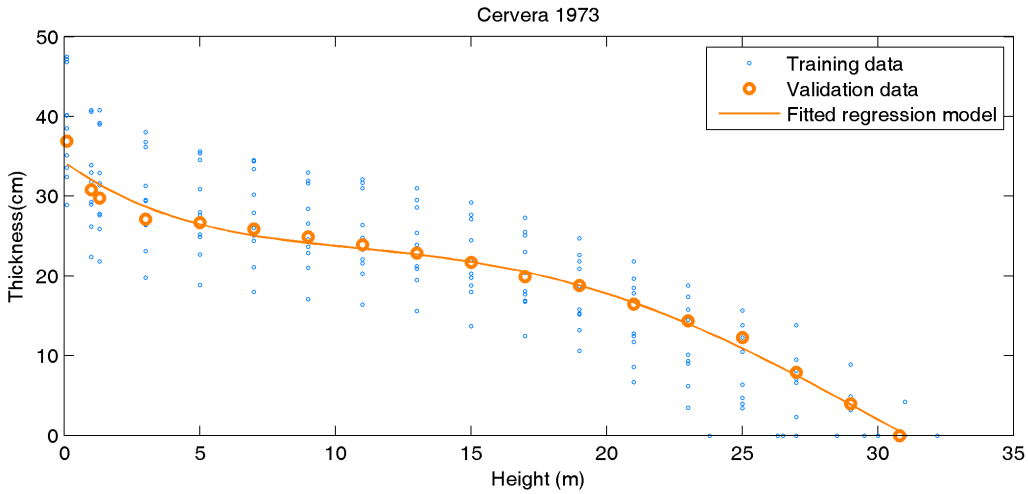


(b) Smoothing spline

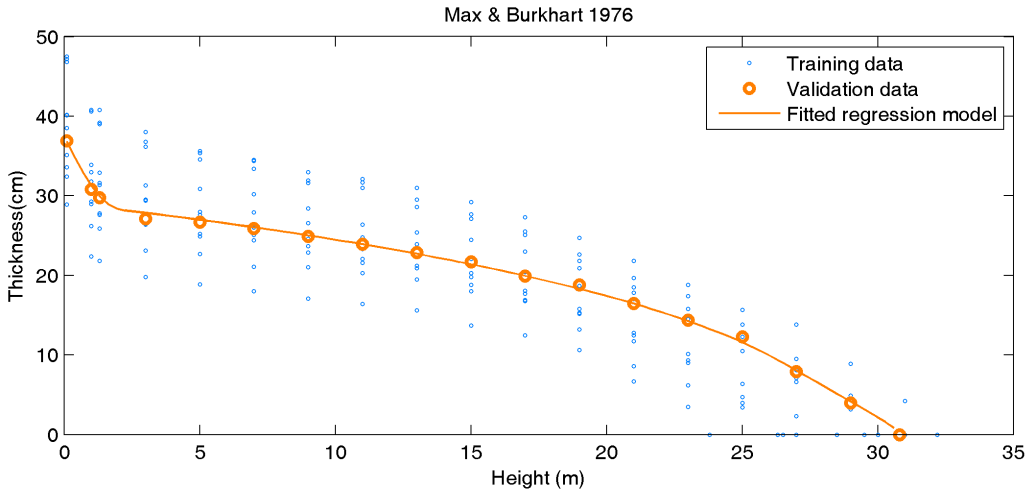


(c) P-spline

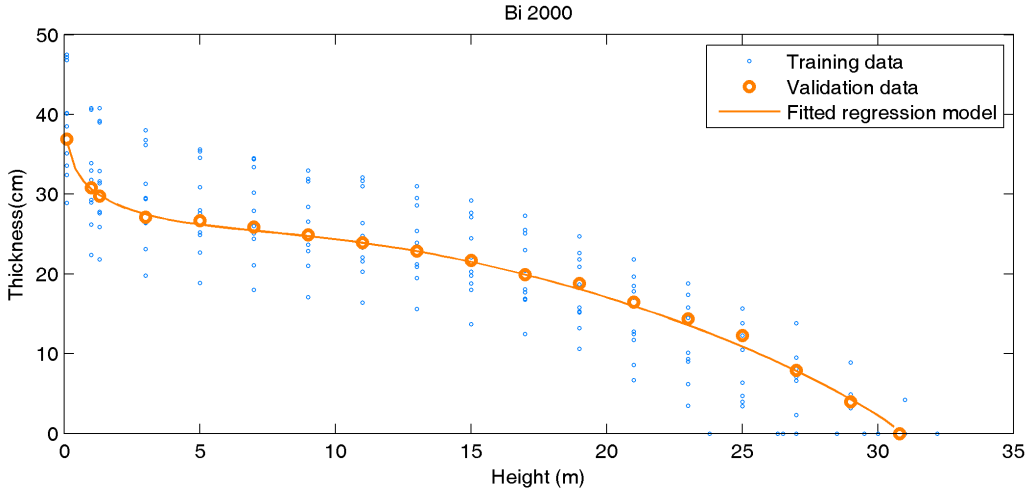
Figure D.3: Selected taper models based on 84 stems and the validation stem fitted by the model



(a) Cervera (1973)

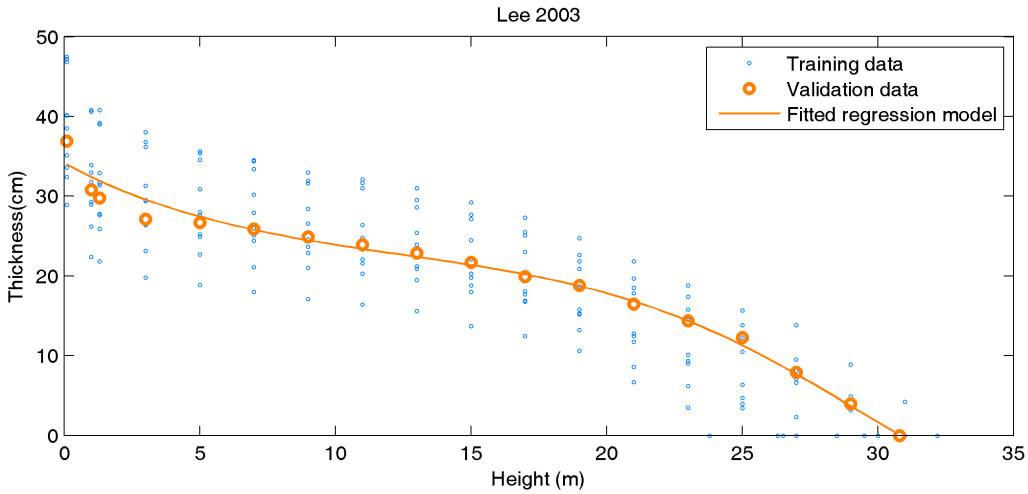


(b) Max and Burkhart [1976]

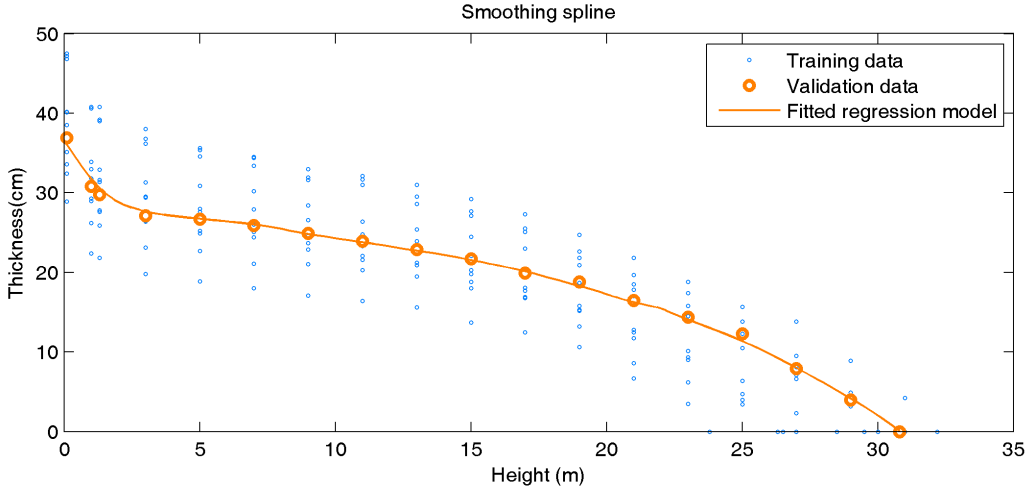


(c) Bi [2000]

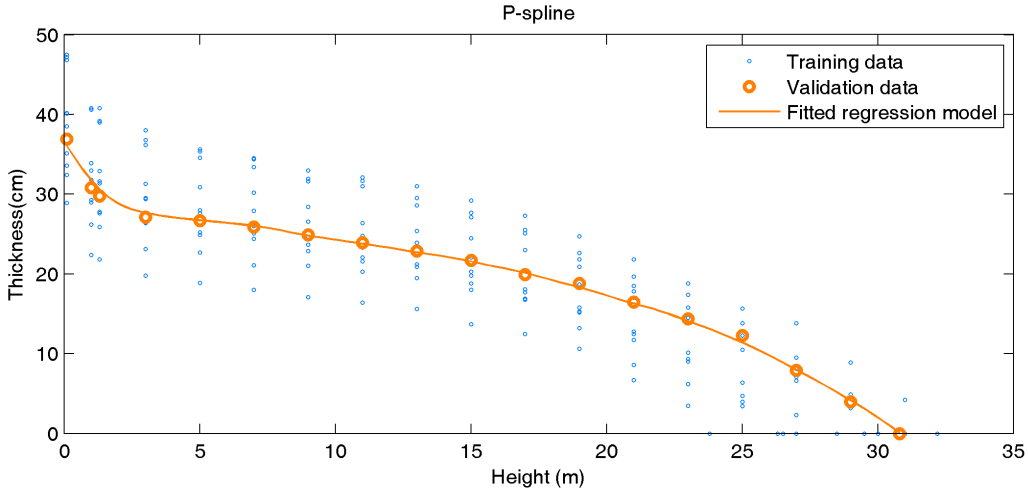
Figure D.4: Selected taper models based on 10 stems and the validation stem fitted by the model



(a) Lee et al. [2003]



(b) Smoothing spline



(c) P-spline

Figure D.5: Selected taper models based on 10 stems and the validation stem fitted by the model